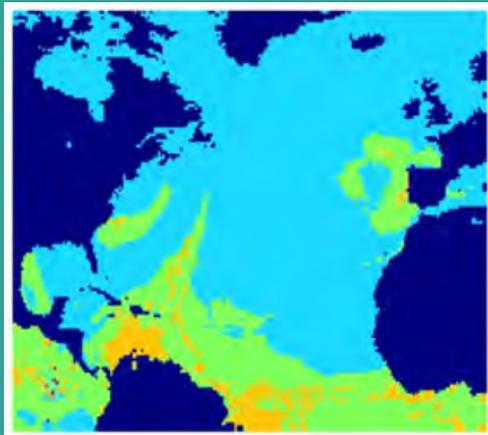


Florentin Smarandache, Jean Dezert, Albena Tchamova

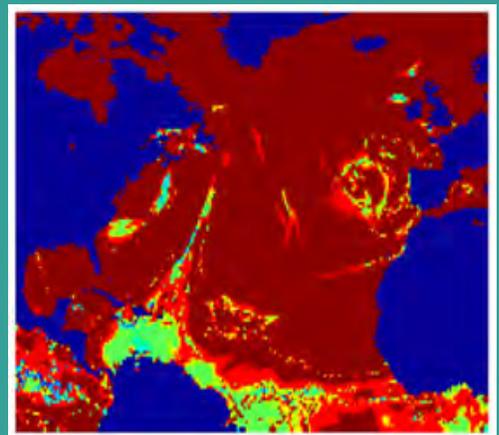
*Editors*

# Advances and Applications of DS<sub>m</sub>T for Information Fusion

*Collected Works. Volume 5*



Storm risk predictions on Atlantic Ocean  
(PCR6 fusion of five real physical criteria).



Confidence of predictions of Soft Electre Tri  
method based on PCR6 fusion.

Biblio Publishing

2023



**Biblio Publishing**  
**1091 West 1st Ave**  
**Grandview Heights, OH 43212**  
**United States of America**  
**Ph. 614.485.0721**  
**Em. [Info@BiblioPublishing.com](mailto:Info@BiblioPublishing.com)**  
**<https://BiblioPublishing.com/>**

*About the cover:* The left image shows the result of the storm risk predictions on Atlantic Ocean based on Soft Electre Tri (SET) method used with five criteria (the 3-h accumulated precipitation, the lifted index, the convective available potential energy, the low-level wind divergence and the divergence of the wind above the top of the convective clouds) with real data, and used with the PCR6 fusion rule of combination of belief functions. The right image shows the quality (confidence level) of the result for each pixel of the left image. For more details, see paper #34 of Part 2 (pages 447-454) of this volume, entitled: *Multi-Criteria Information Fusion for Storm Prediction Based on Belief Functions*, by J. Dezert, A. Bouchard and M. Buguet.

Freely download this book from <http://fs.unm.edu/DSmT-book5.pdf> .

Download books from Digital Library of Science at:  
<http://fs.unm.edu//ScienceLibrary.htm> .

© 2023 by the Editors, the Authors for their articles, and the Publishing House.

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying the content (or some part(s) of it) are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder, and without giving the exact original reference or source.

ISBN: 978-1-59973-773-7

**F. Smarandache, J. Dezert & A. Tchamova (Editors)**

**Advances and Applications  
of DSMT for Information Fusion**

**Collected Works. Volume 5**

**Biblio Publishing**

**2023**



## In memoriam

This book is dedicated to the memory of our good friends and colleagues  
Dr. Ludmil Bojilov (1944-2023) and Prof. Arnaud Martin (1974-2023)  
who have always been active and enthusiastic in tracking,  
and in the development of belief functions.  
We will never forget them.



Ludmil Bojilov.



Albena Tchamova, Ludmil Bojilov,  
Jean Dezert, Florentin Smarandache in Varna.



Arnaud Martin.



Arnaud Martin, Florentin Smarandache in Brest.



Arnaud Martin, Jean Dezert, Grégoire Mercier in Xi'an.



# Table of Contents

F. Smarandache, J. Dezert, A. Tchamova (Editors) Foreword	13
--	----

## PART 1 - THEORETICAL ADVANCES ON DS<sub>m</sub>T

F. Smarandache, J. Dezert Modified PCR Rules of Combination with Degrees of Intersections	17
Y. Dong, X. Li, J. Dezert A Hierarchical Flexible Coarsening Method to Combine BBAs in Probabilities	27
N.R. Mahato, L. Jaulin, S. Chakraverty, J. Dezert Validated Enclosure of Uncertain Nonlinear Equations using SIVIA Monte-Carlo	35
Y. Dong, X. Li, J. Dezert Rough Set Classifier Based on DS <sub>m</sub> T	43
J. Dezert, F. Smarandache Canonical Decomposition of Dichotomous Basic Belief Assignment	51
J. Dezert, F. Smarandache Canonical Decomposition of Basic Belief Assignment for Decision-Making Support	61
J. Dezert, F. Smarandache, A. Tchamova, D. Han Fast Fusion of Basic Belief Assignments Defined on a Dichotomous Frame of Discernment	69
J. Dezert, S. Fidanova, A. Tchamova Fast BF-ICrA Method for the Evaluation of MO-ACO Algorithm for WSN Layout	77
T. Dezert, J. Dezert, F. Smarandache Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments	85
J. Dezert, F. Smarandache Explicit Formulas of PCR5 and PCR6 Fusion Rules for Three Bayesian Basic Belief Assignments	109

## PART 2 - APPLICATIONS OF DS<sub>m</sub>T

J. Tian, P. Reinartz, J. Dezert Building Change Detection in Satellite Stereo Imagery Based on Belief Functions	121
J. Tian, J. Dezert, P. Reinartz Refined Building Change Detection in Satellite Stereo Imagery Based on Belief Functions and Reliabilities	127
M. Liu, X. Li, J. Dezert, C. Luo Generic Object Recognition Based on the Fusion of 2D and 3D SIFT Descriptors	133

<b>J. Dezert, K. Benameur, L. Ratton, J.-F. Grandin</b> <b>On the Quality Estimation of Optimal Multiple Criteria Data Association Solutions</b>	<b>141</b>
<b>J. Dezert, J. Moras, B. Pannetier</b> <b>Environment Perception Using Grid Occupancy Estimation with Belief Functions</b>	<b>149</b>
<b>J. Moras, J. Dezert, B. Pannetier</b> <b>Grid Occupancy Estimation for Autonomous Vehicle Perception</b>	<b>157</b>
<b>S. Carladous, J.-M. Tacnet, J. Dezert, M. Batton-Hubert</b> <b>Belief Function Theory Based Decision Support Methods: Application to Torrent Protection Work Effectiveness and Reliability Assessment</b>	<b>165</b>
<b>J. Dezert, A. Tchamova, P. Konstantinova</b> <b>The Impact of the Quality Assessment of Optimal Assignment for Data Association in a Multitarget Tracking Context</b>	<b>173</b>
<b>J. Dezert, A. Tchamova, P. Konstantinova</b> <b>Multitarget Tracking Performance based on the Quality Assessment of Data Association</b>	<b>179</b>
<b>J. Dezert, P. Konstantinova, A. Tchamova</b> <b>Performance Evaluation of Improved QADA-KF and JPDAF for Multitarget Tracking in Clutter</b>	<b>187</b>
<b>J. Dezert, A. Tchamova, P. Konstantinova, E. Blasch</b> <b>A Comparative Analysis of QADA-KF with JPDAF for Multitarget Tracking in Clutter</b>	<b>195</b>
<b>J. Dezert, D. Han, J.-M. Tacnet</b> <b>Multi-Criteria Decision-Making with Imprecise Scores and BF- TOPSIS</b>	<b>205</b>
<b>H. Anzid, G. Le Goic, A. Bekkari, A. Mansouri, D. Mammass</b> <b>Enhancing Multimodal Image Fusion: A Novel DSMT and SVM Classification Approach with Integrated Saliency Model</b>	<b>213</b>
<b>Y. Dong, X. Li, J. Dezert</b> <b>A New Probabilistic Transformation Based on Evolutionary Algorithm for Decision Making</b>	<b>221</b>
<b>Y. Dong, X. Li, J. Dezert</b> <b>A Hierarchical Flexible Coarsening Method to Combine BBAs in Probabilities</b>	<b>229</b>
<b>Y. Dong, X. Li, Y. Liu</b> <b>A Fast Combination Method in DSMT and its Application to Recommender System</b>	<b>237</b>
<b>T. Dezert, S. Palma Lopes, Y. Fargier, P. Côte</b> <b>Geophysical and Geotechnical Data Fusion for Levee Assessment - Interface Detection with Biased Geophysical Data</b>	<b>251</b>
<b>A. Tchamova, J. Dezert, P. Konstantinova, N. Bocheva, B. Genova, M. Stefanova</b> <b>Human Heading Perception Based on Form and Motion Combination</b>	<b>255</b>
<b>J. Tian, J. Dezert</b> <b>Fusion of Multispectral Imagery and DSMs for Building Change Detection Using Belief Functions and Reliabilities</b>	<b>263</b>
<b>V. Dragos, J. Dezert, K. Rein</b> <b>Tracking Uncertainty Propagation from Model to Formalization: Illustration on Trust Assessment</b>	<b>281</b>

<b>I.-A. Gal, D.Bucur, L. Vlădăreanu</b> <b>DSmT Decision-Making Algorithms for Finding Grasping Configurations of Robot Dexterous Hands</b>	<b>299</b>
<b>T. Dezert, S. Palma Lopes, Y. Fargier, P. Côte</b> <b>Combination of Geophysical and Geotechnical Data Using Belief Functions: Assessment with Numerical and Laboratory Data</b>	<b>315</b>
<b>K.G. Ghalem, F. Hendel</b> <b>Dual Iris Authentication System Using Dezert-Smarandache Theory</b>	<b>331</b>
<b>M. Fidali, W. Jamrozik</b> <b>Method of Classification of Global Machine Conditions Based on Spectral Features of Infrared Images and Classifiers Fusion</b>	<b>339</b>
<b>S. Fidanova, J. Dezert, A. Tchamova</b> <b>Inter-Criteria Analysis Based on Belief Functions for GPS Surveying Problems</b>	<b>349</b>
<b>J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet</b> <b>Simplification of Multi-Criteria Decision-Making Using Inter-Criteria Analysis and Belief Functions</b>	<b>357</b>
<b>Y.-W. Du, W. Zhou</b> <b>DSmT-Based Group DEMATEL Method with Reaching Consensus</b>	<b>365</b>
<b>Y. Dong, X. Li, J. Dezert, S.S. Ge</b> <b>A Novel Multi-Criteria Discounting Combination Approach for Multi-Sensor Fusion</b>	<b>381</b>
<b>Z. Zhang, D. Han, J. Dezert, Y. Yang</b> <b>A New Image Registration Algorithm Based on Evidential Reasoning</b>	<b>393</b>
<b>J. Dezert, A. Tchamova, S. Fidanova, D. Han</b> <b>Two Applications of Inter-Criteria Analysis with Belief Functions</b>	<b>407</b>
<b>Y. Dong, X. Li, J. Dezert, M.O. Khyam, Md. Noor-A-Rahim, S.S. Ge</b> <b>DSmT-Based Fusion Strategy for Human Activity Recognition in Body Sensor Networks</b>	<b>419</b>
<b>A. Tchamova, J. Dezert, N. Bocheva, P. Konstantinova, B. Genova, M. Stefanova</b> <b>A Study on Human Learning Ability During Classification of Motion and Colour Visual Cues and Their Combination</b>	<b>431</b>
<b>T. Dezert, J. Dezert</b> <b>Improvement of Proportional Conflict Redistribution Fusion Rules for Levee Characterization</b>	<b>439</b>
<b>J. Dezert, A. Bouchard, M. Buguet</b> <b>Multi-Criteria Information Fusion for Storm Prediction Based on Belief Functions</b>	<b>447</b>
<b>M. Boumediene, H. Zebiri, J. Dezert</b> <b>Evidential Data Association Based on Dezert-Smarandache Theory</b>	<b>455</b>
<b>T. Dezert, S. Palma Lopes, Y. Fargier, L. Saussaye, P. Côte</b> <b>Data Fusion of In Situ Geophysical and Geotechnical Information for Levee Characterization</b>	<b>465</b>
<b>T. Dezert, Y. Fargier, S. Palma Lopes, V. Guihard</b> <b>Canal Dike Characterization by Means of Electrical Resistivity, Shear Wave Velocity and Particle Size Data Fusion</b>	<b>481</b>

<b>T. Pietkiewicz</b> <b>Application of Dezert-Smarandache Theory Rules in Fusion of Identification Information from ESM Sensors and Radars</b>	<b>495</b>
<b>T. Pietkiewicz, A.Wróbel</b> <b>Maritime Objects Upon FLIR Images Recognition by Fusion of Two SVM Classifiers Using Different Image Descriptors</b>	<b>521</b>
<b>Y. Dong, X. Li, J. Dezert, R. Zhou, C. Zhu, L. Cao, M.O. Khyam, S.S. Ge</b> <b>Multisource Weighted Domain Adaptation With Evidential Reasoning for Activity Recognition</b>	<b>529</b>
<b>F. Dambreville, J. Dezert</b> <b>Flexible User-Oriented Rust Toolbox for Information Fusion</b>	<b>541</b>
<b>Y. Dong, K. Xu, C. Zhu, E. Guan, Y. Liu</b> <b>E-FPN: Evidential Feature Pyramid Network for Ship Classification</b>	<b>557</b>

### **PART 3 - CONTRIBUTIONS TO BELIEF FUNCTIONS**

<b>J. Dezert, A. Tchamova, D. Han</b> <b>A Real Z-box Experiment for Testing Zadeh's Example</b>	<b>575</b>
<b>D. Han, J. Dezert, Y. Yang</b> <b>Two Novel Methods for BBA Approximation Based on Focal Element Redundancy</b>	<b>581</b>
<b>D. Han, J. Dezert, Z. Duan</b> <b>Evaluation of Probability Transformations of Belief Functions for Decision-Making</b>	<b>589</b>
<b>J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang</b> <b>Decision-Making with Belief Interval Distance</b>	<b>607</b>
<b>J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, H. Yin</b> <b>The BF-TOPSIS Approach for Solving Non- Classical MCDM Problems</b>	<b>613</b>
<b>F. Smarandache, J. Dezert, A. Tchamova</b> <b>Examples Where Dempster's Rule is Insensitive to the Conflict Level Between the Sources of Evidence</b>	<b>619</b>
<b>Z. Zhang, D. Han, J. Dezert, Y. Yang</b> <b>Determination of Basic Belief Assignment Using Fuzzy Numbers</b>	<b>623</b>
<b>J. Ding, D. Han, J. Dezert, Y. Yang</b> <b>Comparative Study on BBA Determination Using Different Distances of Interval Numbers</b>	<b>629</b>
<b>J. Tian, J. Dezert, R. Qin</b> <b>Time-series 3D Building Change Detection Based on Belief Functions</b>	<b>635</b>
<b>D. Han, J. Dezert, Y. Yang</b> <b>Belief interval Based Distances Measures in the Theory of Belief Functions</b>	<b>641</b>
<b>J. Dezert, A. Tchamova, D. Han</b> <b>Credibilistic Independence of Two Propositions</b>	<b>659</b>
<b>J. Dezert, A. Tchamova, D. Han</b> <b>Correct Proof of Shafer's Belief Conditioning Formulas</b>	<b>667</b>

<b>J. Dezert, A. Tchamova, D. Han</b> <b>Total Belief Theorem and Generalized Bayes' Theorem</b>	<b>673</b>
<b>J. Dezert, A. Tchamova, D. Han</b> <b>Total Belief Theorem and Conditional Belief Functions</b>	<b>681</b>
<b>J. Dezert, A. Tchamova, D. Han, T. Wickramaratne</b> <b>A Simplified Formulation of Generalized Bayes' Theorem</b>	<b>695</b>
<b>Z. Zhang, D. Han, J. Dezert, Y. Yang</b> <b>A New Adaptive Switching Median Filter for Impulse Noise Reduction with Pre-Detection Based on Evidential Reasoning</b>	<b>705</b>
<b>J. Abellán, E. Bossé</b> <b>Drawbacks of Uncertainty Measures Based on the Pignistic Transformation</b>	<b>725</b>
<b>D. Han, Y. Yang, J. Dezert</b> <b>On Decomposition of Belief Function</b>	<b>733</b>
<b>X. Fan, D. Han, J. Dezert, Y. Yang</b> <b>User-Specified Optimization Based Transformation of Fuzzy Membership Into Basic Belief Assignment</b>	<b>739</b>
<b>Y. Yang, D. Han, J. Dezert</b> <b>Basic Belief Assignment Approximations Using Degree of Non-Redundancy for Focal Element</b>	<b>747</b>
<b>Q. Wei, X. Li, J. Dezert</b> <b>Approximation of Basic Belief Assignment Based on Focal Element Compatibility</b>	<b>761</b>
<b>M. Boumediene, J. Dezert</b> <b>Evaluation of Probabilistic Transformations for Evidential Data Association</b>	<b>769</b>
<b>S. Barhoumi, I.K. Kallel, S.A. Bouhamed, E. Bossé, B. Solaiman</b> <b>Generation of Fuzzy Evidence Numbers for the Evaluation of Uncertainty Measures</b>	<b>775</b>
<b>J. Abellán, E. Bossé</b> <b>Critique of Recent Uncertainty Measures Developed Under the Evidence Theory and Belief Intervals</b>	<b>781</b>
<b>J. Dezert, A. Tchamova</b> <b>On the Effectiveness of Measures of Uncertainty of Basic Belief Assignments</b>	<b>789</b>
<b>J. Dezert</b> <b>An Effective Measure of Uncertainty of Basic Belief Assignments</b>	<b>803</b>
<b>J. Dezert, F. Smarandache, A. Tchamova</b> <b>Analytical Solution of the Simplest Entropiece Inversion Problem</b>	<b>811</b>
<b>J. Dezert, A. Tchamova, D. Han</b> <b>Measure of Information Content of Basic Belief Assignments</b>	<b>817</b>
<b>J. Dezert, A. Tchamova</b> <b>Erratum of paper entitled: On the Validity of Dempster's Fusion Rule and its Interpretation as a Generalization of Bayesian Fusion Rule</b>	<b>823</b>
<b>J. Dezert, A. Tchamova</b> <b>On Inequalities Bounding Imprecision and Nonspecificity Measures of Uncertainty</b>	<b>827</b>

<b>X. Fan, D. Han, J. Dezert, Y. Yang</b> <b>Novel Moderate Transformation of Fuzzy Membership Function into Basic Belief Assignment</b>	<b>833</b>
<b>J. Dezert, F. Dambreville</b> <b>Cross-Entropy and Relative Entropy of Basic Belief Assignments</b>	<b>851</b>
<b>J. Dezert, F. Dambreville</b> <b>On Monotonicity Desideratum for an Efficient Entropy Measure of Basic Belief Assignments</b>	<b>859</b>
<b>J. Dezert, A. Tchamova</b> <b>Involutory Negator of Basic Belief Assignments Applied to Information Fusion</b>	<b>863</b>
<b>J. Dezert, A. Tchamova</b> <b>On Kenn's Rule of Combination Applied to Breast Cancer Precision Therapy</b>	<b>879</b>
<b>K. Zuo, X. Li, J. Dezert, Y. Dong</b> <b>Weighted Fusion of Multiple Classifiers for Human Activity Recognition</b>	<b>883</b>
<b>J. Niu, Z. Liu</b> <b>Imbalanced Data Classification Based on Belief Functions Theory</b>	<b>891</b>
<b>D.-V. Giurgi, M.N. Geletu, T. Josso-Laurain, M. Devanne, J-P. Lauffenburger, J. Dezert</b> <b>Decision Based on Belief Interval for Multi-class Obstacle Perception of Self-driving Cars</b>	<b>897</b>
<b>L. Deregnacourt, A. Lechery, H. Laghmar, S. Ainouz</b> <b>An Evidential Deep Network Based on Dempster- Shafer Theory for Large Dataset</b>	<b>907</b>
<b>Biographies of Contributors</b>	<b>915</b>

# Foreword

This fifth volume on Advances and Applications of DSMT for Information Fusion collects theoretical and applied contributions of researchers working in different fields of applications and in mathematics, and is available in open-access. The collected contributions of this volume have either been published or presented after disseminating the fourth volume in 2015 (available at [fs.unm.edu/DSMT-book4.pdf](http://fs.unm.edu/DSMT-book4.pdf) or [www.onera.fr/sites/default/files/297/2015-DSMT-Book4.pdf](http://www.onera.fr/sites/default/files/297/2015-DSMT-Book4.pdf)) in international conferences, seminars, workshops and journals, or they are new. The contributions of each part of this volume are chronologically ordered.

First Part of this book presents some theoretical advances on DSMT, dealing mainly with modified Proportional Conflict Redistribution Rules (PCR) of combination with degree of intersection, coarsening techniques, interval calculus for PCR thanks to set inversion via interval analysis (SIVIA), rough set classifiers, canonical decomposition of dichotomous belief functions, fast PCR fusion, fast inter-criteria analysis with PCR, and improved PCR5 and PCR6 rules preserving the (quasi-)neutrality of (quasi-)vacuous belief assignment in the fusion of sources of evidence with their Matlab codes.

Because more applications of DSMT have emerged in the past years since the apparition of the fourth book of DSMT in 2015, the second part of this volume is about selected applications of DSMT mainly in building change detection, object recognition, quality of data association in tracking, perception in robotics, risk assessment for torrent protection and multi-criteria decision-making, multi-modal image fusion, coarsening techniques, recommender system, levee characterization and assessment, human heading perception, trust assessment, robotics, biometrics, failure detection, GPS systems, inter-criteria analysis, group decision, human activity recognition, storm prediction, data association for autonomous vehicles, identification of maritime vessels, fusion of support vector machines (SVM), Silx-Furtif RUST code library for information fusion including PCR rules, and network for ship classification.

Finally, the third part presents interesting contributions related to belief functions in general published or presented along the years since 2015. These contributions are related with decision-making under uncertainty, belief approximations, probability transformations, new distances between belief functions, non-classical multi-criteria decision-making problems with belief functions, generalization of Bayes theorem, image processing, data association, entropy and cross-entropy measures, fuzzy evidence numbers, negator of belief mass, human activity recognition, information fusion for breast cancer therapy, imbalanced data classification, and hybrid techniques mixing deep learning with belief functions as well.

We want to thank all the contributors of this fifth volume for their research works and their interests in the development of DSMT, and the belief functions. We are grateful as well to other colleagues for encouraging us to edit this fifth volume, and for sharing with us several ideas and for their questions and comments on DSMT through the years. We thank the International Society of Information Fusion ([www.isif.org](http://www.isif.org)) for diffusing main research works related to information fusion (including DSMT) in the international fusion conferences series over the years.

Florentin Smarandache is grateful to The University of New Mexico, U.S.A., that many times partially sponsored him to attend international conferences, workshops and seminars on Information Fusion.

Jean Dezert is grateful to the Department of Information Processing and Systems (DTIS) of the French Aerospace Lab (Office National d'Études et de Recherches Aéropatiales), Palaiseau, France, for encouraging him to carry on this research and for its financial support.

Albena Tchamova is first of all grateful to Dr. Jean Dezert for the opportunity to be involved during more than 20 years to follow and share his smart and beautiful visions and ideas in the development of the powerful Dezert-Smarandache Theory for data fusion. She is also grateful to the Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, for sponsoring her to attend international conferences on Information Fusion.

## The Editors:

**Prof. Florentin Smarandache**  
Tucson, USA.  
<http://fs.unm.edu/DSMT.htm>  
<http://fs.unm.edu/FS.htm>  
Email: [smarand@unm.edu](mailto:smarand@unm.edu)

**Dr. Jean Dezert**  
Orléans, France.  
<https://www.onera.fr/fr/staff/jean-dezert>  
Email: [jean.dezert@onera.fr](mailto:jean.dezert@onera.fr)

**Dr. Albena Tchamova**  
Sofia, Bulgaria.  
<https://sdp.iict.bas.bg/staff/albenaEN.html>  
Email: [albena.tchamova@iict.bas.bg](mailto:albena.tchamova@iict.bas.bg)



Part 1:

**Theoretical advances  
on DSMT**



# Modified PCR Rules of Combination with Degrees of Intersections

Florentin Smarandache<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Department of Mathematics, University of New Mexico, Gallup, NM, USA.

<sup>b</sup>The French Aerospace Lab, ONERA/DTIS, Palaiseau, France.

Emails: smarand@unm.edu, jean.dezert@onera.fr

Originally published as: F. Smarandache, J. Dezert, *Modified PCR Rules of Combination with Degrees of Intersections*, in Proc. of the 18th Int. Conf. on Information Fusion (Fusion 2015), Washington D.C, USA, July 6–9, 2015, and reprinted with permission.

**Abstract**—In this paper, we propose a modification of PCR5 and PCR6 fusion rules with degrees of intersections for taking into account the cardinality of focal elements of each source of evidence to combine. We show in very simple examples the interest of these new fusion rules w.r.t. classical Dempster-Shafer, PCR6, Zhang’s and Jaccard’s Center rules of combination.

**Keywords:** Information fusion, belief functions, DSMT, PCR6, degrees of intersection.

## I. INTRODUCTION

In this paper, we propose modifications of the Proportional Conflict Redistribution rule no. 6 (PCR6) [1] (Vol. 3) for the combination of basic belief assignments (BBA’s) which integrate the degrees of intersections of focal elements of each source of evidence to combine. Because we consider two possible definitions of degrees of intersections (i.e. Zhang’s and Jaccard’s degrees) and also two normalization methods (simplest and sophisticate), we propose four modified versions of PCR6 rules<sup>1</sup>. After a brief presentation of classical rules of combination and a detailed presentation of our modified PCR6 rules, we evaluate and compare their behaviors in different emblematic examples to guide the choice of the most interesting one.

## II. BELIEF FUNCTIONS AND CLASSICAL FUSION RULES

Belief functions have been introduced by Shafer in 1976 from Dempster’s works [2] in Dempster-Shafer’s theory (DST) of evidence. DST is mainly characterized by a frame of discernment (FoD), sources of evidence represented by basic belief assignment (BBA), belief (Bel) and plausibility (Pl) functions, and Dempster’s rule of combination, denoted as DS rule in the sequel<sup>2</sup> of combination. DST has been modified and extended into Dezert-Smarandache theory [1] (DSMT) to work with quantitative or qualitative BBA and to combine the sources of evidence in a more efficient way thanks to new proportional conflict redistribution (PCR) fusion rules – see [3]–[6] for discussion and examples.

<sup>1</sup>The methodology proposed in this paper is general and can also be applied to modify similarly other PCR rules. Since we consider PCR6 rule the most efficient one [6], we focus our presentation on PCR6 only

<sup>2</sup>DS acronym standing for *Dempster-Shafer* since Dempster’s rule has been widely promoted by Shafer in the development of his mathematical theory of evidence.

More precisely, let’s consider a finite discrete FoD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , of the fusion problem under consideration and its fusion space  $G^\Theta$  which can be chosen either as the power-set  $2^\Theta$ , the hyper-power set<sup>3</sup>  $D^\Theta$ , or the super-power set  $S^\Theta$  depending on the model that fits with the problem [1]. A BBA associated with a given source of evidence is defined as the mapping  $m(\cdot) : G^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in G^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called mass of belief of  $A$  committed by the source of evidence. Belief and plausibility functions are defined by

$$\text{Bel}(A) = \sum_{\substack{B \subseteq A \\ B \in G^\Theta}} m(B), \quad \text{and} \quad \text{Pl}(A) = \sum_{\substack{B \cap A \neq \emptyset \\ B \in G^\Theta}} m(B). \quad (1)$$

If for some  $A \in G^\Theta$ ,  $m(A) > 0$  then  $A$  is called a focal element of the BBA  $m(\cdot)$ . When all focal elements are singletons and  $G^\Theta = 2^\Theta$  then the BBA  $m(\cdot)$  is called a Bayesian BBA [2] and its corresponding belief function  $\text{Bel}(\cdot)$  is homogeneous to a (possibly subjective) probability measure, and one has  $\text{Bel}(A) = P(A) = \text{Pl}(A)$ , otherwise in general one has  $\text{Bel}(A) \leq P(A) \leq \text{Pl}(A)$ ,  $\forall A \in G^\Theta$ . The vacuous BBA, or VBBA for short, representing a totally ignorant source is defined as  $m_v(I_t) = 1$ , where the total ignorance defined as  $I_t \triangleq \theta_1 \cup \theta_2 \cup \dots \cup \theta_n$  if the FoD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . Since in Shafer’s book [2], the total ignorance  $I_t$  is also denoted  $\Theta$ , we will adopt this notation in the sequel.

Many rules have been proposed in the literature over the decades (see [1], Vol. 2 for a detailed list of fusion rules) to combine several distinct sources of evidence represented by the BBA’s  $m_1(\cdot), m_2(\cdot), \dots, m_s(\cdot)$  ( $s \geq 2$ ) defined on same fusion space  $G^\Theta$ . In DST, the combination of  $s \geq 2$  BBA’s is traditionally accomplished with Dempster-Shafer (DS) rule [2] defined by  $m_{1, \dots, s}^{DS}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Theta$

$$m_{1, \dots, s}^{DS}(X) \triangleq \frac{1}{1 - m_{1, \dots, s}(\emptyset)} \sum_{\substack{X_1, \dots, X_s \in 2^\Theta \\ X_1 \cap \dots \cap X_s = X}} \prod_{i=1}^s m_i(X_i), \quad (2)$$

where the numerator of (2) is the mass of belief on the conjunctive consensus on  $X$ . The denominator  $1 - m_{1, \dots, s}(\emptyset)$  is a

<sup>3</sup>which corresponds to a Dedekind’s lattice, see [1] Vol. 1.

normalization constant. The total degree of conflict  $m_{1,\dots,s}(\emptyset)$  between the  $s$  sources of evidences is defined by

$$m_{1,\dots,s}(\emptyset) \triangleq \sum_{\substack{X_1,\dots,X_s \in 2^\Theta \\ X_1 \cap \dots \cap X_s = \emptyset}} \prod_{i=1}^s m_i(X_i). \quad (3)$$

DS rule is associative and commutative and preserves the neutrality of the VBBA.  $s$  sources of evidence are said in total conflict if  $m_{1,\dots,s}(\emptyset) = 1$ . In this case the combination of the sources by DS rule cannot be done because of the mathematical 0/0 indeterminacy in (2). In DS rule,  $m_{1,\dots,s}(\emptyset)$  is redistributed to all focal elements of the conjunctive operator only proportionally to their mass (i.e. without taking care of their cardinalities). So with DS rule and with combination of 2 BBA's, the product  $m_1(X_1)m_2(X_2)$  is transferred to  $X_1 \cap X_2 = X$  only, no matter how the ratio between the cardinality of  $X$  and  $X_1 \cup X_2$  varies. This DS principle of redistribution has been questioned by Zhang in [7] and Fixsen and Malher in [8] because it does not discriminate the case where  $X_1 \cup X_2$  is large but  $X_1 \cap X_2$  is small with respect to the case where  $X_1 \cup X_2$  is small but  $X_1 \cap X_2$  is large. To palliate this problem, Zhang proposed in 1994 a modified version of DS rule [7] including a measure of degree of intersection of focal elements. The general formula of this modified DS rule is defined by  $m_{1,\dots,s}^D(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Theta$

$$m_{1,\dots,s}^D(X) \triangleq \frac{1}{K_{1,\dots,s}^D} \cdot \sum_{\substack{X_1,\dots,X_s \in 2^\Theta \\ X_1 \cap \dots \cap X_s = X}} D(X_1, \dots, X_s) \prod_{i=1}^s m_i(X_i), \quad (4)$$

where  $D(X_1, \dots, X_s)$  denotes a measure of the degree of intersection between the focal elements  $X_1, X_2, \dots, X_s$ , and where  $K_{1,\dots,s}^D$  is a normalization constant allowing to get  $\sum_{X \in 2^\Theta} m_{1,\dots,s}^D(X) = 1$ . Because the measure of degree of intersection  $D(X_1, \dots, X_s)$  can be defined in different ways, this yields to different versions of the modified DS rule above. In [7], Zhang suggested to define  $D(X_1, \dots, X_s)$  as

$$D^Z(X_1, \dots, X_s) \triangleq \frac{|X_1 \cap X_2 \cap \dots \cap X_s|}{|X_1| \cdot |X_2| \cdot \dots \cdot |X_s|}, \quad (5)$$

where  $|X_1 \cap X_2 \cap \dots \cap X_s|$  is the cardinality of the intersection of the focal elements  $X_1, X_2, \dots, X_s$ , and  $|X_1|, |X_2|, \dots, |X_s|$  their cardinalities. Replacing  $D(X_1, \dots, X_s)$  by  $D^Z(X_1, \dots, X_s)$  in the formula (4) defines Zhang's Center Rule (ZCR) of combination [7], denoted  $m_{1,\dots,s}^{ZCR}(\cdot)$  in the sequel. The normalization constant of ZCR is denoted  $K_{1,\dots,s}^{ZCR}$ .

If we use Jaccard's index as measure of the degree of intersection [9] which is defined by

$$D^J(X_1, \dots, X_s) \triangleq \frac{|X_1 \cap X_2 \cap \dots \cap X_s|}{|X_1 \cup X_2 \cup \dots \cup X_s|}, \quad (6)$$

then we obtain Jaccard's center rule (JCR) of combination, and we denote it  $m_{1,\dots,s}^{JCR}(\cdot)$ , in replacing  $D(X_1, \dots, X_s)$  by  $D^J(X_1, \dots, X_s)$  in the formula (4). The normalization constant of JCR is denoted  $K_{1,\dots,s}^{JCR}$ . ZCR and JCR rules are particular instances of Modified DS rule (MDS) proposed by Fixsen and Mahler in [8]. ZCR and JCR are commutative but not

idempotent. It can be proved that Zhang's degree is associative that is  $D^Z(X_1, X_2, \dots, X_s) = D^Z(X_1, D^Z(X_2, \dots, X_s))$ , whereas Jaccard's degree is not associative. If one combines three (or more) BBA's and there is no conflicting mass, then ZCR is associative, whereas JCR is not associative. If there is conflicting masses, then ZCR is still associative, but JCR is not associative. Zhang's and Jaccard's degrees pose a problem because ZCR and JCR become strictly equivalent with DS rule when the cardinality is 1 for all relevant sets, or when  $|X_1 \cap X_2 \cap \dots \cap X_s| = |X_1| \cdot |X_2| \cdot \dots \cdot |X_s|$  in the circumstance of conflicting evidence. Therefore, it inherits the same limitations as DS rule – see example 2 in Section V.

The doubts of the validity of DS rule has been discussed by Zadeh in 1979 [10]–[12] based on a very simple example with two highly conflicting sources of evidences. Since 1980's, many criticisms have been done about the behavior and the justification of such DS rule. More recently, Dezert et al. in [3], [4], [18] have put in light other counter-intuitive behaviors of DS rule even in low conflicting cases and showed serious flaws in logical foundations of DST [5]. To overcome the limitations and problems of DS rule of combination, a new family of PCR rules have been developed in DSMT framework [1]. In PCR rules, we transfer the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is entirely preserved. The general principle of PCR consists: 1) to apply the conjunctive rule, 2) to calculate the total or partial conflicting masses; 3) then redistribute the (total or partial) conflicting mass proportionally on non-empty sets according to the integrity constraints one has for the frame  $\Theta$ . Because the proportional transfer can be done in different ways, there exist several versions of PCR rules of combination. PCR6 fusion rule has been proposed by Martin and Osswald in [1] Vol. 2, Chap. 2, as a serious alternative to PCR5 fusion rule proposed originally by Smarandache and Dezert in [1] Vol. 2, Chap. 1. When only two BBA's are combined, PCR6 and PCR5 fusion rules coincide, but they differ in general as soon as more than two sources have to be combined altogether. Recently, it has been proved in [6] that only PCR6 rule is consistent with the averaging fusion rule which allows to estimate the empirical (frequentist) probabilities involved in a discrete random experiment, and that is why we recommend to use it in applications when possible. For Shafer's model of FoD<sup>4</sup>, the PCR6<sup>5</sup> combination of two BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$  is defined by  $m_{1,2}^{PCR5/6}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Theta$

$$m_{1,2}^{PCR5/6}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right], \quad (7)$$

<sup>4</sup>that is when  $G^\Theta = 2^\Theta$ , and assuming all elements exhaustive and exclusive.

<sup>5</sup>which turns to be equal to PCR5 formula in case of fusion of two BBA's only.

where all denominators in (7) are different from zero. If a denominator is zero, that fraction is discarded. All propositions/sets are in a canonical form [1]. Basic MatLab codes of PCR rules can be found in [1], [13] or from the toolboxes repository on the web [14]. The general and concise formula of PCR6 rule for combining  $s > 2$  sources of evidences is

$$m_{1,2,\dots,s}^{PCR6}(X) = m_{1,2,\dots,s}(X) + CR^{PCR6}(X) \quad (8)$$

where  $m_{1,2,\dots,s}(X)$  corresponds to the conjunctive consensus on  $X$  between  $s$  sources of evidence, which is defined by

$$m_{1,2,\dots,s}(X) \triangleq \sum_{\substack{X_1, \dots, X_s \in 2^\Theta \\ X_1 \cap \dots \cap X_s = X}} \prod_{i=1}^s m_i(X_i), \quad (9)$$

and where  $CR^{PCR6}(X)$  is the part of the conflicting masses redistributed back to the focal element  $X$  according to PCR6 redistribution principle, that is

$$CR^{PCR6}(X) \triangleq \sum_{k=1}^{s-1} \sum_{\substack{X_{i_1}, X_{i_2}, \dots, X_{i_k} \in G^\Theta \setminus X \\ (\bigcap_{j=1}^k X_{i_j}) \cap X = \emptyset}} \sum_{(i_1, i_2, \dots, i_k) \in \mathcal{P}^s(\{1, \dots, s\})} [m_{i_1}(X) + m_{i_2}(X) + \dots + m_{i_k}(X)] \cdot \frac{m_{i_1}(X) \dots m_{i_k}(X) m_{i_{k+1}}(X_{i_{k+1}}) \dots m_{i_s}(X_{i_s})}{m_{i_1}(X) + \dots + m_{i_k}(X) + m_{i_{k+1}}(X_{i_{k+1}}) + \dots + m_{i_s}(X_{i_s})}. \quad (10)$$

In Eq.(10),  $\mathcal{P}^s(\{1, \dots, s\})$  is the set of all permutations of the elements  $\{1, 2, \dots, s\}$ . It should be observed that  $X_{i_1}, X_{i_2}, \dots, X_{i_s}$  may be different from each other, or some of them equal and others different, etc. As discussed and justified in [6], we focus here and in the sequel on PCR6 rule of combination rather than PCR5, but the general formula of PCR5 rule can be found in [1], [6] with examples, and a concise PCR5 general formula similar to (11) is possible. Like the averaging fusion rule, the PCR5 and PCR6 fusion rules are commutative but not associative.

### III. PCR6 RULE WITH DEGREES OF INTERSECTION

As presented in the previous section, the original versions of PCR5 or PCR6 rules of combination (as well as original DS rule) use only part of the whole information available (i.e. the values of the masses of belief only), because they do not exploit the cardinalities of focal elements entering in the fusion process. Because the cardinalities of focal elements are fully taken into account in the computation of the measure of degree of intersection between sets, we propose to improve PCR rules using this measure. The basic idea is to replace any conjunctive product by its discounted version thanks to the measure of degree of intersection  $D$  when the intersection of focal elements is not empty. The product of partial (or total) conflicting masses are not discounted by the measure of degree of intersection because the degree of intersection between two (or more) conflicting focal elements always equals zero, that is if  $X \cap Y = \emptyset$ , then  $D(X, Y) = 0$ . Because there are different ways to define degrees of intersection between set (here we consider only Zhang's and Jaccard' degrees), and

there are different ways to make the normalization because of the weighted conjunctive product involved in formulas, we come up with several versions of modified PCR6 rule of combination. We consider in fact two main modified versions of PCR6. The first modified version uses a classical normalization step based on the division by a normalization factor. The second modified version uses a sophisticate normalization step as shown through the general modified PCR6 formulas.

#### A. Simplest modified PCR6 rule

The simplest modified PCR6 rule including the measure of degree of intersection between sets is defined for  $s \geq 2$  BBA by  $m_{1,2,\dots,s}^{DPCR6}(\emptyset) = 0$  and for any non empty  $X \in 2^\Theta$ , by

$$m_{1,2,\dots,s}^{DPCR6}(X) \triangleq \frac{1}{K_{1,2,\dots,s}^{DPCR6}} \cdot [m_{1,2,\dots,s}^D(X) + CR^{PCR6}(X)], \quad (11)$$

where  $K_{1,2,\dots,s}^{DPCR6}$  is a normalization constant allowing to get  $\sum_{X \in 2^\Theta} m_{1,2,\dots,s}^{DPCR6}(X) = 1$ ;  $CR^{PCR6}(X)$  is the part of the conflicting masses redistributed back to the focal element  $X$  according to PCR6 redistribution principle and defined by (10); and  $m_{1,2,\dots,s}^D(X)$  is the discounted conjunctive consensus by the measure of the degree of intersection, defined by

$$m_{1,2,\dots,s}^D(X) \triangleq \sum_{\substack{X_1, \dots, X_s \in 2^\Theta \\ X_1 \cap \dots \cap X_s = X}} D(X_1, \dots, X_s) \prod_{i=1}^s m_i(X_i). \quad (12)$$

A similar general formula holds for the modified PCR5 rule with degrees of intersection between focal elements. For the fusion of two BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$ , the modified PCR6 and PCR5 formulas coincide and reduce to the formula below

$$m_{1,2}^{DPCR5/6}(X) = \frac{1}{K_{1,2}^{DPCR5/6}} \cdot \left[ \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} D(X_1, X_2) m_1(X_1) m_2(X_2) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right] \right]. \quad (13)$$

Depending on the degree of intersection we take (either  $D^Z$  or  $D^J$ ), we get two versions of this modified PCR6 rule. The result of the fusion for each version will be denoted  $m_{1,2,\dots,s}^{ZPCR6}(\cdot)$  and  $m_{1,2,\dots,s}^{JPCR6}(\cdot)$  in the sequel. ZPCR6 and JPCR6 rules<sup>6</sup> are commutative but not associative.

#### B. Sophisticate modified PCR6 rule

We propose here a more sophisticate modified PCR6 rule which does not use the normalization by the division with a normalization constant but which makes a proportional redistribution of the non conflicting mass missing from the discounted conjunctive rule (after including a degree of intersection). Before providing the general formula of this sophisticate modified PCR6 rule, let's explain how the redistribution that we propose is done in the two BBA's case at first for simplicity.

<sup>6</sup>ZPCR6 and JPCR6 denote the PCR6 rules modified with Zhang's and Jaccard's degrees of intersection respectively.

Let's suppose to have only two BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$  defined on the same FoD  $\Theta$  (assuming Shafer's model for simplicity). When  $X_1 \cap X_2 = X$ , then  $(1 - D(X_1, X_2))m_1(X_1)m_2(X_2)$  will be transferred back to  $X_1$  and  $X_2$  proportionally with respect to their masses (following PCR5/6 principle), that is:

$$\frac{\alpha}{m_1(X_1)} = \frac{\beta}{m_2(X_2)} = \frac{(1 - D(X_1, X_2))m_1(X_1)m_2(X_2)}{m_1(X_1) + m_2(X_2)},$$

whence,

$$\alpha = (1 - D(X_1, X_2)) \cdot \frac{m_1^2(X_1)m_2(X_2)}{m_1(X_1) + m_2(X_2)},$$

$$\beta = (1 - D(X_1, X_2)) \cdot \frac{m_1(X_1)m_2^2(X_2)}{m_1(X_1) + m_2(X_2)}.$$

The formula of this sophisticate modified combination rule, denoted<sup>7</sup> SDPCR5/6, is given by  $m_{1,2}^{SDPCR5/6}(\emptyset) = 0$  and by

$$m_{1,2}^{SDPCR5/6}(X) \triangleq \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} D(X_1, X_2)m_1(X_1)m_2(X_2)$$

$$+ \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right]$$

$$+ \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y \neq \emptyset}} (1 - D(X, Y)) \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} \right.$$

$$\left. + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right]. \quad (14)$$

The third sum of Eq.(14) represents the non-conflicting mass missing from the conjunctive rule including a degree of intersection. As for ZPCR6 or JPCR6 rules, we can choose Zhang's or Jaccard's degrees (or any other measures of degree of intersection if preferred). The generalization of this principle of redistribution of missing discounting conjunctive masses yields the following general sophisticate modified PCR6 rule of combination.

$$m_{1,2,\dots,s}^{SDPCR6}(X) = m_{1,2,\dots,s}^D(X) + CR^{PCR6}(X) + MR^{PCR6}(X), \quad (15)$$

where  $MR^{PCR6}(X)$  is the part of the missing conjunctive masses due to discounting back to the focal element involved in the conjunction which is redistributed according to PCR6 redistribution principle.  $MR^{PCR6}(X)$  is defined by

$$MR^{PCR6}(X) \triangleq \sum_{k=1}^{s-1} \sum_{\substack{X_{i_1}, X_{i_2}, \dots, X_{i_k} \in 2^\Theta \setminus X \\ (\cap_{j=1}^k X_{i_j}) \cap X \neq \emptyset}} \sum_{(i_1, i_2, \dots, i_k) \in \mathcal{P}^s(\{1, \dots, s\})} (1 - D(X, \dots, X, X_{i_{k+1}}, \dots, X_{i_s})) \cdot \sum_{j=1}^k m_{i_j}(X)$$

$$\cdot \frac{m_{i_1}(X) \dots m_{i_k}(X) m_{i_{k+1}}(X_{i_{k+1}}) \dots m_{i_s}(X_{i_s})}{m_{i_1}(X) + \dots + m_{i_k}(X) + m_{i_{k+1}}(X_{i_{k+1}}) + \dots + m_{i_s}(X_{i_s})}. \quad (16)$$

SZPCR6 and SJPCR6 rules<sup>8</sup> are commutative but not associative.

<sup>7</sup>S letter in this acronym stands for *Sophisticate*.

<sup>8</sup>SZPCR6 and SJPCR6 denote the PCR6 rules modified with Zhang's and Jaccard's degrees of intersection respectively.

#### IV. ANALYSIS OF THE NEUTRALITY OF VBBA

When there is no conflict between BBA's, DS, PCR5 or PCR6 rules reduce to the conjunctive rule which preserves the neutrality of VBA. When there is conflict between BBA's only DS preserves neutrality of VBA because DS is associative. In general, PCR5 and PCR6 do not preserve the neutrality of the VBA if more than two conflicting BBA's (including the VBA) are combined altogether<sup>9</sup>. In general, the VBA  $m_v(\cdot)$  is not a neutral element for the conjunctive rule of combination discounted with Jaccard's degree of intersection when combining two (or more) BBA's as shown in the following counterexample. If we take  $\Theta = \{A, B\}$ , with  $A \cap B = \emptyset$ , and  $m_1(\cdot)$  defined as  $m_1(A) = 0.5$ ,  $m_1(B) = 0.3$  and  $m_v(A \cup B) = 0.2$ . Then the result of the JCR fusion is  $m_{1,v}^{JCR}(A) \approx 0.4167$ ,  $m_{1,v}^{JCR}(B) = 0.25$  and  $m_{1,v}^{JCR}(A \cup B) \approx 0.3333$ , which shows that  $m_{1,v}^{JCR}(\cdot) \neq m_1(\cdot)$ . The VBA  $m_v(\cdot)$  is a neutral element for the ZCR combination of  $m_1(\cdot)$  with the VBA  $m_v(\cdot)$ , because the discounted conjunctive mass for any focal element  $X$  is  $m_{1,v}(X) = \frac{|X \cap \Theta|}{|X| \cdot |\Theta|} m_1(X) m_v(\Theta) = \frac{|X|}{|X| \cdot |\Theta|} m_1(X) \cdot 1 = \frac{1}{n} m_1(X)$ , where  $n = |\Theta|$ . The normalization constant equals  $K_{1,v}^{ZCR} = \sum_X \frac{1}{n} m_1(X) = 1/n$ . Therefore, after dividing by  $K_{1,v}^{ZCR}$ , we always gets  $m_{1,v}^{ZCR}(X) = m_1(X)$  for any focal element  $X$  of  $m_1(\cdot)$ . Same property holds if we combine three (or more) BBA's with the VBA and even if these BAA's are in conflict or not. Because  $D^Z(X_1, \dots, X_n, \Theta) = D^Z(X_1, \dots, X_n)/|\Theta|$  and  $m_v(\Theta) = 1$ , the constant  $|\Theta|$  always simplifies in normalization step of ZCR and because conjunctive rule and Zhang's degree are associative. In the general case, ZPCR6, SZPCR6, JPCR6 and SJPCR6 do not preserve the neutrality of the VBA. This can be verified using the simple example of the footnote no 9. More precisely, the combination  $[m_1 \oplus m_2 \oplus \dots \oplus m_n \oplus m_v](\cdot)$  is not equal to  $[m_1 \oplus m_2 \oplus \dots \oplus m_n](\cdot)$ . In the very specific case when there is no conflict between the BBA's, only ZPCR6 rule preserves the neutrality of VBA because it coincides with ZCR.

#### V. EXAMPLES

Here we analyze the behavior of the different rules (DS, PCR6, ZCR, JCR, ZPCR6, JPCR6, SZPCR6 and SJPCR6) in emblematic examples to determinate which one presents the best interest for the combination of BBA's.

##### Example 1: (No conflicting case)

Let's consider the FoD  $\Theta = \{A_1, A_2, \dots, A_{10}\}$  with Shafer's model, and the following two BBA's to combine  $m_1(A_1) = 0.9$ ,  $m_1(\Theta) = 0.1$ ,  $m_2(X) = 0.9$  and  $m_2(\Theta) = 0.1$  where the focal element  $X$  of  $m_2(\cdot)$  can take the values  $A_1, A_1 \cup A_2, A_1 \cup A_2 \cup A_3, \dots$ , or  $\Theta$ .

In this case, the DS and PCR5/6 rules coincide with the conjunctive rule of combination because there is no

<sup>9</sup>For example, if one considers  $\Theta = \{A, B\}$  with Shafer's model, and the BBA's  $\{m_1(A) = a_1, m_1(B) = b_1, m_1(\Theta) = c_1\}$ ,  $\{m_2(A) = a_2, m_2(B) = b_2, m_2(\Theta) = c_2, m_v(\Theta) = 1\}$ . Then  $[m_1 \oplus m_2](\cdot) \neq [m_1 \oplus m_2 \oplus m_v](\cdot)$  (where  $\oplus$  denotes the PCR5 or PCR6 fusion rule) because in  $m_1 \oplus m_2$  nothing from the redistribution of the conflicting mass goes to ignorance, contrarily to what happens in  $[m_1 \oplus m_2 \oplus m_v](\cdot)$ .

conflicting mass to redistribute because  $m_{1,2}(\emptyset) = 0$ . If  $X = A_1$ , then  $m_{1,2}^{DS}(A_1) = m_{1,2}^{PCR6}(A_1) = m_1(A_1)m_2(A_1) + m_1(A_1)m_2(\Theta) + m_1(\Theta)m_2(A_1) = 0.99$  and  $m_{1,2}^{DS}(\Theta) = m_{1,2}^{PCR6}(\Theta) = m_1(\Theta)m_2(\Theta) = 0.01$ , which is a reasonable result since the belief in  $A_1$  is reinforced because each source does strongly support the same hypothesis  $A_1$ . When  $X \supset A_1$  and  $|X| > 1$ , the behavior of the conjunctive rule becomes questionable because one always gets

$$\begin{aligned} m_{1,2}^{DS}(A_1) &= m_{1,2}^{PCR5/6}(A_1) = m_1(A_1)(m_2(X) + m_2(\Theta)) = 0.9, \\ m_{1,2}^{DS}(X) &= m_{1,2}^{PCR5/6}(X) = m_1(\Theta)m_2(X) = 0.09, \\ m_{1,2}^{DS}(\Theta) &= m_{1,2}^{PCR5/6}(\Theta) = m_1(\Theta)m_2(\Theta) = 0.01. \end{aligned}$$

When  $X \rightarrow \Theta$ ,  $m_2(\cdot)$  tends to become a fully ignorant source of evidence, and the combination of  $m_1(\cdot)$  with  $m_2(\cdot)$  tends towards  $m_1(\cdot)$  because  $m_2(\cdot)$  brings none useful information at all in this limit case. This behavior of conjunctive rule is then conform with what we intuitively expect. However, when  $|X|$  decreases from  $r = 10$  to  $r = 2$ , the behavior of conjunctive rule (and in this case DS and PCR6 rules also) is not very satisfactory, because we obtain same results on the mass of  $A_1$  whatever the cardinality of  $X$  is. In fact, it is rather intuitively expected that after the combination, the mass of  $A_1$  should substantially increase if the cardinality of  $X$  decreases because  $m_2(\cdot)$  becomes more and more specific (and focused towards  $A_1$ ). When  $m_2(\cdot)$  is more in agreement with  $m_1(\cdot)$ , the combination of  $m_1(\cdot)$  with  $m_2(\cdot)$  should reinforce the belief on  $A_1$  when  $|X|$  decreases, which is not what happens with the pure (strict) conjunctive rule.

Let's examine how ZCR, JCR rules work in this example. Let  $|X| = r \geq 1$ , and  $r \leq 10$ . Also  $|\Theta| = |A_1 \cup A_2 \cup \dots \cup A_{10}| = 10$ . If we compute the (unnormalized) discounted conjunctive fusion with Zhang's degree of intersection, we get

$$\begin{aligned} m_{1,2}^Z(A_1) &= \frac{|A_1 \cap X|}{|A_1| \cdot |X|} m_1(A_1)m_2(X) + \frac{|A_1 \cap \Theta|}{|A_1| \cdot |\Theta|} m_1(A_1)m_2(\Theta) \\ &= \frac{1}{r}(0.9)(0.9) + \frac{1}{10}(0.9)(0.1) = \frac{0.81}{r} + 0.009, \\ m_{1,2}^Z(X) &= \frac{|\Theta \cap X|}{|\Theta| \cdot |X|} m_1(\Theta)m_2(X) = \frac{1}{10}(0.1)(0.9) = 0.009, \\ m_{1,2}^Z(\Theta) &= \frac{|\Theta \cap \Theta|}{|\Theta| \cdot |\Theta|} m_1(\Theta)m_2(\Theta) = \frac{1}{10}(0.1)(0.1) = 0.001. \end{aligned}$$

If we compute the (unnormalized) discounted conjunctive fusion with Jaccard's degree of intersection, we get

$$\begin{aligned} m_{1,2}^J(A_1) &= \frac{|A_1 \cap X|}{|A_1 \cup X|} m_1(A_1)m_2(X) + \frac{|A_1 \cap \Theta|}{|A_1 \cup \Theta|} m_1(A_1)m_2(\Theta) \\ &= \frac{1}{r}(0.9)(0.9) + \frac{1}{10}(0.9)(0.1) = \frac{0.81}{r} + 0.009, \\ m_{1,2}^J(X) &= \frac{|\Theta \cap X|}{|\Theta \cup X|} m_1(\Theta)m_2(X) = \frac{r}{10}(0.1)(0.9) = 0.009 \cdot r, \\ m_{1,2}^J(\Theta) &= \frac{|\Theta \cap \Theta|}{|\Theta \cup \Theta|} m_1(\Theta)m_2(\Theta) = \frac{10}{10}(0.1)(0.1) = 0.01. \end{aligned}$$

After normalization of  $m_{1,2}^Z(\cdot)$  by  $K_{1,2}^Z = \frac{0.81}{r} + 0.019$ , and  $m_{1,2}^J(\cdot)$  by  $K_{1,2}^J = \frac{0.81}{r} + 0.009 \cdot r + 0.010$  we get the result

of ZCR and JCR rules, which are

$$\begin{aligned} m_{1,2}^{ZCR}(A_1) &= [\frac{0.81}{r} + 0.009]/K_{1,2}^Z, & m_{1,2}^{JCR}(A_1) &= [\frac{0.81}{r} + 0.009]/K_{1,2}^J, \\ m_{1,2}^{ZCR}(X) &= 0.009/K_{1,2}^Z, & m_{1,2}^{JCR}(X) &= 0.009 \cdot r/K_{1,2}^J, \\ m_{1,2}^{ZCR}(\Theta) &= 0.001/K_{1,2}^Z, & m_{1,2}^{JCR}(\Theta) &= 0.01/K_{1,2}^J. \end{aligned}$$

In the limit case when  $r = 1$  we get

$$\begin{aligned} m_{1,2}^{ZCR}(A_1) &= 0.988, & m_{1,2}^{JCR}(A_1) &= 0.988, \\ m_{1,2}^{ZCR}(\Theta) &= 0.012, & m_{1,2}^{JCR}(\Theta) &= 0.012. \end{aligned}$$

In the limit case when  $r = 10$  we get

$$\begin{aligned} m_{1,2}^{ZCR}(A_1) &= 0.90, & m_{1,2}^{JCR}(A_1) &= 0.4337, \\ m_{1,2}^{ZCR}(\Theta) &= 0.10, & m_{1,2}^{JCR}(\Theta) &= 0.5263. \end{aligned}$$

Clearly, one sees that both ZCR and JCR have now a good expected behavior when  $|X|$  decreases, but only ZCR provides also a good behavior when  $r = 10$  because in this case one gets  $m_{1,2}^{ZCR}(\cdot) = m_1(\cdot)$  which is normal because  $m_2(\cdot)$  is the VBA (fully ignorant source). With JCR, the result we obtain when  $|X| = r = 10$  is not good because  $m_{1,2}^{ZCR}(\cdot) \neq m_1(\cdot)$ . Because there is no conflict, ZPCR6 rule coincides with ZCR rule in this example, and JPCR6 rule coincides with JCR rule. Therefore, JPCR6 rule does not work well (at least for this example) as explained previously. The evaluation of masses of  $A_1$  and of  $\Theta$  after the combination of  $m_1(\cdot)$  with  $m_2(\cdot)$  for the different rules is shown in Fig. 1 and Fig. 2 respectively and for different values of  $r = |X|$ .

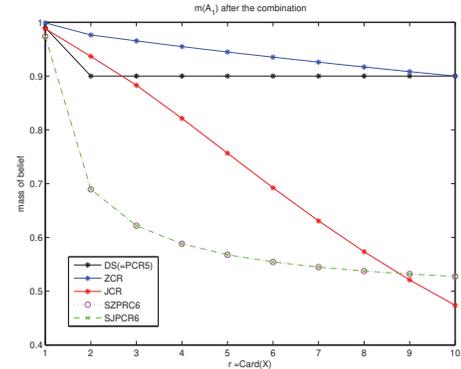


Figure 1.  $m(A_1)$  after combination of  $m_1(\cdot)$  with  $m_2(\cdot)$ .

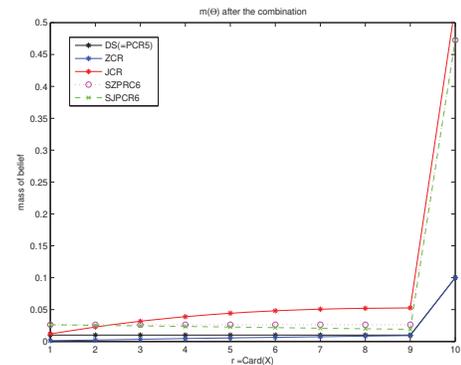


Figure 2.  $m(\Theta)$  after combination of  $m_1(\cdot)$  with  $m_2(\cdot)$ .

If we apply sophisticate normalization procedures we obtain<sup>10</sup> with SZPCR6 and SJPCR6

$$\begin{aligned} m_{1,2}^{SZPCR6}(A_1) &= 0.0819 + 0.81 \cdot \frac{1}{r} + 0.405 \cdot \frac{r-1}{r}, \\ m_{1,2}^{SZPCR6}(X) &= 0.0819 + 0.405 \cdot \frac{r-1}{r}, \\ m_{1,2}^{SZPCR6}(\Theta) &= 0.0262, \end{aligned}$$

$$\begin{aligned} m_{1,2}^{SJPCR6}(A_1) &= 0.0819 + 0.81 \cdot \frac{1}{r} + 0.405 \cdot \frac{r-1}{r}, \\ m_{1,2}^{SJPCR6}(X) &= 0.009 \cdot r + 0.405 \cdot \frac{r-1}{r} + (10-r) \cdot 0.0081, \\ m_{1,2}^{SJPCR6}(\Theta) &= 0.0181 + (10-r) \cdot 0.0081. \end{aligned}$$

In the limit case, when  $r = 1$  we get

$$\begin{aligned} m_{1,2}^{SZPCR6}(A_1) &= 0.9738, & m_{1,2}^{SJPCR6}(A_1) &= 0.9738, \\ m_{1,2}^{SZPCR6}(\Theta) &= 0.0262, & m_{1,2}^{SJPCR6}(\Theta) &= 0.0262. \end{aligned}$$

In the limit case, when  $r = 10$  we get

$$\begin{aligned} m_{1,2}^{SZPCR6}(A_1) &= 0.5274, & m_{1,2}^{SJPCR6}(A_1) &= 0.5274, \\ m_{1,2}^{SZPCR6}(\Theta) &= 0.4726, & m_{1,2}^{SJPCR6}(\Theta) &= 0.4726. \end{aligned}$$

This result shows clearly that SZPCR6 and SJPCR6 rules behave better than conjunctive rule (and so better than DS and PCR6 rules) in the limit case when  $X = A_1$  because after the combination the mass committed to  $A_1$  is reinforced (as it is naturally expected). But the reinforcement of mass of  $A_1$  is lower than with ZPCR6 or JPCR6 rules<sup>11</sup> based on simple normalization because the sophisticate normalization procedure degrades the specificity of the information. In the other limit case when  $r = 10$ , (i.e.  $X = \Theta$ , and  $m_2(\cdot)$  equals the VBA) SZPCR6 and SJPCR6 rules do not work well because clearly one has  $m_{1,2}^{SZPCR6}(\cdot) \neq m_1(\cdot)$  and  $m_{1,2}^{SJPCR6}(\cdot) \neq m_1(\cdot)$  also. So we at least have shown one example where SZPCR6 and SJPCR6 are not very efficient and consequently, we do not recommend to use them. In summary, only ZCR and ZPCR6 (equivalent to ZCR in this example) allow to get an acceptable behavior for combining the two BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$  for any focal element  $X \supseteq A_1$ .

### Example 2 (Zadeh [10], [12]): (Conflicting case)

Let's  $\Theta = \{A, B, C\}$  with Shafer's model, and the two BBA's to combine  $m_1(A) = 0.9$ ,  $m_1(C) = 0.1$ ,  $m_2(B) = 0.9$  and  $m_2(C) = 0.1$ .

In this case, Shafer's conflict is  $m_{1,2}(\emptyset) = m_1(A)(m_2(B) + m_2(C)) + m_1(C)m_2(B) = 0.9 + 0.1 \cdot 0.9 = 0.99$ . If we use DS rule (2), we get  $m_{1,2}^{DS}(C) = 1$ . The discounted conjunctive consensus  $D(C, C)m_1(C)m_2(C)$  (with Zhang's or Jaccard's degree) is always equal to the un-discounted conjunctive consensus  $m_1(C)m_2(C) = 0.01$  because  $D^Z(C, C) = \frac{|C \cap C|}{|C| \cdot |C|} = 1$  and  $D^J(C, C) = \frac{|C \cap C|}{|C \cup C|} = 1$ . Therefore the degree of intersection does not impact the conjunctive combination result

<sup>10</sup>Here there is no conflicting mass to redistribute which makes the derivation more easier.

<sup>11</sup>which coincide here with ZCR and JCR rule because there is no conflicting mass to redistribute.

and ZCR and JCR rules (4) give same counter-intuitive result as DS rule, that is  $m_{1,2}^{ZCR}(C) = m_{1,2}^{JCR}(C) = m_{1,2}^{DS}(C) = 1$ .

Because the degree of intersection does not impact the conjunctive combination part of PCR6 rule in this example, modified PCR6 rules (ZPCR6, JPCR6, SZPCR6 and SJPCR6) give the same result as PCR6 rule which is  $m_{1,2}^{PCR5/6}(A) = 0.486$ ,  $m_{1,2}^{PCR5/6}(B) = 0.486$  and  $m_{1,2}^{PCR5/6}(C) = 0.028$ .

In summary, ZCR and JCR rules do not help to modify the result obtained by DS rule in Zadeh's example and cannot be viewed as real alternatives to DS rule for this example. Conversely, ZPCR6, JPCR6, SZPCR6 and SJPCR6 rule (which coincide with PCR6 rule in this example) remain good alternatives to DS rule.

### Example 3 (Voorbraak [15]): (Conflicting case)

Let's consider the FoD  $\Theta = \{A, B, C\}$  with Shafer's model, and the following two BBA's to combine  $m_1(A) = 0.5$ ,  $m_1(B \cup C) = 0.5$ ,  $m_2(C) = 0.5$ , and  $m_2(A \cup B) = 0.5$ .

One has  $m_{1,2}(\emptyset) = m_1(A)m_2(C) = 0.25$ , and DS rule gives  $m_{1,2}^{DS}(A) = m_{1,2}^{DS}(B) = m_{1,2}^{DS}(C) = 1/3$ . As reported by Voorbraak [15], this result is counterintuitive, since intuitively  $B$  seems to *share* twice a probability mass of 0.5, while both  $A$  and  $C$  only have to share once 0.5 with  $B$  and are once assigned 0.5 individually. This counterintuitive result comes from the fact that DS rule implicitly assumes that all possible pairs of focal elements are equally confirmed by the combined evidence, while intuitively, in this example  $B = (B \cup C) \cap (A \cup B)$  is less confirmed than  $A = A \cap (A \cup B)$  and  $C = (B \cup C) \cap C$ . With ZCR and JCR rules, we get

$$\begin{aligned} m_{1,2}^{ZCR}(A) &= 0.40, & m_{1,2}^{JCR}(A) &= 0.375, \\ m_{1,2}^{ZCR}(B) &= 0.20, & m_{1,2}^{JCR}(B) &= 0.250, \\ m_{1,2}^{ZCR}(C) &= 0.40, & m_{1,2}^{JCR}(C) &= 0.375. \end{aligned}$$

Contrarily to DS rule, with ZCR or JCR rules one sees that the mass committed to  $B$  is less than of  $A$  and of  $C$  which is a more reasonable result. In applying PCR6 rule, we also circumvent this problem because we get from Eq. (13),  $m_{1,2}^{PCR6}(A) = 0.375$ ,  $m_{1,2}^{PCR6}(B) = 0.25$  and  $m_{1,2}^{PCR6}(C) = 0.375$  (same as with JCR results for this particular example).

With ZPCR6 rule, we compute at first the following (un-normalized) discounted conjunctive masses added with proportional conflict redistribution

$$\begin{aligned} m_{1,2}^Z(A) &= \frac{|A \cap (A \cup B)|}{|A| \cdot |A \cup B|} m_1(A)m_2(A \cup B) + \frac{1}{2} m_{1,2}(\emptyset) = 0.25, \\ m_{1,2}^Z(B) &= \frac{|(B \cup C) \cap (A \cup B)|}{|B \cup C| \cdot |A \cup B|} m_1(B \cup C)m_2(A \cup B) = 0.0625, \\ m_{1,2}^Z(C) &= \frac{|(B \cup C) \cap C|}{|B \cup C| \cdot |C|} m_1(B \cup C)m_2(C) + \frac{1}{2} m_{1,2}(\emptyset) = 0.25. \end{aligned}$$

After a simple normalization (dividing by  $K_{1,2}^Z = 0.25 + 0.0625 + 0.25 = 0.5625$ ), we get finally

$$\begin{aligned} m_{1,2}^{ZPCR6}(A) &= 0.25/0.5625 \approx 0.4444, \\ m_{1,2}^{ZPCR6}(B) &= 0.0625/0.5625 \approx 0.1112, \\ m_{1,2}^{ZPCR6}(C) &= 0.25/0.5625 \approx 0.4444. \end{aligned}$$

Similarly, if we apply JPCR6 rule based on Jaccard's index and simple normalization step, we will get the following result

$$\begin{aligned} m_{1,2}^{JPCR6}(A) &= [(0.25/2) + 0.125]/K_{1,2}^J \approx 0.4286, \\ m_{1,2}^{JPCR6}(B) &= (0.25/3)/K_{1,2}^J = 0.1428, \\ m_{1,2}^{JPCR6}(C) &= [(0.25/2) + 0.125]/K_{1,2}^J \approx 0.4286, \end{aligned}$$

where the normalization factor equals  $K_{1,2}^J = (0.25/2) + 0.125 + (0.25/3) + (0.25/2) + 0.125 \approx 0.5833$ .

These results show that ZPCR6 and JPCR6 rules diminish substantially the mass committed to  $B$  (as expected) and reinforce more strongly the masses of  $A$  and  $C$  than with ZCR, JCR or PCR6 rules.

If we apply the sophisticate normalization for SZPCR6, the lost discounted mass  $(1 - \frac{|A \cap (A \cup B)|}{|A| \cdot |A \cup B|})m_1(A)m_2(A \cup B) = 0.125$  is redistributed to  $A$  and to  $A \cup B$  proportionally<sup>12</sup> to  $m_1(A) = 0.5$  and  $m_2(A \cup B) = 0.5$ . Similarly, the second lost discounted mass  $(1 - \frac{|B \cup C \cap (A \cup B)|}{|B \cup C| \cdot |A \cup B|})m_1(B \cup C)m_2(A \cup B) = 0.1875$  is redistributed to  $B \cup C$  and to  $A \cup B$  proportionally to  $m_1(B \cup C) = 0.5$  and  $m_2(A \cup B) = 0.5$ , and the third lost discounted mass  $(1 - \frac{|B \cup C \cap C|}{|B \cup C| \cdot |C|})m_1(B \cup C)m_2(C) = 0.125$  is redistributed to  $B \cup C$  and to  $C$  proportionally to  $m_1(B \cup C) = 0.5$  and  $m_2(C) = 0.5$ . Similar computations are done for SJPCR6 in replacing Zhang's degree by Jaccard's degree of intersection. Finally we obtain with SZPCR6 and SJPCR6 the following combined masses:

$$\begin{aligned} m_{1,2}^{SZPCR6}(A) &= 0.3125, & m_{1,2}^{SZPCR6}(A \cup B) &= 0.15625, \\ m_{1,2}^{SZPCR6}(B) &= 0.0625, & m_{1,2}^{SZPCR6}(B \cup C) &= 0.15625, \\ m_{1,2}^{SZPCR6}(C) &= 0.3125, \end{aligned}$$

and

$$\begin{aligned} m_{1,2}^{SJPCR6}(A) &= 0.3125, & m_{1,2}^{SJPCR6}(A \cup B) &\approx 0.14585, \\ m_{1,2}^{SJPCR6}(B) &\approx 0.0833, & m_{1,2}^{SJPCR6}(B \cup C) &\approx 0.14585, \\ m_{1,2}^{SJPCR6}(C) &= 0.3125. \end{aligned}$$

Of course, these results are a bit less specific than with ZPCR6 and JPCR6, which is normal. As shown, SZPCR6 and SJPCR6 rules diminish also the mass committed to  $B$  (as expected) but reinforce less strongly the masses of  $A$  and  $C$  because the specificity of the result is degraded because one gets positive masses committed to new uncertainties  $A \cup B$  and  $B \cup C$ . For this example, ZPCR6 and JPCR6 are the most interesting rules for combining BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$ .

<sup>12</sup>equally in fact in this case.

#### Example 4 (Dezert et al. [3]): (Conflicting case)

This emblematic example is very interesting to analyze because for in this case the DS rule does not respond to level of conflict between the sources. This *anomaly* has been analyzed and discussed in details in [3].

Let's consider the FoD  $\Theta = \{A, B, C\}$  with Shafer's model, and the following two BBA's to combine

$$\begin{aligned} m_1(A) &= 0.9, & m_1(A \cup B) &= 0.1, & m_2(A \cup B) &= 0.1, \\ m_2(C) &= 0.7, & m_2(A \cup B \cup C) &= 0.2. \end{aligned}$$

In this example, the two sources are not vacuous (they are truly informative), they are in conflict because  $m_{1,2}(\emptyset) = 0.7$  but DS rule does not respond to the level of conflict because one gets  $m_{1,2}(\cdot) = m_1(\cdot)$ . In fact, the second source has no impact in the DS fusion as if it is equivalent to the VBA.

If we apply PCR6 rule of combination the first partial conflict  $m_1(A)m_2(C) = 0.72$  is redistributed to  $A$  and  $C$  proportionally to  $m_1(A)$  and  $m_2(C)$ , and the second conflict  $m_1(A \cup B)m_2(C) = 0.08$  is redistributed to  $A \cup B$  and to  $C$  proportionally to  $m_1(A \cup B)$  and  $m_2(C)$ . So with PCR6 rule (7), we obtain  $m_{1,2}^{PCR6}(A) = 0.6244$ ,  $m_{1,2}^{PCR6}(A \cup B) = 0.0388$  and  $m_{1,2}^{PCR6}(C) = 0.3369$ . One sees that the PCR6 fusion result now reacts with the value of second sources because  $m_{1,2}^{PCR6}(\cdot) \neq m_1(\cdot)$  which makes sense if both sources are equireliable, truly informative and in some disagreement.

In discounting with Zhang's degree, one gets the (unnormalized) discounted conjunctive BBA

$$\begin{aligned} m_{1,2}^Z(A) &= \frac{1}{2}(0.9)(0.1) + \frac{1}{3}(0.9)(0.2) = 0.1050, \\ m_{1,2}^Z(A \cup B) &= \frac{2}{2 \cdot 2}(0.1)(0.1) + \frac{2}{2 \cdot 3}(0.1)(0.2) \approx 0.0117. \end{aligned}$$

After the normalization by the factor  $K_{1,2}^Z = 0.1050 + 0.0117 = 0.1167$ , we get finally  $m_{1,2}^{ZCR}(A) = 0.1050/0.1167 \approx 0.9$  and  $m_{1,2}^{ZCR}(A \cup B) \approx 0.0117/0.1167 \approx 0.1$ . Therefore as with DS rule, we get same behavior with ZCR rule that is  $m_{1,2}^{ZCR}(\cdot) = m_1(\cdot)$  as if the second informative source does not count in the fusion process, which is abnormal.

If we use Jaccard's degree, one gets

$$\begin{aligned} m_{1,2}^J(A) &= \frac{1}{2}(0.9)(0.1) + \frac{1}{3}(0.9)(0.2) = 0.1050, \\ m_{1,2}^J(A \cup B) &= \frac{2}{2}(0.1)(0.1) + \frac{2}{3}(0.1)(0.2) \approx 0.0233. \end{aligned}$$

After the normalization by the factor  $K_{1,2}^J = 0.1050 + 0.0233 = 0.12833$ , we get finally  $m_{1,2}^{JCR}(A) \approx 0.8182$  and  $m_{1,2}^{JCR}(A \cup B) \approx 0.1818$ . One sees that JCR fusion result is not equal to the BBA  $m_1(\cdot)$ , which means that  $m_2(\cdot)$  has had some impact in the fusion process with JCR (as expected). However, it is not clear why such JCR result will really make sense or not. Because we have already shown in Example 1, that it can happen than JCR does not work well, we have serious

doubt on the interest of using JCR result in such emblematic example.

With ZPCR6 rule of combination, we obtain  $m_{1,2}^{ZPCR6}(A) = \frac{1}{K_{1,2}^{ZPCR6}}[0.1050 + x(A)] = 0.56250$ ,  $m_{1,2}^{ZPCR6}(A \cup B) = \frac{1}{K_{1,2}^{ZPCR6}}[0.0117 + x(A \cup B)] = 0.0250$ , and  $m_{1,2}^{ZPCR6}(C) = \frac{1}{K_{1,2}^{ZPCR6}}[x_1(C) + x_2(C)] = 0.4125$ , where  $K_{1,2}^{ZPCR6}$  is the normalization constant, and where  $x(A) = m_1(A) \frac{m_1(A)m_2(C)}{m_1(A)+m_2(C)} = 0.354375$  is the part of the conflicting mass  $m_1(A)m_2(C) = 0.63$  transferred to  $A$ ;  $x_1(C) = m_1(C) \frac{m_1(A)m_2(C)}{m_1(A)+m_2(C)} = 0.275625$  is the part of the conflicting mass  $m_1(A)m_2(C) = 0.63$  transferred to  $C$ ;  $x(A \cup B) = m_1(A \cup B) \frac{m_1(A \cup B)m_2(C)}{m_1(A \cup B)+m_2(C)} = 0.00875$  is the part of the conflicting mass  $m_1(A \cup B)m_2(C) = 0.07$  transferred to  $A \cup B$ ; and  $x_2(C) = m_1(C) \frac{m_1(A \cup B)m_2(C)}{m_1(A \cup B)+m_2(C)} = 0.06125$  is the part of the conflicting mass  $m_1(A \cup B)m_2(C) = 0.07$  transferred to  $C$ .

With JPCR6 rule of combination, we obtain  $m_{1,2}^{JPCR6}(A) \approx 0.55458$ ,  $m_{1,2}^{JPCR6}(A \cup B) = 0.03873$  and  $m_{1,2}^{JPCR6}(C) = 0.40669$ , which is close to ZPCR6 result. Comparatively to PCR6, we diminish the mass of belief committed to  $A$  and to  $A \cup B$  and we reinforce the mass committed to  $C$  using ZPCR6 and JPCR6 rules. We do not give results with SZPCR6 and SJPCR6 due to space constraint and because we know that these rules do not perform so well as shown in the previous examples.

**Example 5 (Sebbak [16]):** (Conflicting case with 3 sources)

Let's consider the FoD  $\Theta = \{A, B, C\}$  with Shafer's model, and the following three BBA's to combine

$$\begin{aligned} m_1(A) &= 0.8, & m_1(A \cup B \cup C) &= 0.2, \\ m_2(A) &= 0.1, & m_2(C) &= 0.9, \\ m_3(A) &= 0.4, & m_3(A \cup B \cup C) &= 0.6. \end{aligned}$$

The conjunctive rule gives

$$\begin{aligned} m_{1,2,3}(A) &= m_1(A)m_2(A)m_3(A) + m_1(A)m_2(A)m_3(\Theta) \\ &\quad + m_1(\Theta)m_2(A)m_3(\Theta) + m_1(\Theta)m_2(A)m_3(A) = 0.10 \\ m_{1,2,3}(C) &= m_1(\Theta)m_2(C)m_3(\Theta) = 0.108 \end{aligned}$$

with the total conflicting mass

$$\begin{aligned} m_{1,2,3}(\emptyset) &= m_1(A)m_2(C)m_3(A) + m_1(A)m_2(C)m_3(\Theta) \\ &\quad + m_1(\Theta)m_2(C)m_3(A) = 0.792. \end{aligned}$$

With DS rule we get  $m_{12}^{DS}(A) \approx 0.4808$  and  $m_{12}^{DS}(C) \approx 0.5192$ , and With PCR5 and PCR6 rules [17]

$$\begin{aligned} m_{1,2,3}^{PCR5}(A) &= 0.3450, & m_{1,2,3}^{PCR6}(A) &= 0.4340, \\ m_{1,2,3}^{PCR5}(C) &= 0.5327, & m_{1,2,3}^{PCR6}(C) &= 0.4437, \\ m_{1,2,3}^{PCR5}(\Theta) &= 0.1223, & m_{1,2,3}^{PCR6}(\Theta) &= 0.1223. \end{aligned}$$

Note that with PCR5 one gets  $0.4247/0.7920 \approx 53.62\%$  of the total conflicting mass redistributed to  $C$ , but not *almost all conflicting mass*. Using PCR6,  $C$  actually gained from the total conflicting mass only  $0.3357/0.7920 \approx 42.3864\%$ , not even

half of it, not *almost all of the conflicting mass (the majority)* as the authors wrongly claimed in [16].

With ZCR and JCR rules, one gets

$$\begin{aligned} m_{1,2,3}^{ZCR}(A) &= 0.8125, & m_{1,2,3}^{JCR}(A) &= 0.6032, \\ m_{1,2,3}^{ZCR}(\Theta) &= 0.1875, & m_{1,2,3}^{JCR}(\Theta) &= 0.3968. \end{aligned}$$

With ZPCR6, JPCR6, SZPCR6 and SJPCR6 rules<sup>13</sup> one gets

$$\begin{aligned} m_{1,2,3}^{ZPCR6}(A) &= 0.4511, & m_{1,2,3}^{JPCR6}(A) &= 0.4405, \\ m_{1,2,3}^{ZPCR6}(C) &= 0.4061, & m_{1,2,3}^{JPCR6}(C) &= 0.4210, \\ m_{1,2,3}^{ZPCR6}(\Theta) &= 0.1428, & m_{1,2,3}^{JPCR6}(\Theta) &= 0.1385. \\ m_{1,2,3}^{SZPCR6}(A) &= 0.4102955, & m_{1,2,3}^{SJPCR6}(A) &= 0.412699, \\ m_{1,2,3}^{SZPCR6}(C) &= 0.3984240, & m_{1,2,3}^{SJPCR6}(C) &= 0.409718, \\ m_{1,2,3}^{SZPCR6}(\Theta) &= 0.1912805, & m_{1,2,3}^{SJPCR6}(\Theta) &= 0.177616. \end{aligned}$$

One sees that  $C$  gained  $(0.4061 - 0.108)/0.7920 \approx 37.64\%$  using ZPCR6,  $(0.4210 - 0.108)/0.7920 \approx 39.52\%$  using JPCR6,  $(0.398424 - 0.108)/0.7920 \approx 36.67\%$  using SZPCR6, and  $(0.409718 - 0.108)/0.7920 \approx 38.10\%$  using SJPCR6.

## VI. CONCLUSIONS

The modifications of the PCR6 rule of combination presented exploit judiciously Zhang's and Jaccard's degrees of intersections of focal elements. Our analysis shows that ZPCR6 rule is in fact the most interesting modified PCR6 rule because it behaves well in all emblematic examples contrarily to other rules. SZPCR6 and SJPCR6 rules are more complicate to implement and they increase the non-specificity of the result in general which is not good for helping the decision-making. So we do not recommend them for applications. All these rules are not associative and do not preserve the neutrality of VBA when some sources are in conflict.

## REFERENCES

- [1] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSmt for information fusion*, American Research Press, Rehoboth, NM, U.S.A., Vol. 1-4, 2004-2015. <http://fs.gallup.unm.edu/DSmt.htm>
- [2] G. Shafer, *A Mathematical Theory of Evidence*. Princeton: Princeton University Press, 1976.
- [3] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Proc. of Fusion 2012 Int. Conf., Singapore, July 9-12, 2012.
- [4] A. Tchamova, J. Dezert, *On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory*, (Best paper awards), 6th IEEE Int. Conf. on Int. Syst., Sofia, Bulgaria, Sept. 6-8, 2012.
- [5] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intelligent Syst., Vol. 29, Issue 3, pages 223-252, March 2014.
- [6] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, Proc. of Fusion 2013, Istanbul, Turkey, July 2013.
- [7] L. Zhang, *Representation, independence and combination of evidence in Dempster-Shafer theory*, in *Advances in the Dempster-Shafer theory of evidence*, John Wiley & Sons, New York, pp. 51-95, 1994.
- [8] D. Fixsen, R. Mahler, *The modified Dempster-Shafer approach to classification*, IEEE Trans. on SMC - Part A: Systems and Humans, Vol. 27, No. 1, pp. 96-104, Jan. 1997.

<sup>13</sup>The derivations are not included in this paper due to space restriction.

- [9] P. Jaccard, *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*, Bulletin de la Société Vaudoise des Sciences Naturelles, Vol. 37, pp. 547–579, 1901.
- [10] L.A. Zadeh, *On the validity of Dempster's rule of combination*, Memo M79/24, Univ. of California, Berkeley, CA, U.S.A., 1979.
- [11] L.A. Zadeh, *Book review: A mathematical theory of evidence*, The AI Magazine, Vol. 5, No. 3, pp. 81-83, 1984.
- [12] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, Vol. 7, No. 2, 1986.
- [13] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, Proc. of Fusion 2010 Int. Conf., Edinburgh, UK, July 26–29, 2010.
- [14] <http://bfas.iutlan.univ-rennes1.fr/wiki/index.php/Toolboxes>
- [15] F. Voorbraak, *On the justification of Dempster rule of combination*, Artificial Intelligence, Vol. 48, No. 2, pp. 171–197, March 1991.
- [16] F. Sebbak et al., *An Alternative Combination Rule for Evidential Reasoning*, Proc. of Fusion 2014, Salamanca, Spain, 7-10 July 2014.
- [17] F. Smarandache, J. Dezert, A. Martin A., *Comments on the paper [16]*, *Bulletin of Pure and Applied Sciences*, Volume 33 E (Math & Stat.) Issue (No.2) 2014, pp. 91–94.
- [18] F. Smarandache, V. Kroumov, J. Dezert, *Examples where the conjunctive and Dempster's rules are insensitive*, Proc. of 2013 Int. Conf. on Advanced Mechatronic Systems, Luoyang, China, Sept. 25–27, 2013.



# A Hierarchical Flexible Coarsening Method to Combine BBAs in Probabilities

Yilin Dong<sup>a</sup>, Xinde Li<sup>b</sup>, Jean Dezert<sup>c</sup>

<sup>a</sup>College of Information Engineering, Shanghai Maritime University, Shanghai, China

<sup>b</sup>Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing, China

<sup>c</sup>The French Aerospace Lab, ONERA, Palaiseau, France

Emails: yldong@shmtu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr

Originally published as: Y. Dong, X. Li, J. Dezert, *A Hierarchical Flexible Coarsening Method to Combine BBAs in Probabilities*, in Proc. of 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10-13, 2017, and reprinted with permission.

**Abstract**—In many applications involving epistemic uncertainties usually modeled by belief functions, it is often necessary to approximate general (non-Bayesian) basic belief assignments (BBAs) to subjective probabilities (called Bayesian BBAs). This necessity occurs if one needs to embed the fusion result in a system based on the probabilistic framework and Bayesian inference (e.g. tracking systems), or if one wants to use classical decision theory to make a decision. There exists already several methods (probabilistic transforms) to approximate any general BBA to a Bayesian BBA. From a fusion standpoint, two approaches are usually adopted: 1) one can approximate at first each BBA in subjective probabilities and use Bayes fusion rule to get the final Bayesian BBA, or 2) one can fuse all the BBAs with a fusion rule, typically Dempster-Shafer's, or PCR6 rules (which is very costly in computations), and convert the combined BBA in a subjective probability measure. The former method is the simplest method but it generates a high loss of information included in original BBAs, whereas the latter is intractable for high dimension problems. This paper presents a new method to achieve this task based on hierarchical decomposition (coarsening) of the frame of discernment, which can be seen as an intermediary approach between the two aforementioned methods. After the presentation of this new method, we show through simulations how it performs with respect to other methods.

**Keywords:** Information fusion, belief functions, DST, DSMT, PCR6 rule, coarsening.

## I. INTRODUCTION

The theory of belief functions, known as Dempster-Shafer Theory (DST) has been developed by Shafer [1] in 1976 from Dempster's works [2]. Belief functions allow to model epistemic uncertainty and they have been already used in many applications since the 1990's [3], mainly those related to expert systems, decision-making support and information fusion. To palliate some limitations of DST, Dezert and Smarandache have proposed an extended mathematical framework of belief functions with new efficient quantitative and qualitative rules of combinations, which is called DSMT (Dezert and Smarandache Theory) in the literature [4], [5] with applications listed in [6]. One of the major drawbacks of DST and DSMT is their high computational complexities, as soon as the fusion space (i.e. frame of discernment - FoD) and the number of sources to combine are large<sup>1</sup>.

<sup>1</sup>DSMT is more complex than DST, and the Proportional Conflict Redistribution rule #6 (PCR6 rule) becomes computationally intractable in the worst case as soon as the frame of discernment has at least six elements.

To reduce the computational cost of operations with belief functions when the number of focal elements is very large, several approaches have been proposed by different authors. Basically, the existing approaches rely either on efficient implementations of computations as proposed for instance in [7], [8], or on approximation techniques of original Basic Belief Assignment (BBA) to combine [9]–[12], or both. In many applications involving epistemic uncertainties usually modeled by belief functions, it is often necessary to approximate general (non-Bayesian) basic belief assignments (BBAs) to subjective probabilities (called Bayesian BBAs). This necessity occurs if one needs to embed the fusion result in a system based on the probabilistic framework and Bayesian inference (e.g. tracking systems), or if one wants to use classical decision theory to make a decision. From a fusion standpoint, two approaches are usually adopted: 1) one can approximate at first each BBA in subjective probabilities and use Bayes fusion rule to get the final Bayesian BBA, or 2) one can fuse all the BBAs with a fusion rule, typically Dempster-Shafer's, or PCR6 rules (which is very costly in computations), and convert the combined BBA in a subjective probability measure. The former method is the simplest method but it generates a high loss of information included in original BBAs, whereas the latter direct method is intractable for high dimension problems. This paper presents a new method to achieve this task based on hierarchical decomposition (coarsening) of the frame of discernment, which can be seen as an intermediary approach between the two aforementioned methods.

This paper presents a new approach to fuse BBAs into a Bayesian BBA in order to reduce computational burden and keep the fusion tractable even for large dimension problems. This method is based on a hierarchical decomposition (coarsening) framework which allows to keep as much as possible information of original BBAs in preserving lower complexity. The main contributions of this paper are:

- 1) the presentation of the FoD bintree decomposition on which will be done the BBAs approximations;
- 2) the presentation of the fusion of approximate BBAs from bintree representation.

This hierarchical structure allows to encompass bintree decomposition and BBAs approximations on it to obtain the

final approximate fused Bayesian BBA.

This paper is organized as follows. In section II, we recall some basics of DST and DSMT that are relevant to the new method presented in this paper. More details with examples can easily be found in [1], [5]. We will also briefly recall our preliminary works about hierarchical coarsening of FoD. Section III presents the novel hierarchical flexible (adaptive) coarsening method which can be regarded as the extension of our previous works. Two simple examples are given in section IV to illustrate the detailed calculation steps. Simulation experiments are presented in section V to show the rationality of this new approach. Finally, Sect.VI concludes the paper with future works perspectives.

## II. MATHEMATICAL BACKGROUND

This section provides a brief reminder of basics of DST and DSMT, and of original hierarchical coarsening method which are necessary for the presentation and the understanding of the more general flexible coarsening approximate method of section III.

### A. Basics of DST and DSMT

In DST framework, the frame of discernment<sup>2</sup>  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$  ( $n \geq 2$ ) is a set of exhaustive and exclusive elements (hypotheses) which represent the possible solutions of the problem under consideration and thus Shafer's model assumes  $\theta_i \cap \theta_j = \emptyset$  for  $i \neq j$  in  $\{1, \dots, n\}$ . A basic belief assignment (BBA)  $m(\cdot)$  is defined by the mapping:  $2^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . In DSMT, one can abandon Shafer's model (if Shafer's model doesn't fit with the problem) and refute the principle of the third excluded middle<sup>3</sup>. Instead of defining the BBAs on the power set  $2^\Theta \triangleq (\Theta, \cup)$  of the FoD, the BBAs are defined on the so-called *hyper-power set* (or Dedekind's lattice) denoted  $D^\Theta \triangleq (\Theta, \cup, \cap)$  whose cardinalities follows Dedekind's numbers sequence, see [5], Vol.1 for details and examples. A (generalized) BBA, called a mass function,  $m(\cdot)$  is defined by the mapping:  $D^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in D^\Theta} m(A) = 1$ . DSMT framework encompasses DST framework because  $2^\Theta \subset D^\Theta$ . In DSMT we can take into account also a set of *integrity constraints* on the FoD (if known), by specifying all the pairs of elements which are really disjoint. Stated otherwise, Shafer's model is a specific DSMT model where all elements are known to be disjoint.  $A \in D^\Theta$  is called a focal element of  $m(\cdot)$  if  $m(A) > 0$ . A BBA is called a Bayesian BBA if all of its focal elements are singletons and Shafer's model is assumed, otherwise it is called non-Bayesian [1]. A full ignorance source is represented by the vacuous BBA  $m_v(\Theta) = 1$ . The belief (or credibility) and plausibility functions are respectively defined by  $Bel(X) \triangleq \sum_{Y \in D^\Theta | Y \subseteq X} m(Y)$  and  $Pl(X) \triangleq \sum_{Y \in D^\Theta | Y \cap X \neq \emptyset} m(Y)$ .  $BI(X) \triangleq [Bel(X), Pl(X)]$  is called the belief interval of

$X$ . Its length  $U(X) \triangleq Pl(X) - Bel(X)$  measures the degree of uncertainty of  $X$ .

In 1976, Shafer did propose Dempster's rule<sup>4</sup> to combine BBAs in DST framework. DS rule is defined by  $m_{DS}(\emptyset) = 0$  and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ ,

$$m_{DS}(A) = \frac{\sum_{B, C \in 2^\Theta | B \cap C = A} m_1(B)m_2(C)}{1 - \sum_{B, C \in 2^\Theta | B \cap C = \emptyset} m_1(B)m_2(C)}. \quad (1)$$

DS rule formula is commutative and associative and can be easily extended to the fusion of  $S > 2$  BBAs. Unfortunately, DS rule has been highly disputed during the last decades by many authors because of its counter-intuitive behavior in high or even low conflict situations, and that is why many rules of combination have been proposed in the literature to combine BBAs [13]. To palliate DS rule drawbacks, the very interesting PCR6 (Proportional Conflict redistribution rule #6) has been proposed in DSMT and it is usually adopted<sup>5</sup> in recent applications of DSMT. The fusion of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  by the PCR6 rule is obtained by  $m_{PCR6}(\emptyset) = 0$  and  $\forall A \in D^\Theta \setminus \{\emptyset\}$

$$m_{PCR6}(A) = m_{12}(A) + \sum_{B \in D^\Theta \setminus \{A\} | A \cap B = \emptyset} \left[ \frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right], \quad (2)$$

where  $m_{12}(A) = \sum_{B, C \in D^\Theta | B \cap C = A} m_1(B)m_2(C)$  is the conjunctive operator, and each element  $A$  and  $B$  are expressed in their disjunctive normal form. If the denominator involved in the fraction is zero, then this fraction is discarded. The general PCR6 formula for combining more than two BBAs altogether is given in [5], Vol. 3. We adopt the generic notation  $m_{12}^{PCR6}(\cdot) = PCR6(m_1(\cdot), m_2(\cdot))$  to denote the fusion of  $m_1(\cdot)$  and  $m_2(\cdot)$  by PCR6 rule. PCR6 is not associative and PCR6 rule can also be applied in DST framework (with Shafer's model of FoD) by replacing  $D^\Theta$  by  $2^\Theta$  in Eq. (2).

### B. Hierarchical coarsening for fusion of Bayesian BBAs

Here, we briefly recall the principle of hierarchical coarsening of FoD to reduce the computational complexity of PCR6 combination of original Bayesian BBAs. The fusion of original non-Bayesian BBAs will be presented in the next section.

This principle was called *rigid grouping* in our previous works [17]–[19]. The goal of this coarsening is to replace the original (refined) Frame of Discernment (FoD)  $\Theta$  by a set of coarsened ones to make the computation of PCR6 rule tractable. Because we consider here only Bayesian BBA to combine, their focal elements are only singletons of the FoD  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$ , with  $n \geq 2$ , and we assume Shafer's model of the FoD  $\Theta$ .

A coarsening of the FoD  $\Theta$  means to replace it with another FoD less specific of smaller dimension  $\Omega = \{\omega_1, \dots, \omega_k\}$  with  $k < n$  from the elements of  $\Theta$ . This can be done in many

<sup>2</sup>We use the symbol  $\triangleq$  to mean equals by definition.

<sup>3</sup>The third excluded middle principle assumes the existence of the complement for any elements/propositions belonging to the power set  $2^\Theta$ .

<sup>4</sup>We use DS index to refer to Dempster-Shafer's rule (DS rule) because Shafer did really promote Dempster's rule in his milestone book [1].

<sup>5</sup>PCR6 rule coincides with PCR5 when combining only two BBAs [5].

ways depending the problem under consideration. Generally, the elements of  $\Omega$  are singletons of  $\Theta$ , and disjunctions of elements of  $\Theta$ . For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , then the possible coarsened frames built from  $\Theta$  could be, for instance,  $\Omega = \{\omega_1 = \theta_1, \omega_2 = \theta_2, \omega_3 = \theta_3 \cup \theta_4\}$ , or  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2, \omega_2 = \theta_3 \cup \theta_4\}$ , etc. When dealing with Bayesian BBAs, the projection<sup>6</sup>  $m^\Omega(\cdot)$  of the original BBA  $m^\Theta(\cdot)$  is simply obtained by taking

$$m^\Omega(\omega_i) = \sum_{\theta_j \subseteq \omega_i} m^\Theta(\theta_j). \quad (3)$$

The hierarchical coarsening process (or rigid grouping) is a simple dichotomous approach of coarsening obtained as follows:

- If  $n = |\Theta|$  is an even number:

The disjunction of the  $n/2$  first elements  $\theta_1$  to  $\theta_{n/2}$  of  $\Theta$  define the element  $\omega_1$  of  $\Omega$ , and the last  $n/2$  elements  $\theta_{n/2+1}$  to  $\theta_n$  of  $\Theta$  define the element  $\omega_2$  of  $\Omega$ , that is

$$\Omega \triangleq \{\omega_1 = \theta_1 \cup \dots \cup \theta_{n/2}, \omega_2 = \theta_{n/2+1} \cup \dots \cup \theta_n\},$$

and based on (3), one has

$$m^\Omega(\omega_1) = \sum_{j=1, \dots, \frac{n}{2}} m^\Theta(\theta_j), \quad (4)$$

$$m^\Omega(\omega_2) = \sum_{j=\frac{n}{2}+1, \dots, n} m^\Theta(\theta_j). \quad (5)$$

For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , and one considers the Bayesian BBA  $m^\Theta(\theta_1) = 0.1$ ,  $m^\Theta(\theta_2) = 0.2$ ,  $m^\Theta(\theta_3) = 0.3$  and  $m^\Theta(\theta_4) = 0.4$ , then  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2, \omega_2 = \theta_3 \cup \theta_4\}$  and  $m^\Omega(\omega_1) = 0.1 + 0.2 = 0.3$  and  $m^\Omega(\omega_2) = 0.3 + 0.4 = 0.7$ .

- If  $n = |\Theta|$  is an odd number:

In this case, the element  $\omega_1$  of the coarsened frame  $\Omega$  is the disjunction of the  $\lceil n/2 + 1 \rceil$ <sup>7</sup> first elements of  $\Theta$ , and the element  $\omega_2$  is the disjunction of other elements of  $\Theta$ . That is

$$\Omega \triangleq \{\omega_1 = \theta_1 \cup \dots \cup \theta_{\lceil n/2 + 1 \rceil}, \omega_2 = \theta_{\lceil n/2 + 1 \rceil + 1} \cup \dots \cup \theta_n\},$$

and based on (3), one has

$$m^\Omega(\omega_1) = \sum_{j=1, \dots, \lceil \frac{n}{2} + 1 \rceil} m^\Theta(\theta_j), \quad (6)$$

$$m^\Omega(\omega_2) = \sum_{j=\lceil \frac{n}{2} + 1 \rceil + 1, \dots, n} m^\Theta(\theta_j). \quad (7)$$

For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and one considers the Bayesian BBA  $m^\Theta(\theta_1) = 0.1$ ,  $m^\Theta(\theta_2) = 0.2$ ,  $m^\Theta(\theta_3) = 0.3$ ,  $m^\Theta(\theta_4) = 0.3$  and  $m^\Theta(\theta_5) = 0.1$ , then  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2 \cup \theta_3, \omega_2 = \theta_4 \cup \theta_5\}$  and  $m^\Omega(\omega_1) = 0.1 + 0.2 + 0.3 = 0.6$  and  $m^\Omega(\omega_2) = 0.3 + 0.1 = 0.4$ .

<sup>6</sup>For clarity and convenience, we put explicitly as upper index the FoD for which the belief mass refers.

<sup>7</sup>The notation  $\lceil x \rceil$  means the integer part of  $x$ .

Of course, the same coarsening applies to all original BBAs  $m_s^\Theta(\cdot)$ ,  $s = 1, \dots, S$  of the  $S > 1$  sources of evidence to work with less specific BBAs  $m_s^\Omega(\cdot)$ ,  $s = 1, \dots, S$ . The less specific BBAs (called coarsened BBAs by abuse of language) can then be combined with PCR6 rule of combination according to formula (2). This dichotomous coarsening method is repeated iteratively  $l$  times as schematically represented by a bintree<sup>8</sup>. The last step of this hierarchical process is to calculate the combined (Bayesian) BBA of all focal elements according to the connection weights of the bintree structure, where the number of iterations (or layers)  $l$  of the tree depends on the cardinality  $|\Theta|$  of the original FoD  $\Theta$ . Specifically, the assignment of each focal element is updated according to the connection weights of link paths from root to terminal nodes. This principle is illustrated in details in the following example.

**Example 1:** Let's consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and the following three Bayesian BBAs

Focal elem.	$m_1^\Theta(\cdot)$	$m_2^\Theta(\cdot)$	$m_3^\Theta(\cdot)$
$\theta_1$	0.1	0.4	0
$\theta_2$	0.2	0	0.1
$\theta_3$	0.3	0.1	0.5
$\theta_4$	0.3	0.1	0.4
$\theta_5$	0.1	0.4	0

The hierarchical coarsening and fusion of BBAs is obtained from the following steps:

**Step 1:** We define the bintree structure based on iterative half split of FoD as shown in Fig. 1.

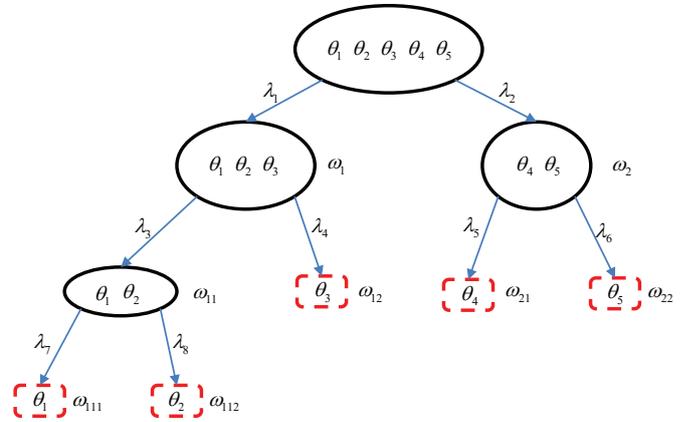


Figure 1: Fusion of Bayesian BBAs using bintree coarsening for Example 1.

The connecting weights are denoted as  $\lambda_1, \dots, \lambda_8$ . The elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$ :  $\Omega_1 = \{\omega_1 \triangleq \theta_1 \cup \theta_2 \cup \theta_3, \omega_2 \triangleq \theta_4 \cup \theta_5\}$
- At layer  $l = 2$ :

$$\Omega_2 = \{\omega_{11} \triangleq \theta_1 \cup \theta_2, \omega_{12} \triangleq \theta_3, \omega_{21} \triangleq \theta_4, \omega_{22} \triangleq \theta_5\}$$

<sup>8</sup>Here we consider bintree only for simplicity, which means that the coarsened frame  $\Omega$  consists of two elements only. Of course a similar method can be used with tri-tree, quad-tree, etc.

- At layer  $l = 3$ :  $\Omega_3 = \{\omega_{111} \triangleq \theta_1, \omega_{112} \triangleq \theta_2\}$

**Step 2:** The BBAs of elements of the (sub-)frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use (6)-(7) because  $|\Theta| = 5$  is an odd number. Therefore, we get

Focal elem.	$m_1^{\Omega_1}(\cdot)$	$m_2^{\Omega_1}(\cdot)$	$m_3^{\Omega_1}(\cdot)$
$\omega_1 \triangleq \theta_1 \cup \theta_2 \cup \theta_3$	0.6	0.5	0.6
$\omega_2 \triangleq \theta_4 \cup \theta_5$	0.4	0.5	0.4

- At layer  $l = 2$ : We work with the two subframes  $\Omega_{21} \triangleq \{\omega_{11}, \omega_{12}\}$  and  $\Omega_{22} \triangleq \{\omega_{21}, \omega_{22}\}$  of  $\Omega_2$  with the BBAs:

Focal elem.	$m_1^{\Omega_{21}}(\cdot)$	$m_2^{\Omega_{21}}(\cdot)$	$m_3^{\Omega_{21}}(\cdot)$
$\omega_{11} \triangleq \theta_1 \cup \theta_2$	$\frac{1}{2}$	$\frac{4}{5}$	$\frac{1}{6}$
$\omega_{12} \triangleq \theta_3$	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{5}{6}$

Focal elem.	$m_1^{\Omega_{22}}(\cdot)$	$m_2^{\Omega_{22}}(\cdot)$	$m_3^{\Omega_{22}}(\cdot)$
$\omega_{21} \triangleq \theta_4$	$\frac{3}{4}$	$\frac{1}{5}$	1
$\omega_{22} \triangleq \theta_5$	$\frac{1}{4}$	$\frac{4}{5}$	0

These mass values are obtained by the proportional redistribution of the mass of each focal element with respect to the mass of its parent focal element in the bin tree. For example, the value  $m_2^{\Omega_{21}}(\omega_{11}) = 4/5$  is derived by taking

$$m_2^{\Omega_{21}}(\omega_{11}) = \frac{m_2^\Theta(\theta_1) + m_2^\Theta(\theta_2)}{m_2^\Theta(\theta_1) + m_2^\Theta(\theta_2) + m_2^\Theta(\theta_3)} = \frac{0.4}{0.5} = \frac{4}{5}.$$

Other mass values are computed similarly using this proportional redistribution method.

- At layer  $l = 3$ : We use again the proportional redistribution method which gives us

Focal elem.	$m_1^{\Omega_3}(\cdot)$	$m_2^{\Omega_3}(\cdot)$	$m_3^{\Omega_3}(\cdot)$
$\omega_{111} \triangleq \theta_1$	$\frac{1}{3}$	1	0
$\omega_{112} \triangleq \theta_2$	$\frac{2}{3}$	0	1

**Step 3:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening elements. In each layer  $l$ , we fuse sequentially<sup>9</sup> the three BBAs using PCR6 formula (2). More precisely, we compute at first  $m_{(12)3}^{PCR6, \Omega_1}(\cdot) = PCR6(m_1^{\Omega_1}(\cdot), m_2^{\Omega_1}(\cdot))$  and then  $m_{(12)3}^{PCR6, \Omega_2}(\cdot) = PCR6(m_{(12)3}^{PCR6, \Omega_1}(\cdot), m_3^{\Omega_2}(\cdot))$ . Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :

$$\lambda_1 = m_{(12)3}^{PCR6, \Omega_1}(\omega_1) = 0.6297,$$

$$\lambda_2 = m_{(12)3}^{PCR6, \Omega_1}(\omega_2) = 0.3703.$$

- At layer  $l = 2$ :

$$\lambda_3 = m_{(12)3}^{PCR6, \Omega_{21}}(\omega_{11}) = 0.4137,$$

$$\lambda_4 = m_{(12)3}^{PCR6, \Omega_{21}}(\omega_{12}) = 0.5863,$$

$$\lambda_5 = m_{(12)3}^{PCR6, \Omega_{22}}(\omega_{21}) = 0.8121,$$

$$\lambda_6 = m_{(12)3}^{PCR6, \Omega_{22}}(\omega_{22}) = 0.1879.$$

<sup>9</sup>Because PCR6 fusion is not associative, we should apply the general PCR6 formula to get best results. Here we use sequential fusion to reduce the computational complexity even if the fusion result is approximate.

- At layer  $l = 3$ :

$$\lambda_7 = m_{(12)3}^{PCR6, \Omega_3}(\omega_{111}) = 0.3103,$$

$$\lambda_8 = m_{(12)3}^{PCR6, \Omega_3}(\omega_{112}) = 0.6897.$$

**Step 4:** The final assignment of belief mass to the elements of original FoD  $\Theta$  are calculated using the product of the connection weights of link paths from root (top) node to terminal nodes (leaves). We finally get the following *resulting combined and normalized Bayesian BBA*

$$m^\Theta(\theta_1) = \lambda_1 \cdot \lambda_3 \cdot \lambda_7 = 0.6297 \cdot 0.4137 \cdot 0.3103 = 0.0808,$$

$$m^\Theta(\theta_2) = \lambda_1 \cdot \lambda_3 \cdot \lambda_8 = 0.6297 \cdot 0.4137 \cdot 0.6897 = 0.1797,$$

$$m^\Theta(\theta_3) = \lambda_1 \cdot \lambda_4 = 0.6297 \cdot 0.5863 = 0.3692,$$

$$m^\Theta(\theta_4) = \lambda_2 \cdot \lambda_5 = 0.3703 \cdot 0.8121 = 0.3007,$$

$$m^\Theta(\theta_5) = \lambda_2 \cdot \lambda_6 = 0.3703 \cdot 0.1879 = 0.0696.$$

### III. NEW HIERARCHICAL FLEXIBLE COARSENING METHOD

Contrary to the (rigid) hierarchical coarsening method presented in section II, in our new flexible coarsening approach the elements  $\theta_i$ ,  $i = 1, \dots, n$  in FoD  $\Theta$  will not be half split to build coarsening focal elements  $\omega_j$ ,  $j = 1, \dots, k$  of the FoD  $\Omega_l$ . In the hierarchical flexible (adaptive) coarsening method, the elements  $\theta_i$  chosen to belong to the same group are determined using the consensus information drawn from the BBAs provided by the sources. Specifically, the degrees of disagreement between the provided sources on decisions  $(\theta_1, \theta_2, \dots, \theta_n)$  are first calculated using the belief-interval based distance  $d_{BI}$  [16], [20] to obtain *disagreement vector*. Then, the k-means algorithm is applied for clustering elements  $\theta_i$ ,  $i = 1, \dots, n$  based on the corresponding value in consensus vector. It is worth noting that values of disagreement reflect the preferences of independent sources of evidence for the same focal element. If they are small, it means that all sources have a consistent opinion and these elements should be clustered in the same group. Conversely, if disagreement values are large, it means that the sources have strong disagreement on these focal elements, and these focal elements need to be clustered in another group.

#### A. Calculating the disagreement vector

Let us consider several BBAs  $m_s^\Theta(\cdot)$ , ( $s = 1, \dots, S$ ) defined on same FoD  $\Theta$  of cardinality  $|\Theta| = n$ . The specific BBAs  $m_{\theta_i}(\cdot)$ ,  $i = 1, \dots, n$  entirely focused on  $\theta_i$  are defined by  $m_{\theta_i}(\theta_i) = 1$ , and for  $X \neq \theta_i$   $m_{\theta_i}(X) = 0$ . The disagreement of opinions of two sources about  $\theta_i$  is defined as the  $L_1$ -distance between the  $d_{BI}$  distances of the BBAs  $m_s^\Theta(\cdot)$ ,  $s = 1, 2$  to  $m_{\theta_i}(\cdot)$ , which is expressed by

$$D_{12}(\theta_i) \triangleq |d_{BI}(m_1^\Theta(\cdot), m_{\theta_i}(\cdot)) - d_{BI}(m_2^\Theta(\cdot), m_{\theta_i}(\cdot))|. \quad (8)$$

The disagreement of opinions of  $S \geq 3$  sources about  $\theta_i$ , is defined as

$$D_{1-S}(\theta_i) \triangleq \frac{1}{2} \sum_{i=1}^S \sum_{j=1}^S |d_{BI}(m_i^\Theta(\cdot), m_{\theta_i}(\cdot)) - d_{BI}(m_j^\Theta(\cdot), m_{\theta_i}(\cdot))|, \quad (9)$$

where  $d_{BI}$  distance is defined by<sup>10</sup> [20]

$$d_{BI}^E(m_1, m_2) \triangleq \sqrt{n_c \cdot \sum_{i=1}^{2^n-1} [d^I(BI_1(\theta_i), BI_2(\theta_i))]^2}. \quad (10)$$

Here,  $n_c = 1/2^{n-1}$  is the normalization constant and  $d^I([a, b], [c, d])$  is the Wasserstein's distance defined by  $d^I([a, b], [c, d]) = \sqrt{[\frac{a+b}{2} - \frac{c+d}{2}]^2 + \frac{1}{3}[\frac{b-a}{2} - \frac{d-c}{2}]^2}$ . And  $BI(\theta_i) = [Bel(\theta_i), Pl(\theta_i)]$ .

The disagreement vector  $\mathbf{D}_{1-S}$  is defined by

$$\mathbf{D}_{1-S} \triangleq [D_{1-S}(\theta_1), \dots, D_{1-S}(\theta_n)]. \quad (11)$$

### B. Clustering focal elements

Once  $\mathbf{D}_{1-S}$  is derived, a clustering algorithm is used to coarsen focal elements according to their corresponding values in  $\mathbf{D}_{1-S}$ . In this paper, we have used the k-means algorithm<sup>11</sup> to cluster focal elements. For each source  $s = 1, \dots, S$ , the mass assignments of focal elements in two<sup>12</sup> different clusters are added up according to formulas (12)–(13).

$$m_s^\Omega(\omega_1) = \sum_{\theta_i \in \omega_1} m_s^\Theta(\theta_i), \quad (12)$$

$$m_s^\Omega(\omega_2) = \sum_{\theta_j \in \omega_2} m_s^\Theta(\theta_j). \quad (13)$$

### C. Combination of the BBAs

Based on the disagreement vector and k-means algorithm, a new adaptive bintree structure based on this flexible coarsening decomposition is obtained (see example in the next section) and the elements in FoD  $\Theta$  are grouped more reasonably in each layer of the decomposition. Once the adaptive bintree structure is derived, other steps (multiplications of link weights) can be implemented which are identical to hierarchical (rigid) coarsening method presented in section II to get the final combined Bayesian BBA.

### D. Summary of the method

The fusion method of BBAs to get a combined Bayesian BBA based on hierarchical flexible decomposition of the FoD consists of the four steps below illustrated in Fig. 2.

- **Step 1 (pre-processing):** At first, all input BBAs to combine are approximated to Bayesian BBAs with DSMT transform.
- **Step 2 (disagreement vector):**  $\mathbf{D}_{1-S}(\cdot)$  is calculated using  $d_{BI}$  distances to estimate the degree of disagreement of BBAs  $m_1^\Theta, \dots, m_S^\Theta$  on potential decisions  $\theta_1, \dots, \theta_n$ .
- **Step 3 (adaptive bintree):** The adaptive bintree decomposition of the FoD  $\Theta$  is obtained using k-Means algorithm to get elements of subframes  $\Omega_l$ .
- **Step 4 (assignments and connection weights):** For each source  $m_s^\Theta(\cdot)$  to combine, the mass assignment of

<sup>10</sup>For simplicity, we assume Shafer's model so that  $|2^\Theta| = 2^n$ , otherwise the number of elements in the summation of (10) should be  $|D^\Theta| - 1$  with another normalization constant  $n_c$ .

<sup>11</sup>which is implemented in Matlab™

<sup>12</sup>because we use here the bisection decomposition.

each element of subframe  $\Omega_l$  is computed by (12)–(13). The weight of links between two layers of the bintree decomposition are obtained with PCR6 rule<sup>13</sup>.

- **Step 5 (fusion):** The final result (combined Bayesian BBA) is computed by the product of weights of link paths from root to terminal nodes.

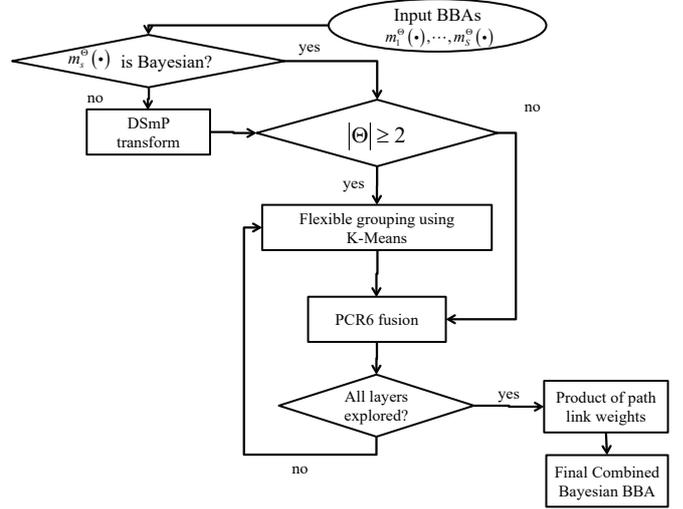


Figure 2: Hierarchical flexible decomposition of FoD for fusion.

## IV. TWO SIMPLE EXAMPLES

### A. Example 1 (fusion of Bayesian BBAs)

Let us revisit example 1 presented in section II-B. It can be verified in applying formula (9) that the disagreement vector  $\mathbf{D}_{1-3}$  for this example is equal to

$$\mathbf{D}_{1-3} = [0.4085, 0.2156, 0.3753, 0.2507, 0.4086]$$

The derivation of  $D_{1-3}(\theta_1)$  is given below for convenience.

$$\begin{aligned} D_{1-3}(\theta_1) &= |d_{BI}(m_1^\Theta(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_2^\Theta(\cdot), m_{\theta_1}(\theta_1))| \\ &\quad + |d_{BI}(m_2^\Theta(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_3^\Theta(\cdot), m_{\theta_1}(\theta_1))| \\ &\quad + |d_{BI}(m_1^\Theta(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_3^\Theta(\cdot), m_{\theta_1}(\theta_1))| \\ &= 0.4085. \end{aligned}$$

Based on the disagreement vector and k-means algorithm, a new adaptive bintree structure is obtained and shown in Fig. 3. Compared to Fig. 1, the elements in FoD  $\Theta$  are grouped more reasonably. In vector  $\mathbf{D}_{1-3}$ ,  $\theta_1$  and  $\theta_5$  lie in similar degree of disagreement so that they are put in the same group. Similarly for  $\theta_2$  and  $\theta_4$ . However, element  $\theta_3$  seems *weird*, which is put alone at the beginning of flexible coarsening. Once this adaptive bintree decomposition is obtained, other steps can be implemented which are identical to hierarchical coarsening method of section II to get the final combined BBA.

The flexible coarsening and fusion of BBAs is obtained from the following steps:

<sup>13</sup>general formula preferred, or applied sequentially to reduce complexity.

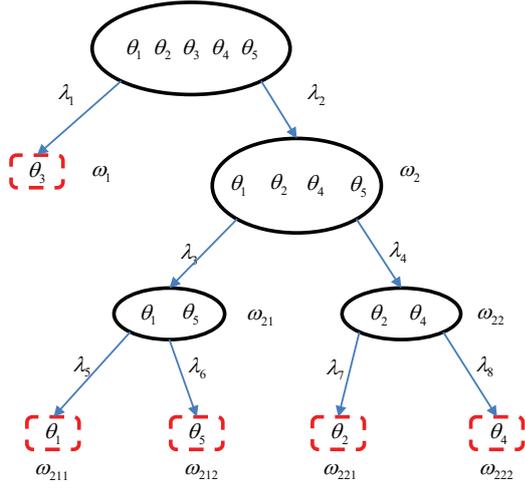


Figure 3: Example 1: Flexible bintree decomposition of FoD.

**Step 1:** According to Fig.3, the elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$ :  $\Omega_1 = \{\omega_1 \triangleq \theta_3, \omega_2 \triangleq \theta_1 \cup \theta_2 \cup \theta_4 \cup \theta_5\}$
- At layer  $l = 2$ :  $\Omega_2 = \{\omega_{21} \triangleq \theta_1 \cup \theta_5, \omega_{22} \triangleq \theta_2 \cup \theta_4\}$
- At layer  $l = 3$ :  $\Omega_3 = \{\omega_{211} \triangleq \theta_1, \omega_{212} \triangleq \theta_5, \omega_{221} \triangleq \theta_2, \omega_{222} \triangleq \theta_4\}$

**Step 2:** The BBAs of elements of the (sub-)frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use (12)-(13) and we get

Focal elem.	$m_1^{\Omega_1}(\cdot)$	$m_2^{\Omega_1}(\cdot)$	$m_3^{\Omega_1}(\cdot)$
$\omega_1 \triangleq \theta_3$	0.3	0.1	0.5
$\omega_2 \triangleq \theta_1 \cup \theta_2 \cup \theta_4 \cup \theta_5$	0.7	0.9	0.5

- At layer  $l = 2$ : We use again the proportional redistribution method which gives us:

Focal elem.	$m_1^{\Omega_2}(\cdot)$	$m_2^{\Omega_2}(\cdot)$	$m_3^{\Omega_2}(\cdot)$
$\omega_{21} \triangleq \theta_1 \cup \theta_5$	$\frac{3}{7}$	$\frac{4}{9}$	$\frac{1}{5}$
$\omega_{22} \triangleq \theta_2 \cup \theta_4$	$\frac{4}{7}$	$\frac{5}{9}$	$\frac{4}{5}$

- At layer  $l = 3$ : We work with the two subframes  $\Omega_{31} \triangleq \{\omega_{211}, \omega_{212}\}$  and  $\Omega_{32} \triangleq \{\omega_{221}, \omega_{222}\}$  of  $\Omega_3$  with the BBAs

Focal elem.	$m_1^{\Omega_{31}}(\cdot)$	$m_2^{\Omega_{31}}(\cdot)$	$m_3^{\Omega_{31}}(\cdot)$
$\omega_{211} \triangleq \theta_1$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\omega_{212} \triangleq \theta_5$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Focal elem.	$m_1^{\Omega_{32}}(\cdot)$	$m_2^{\Omega_{32}}(\cdot)$	$m_3^{\Omega_{32}}(\cdot)$
$\omega_{221} \triangleq \theta_2$	$\frac{2}{5}$	0	$\frac{1}{5}$
$\omega_{222} \triangleq \theta_4$	$\frac{3}{5}$	1	$\frac{4}{5}$

**Step 3:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening elements. Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :

$$\lambda_1 = 0.2226; \quad \lambda_2 = 0.7774.$$

- At layer  $l = 2$ :

$$\lambda_3 = 0.2200; \quad \lambda_4 = 0.7800.$$

- At layer  $l = 3$ :

$$\lambda_5 = 0.5; \quad \lambda_6 = 0.5; \quad \lambda_7 = 0.0669; \quad \lambda_8 = 0.9331.$$

**Step 4:** We finally get the following *resulting combined and normalized Bayesian BBA*

$$m^\Theta(\cdot) = \{0.0855, 0.0406, 0.2226, 0.5658, 0.0855\}.$$

**B. Example 2 (with non-Bayesian BBAs)**

**Example 1bis:** Let's consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and the following BBAs given by

Focal elem.	$m_1^\Theta(\cdot)$	$m_2^\Theta(\cdot)$	$m_3^\Theta(\cdot)$
$\theta_1$	0.1	0.4	0
$\theta_2$	0.2	0	0
$\theta_3$	0.3	0.05	0
$\theta_4$	0.03	0.05	0
$\theta_5$	0.1	0.04	0
$\theta_1 \cup \theta_2$	0.1	0.04	0
$\theta_2 \cup \theta_3 \cup \theta_5$	0	0.02	0.1
$\theta_3 \cup \theta_4$	0.02	0.1	0.2
$\theta_1 \cup \theta_5$	0.1	0.3	0.2
$\Theta$	0.05	0	0.5

**Step 1 (Pre-Processing):** All these three BBAs are transformed into Bayesian BBAs with DSMP transform and the generated BBAs are illustrated as

Focal elem.	$m_1^{\Theta}(\cdot)$	$m_2^{\Theta}(\cdot)$	$m_3^{\Theta}(\cdot)$
$\theta_1$	0.1908	0.7127	0.2000
$\theta_2$	0.2804	0	0.1334
$\theta_3$	0.3387	0.1111	0.2333
$\theta_4$	0.0339	0.1	0.2000
$\theta_5$	0.1562	0.0761	0.2333

It can be verified in applying formula (9) that the disagreement vector  $\mathbf{D}_{1-3}$  for this example is equal to

$$\mathbf{D}_{1-3} = [0.5385, 0.3632, 0.3453, 0.2305, 0.2827].$$

**Step 2:** According to the clustering algorithm, the elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$ :  $\Omega_1 = \{\omega_1 \triangleq \theta_1, \omega_2 \triangleq \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5\}$
- At layer  $l = 2$ :  $\Omega_2 = \{\omega_{21} \triangleq \theta_2 \cup \theta_3, \omega_{22} \triangleq \theta_4 \cup \theta_5\}$
- At layer  $l = 3$ :  $\Omega_3 = \{\omega_{211} \triangleq \theta_2, \omega_{212} \triangleq \theta_3, \omega_{221} \triangleq \theta_4, \omega_{222} \triangleq \theta_5\}$

**Step 3:** The BBAs of elements of the (sub-)frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use (12)-(13) and we get

Focal elem.	$m_1^{\Omega_1}(\cdot)$	$m_2^{\Omega_1}(\cdot)$	$m_3^{\Omega_1}(\cdot)$
$\omega_1 \triangleq \theta_1$	0.1908	0.7127	0.2000
$\omega_2 \triangleq \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5$	0.8092	0.2873	0.8000

- At layer  $l = 2$ : We use again the proportional redistribution method which gives us:

Focal elem.	$m_1^{\Omega_2}(\cdot)$	$m_2^{\Omega_2}(\cdot)$	$m_3^{\Omega_2}(\cdot)$
$\omega_{21} \triangleq \theta_2 \cup \theta_3$	0.7651	0.3867	0.4584
$\omega_{22} \triangleq \theta_4 \cup \theta_5$	0.2349	0.6133	0.5416

- At layer  $l = 3$ : We work with the two subframes  $\Omega_{31} \triangleq \{\omega_{211}, \omega_{212}\}$  and  $\Omega_{32} \triangleq \{\omega_{221}, \omega_{222}\}$  of  $\Omega_3$  with the BBAs:

Focal elem.	$m_1^{\Omega_{31}}(\cdot)$	$m_2^{\Omega_{31}}(\cdot)$	$m_3^{\Omega_{31}}(\cdot)$
$\omega_{211} \triangleq \theta_2$	0.4529	0	0.3638
$\omega_{212} \triangleq \theta_3$	0.5471	1	0.6362
Focal elem.	$m_1^{\Omega_{32}}(\cdot)$	$m_2^{\Omega_{32}}(\cdot)$	$m_3^{\Omega_{32}}(\cdot)$
$\omega_{221} \triangleq \theta_4$	0.1783	0.5679	0.4616
$\omega_{222} \triangleq \theta_5$	0.8217	0.4321	0.5384

**Step 4:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening elements. Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :

$$\lambda_1 = 0.2345; \quad \lambda_2 = 0.7655.$$

- At layer  $l = 2$ :

$$\lambda_3 = 0.5533; \quad \lambda_4 = 0.4467.$$

- At layer  $l = 3$ :

$$\lambda_5 = 0.1606; \quad \lambda_6 = 0.8394;$$

$$\lambda_7 = 0.3349; \quad \lambda_8 = 0.6651.$$

**Step 5:** We finally get the following *resulting combined and normalized Bayesian BBA*

$$m^\ominus(\cdot) = \{0.2345, 0.0681, 0.3555, 0.1145, 0.2274\}.$$

## V. SIMULATION RESULTS AND PERFORMANCES

### A. Flexible Grouping of Singletons

1) *Similarity*:<sup>14</sup> Assuming that  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}\}$  and first, we randomly generate 2 BBAs, denoted as  $m_1^\ominus(\cdot)$  and  $m_2^\ominus(\cdot)$ , which can be seen in Table I.

Table I: BBAs for Two Sources  $m_1^\ominus(\cdot)$  and  $m_2^\ominus(\cdot)$

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$
$m_1^\ominus(\cdot)$	0.1331	0.0766	0.0175	0.0448	0.0229
$m_2^\ominus(\cdot)$	0.1020	0.0497	0.1094	0.0612	0.0612
	$\theta_6$	$\theta_7$	$\theta_8$	$\theta_9$	$\theta_{10}$
$m_1^\ominus(\cdot)$	0.1142	0.0023	0.2254	0.1583	3.4959e-04
$m_2^\ominus(\cdot)$	0.0069	0.0070	0.0128	0.0833	0.0338
	$\theta_{11}$	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{15}$
$m_1^\ominus(\cdot)$	0.0075	0.0514	0.1121	0.0314	0.0021
$m_2^\ominus(\cdot)$	0.1180	0.1202	0.1351	0.0686	0.0309

In order to fully verify the similarity between hierarchical flexible coarsening method and PCR6 in DSMT, a new strict

<sup>14</sup>Similarity represents the approximate degree between fusion results using flexible coarsening and PCR6.

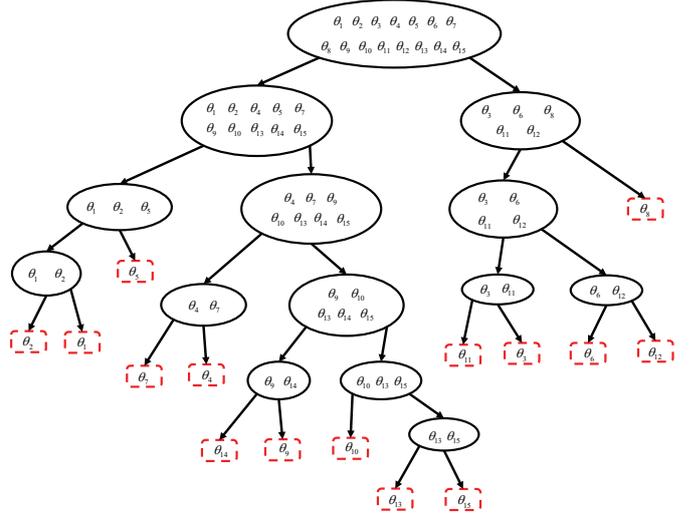


Figure 4: Structure of Hierarchical Flexible Coarsening.

distance metric between two BBAs, denoted  $d_{BI}^E$ , was recently proposed in [20], [16] and it will be used in this paper.

In this paper, we regard  $d_{BI}^E$  as one criteria for evaluating the degree of similarity between the fusion results obtained from flexible coarsening and PCR6.

Based on (8) and (10), the disagreement vector  $\mathbf{D}(\cdot)$  is obtained:

$$\mathbf{D}(\cdot) = (0.0032, 0.0020, 0.0290, 0.0092, 0.0147, 0.0228, 0.0059, 0.0537, 0.0154, 0.0131, 0.0338, 0.0235, 0.0118, 0.0145, 0.0120).$$

Thus, bintree structure of hierarchical flexible coarsening is illustrated in Fig. 4 and the similarity between fusion results of hierarchical flexible coarsening and PCR6 is 0.9783. And the similarity between hierarchical coarsening method and PCR6 is 0.9120. In particular, terminal nodes (the red small box in Fig. 4) of flexible grouping are not in accordance with the original order  $\theta_1, \theta_2, \dots, \theta_{15}$ . This is quite different compared to original hierarchical coarsening method.

From the point of view of statistics, 100 BBAs are randomly generated to be fused with three methods: hierarchical flexible coarsening, hierarchical coarsening and also PCR6. Comparisons are made in Fig. 5, which show the superiority of our new approach proposed in this paper (Average value of new method is **97%** and the old method is **93.5%**).

### B. Flexible Grouping of Conflicting Focal Elements

Assuming that there are five sources of evidence  $m_1^\ominus(\cdot), m_2^\ominus(\cdot), m_3^\ominus(\cdot), m_4^\ominus(\cdot), m_5^\ominus(\cdot)$ , and the restricted hype-power set  $D^\ominus = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_1 \cap \theta_2, \theta_5 \cap \theta_6 \cap \theta_7, \theta_1 \cap \theta_5 \cap \theta_9 \cap \theta_{10}\}$ . And then we randomly generate 1000 BBAs for each source to calculate the similarity using (10). From Fig. 6, we can find that hierarchical flexible coarsening method can also maintain high degree of similarity which performs better than hierarchical coarsening.

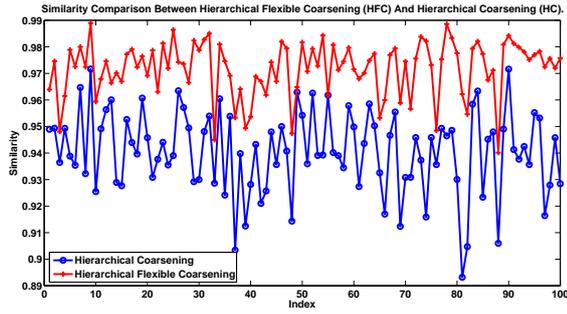


Figure 5: Comparisons Between HFC and HC (Only Singletons).

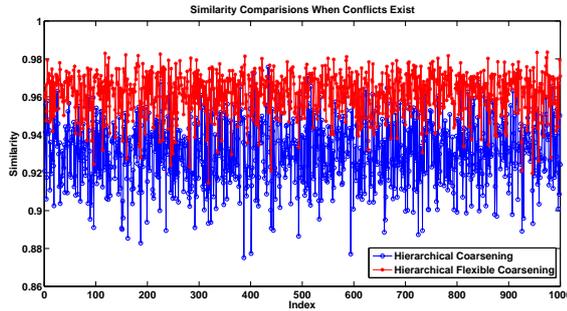


Figure 6: Comparisons Between HFC and HC (Singletons and Conflicting Focal Elements).

### C. Flexible Grouping of Uncertain and Hybrid Focal Elements

We can also deal with uncertain and hybrid focal elements. Assuming that there are also five sources of evidence  $m_1^\ominus(\cdot), m_2^\ominus(\cdot), m_3^\ominus(\cdot), m_4^\ominus(\cdot), m_5^\ominus(\cdot)$  and  $D_1^\ominus = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_1 \cup \theta_2, \theta_5 \cup \theta_6 \cup \theta_7, \theta_1 \cup \theta_5 \cup \theta_9 \cup \theta_{10}\}$ ;  $D_2^\ominus = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_2 \cap \theta_4 \cup \theta_6, \theta_1 \cup \theta_3 \cap \theta_5 \cup \theta_7 \cap \theta_9\}$ <sup>15</sup>. And then we respectively and randomly generate 1000 BBAs for these two cases  $D_1^\ominus$  and  $D_2^\ominus$ . Finally, we calculate the average similarity degree of HFC and HC with PCR6 in Table II, which illustrates HFC performs better than old method. However, there exist the extra time cost of HFC compared to HC due to the clustering steps in coarsening process.

Table II: Similarity Comparisons

	Hierarchical Flexible Coarsening	Hierarchical Coarsening
$D_1^\ominus$	98%	91%
$D_2^\ominus$	97%	93%

## VI. CONCLUSION AND PERSPECTIVES

A novel hierarchical flexible approximate method in DSMT is proposed here. Compared to original hierarchical coarsen-

<sup>15</sup>In this case,  $D_1^\ominus$  represents uncertain focal elements and  $D_2^\ominus$  represents hybrid focal elements.

ing, flexible strategy guarantees higher similarity with PCR6 rules in fusion process. Besides, whether focal elements in hyper power set are singletons, conflicting focal elements, uncertain or even hybrid focal elements, the new method works well. In the future work, we will focus on the general framework of hierarchical coarsening, which could generate final non-Bayesian BBAs in order to avoid loss of information. Furthermore, other advantages or disadvantages of our proposed methods such as computational efficiency and time consumption need to be further investigated.

## ACKNOWLEDGMENT

This work was supported by NNSF of China (No.61175091), Qing Lan Project of Jiangsu Province, Aeronautical Science Foundation of China (20140169002), and Six Major Top-talent Plan of Jiangsu Province.

## REFERENCES

- [1] G. Shafer, *A mathematical theory of evidence*, Princeton Univ. Press, 1976.
- [2] A. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, in *Annals of Mathematical Statistics*, Vol. 38, pp. 325–339, 1967.
- [3] P. Smets, *Practical uses of belief functions*, in K.B. Laskey and H. Prade Editors, 15th Conf. on Uncertainty in Artificial Intelligence, pp. 612–621, Stockholm, Sweden, 1999.
- [4] J. Dezert, *Foundations for a new theory of plausible and paradoxical reasoning*, in *Information & Security: An Int. Journal*, Vol. 9, 2002.
- [5] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, American Research Press, Rehoboth, NM, U.S.A., Vol. 1–4, 2004–2015. <https://www.onera.fr/fr/staff/jean-dezert/references>
- [6] <http://www.onera.fr/staff/jean-dezert>
- [7] R. Kennes, *Computational aspects of the Möbius transform of graphs*, in *IEEE Trans. on SMC*, Vol. 22, pp. 201–223, 1992.
- [8] G. Shafer, R. Logan, *Implementing Dempster's rule for hierarchical evidence*, in *Artificial Intelligence*, Vol. 33, pp. 271–298, 1987.
- [9] Y. Yang, Y.L. Liu, *Iterative approximation of basic belief assignment based on distance of evidence*, in *Plos One*, Vol. 11, No. 2, 2016.
- [10] T. Denœux, *Inner and outer approximation of belief structures using a hierarchical clustering approach*, in *Int. J. of Uncertainty, Fuzziness and Knowledge-Based System*, Vol. 9, No 4, pp. 437–460, 2001.
- [11] Y. Yang, D. Han, C. Han, F. Cao, *A novel approximation of basic probability assignment based on rank-level fusion*, in *Chinese Journal of Aeronautics*, Vol. 26, No. 4, pp. 993–999, 2013.
- [12] D. Han, Y. Yang, J. Dezert, *Two novel methods of BBA approximation based on focal element redundancy*, in *Proc. of Fusion 2015*, Washington, D.C., USA, July 2015.
- [13] P. Smets, *Analyzing the combination of conflicting belief functions*, in *Information Fusion*, Vol. 8, pp. 387–412, 2006.
- [14] P. Smets, *Decision making in the TBM: the necessity of the pignistic transformation*, in *Int. J. of Approx. reasoning*, Vol. 38, 2005.
- [15] J. Dezert, F. Smarandache, *A new probabilistic transformation of belief mass assignment*, in *Proc. of Fusion 2008*.
- [16] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-making with belief interval distance*, in *Proc. of Belief 2016 Int. Conf.*, Prague, Czech Republic, 2016.
- [17] X.D. Li, J. Dezert, X.H. Huang, Z.D. Meng, X.J. Wu, *A fast approximate reasoning method in hierarchical DSMT (A)*, in *Acta Electronica Sinica*, Vol. 38, No. 11, pp. 2567–2572, 2010.
- [18] X.D. Li, W.D. Yang, X.J. W, J. Dezert, *A fast approximate reasoning method in hierarchical DSMT (B)*, in *Acta Electronica Sinica*, Vol. 39, No. 3A, pp. 32–36, 2011.
- [19] X.D. Li, W.D. Yang, J. Dezert, *A fast approximate reasoning method in hierarchical DSMT (C)*, in *J. Huazhong Univ. of Sci. and Tech. (Natural Science Edition)*, Vol. 39, pp. 151–156, 2011.
- [20] D. Han, J., Dezert, Y., Yang, *New distance measures of evidence based on belief intervals*, in *Proc. of Belief 2014*, Oxford, UK, 2014.

# Validated Enclosure of Uncertain Nonlinear Equations using SIVIA Monte-Carlo

Nisha Rani Mahato<sup>a</sup>, Luc Jaulin<sup>b</sup>, Snehashish Chakraverty<sup>a</sup>, Jean Dezert<sup>c</sup>

<sup>a</sup>Department of Mathematics, National Institute of Technology Rourkela, Rourkela-769008, Odisha, India.

<sup>b</sup>ENSTA-Bretagne, LabSTICC, CNRS 6285, 2 rue François Verny, 29806 Brest, France.

<sup>c</sup>The French Aerospace Lab, ONERA/DTIS, 91120 Palaiseau, France.

Emails: nisha.mahato1@gmail.com, lucjaulin@gmail.com, sne\_chak@yahoo.com, jean.dezert@onera.fr

Originally published as: N.R. Mahato, L. Jaulin, S. Chakraverty, J. Dezert, *Validated Enclosure of Uncertain Nonlinear Equations using SIVIA Monte-Carlo*, in Proc. of 8th National Conference on Wave Mechanics and Vibrations (WMVC 2018), National Institute of Technology Rourkela (NITR), Odisha, India, Nov. 26-28, 2018, and reprinted with permission.

**Abstract**—The dynamical systems in various science and engineering problems are often governed by nonlinear equations (differential equations). Due to insufficiency and incompleteness of system information, the parameters in such equations may have uncertainty. Interval analysis serves as an efficient tool for handling uncertainties in terms of closed intervals. One of the major problem with interval analysis is handling “dependency problems” for computation of tightest range of solution enclosure or exact enclosure. Such dependency problems are often observed while dealing with complex nonlinear equations. In this regard, initially two test problems comprising of interval nonlinear equations are considered. The Set Inversion via Interval Analysis (SIVIA) along with Monte-Carlo approach is used to compute the exact enclosure of the test problems. Further, the efficiency of the proposed approach has also been verified for solving nonlinear differential equations (Van der Pol oscillator) subject to interval initial conditions.

**Keywords:** uncertain nonlinear equations, nonlinear oscillator, dependency problem, SIVIA Monte-Carlo, contractor.

## I. INTRODUCTION

Various vibration problems in science and engineering disciplines viz. structural mechanics, control theory, seismology, physics, biology etc. may be expressed in terms of nonlinear equations, system of nonlinear equations and nonlinear differential equations. Generally, the parameters in such equations deal with precise variables. But, the insufficiency and incompleteness of the system information often led to parameters or variables with imprecision or uncertainty. For instance, let us consider a nonlinear damped spring-mass system as given in Fig. 1 governed by the equation,

$$m\ddot{x} + c\dot{x} + \alpha x^2 + kx + \beta x^3 = f(t), \quad (1)$$

where,  $m$ ,  $c$  and  $k$  are respectively mass, damping and stiffness of the nonlinear system. Here, the external force applied on the system is  $f(t)$  with damping force  $f_d = c\dot{x} + \alpha x^2$  and spring force  $f_s = kx + \beta x^3$ .

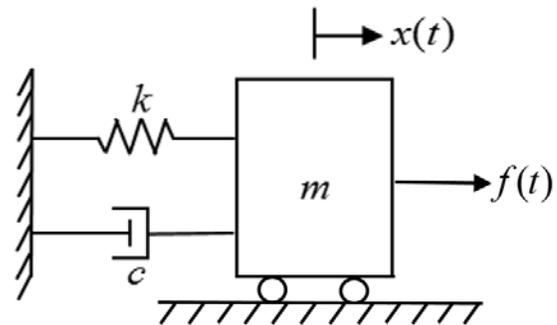


Figure 1. Damped spring-mass system.

The uncertainty of the material properties in Eq. (1) led to uncertain nonlinear differential equation. Such uncertainties may be modeled either using probabilistic approach, interval computation or fuzzy set theory. In case of non-availability of sufficient experimental data, probabilistic methods may not be able to deliver reliable results. Moreover, in fuzzy set theory a fuzzy number is expressed in terms of closed intervals through  $\alpha$ -cut approach. As such, interval analysis have emerged as a powerful tool for various practical problems in handling the uncertainties.

In early 1960s the pioneer concept related to interval computations, functions, matrices, integral and differential equations has been started by R. E. Moore [12]–[14]. System of equations, algebraic eigenvalue problems, second order initial and boundary value problems has been discussed by Alefeld and Herzberger [3]. Guaranteed interval computations with respect to set approximations, parameter and state estimation with applications in robust control and robotics are addressed by Jaulin et al. [10]. While dealing with interval computations, one of the major obstacle is to handle the ‘dependency problems’ effectively such that the tightest enclosure of solution bound may be obtained. Such dependency problems often occur in dealing with systems governed by complex nonlinear equations which often lead to over-estimation of solution bound. The dependency problem due to overestimation (*wrapping effect*) has been studied by Krämer [11] with respect to generalized interval arithmetic proposed by Hansen [9].

Other approach for reduction of overestimation while handling dependency problem may be performed using contractors [10], affine arithmetic [15] and/or parametric forms. As such, the present work proceeds with the introduction section. The preliminaries of classical arithmetic of Interval Analysis (IA) along with its application for two complex nonlinear equations comprising of imprecise variables are considered in Section II. The Set Inversion via Interval Analysis (SIVIA) along with Monte-Carlo approach is then used to compute the exact enclosure of the two test problems in Section III. Further, the proposed approach has also been verified for computing validated enclosure of nonlinear differential equations (Van der Pol oscillator) subject to interval initial conditions in Section IV.

## II. CLASSICAL INTERVAL COMPUTATIONS

Interval analysis deals with interval computations on a set of closed intervals  $\mathbb{IR}$  of real line  $\mathbb{R}$ , in order to obtain the tightest bound or enclosure for uncertain systems. A closed interval  $[x] \subset \mathbb{IR}$  is denoted by  $[x] = [\underline{x}, \bar{x}]$  such that

$$[x] = [\underline{x}, \bar{x}] = \{t \mid \underline{x} \leq t \leq \bar{x}, \text{ where } \underline{x}, \bar{x} \in \mathbb{R}\}.$$

Here,  $\underline{x} = \inf[x]$  is the infimum or lower bound of  $[x]$  and  $\bar{x} = \sup[x]$  is the supremum or upper bound of  $[x]$ . The width and center of  $[x]$  may be referred as  $[x]^w = \bar{x} - \underline{x}$  and  $[x]^c = \frac{\underline{x} + \bar{x}}{2}$  respectively.

Basic operations using classical interval arithmetic given in Moore et al. [14] are illustrated as follows:

- **Addition:**

$$[x] + [y] = [\underline{x} + \underline{y}, \bar{x} + \bar{y}],$$

- **Subtraction:**

$$[x] - [y] = [\underline{x} - \bar{y}, \bar{x} - \underline{y}],$$

- **Multiplication:**

$$[x] \cdot [y] = [\min\{S.([x], [y])\}, \max\{S.([x], [y])\}],$$

where  $S.([x], [y]) = \{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}$ .

- **Division:**

$$[x]/[y] = \begin{cases} [\underline{x}, \bar{x}] \cdot \left[\frac{1}{\bar{y}}, \frac{1}{\underline{y}}\right], & 0 \notin [y, \bar{y}], \\ (-\infty, \infty), & 0 \in [y, \bar{y}] \end{cases}$$

- **Power:**

– If  $n > 0$  is an odd number, then

$$[\mathbf{x}]^n = [\underline{x}^n, \bar{x}^n];$$

– If  $n > 0$  is an even number, then

$$[\mathbf{x}]^n = \begin{cases} [\underline{x}^n, \bar{x}^n], & [\mathbf{x}] > 0, \\ [\bar{x}^n, \underline{x}^n], & [\mathbf{x}] < 0, \\ [0, \max\{\underline{x}^n, \bar{x}^n\}], & 0 \in [\mathbf{x}]. \end{cases}$$

Then, we have illustrated two test examples for the implementation of basic interval arithmetic in Examples 1 and 2.

**Example 1:** Compute the bound  $[z_1]$  satisfying constraint

$$z_1 = x_1y_1 + x_1y_3 + x_3y_1. \quad (2)$$

such that  $x_1 + x_2 + x_3 = 1$  and  $y_1 + y_2 + y_3 = 1$ . Here,  $x_1 \in [x_1] = [0.2, 0.3]$ ,  $x_2 \in [x_2] = [0.1, 0.2]$ ,  $y_1 \in [y_1] = [0.4, 0.6]$  and  $y_2 \in [y_2] = [0.2, 0.3]$ .

Using classical IA, the bounds  $[x_3]$  and  $[y_3]$  are initially estimated as  $[x_3] \sim 1 - [x_1] - [x_2] = [0.5, 0.7]$  and  $[y_3] \sim 1 - [y_1] - [y_2] = [0.1, 0.4]$  respectively with respect to the constraints  $x_1 + x_2 + x_3 = 1$  and  $y_1 + y_2 + y_3 = 1$ . Then, the bound  $[z_1]$  is obtained as

$$[z_1]^{IA} \sim [x_1] \cdot [y_1] + [x_1] \cdot [y_3] + [x_3] \cdot [y_1] = [0.30, 0.72]. \quad (3)$$

Further, we have considered a more complicated nonlinear constraint in Example 2, related to problems of multi-criteria decision-making under imprecise scores given in Dezert et al. [7].

**Example 2:** [7] Compute the bound  $[z_2]$  satisfying constraint given by the imprecise proportional conflict redistribution (PCR) fusion rule

$$z_2 = z_1 + \frac{x_1^2 y_2}{x_1 + y_2} + \frac{y_1^2 x_2}{y_1 + x_2} \quad (4)$$

such that  $x_1 \in [0.2, 0.3]$ ,  $x_2 \in [0.1, 0.2]$ ,  $y_1 \in [0.4, 0.6]$  and  $y_2 \in [0.2, 0.3]$ .

Here, the bound of  $[z_2]$  is obtained as

$$[z_2]^{IA} \sim [z_1]^{IA} + \frac{[x_1]^2 [y_2]}{[x_1] + [y_2]} + \frac{[y_1]^2 [x_2]}{[y_1] + [x_2]} = [0.3333, 0.9315]. \quad (5)$$

The enclosures obtained in Eqs. (3) and (5) have been compared with enclosures obtained using Monte-Carlo simulation in Table I.

Table I  
INTERVAL BOUNDS OF  $z_1$  AND  $z_2$ .

i	Interval bounds	
	$[z_i]^{IA}$	$[z_i]^{MC}$
1	[0.30, 0.72]	[0.3850, 0.5935]
2	[0.3333, 0.9315]	[0.4617, 0.6825]

Here, the Monte-Carlo simulation approach using uniformly distributed 100000 independent random sample values of variables  $x_1, x_2, y_1$  and  $y_2$  have been considered, where  $x_1 \sim U([x_1])$ ,  $x_2 \sim U([x_2])$ ,  $y_1 \sim U([y_1])$  and  $y_2 \sim U([y_2])$ . From Table I, it is worth mentioning that the bounds for  $i = 1, 2$  satisfy

$$[z_i]^{MC} \subset [z_i]^{IA}.$$

In case of more sample values, the Monte-Carlo simulation may yield better interval enclosure with respect to the constraints (2) and (4), but such approach is inefficient with respect to computational time. So, we may consider the

problem in handling interval computations as to interpret the tightest or the exact enclosure  $[z_i]$  of  $z_i$  that satisfies

$$[z_i]^{MC} \subset [z_i] \subset [z_i]^{IA}, \quad (6)$$

such that

$$\begin{cases} \inf [z_i]^{IA} \leq \inf [z_i] \leq \inf [z_i]^{MC}, \\ \sup [z_i]^{MC} \leq \sup [z_i] \leq \sup [z_i]^{IA}. \end{cases} \quad (7)$$

or

$$\begin{cases} \underline{z}_i^{IA} \leq \underline{z}_i \leq \underline{z}_i^{MC}, \\ \bar{z}_i^{MC} \leq \bar{z}_i \leq \bar{z}_i^{IA}. \end{cases} \quad (8)$$

Although in the above computations, interval arithmetic looks simple for basic operations with intervals and seems appealing. But, the ‘‘dependency problem’’ is a major obstacle when complicated expressions have to be computed in order to find tightest enclosure. In this regard, the dependency effect has been discussed in detail in next section.

#### A. Dependency Problem in IA

Variable or parameter dependency problem in IA is generally exhibited when we have more than one occurrence of imprecise parameter in the governing constraint. For instance, in case of the nonlinear constraint

$$z = x^2 + y^2 \text{ for } x \in [0.1, 0.5] \text{ and } y \in [-0.6, 0.1],$$

the occurrence of each imprecise variable  $x$  and  $y$  is once. The computation of enclosure with respect to constraint  $z = x^2 + y^2$  using classical IA results to  $[z]^{IA} = [0.01, 0.61]$  which is found equivalent to the Monte-Carlo simulation of  $x \sim U([0.1, 0.5])$ ,  $y \sim U([-0.6, 0.1])$  for 100000 sample values yield  $[z]^{MC} = [0.01, 0.61]$ . But, the complexity occurs while dealing with complex nonlinear constraints as given in Examples 1 and 2, where the dependency effect is exhibited due to multiple occurrence on imprecise variables.

The dependency effect may be reduced by replacing the constraint given in Eq. (2) with an equivalent simpler constraint having less (or none) redundant variables. For instance, the equivalent constraint

$$z_1 = (1 - x_2)y_2 + x_3y_1 \quad (9)$$

results to a better enclosure approximation  $[z_1]^{IA} = [0.34, 0.66]$ . Here, the interval bound  $[0.34, 0.66]$  is contained in the bound  $[0.30, 0.72]$  obtained using the equivalent constraint given in Eq. (2). But, on the other-hand an equivalent constraint

$$z_1 = (1 - y_2)x_1 + (1 - x_2)y_1 - x_1y_1 \quad (10)$$

results to an overestimated bound  $[z_1]^{IA} = [0.28, 0.70]$ . Due to such dependency, the interval bounds often yield overestimation of the tightest enclosure. Similar, dependency effect is exhibited while computing  $[z_2]^{IA}$  for constraints  $z_2 = z_1 + \left(\frac{1}{x_1y_2} + \frac{1}{x_1^2}\right)^{-1} + \frac{y_1^2x_2}{y_1+x_2}$  and  $z_2 = z_1 + \frac{x_1^2y_2}{x_1+y_2} + \left(\frac{1}{y_1x_2} + \frac{1}{y_1^2}\right)^{-1}$  with respect to (4). As such, identification of constraint yielding tightest enclosure is cumbersome. In this regard, the problem formulation for reduction of dependency effect has been carried out in the next section.

1) *Problem Formulation:* The main aim in the present work is to compute tightest enclosure  $[\underline{z}_i, \bar{z}_i]$  or exact enclosure such that  $[z_i]^{MC} \sim [z_i]^{IA}$  or

$$\begin{cases} \underline{z}_i^{IA} = \underline{z}_i = \underline{z}_i^{MC}, \\ \bar{z}_i^{MC} = \bar{z}_i = \bar{z}_i^{IA}, \end{cases} \quad (11)$$

associated with some nonlinear constraint  $z_i = f(x_1, x_2, y_1, y_2)$ , where  $x_i \in [x_i]$  and  $y_i \in [y_i]$  for  $i = 1, 2$ . In this regard, SIVIA Monte-Carlo approach based on set inversion via interval computations and Monte-Carlo simulation have been proposed to estimate exact bounds in next section.

### III. SIVIA MONTE-CARLO APPROACH

Initially, the general procedure of SIVIA has been incorporated in Section III-A followed by contractors in Section III-B. Finally, the combination of SIVIA with Monte-Carlo approach (i.e SIVIA-MC) has been performed in Section III-C.

#### A. SIVIA

Set inversion of a typical set  $X \subset \mathbb{R}^m$  with respect to function  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$  is expressed as

$$X = f^{-1}(Y) = \{x \in \mathbb{R}^m \mid f(x) \in Y\},$$

where,  $Y \subset \mathbb{R}^n$ . In case of SIVIA [10], an initial search set  $[x_0]$  is assumed containing the required set  $X$ . Then, using sub-pavings as given in Fig. 2, the desired enclosure of solution set  $X$  is obtained based on the inclusion properties:

- 1) *Case I:*  $[f]([x]) \subset Y \implies [x] \subset X$ , then  $[x]$  is a solution,
- 2) *Case II:*  $[f]([x]) \cap Y = \phi \implies [x] \cap X = \phi$ , then  $[x]$  is not a solution,
- 3) *Case III:*  $[f]([x]) \cap Y \neq \phi$  and  $[f]([x]) \not\subset Y$  then,  $[x]$  is an undetermined solution.

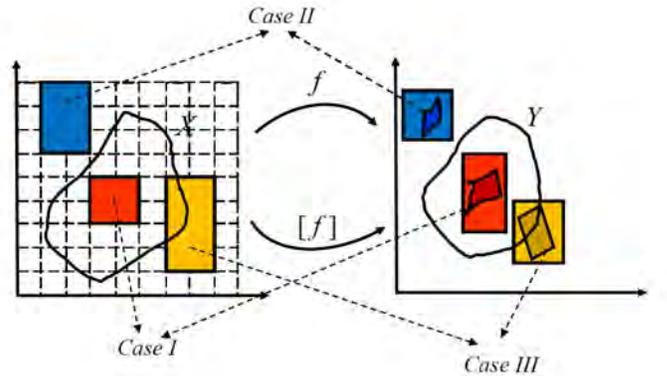


Figure 2. Set inversion via interval analysis.

The detailed illustration of set computation using SIVIA based on regular sub-pavings, bisections etc. may be found in [10]. The sub-pavings in SIVIA may be improved with the usage of contractors discussed in next section.

### B. Contractor

**Contractor:** ([4], [10]) A contractor  $\mathcal{C}$  associated with a set  $\mathbb{X} \subset \mathbb{R}^n$  over domain  $\mathbb{D}$  is an operator

$$\mathcal{C} : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

satisfying the following properties:

- **Contraction:**  $\mathcal{C}([x]) \subset [x], \forall [x] \in \mathbb{R}^n$ ,
- **Completeness:**  $\mathcal{C}([x]) \cap \mathbb{X} = [x] \cap \mathbb{X}, \forall [x] \in \mathbb{R}^n$ .

The pictorial representation of implementation of contractor over set  $\mathbb{X} \subset \mathbb{R}^2$  is illustrated in Fig. 3

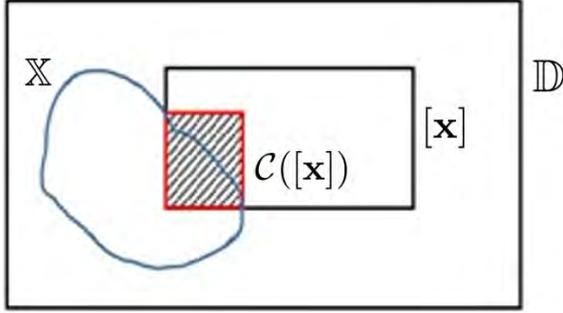


Figure 3. Contraction of  $[x]$ .

There exist various types of contractors viz. fixed-point, forward-backward, Newton, Gauss-Seidel contractors etc. Contractor based set inversion of lemniscate curve  $(x^2 + y^2)^2 + a^2(x^2 - y^2) = 0$  having width  $a \in [2, 3]$  has been obtained based on the PyIbex library [6] and depicted in Fig. 4, where the initial search set is  $[-4, 4] \times [-4, 4]$ .

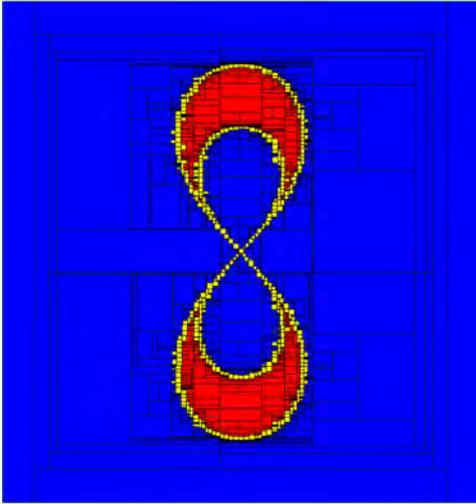


Figure 4. SIVIA of lemniscate curve with width  $[2, 3]$ .

In order to perform SIVIA Monte-Carlo approach, we have used forward-backward and fixed-point contractors. Detailed implementation of forward-backward and fixed-point contractors have been incorporated in Appendix.

### C. SIVIA Monte-Carlo

SIVIA Monte-Carlo (or SIVIA-MC) is two form iterative methodology that includes implementation of SIVIA using contractor programming and the Monte-Carlo simulation till the exact enclosure is obtained satisfying (11). In this regard, the iterative procedure is incorporated in Algorithm 1 with respect to constraint  $z = f(x_1, x_2, \dots, x_n)$  such that each  $x_i \in [x_i] \in \mathbb{R}$  for  $i = 1, 2, \dots, n$ . Here, the initial search set containing the exact enclosure is assumed as  $[z_0]$ .

---

#### Algorithm 1: Implementation of SIVIA-MC approach

---

Input:  $[x_i]$  for  $i = 1, 2, \dots, n$ ; Initial domain  $\mathbb{X} = [x_1], [x_2], \dots, [x_n]$ ;

Initial search set  $[z_0]$

Step 1: Compute enclosure using Monte-Carlo

$$\underline{z}^{\text{MC}} = mcl(\mathbb{X}) \text{ and } \bar{z}^{\text{MC}} = mcu(\mathbb{X})$$

Step 2: Compute enclosure using contractors

$$\underline{z}^{\text{IA}} = Ctcl(\mathbb{X}, [z_0]) \text{ and } \bar{z}^{\text{IA}} = Ctcu(\mathbb{X}, [z_0])$$

Step 3: Improve lower and upper range of  $z$

$$\underline{z} \in [\underline{z}^{\text{IA}}, \underline{z}^{\text{MC}}] \text{ and } \bar{z} \in [\bar{z}^{\text{MC}}, \bar{z}^{\text{IA}}]$$

Step 4: Compute improved lower  $\underline{\mathbb{X}} = [f]^{-1}([\underline{z}^{\text{IA}}, \underline{z}^{\text{MC}}])$  and

upper  $\bar{\mathbb{X}} = [f]^{-1}([\bar{z}^{\text{MC}}, \bar{z}^{\text{IA}}])$  domains using SIVIA

$$\underline{\mathbb{X}}, \bar{\mathbb{X}} = \text{SIVIA}(\mathbb{X}, [f], [z_0], \epsilon)$$

Step 5: Repeat steps 1 to 3 for domains  $\underline{\mathbb{X}}$  and  $\bar{\mathbb{X}}$

Step 6: Repeat step 4 for different domains  $\underline{\mathbb{X}}$  and  $\bar{\mathbb{X}}$

Step 7: Iterate steps 4 and 5 till  $\underline{z} = \underline{z}^{\text{IA}} \sim \underline{z}^{\text{MC}}$  and  $\bar{z} = \bar{z}^{\text{MC}} \sim \bar{z}^{\text{IA}}$

Output:  $[\underline{z}, \bar{z}]$

---

In Algorithm 1,  $mcl(\cdot), mcu(\cdot)$  are functions that compute the minimum and maximum function value with respect to domain  $\mathbb{X}$ . Then,  $Ctcl(\cdot), Ctcu(\cdot)$  uses forward-backward contractor along with fixed-point contractor for computing interval enclosure based on classical IA. Further,  $\text{SIVIA}(\cdot)$  computes the set inversion for domain  $\mathbb{X}$  based on constraint function  $f$  with precision  $\epsilon$ .

Let us again consider the Examples 1 and 2 in order to compute the exact enclosure using SIVIA Monte-Carlo in Example 3.

**Example 3:** Compute the interval bounds for the constraints

$$z_1 = x_1 y_1 + x_1 y_3 + x_3 y_1 \text{ and } z_2 = z_1 + \frac{x_1^2 y_2}{x_1 + y_2} + \frac{y_1^2 x_2}{y_1 + x_2}$$

using SIVIA Monte-Carlo such that  $x_1 + x_2 + x_3 = 1$  and  $y_1 + y_2 + y_3 = 1$ . Again,  $x_1 \in [x_1] = [0.2, 0.3]$ ,  $x_2 \in [x_2] = [0.1, 0.2]$ ,  $y_1 \in [y_1] = [0.4, 0.6]$  and  $y_2 \in [y_2] = [0.2, 0.3]$ . Using Algorithm 1 for SIVIA precision  $\epsilon = 0.001$  and different sample values viz. 100000, 1000, 100, 10, the tightest enclosures with respect to constraints  $z_1 = x_1 y_1 + x_1 y_3 + x_3 y_1$  and  $z_2 = z_1 + \frac{x_1^2 y_2}{x_1 + y_2} + \frac{y_1^2 x_2}{y_1 + x_2}$  for different sample values are obtained and incorporated in Tables II and III respectively.

Table II  
 INTERVAL ENCLOSURE OF  $z_1$ 

SIVIA (0.001 precision) and Monte Carlo samples				
Iterations	100000 samples		1000 samples	
	$\underline{z}_1 \in$	$\bar{z}_1 \in$	$\underline{z}_1 \in$	$\bar{z}_1 \in$
1	[0.3796, 0.385]	[0.5935, 0.6007]	[0.3796, 0.3850]	[0.5935, 0.6007]
2	[0.3796, 0.3807]	[0.5993, 0.6007]	[0.3796, 0.3822]	[0.5971, 0.6007]
3	[0.3796, 0.3801]	[0.5999, 0.6006]	[0.3796, 0.3808]	[0.5987, 0.6008]
4	—	—	[0.3797, 0.3803]	[0.5995, 0.6007]
$[z_1]$	[0.38, 0.6]		[0.38, 0.6]	
Time (s)	5.1388		5.5936	
Iterations	100 samples		10 samples	
	$\underline{z}_1 \in$	$\bar{z}_1 \in$	$\underline{z}_1 \in$	$\bar{z}_1 \in$
1	[0.3796, 0.385]	[0.5935, 0.6007]	[0.3796, 0.385]	[0.5935, 0.6007]
2	[0.3796, 0.3833]	[0.5965, 0.6007]	[0.3796, 0.384]	[0.5945, 0.6007]
3	[0.3797, 0.3817]	[0.5982, 0.6007]	[0.3796, 0.3836]	[0.595, 0.6007]
4	[0.3797, 0.3814]	[0.5989, 0.6007]	[0.3797, 0.3829]	[0.5971, 0.6007]
5	[0.3797, 0.3808]	[0.5995, 0.6006]	[0.3797, 0.3811]	[0.5977, 0.6007]
6	[0.3797, 0.3805]	[0.5996, 0.6006]	[0.3797, 0.3808]	[0.5978, 0.6006]
7	—	—	[0.3797, 0.3805]	[0.5988, 0.6006]
$[z_1]$	[0.38, 0.6]		[0.38, 0.6]	
Time (s)	6.0616		9.511	

 Table III  
 INTERVAL ENCLOSURE OF  $z_2$ 

SIVIA (0.001 precision) and Monte Carlo samples								
Iterations	100000 samples		1000 samples		100 samples		10 samples	
	$\underline{z}_2 \in$	$\bar{z}_2 \in$						
1	[0.4565, 0.4617]	[0.6825, 0.6889]	[0.4565, 0.4617]	[0.6825, 0.6889]	[0.4565, 0.4617]	[0.6825, 0.6889]	[0.3796, 0.385]	[0.5935, 0.6007]
2	[0.4565, 0.4581]	[0.6869, 0.6887]	[0.4565, 0.4591]	[0.6859, 0.6887]	[0.4565, 0.4609]	[0.6826, 0.6887]	[0.4565, 0.4609]	[0.6854, 0.6887]
3	[0.4566, 0.4575]	[0.6872, 0.6886]	[0.4565, 0.4579]	[0.6864, 0.6887]	[0.4565, 0.4599]	[0.6858, 0.6887]	[0.4565, 0.4601]	[0.6828, 0.6886]
4	—	—	[0.4565, 0.4577]	[0.6867, 0.6887]	[0.4566, 0.4578]	[0.6863, 0.6885]	[0.4565, 0.4589]	[0.6836, 0.6886]
5	—	—	[0.4565, 0.4575]	[0.6869, 0.6887]	[0.4566, 0.4577]	[0.6864, 0.6891]	[0.4565, 0.4584]	[0.6839, 0.6885]
6	—	—	[0.4565, 0.4574]	[0.687, 0.6889]	[0.4566, 0.4576]	[0.6866, 0.689]	[0.4565, 0.4581]	[0.685, 0.6885]
7	—	—	—	—	[0.4566, 0.4574]	[0.6867, 0.689]	[0.4565, 0.4577]	[0.6856, 0.6885]
8	—	—	—	—	—	—	[0.4565, 0.4575]	[0.686, 0.6885]
9	—	—	—	—	—	—	[0.4566, 0.4574]	[0.6865, 0.6884]
$[z_2]$	[0.46, 0.69]		[0.46, 0.69]		[0.46, 0.69]		[0.46, 0.69]	
Time (s)	6.7797		7.5464		9.2454		24.847	

It may be observed from Table II that the SIVIA Monte-Carlo method iteratively converge to exact enclosure  $[0.38, 0.6]$  (up to two decimals) even for less sample values viz. 100 and 10 respectively. Also, it may be noted that the iterative enclosures converge to exact bound though the computational time increases from 5.1388 to 9.511 seconds for different samples ranging from 100000 to 10 respectively. From Table II, the proposed method seems appealing as even for less sample values the convergent or exact solution bound is achieved. Many practical application problems do not yield sufficient data and sometimes availability of large data are cost effective, in such cases the proposed method may be used to obtain exact enclosure and the increase in computational time may be neglected.

Similar observations of exact enclosure convergence may be found in Table III with respect to different sample values. Moreover, due to complexity of the constraint (4), the required computational time 24.847 seconds for  $[z_2]$  is comparatively higher than time 9.511 seconds required for  $[z_1]$ . Further, a nonlinear differential equations with respect to dynamic

problems has been considered in next section for verification and effectiveness of SIVIA Monte-Carlo approach.

#### IV. NONLINEAR OSCILLATOR

Sometimes, dynamic problems are governed by  $m\ddot{x} + c\dot{x} + kx = f(t)$  having nonlinear stiffness ( $k_1x + k_2x^2 + \dots$ ) which result to nonlinear differential equations (nonlinear oscillators). In case of uncertain nonlinear oscillators, the SIVIA Monte-Carlo method has been implemented using nonlinear equations obtained based on Runge-Kutta 4<sup>th</sup> order [5], [8]. As such, the enclosure obtained in present section yield a validated enclosure rather than the tightest bound. There exists several validated interval methods and solvers viz. DynIbex [16] and CAPD [1] libraries for obtaining validated bounds.

**Example 4:** Consider Van der Pol equation (crisp or precise case given in Akbari et al. [2]),

$$\ddot{x}(t) + 0.15(1 - x^2)\dot{x} + 1.44x = 0, \quad (12)$$

subject to uncertain initial conditions  $x(0) \in [0.1, 0.3]$  and  $\dot{x}(0) = 0$ .

The system of first-order differential equation corresponding to (12) is obtained as

$$\dot{u} = v = f_u(t, u, v), \quad \dot{v} = 0.15(u^2 - 1)v - 1.44u = f_v(t, u, v),$$

subject to initial conditions  $u(0) \in [0.1, 0.3]$  and  $v(0) = 0$ . Using Runge-Kutta fourth-order (RK4), the nonlinear constraints involved in computation of (12) are

$$u_{n+1} = u_n + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4), \quad (13)$$

$$v_{n+1} = v_n + \frac{h}{6} (l_1 + 2l_2 + 2l_3 + l_4), \quad (14)$$

where,

$$k_1 = hf_u(t_n, u_n, v_n), \quad l_1 = hf_v(t_n, u_n, v_n),$$

$$k_2 = hf_u\left(t_n + \frac{h}{2}, u_n + \frac{k_1}{2}, v_n + \frac{l_1}{2}\right),$$

$$l_2 = hf_v\left(t_n + \frac{h}{2}, u_n + \frac{k_1}{2}, v_n + \frac{l_1}{2}\right),$$

$$k_3 = hf_u\left(t_n + \frac{h}{2}, u_n + \frac{k_2}{2}, v_n + \frac{l_2}{2}\right),$$

$$l_3 = hf_v\left(t_n + \frac{h}{2}, u_n + \frac{k_2}{2}, v_n + \frac{l_2}{2}\right),$$

$$k_4 = hf_u(t_n + h, u_n + k_3, v_n + l_3),$$

$$l_4 = hf_v(t_n + h, u_n + k_3, v_n + l_3).$$

Using Algorithm 1 with respect to constraints (13) and (14), the validated enclosure of  $x(t)|_{t=T}$  is obtained and incorporated in Table IV and Fig. 5.

Table IV  
INSTANTANEOUS SOLUTION ENCLOSURE OF  $x(t)|_{t=T}$ .

T	Enclosures	
	$[x](T) = [u](T)$	$[v](T)$
0.1	[0.0992, 0.2978]	[-0.0428, -0.0143]
0.2	[0.0971, 0.2917]	[-0.0843, -0.0282]
0.3	[0.0898, 0.2828]	[-0.1242, -0.0414]

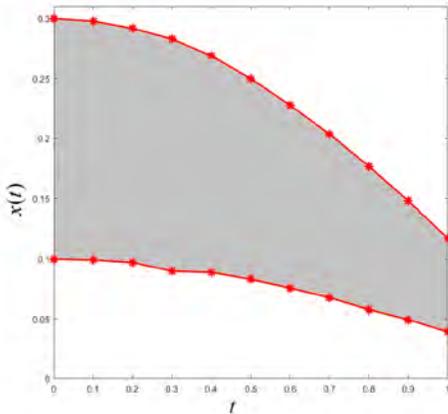


Figure 5. Enclosure of  $x(t)$  for  $t \in [0, 1]$ .

## V. CONCLUSION

Generally, dynamical systems occurring in various science and engineering problems are governed by nonlinear equations or nonlinear differential equations. An iterative procedure based on set inversion via interval analysis and Monte-Carlo method has been proposed for computation of exact enclosure of nonlinear equations having imprecise or uncertain variables. The effectiveness of SIVIA Monte-Carlo method has also been verified based on the considered test problems that yield exact enclosures even with respect to very less sample values. So, the method may be well implemented in computation of exact enclosures of various nonlinear equations irrespective of the dependency problem. Further, the method has also been implemented to compute validated enclosure in case of Van der Pol oscillator. Accordingly, the method may be applied to other practical nonlinear system of equations involving uncertain parameters.

## APPENDIX

**Forward-backward contractor:** The forward-backward contractor is based on constraint  $f(x) = 0$  where  $x \in [x]$  and  $[x] \in \mathbb{IR}^n$  which is illustrated using an example problem.

**Example A1:** Perform forward-backward contractor subject to constraint  $w = 2u + v$  where,  $[w] = [3, 20]$ ,  $[u] = [-10, 5]$  and  $[v] = [0, 4]$ .

Here, the constraint  $w = 2u + v$  may be expressed in terms of function  $f$  as  $f(u, v, w) = w - 2u - v$ . Further, the possible different forms of the constraint may be written are:

$$u = \frac{w - v}{2},$$

$$v = w - 2u,$$

$$w = 2u + v.$$

The forward-backward steps are then followed with respect to classical interval computations mentioned in Section II as:

$$[u] \cap \left( \frac{[w] - [v]}{2} \right) = [-10, 5] \cap \left( \frac{[3, 20] - [0, 4]}{2} \right) = [-0.5, 5],$$

$$[v] \cap ([w] - 2[u]) = [0, 4] \cap ([3, 20] - 2[-0.5, 5]) = [0, 4],$$

$$[w] \cap (2[u] + [v]) = [3, 20] \cap (2[-0.5, 5] + [0, 4]) = [3, 14].$$

As such, the new interval bounds are  $[z] = [3, 14]$ ,  $[x] = [-0.5, 5]$  and  $[y] = [0, 4]$ .

**Fixed-point contractor:** A fixed-point contraction associated with  $\psi$  is implemented with respect to the constraint  $f(x) = 0$  as  $x = \psi(x)$ , where  $x \in [x] \in \mathbb{IR}^n$ . The fixed-point contractor with respect to constraint  $u^2 + 2u + 1 = 0$  is performed as

$$\begin{aligned} u \in [u] \text{ and } u = \psi(u) &\implies u \in [u] \text{ and } u \in \psi([u]), \\ &\implies u \in [u] \cap [\psi]([u]). \end{aligned}$$

In case of implementation of forward-backward contractor along with fixed point contractor helps in computation of forward-backward contractor until the fixed interval is reached.

## REFERENCES

- [1] Computer Assisted Proofs in Dynamics group (CAPD), <http://capd.ii.uj.edu.pl/>
- [2] M. Akbari, D. Ganji, A. Majidian, A. Ahmadi, *Solving nonlinear differential equations of vanderpol, rayleigh and duffing by agm*, Frontiers of Mechanical Engineering, Vol. 9(2), pp. 177–190, 2014.
- [3] G. Alefeld, J. Herzberger, *Introduction to Interval Computation*. Academic Press, London, 2012.
- [4] G. Chabert, L. Jaulin, *Contractor programming*, Artificial Intelligence, Vol. 173, pp. 1079–1100, 2009.
- [5] S. Chakraverty, N.R. Mahato, P. Karunakar, T.D. Rao, *Advanced numerical and semi-analytical methods for differential equations*, Wiley Telecom, 256 pages, 2019.
- [6] B. Desrochers, *Pyibex*, <http://benensta.github.io/pyIbex/sphinx/index.html>
- [7] J. Dezert, D. Han, J.-M. Tacnet, *Multi-criteria decision-making with imprecise scores and BF-TOPSIS*, In Proc. of 20th International Conference on Information Fusion (Fusion 2017), pp. 1–8, Xi'an, China, July 2017.
- [8] C.F. Gerald, *Applied numerical analysis*, Pearson Education India, 2004.
- [9] E.R. Hansen, *A generalized interval arithmetic*, in Interval mathematics, pp. 7–18. Springer, 1975.
- [10] L. Jaulin, M. Kieffer, O. Didrit, E. Walter, *Applied interval analysis: with examples in parameter and state estimation, robust control and robotics*, Vol. 1, Springer-Verlag, London, 2001.
- [11] W. Krämer, *Generalized intervals and the dependency problem*, in PAMM: Proceedings in Applied Mathematics and Mechanics, Vol. 6, pp. 683–684, Wiley Online Library, 2006.
- [12] R.E. Moore, *Interval arithmetic and automatic error analysis in digital computing*, Ph. D. Dissertation, Department of Mathematics, Stanford University, Stanford, USA, 1962.
- [13] R.E. Moore, *Methods and applications of interval analysis*, Vol. 2., Siam, 1979.
- [14] R.E. Moore, R.B. Kearfott, M.J. Cloud, *Introduction to Interval Analysis*, SIAM Publications, Philadelphia, PA, USA, 2009.
- [15] J. Alexandre-dit-Sandretto, A. Chapoutot, O. Mullier, *DynIbex* <http://perso.ensta-paristech.fr/~chapoutot/dynibex/>
- [16] J. Stolfi, L.H. De Figueiredo, *An introduction to affine arithmetic*, Trends in Applied and Computational Mathematics, Vol. 4(3), pp. 297–312, 2003.



# Rough Set Classifier Based on DSMT

Yilin Dong<sup>a</sup>, Xinde Li<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE, School of Automation,  
Southeast University, Nanjing, China

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France  
Emails: dyl@seu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr

Originally published as: Y. Dong, X. Li, J. Dezert, *Rough Set Classifier Based on DSMT*, in Proc. of the 21th International Conference on Information Fusion (Fusion 2018), Cambridge, UK, July 10-13, 2018, and reprinted with permission.

**Abstract**—The classifier based on rough sets is widely used in pattern recognition. However, in the implementation of rough set-based classifiers, there always exist the problems of uncertainty. Generally, information decision table in Rough Set Theory (RST) always contains many attributes, and the classification performance of each attribute is different. It is necessary to determine which attribute needs to be used according to the specific problem. In RST, such problem is regarded as attribute reduction problems which aims to select proper candidates. Therefore, the uncertainty problem occurs for the classification caused by the choice of attributes. In addition, the voting strategy is usually adopted to determine the category of target concept in the final decision making. However, some classes of targets cannot be determined when multiple categories cannot be easily distinguished (for example, the number of votes of different classes is the same). Thus, the uncertainty occurs for the classification caused by the choice of classes. In this paper, we use the theory of belief functions to solve two above mentioned uncertainties in rough set classification and rough set classifier based on Dezert-Smarandache Theory (DSMT) is proposed. It can be experimentally verified that our proposed approach can deal efficiently with the uncertainty in rough set classifiers.

**Keywords:** Classification, rough set, uncertainty, evidence reasoning, DSMT, belief functions..

## I. INTRODUCTION

*a) Motivation:* In recent years, we have witnessed the rapid development of Rough Set Theory (RST) [1]. There are many practical applications of this theory [2], [3], [4], [5]. Among these, Rough Set Classifier (RSC) has been widely used in the real classification problems [6], [7], [8], [9].

*b) Challenges:* However, in the practical use of RSC, there always exists uncertainty. In the literature [10] and [11], the discussions of the uncertainty in RST mainly focus on the following points of view: Chen [10] proposed several uncertainty measures of neighborhood granules, which are neighborhood accuracy, information quantity, neighborhood entropy and information granularity in the neighborhood RST; Zheng [11] estimated the uncertainty of rough set originated from two parts of boundary region. Although the uncertainties discussed in the above literature are of certain significance, however, the uncertainties discussed in this paper are shown in two aspects:

1) The choice of attributes: for example, in the decision information table, some attributes are not significant in a representation and deleting of these attributes has no real impact on the classification results. However, such concept of significancy is relative, for different problems, the role of

each attribute is quite different. Thus, the problems of attribute selection are always ad hoc and depending on the user's preference. Obviously, different attribute selections correspond to different strategies, which generally yield different results. For example, in [12], authors attempted to select the most information-rich attributes from a dataset by incorporating a controlled degree of misclassification into approximations of rough sets. Gao et.al [13] proposed a new uncertainty measure, named maximum decision entropy, for attribute reduction in the decision-theoretic rough set model. Although many robust and efficient reduction algorithms have been proposed, most of them concentrate on the properties of data or user preference in the definition of attribute reduction, which result in the difficulties of choosing appropriate attribute reductions for specific applications. For the same data, different users can define different reductions and obtain their interested results according to their applications. Jia et.al [14] reviewed nearly twenty two different attribute reduction methods, but to design of a robust attribute reduction method is not the focus of this paper. We emphasize the uncertainty caused by the choice of attributes, which is not discussed in details in the recent development of RST. For this aim, one typically seeks a policy for avoiding choosing attributes, and we propose to emphasize the importance of each attribute for the specific problems.

2) The choice of classes: besides, in RST, the category of target concept is determined according to the element composition of its corresponding approximate set: if the number of elements belonging to one class is the largest, the concept of target is labelled as this class. However, this kind of voting method often leads to uncertainty in making decisions, which affects the final precision of RSC. In order to illustrate this problem more vividly, we explain it through Figure 1: in case one, the approximate set of target concept (**red five-pointed star**) has four elements (**plus**) belonging to class 1, three elements (**plus**) belonging to class 2 and two elements (**plus**) belonging to class 3. Thus, in case one, we can easily draw the conclusion that the target belongs to class 1. However, in case two or case three, the target cannot be labelled with single category because there are some classes (class 1 and class 2 in case two, class 1, class 2 and class 3 in case three) that have the same number of votes. More specifically, if the approximate set of such target is empty-set (case four), which category should be allocated to the target concept? As aforementioned, for the RSCs, there are two mentioned neglected uncertainty issues. The theory of belief functions

[15] is widely used in uncertainty management and uncertainty reasoning for decision-making. In this paper, we attempt to use it to model and manage the uncertainty incorporated in RSCs.

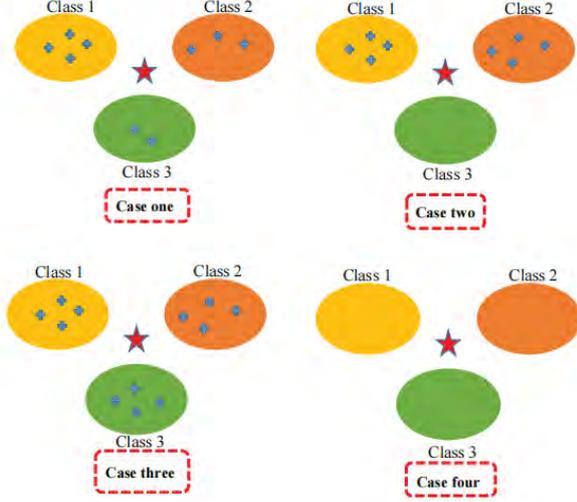


Figure 1: Uncertainty in Voting Strategy.

*c) Contributions:* Because a certain attribute does not have the ability to distinguish items on a particular problem, but there may be a discriminative performance on another problem. Thus, according to the classification performance of each attribute, the corresponding weights of all attributes in information decision table are calculated, which are used as the evaluation index of the importance of an attribute. At the same time, we do not directly delete unimportant attributes which the classical reduction algorithms have done. We just consider all the attributes in the final classification, after all, we consider that all existing attributes must play a role in the decision. For the uncertainty of the voting strategy in traditional RSC, we have no statistics of the number of votes of each class in approximate sets. Instead, we first calculate the coordinate of each class with respect to each attribute and then get the distance between the target concept and each class in every attribute, in order to calculate the Basic Belief Assignment (BBA) of the target in each attribute. Then, we use the classical combination rule (PCR5 is used in this paper) proposed in DSMT [16] to sequentially<sup>1</sup> combine all BBAs (each attribute has a corresponding BBA). Finally, according to the principle of maximum belief mass, we can obtain the final class of the target concept.

This paper is organized as follows. Section II reviews some basic concepts of Dempster-Shafer Theory (DST), and DSMT. The new rough set classifier based on DSMT (RSCD) is proposed in section III. Section IV gives the summary of the

<sup>1</sup>Because PCR5 rule is not associative, which means that the fusion results depend on the order you have chosen. Here, our default way of combination is to combine BBAs in order from small to large. For example, if there are three BBAs:  $m_1, m_2, m_3$ , the way of fusion is  $m_1 = PCR5(m_1, m_2) \rightarrow m_{123} = PCR5(m_{12}, m_3) \rightarrow m_{fusion} = m_{123}$ .

proposed classifier. In section V, we give some experimental results to show the performances of our new method. Also, some meaningful discussions about the extension of RSCD are given in section VI. Section VII concludes the paper with a summary and direction for future.

## II. PRELIMINARIES

This section provides a brief reminder of the basics of DST and DSMT, which is necessary for the presentation and understanding of the more general fusion of evidence.

In DST framework, the Frame of Discernment (FoD)<sup>2</sup>  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$  ( $n \geq 2$ ) is a set of exhaustive and exclusive elements (hypotheses) which represent the possible solutions of the problem under consideration and thus Shafer's model assumes  $\theta_i \cap \theta_j = \emptyset$  for  $i \neq j$  in  $\{1, \dots, n\}$ . A BBA  $m(\cdot)$  is defined by the mapping:  $2^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . In DSMT, one can abandon Shafer's model (if Shafer's model doesn't fit with the problem) and refute the principle of the third excluded middle. The third excluded middle principle assumes the existence of the complement for any elements/propositions belonging to the power set  $2^\Theta$ . Instead of defining the BBAs on the power set  $2^\Theta \triangleq (\Theta, \cup)$  of the FoD, the BBAs are defined on the so-called *hyper-power set* (or Dedekind's lattice) denoted  $D^\Theta \triangleq (\Theta, \cup, \cap)$  whose cardinalities follows Dedekind's numbers sequence, see [17], Vol.1 for details and examples. A (generalized) BBA, called a mass function,  $m(\cdot)$  is defined by the mapping:  $D^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in D^\Theta} m(A) = 1$ . The DSMT framework encompasses DST framework because  $2^\Theta \subset D^\Theta$ . In DSMT, we can take into account also a set of integrity constraints on the FoD (if known), by specifying all the pairs of elements which are really disjoint. Stated otherwise, Shafer's model is a specific DSMT model where all elements are deemed to be disjoint.  $A \in D^\Theta$  is called a focal element of  $m(\cdot)$  if  $m(A) > 0$ . A BBA is called a Bayesian BBA if all of its focal elements are singletons and Shafer's model is assumed, otherwise it is called non-Bayesian [18]. A full ignorance source is represented by the vacuous BBA  $m_v(\Theta) = 1$ . The belief (or credibility) and plausibility functions are respectively defined by  $Bel(X) \triangleq \sum_{Y \in D^\Theta | Y \subseteq X} m(Y)$  and  $Pl(X) \triangleq \sum_{Y \in D^\Theta | Y \cap X \neq \emptyset} m(Y)$ .  $BI(X) \triangleq [Bel(X), Pl(X)]$  is called the belief interval of  $X$ . Its length  $U(X) \triangleq Pl(X) - Bel(X)$  measures the degree of uncertainty of  $X$ .

In 1976, Shafer did propose Dempster's rule and we use DS index to refer to Dempster-Shafer's rule (DS rule) because Shafer did really promote Dempster's rule in his milestone book [18]) to combine BBAs in DST framework. DS rule for combining two distinct sources of evidence characterized by BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  is defined by  $m_{DS}(\emptyset) = 0$  and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ :

$$m_{DS}(A) = \frac{\sum_{B, C \in 2^\Theta | B \cap C = A} m_1(B)m_2(C)}{1 - \sum_{B, C \in 2^\Theta | B \cap C = \emptyset} m_1(B)m_2(C)}. \quad (1)$$

<sup>2</sup>Here, we use the symbol  $\triangleq$  to mean equals by definition.

The DS rule formula is commutative and associative and can be easily extended to the fusion of  $S > 2$  BBAs. Unfortunately, DS rule has been highly disputed during the last decades by many authors because of its counter-intuitive behavior in high or even low conflict situations, and that is why many rules of combination were proposed in literature to combine BBAs [19]. To palliate DS rule drawbacks, the very interesting PCR5 was proposed in DSMT and it is usually adopted<sup>3</sup> in recent applications of DSMT. The fusion of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  by the PCR5 rule is obtained by  $m_{PCR5}(\emptyset) = 0$  and  $\forall A \in D^\ominus \setminus \{\emptyset\}$

$$m_{PCR5}(A) = m_{12}(A) + \sum_{B \in D^\ominus \setminus \{A\} | A \cap B = \emptyset} \left[ \frac{m_1^2(A)m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2^2(A)m_1(B)}{m_2(A) + m_1(B)} \right], \quad (2)$$

where  $m_{12}(A) = \sum_{B,C \in D^\ominus | B \cap C = A} m_1(B)m_2(C)$  is the conjunctive operator, and each element  $A$  and  $B$  are expressed in their disjunctive normal form. If the denominator involved in the fraction is zero, then this fraction is discarded. The general PCR5 formula for combining more than two BBAs altogether is given in [17], Vol. 3. We adopt the generic notation  $m_{12}^{PCR5}(\cdot) = PCR5(m_1(\cdot), m_2(\cdot))$  to denote the fusion of  $m_1(\cdot)$  and  $m_2(\cdot)$  by PCR5 rule. PCR5 is not associative and PCR5 rule can also be applied in DST framework (with Shafer's model of FoD) by replacing  $D^\ominus$  by  $2^\ominus$  in Eq (2).

### III. NEW ROUGH SET CLASSIFIER BASED ON DSMT (RSCD)

#### A. Weights of each attribute

RST is a mathematical tool to deal with vagueness and uncertainty [1], which can effectively analyse the incomplete information and does not need additional data beyond the prior information. Next, we briefly give several relevant definitions to show how to calculate the weights of attributes:

**Definition 1:** An information decision system  $\mathbf{S}$  is  $\mathbf{S} = (U, A, D)$ , where  $U = \{x_1, x_2, \dots, x_n\}$  is non-empty finite set of samples,  $A = \{a_1, a_2, \dots, a_m\}$  is a non-empty finite set of attributes,  $D$  is a non-empty set of finite decision classes.

**Definition 2:** Each attribute  $a \in A$  defines an information function  $f_a : U \rightarrow V_a$ , and  $V_a$  is the set value of the attribute  $a$ . We further extend these notations for a set of attributes  $B \subseteq A$ , an indiscernibility relation  $Ind(B)$  can be defined as follows:

$$Ind(B) = \{(x_i, x_j) \in U^2 \mid f_i(a) = f_j(a), \forall a \in B\}, \quad (3)$$

where  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are indiscernible when  $(\mathbf{x}_i, \mathbf{x}_j) \in Ind(B)$ . Some equivalence classes or elementary sets are generated by  $Ind(B)$ . The elementary set of  $\mathbf{x}_i$  is represented by  $[\mathbf{x}_i]_B$ . Any finite union of elementary sets is called a B-definable set. For pattern classification, elements have the same class

<sup>3</sup>Recently, a new combination rule PCR6 was proposed to combine all the BBAs altogether in a single fusion step, which can be found in [20]. Because PCR6 rule coincides with PCR5 when combining only two BBAs [17], we just use PCR5 rule to combine BBAs in this paper.

label consisting of a concept  $X$  so that  $X \in U/D$ , where  $U/D = \{[\mathbf{x}_i]_D \mid \mathbf{x}_i \in U\}$  and  $[\mathbf{x}_i]_D$  represents the elementary sets of  $\mathbf{x}_i$  with respect to decision attribute  $D$ . Sometimes  $X \subseteq U$  is not B-definable. In other words, there exists elements that are in the same elementary set, but have different class labels, so that  $X$  becomes a vague concept. For this, we give the following definitions of approximation sets of such vague concept:

**Definition 3:** The B-upper approximation  $\overline{B}X$  and the B-lower approximation  $\underline{B}X$  of the vague concept  $X$  is defined as follows:

$$\underline{B}X = \{x_i \in U \mid [x_i]_B \subseteq X\}, \quad (4)$$

$$\overline{B}X = \{x_i \in U \mid [x_i]_B \cap X \neq \emptyset\}. \quad (5)$$

$\underline{B}X \subseteq \overline{B}X$ , and  $\underline{B}X$  consists of elements that certainly belong to  $X$ , whereas  $\overline{B}$  consists of elements that possibly belong to  $X$ . The set  $BN_B(X) = \overline{B}X - \underline{B}X$  is called the B-boundary region of  $X$ , and thus consists of those objects that we cannot decisively classify into  $X$  on the basis of knowledge in  $B$ .

**Definition 4:**  $POS_B(D)$  is a positive region of the partition  $U/D$  with respect to  $B$  and is defined as follows:

$$POS_B(D) = \bigcup_{X \in U/D} \underline{B}X \quad (6)$$

$$= \bigcup \{Y \mid Y \subseteq X, Y \subseteq U/B, X \in U/D\}. \quad (7)$$

**Definition 5:** The degree of support of the condition attributes  $B$  with respect to the decision attribute  $D$  is defined as follows:

$$\zeta_D^B = \frac{|POS_B(D)|}{|U|}. \quad (8)$$

Here,  $\zeta$  is regarded as the degree of importance of each attribute in the information decision table  $\mathbf{S}$ . In order to illustrate how to calculate the weight of a particular attribute based on the aforementioned five definitions, we give a simple example below:

**Example 1:** Table I is an information decision table with  $U = \{x_1, x_2, \dots, x_{12}\}$ ,  $A = \{a_1, a_2, a_3, a_4\}$ ,  $D = \{d_1 = 1, d_2 = 2, d_3 = 3\}$ . According to the decision attribute  $d$  and Eq.(3), if  $\mathbf{x}_i$  is set to  $U$  and  $B$  is equal to  $d$ , we can get the  $[\mathbf{x}_i]_B = [U]_d = U/D = \{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$ . Meanwhile, we can also partition  $U$  by using each attribute  $a_i, i = 1, \dots, m$  based on the indiscernibility relation  $Ind(B)$ , which are illustrated in Table II.

Thus, each element  $X$  in  $[U]_d$  can be approximated by each condition attribute  $a_i, i = 1, \dots, m$ , and then we can obtain  $\underline{a_i}X$  in Table III according to Definition 3. Based on Eq.(7), we can get the positive domain of  $D$  with respect to each attribute  $a_i$ , which is also given in Table IV.

In order to explain how positive domains are calculated in detail, we take  $POS_{a_1}(D)$  as an example:  $U/D = [U]_D = \{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$ ,  $U/a_1 = [U]_{a_1} = \{\{x_1, x_4\}, \{x_2\}, \{x_3\}, \{x_5\}, \{x_6\}, \{x_7\},$

Table I: Information decision table.

U	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$
$x_1$	5.1	3.5	1.4	0.2	1
$x_2$	6.6	2.9	4.6	1.3	2
$x_3$	5.2	2.7	3.9	1.4	2
$x_4$	5.1	3.8	1.5	0.3	1
$x_5$	6.4	2.7	5.3	1.9	3
$x_6$	6.8	3.0	5.5	2.1	3
$x_7$	5.5	4.2	1.4	0.2	1
$x_8$	5.0	3.3	1.4	0.2	1
$x_9$	5.0	2.0	3.5	1.0	2
$x_{10}$	5.9	3.0	4.2	1.5	2
$x_{11}$	5.7	2.6	3.5	1.0	2
$x_{12}$	4.6	3.6	1.0	0.2	1

 Table II: Results of partitioning the domain  $U$  using each attribute.

	The partitioning domain
$[U]_{a_1}$	$\{\{x_1, x_4\}, \{x_2\}, \{x_3\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8, x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}\}$
$[U]_{a_2}$	$\{\{x_1\}, \{x_2\}, \{x_3, x_5\}, \{x_4\}, \{x_6, x_{10}\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{11}\}, \{x_{12}\}\}$
$[U]_{a_3}$	$\{\{x_1, x_7, x_8\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_9, x_{11}\}, \{x_{10}\}, \{x_{12}\}\}$
$[U]_{a_4}$	$\{\{x_1, x_7, x_8, x_{12}\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}, \{x_6\}, \{x_9, x_{11}\}, \{x_{10}\}\}$

 Table III: The lower approximation of elements in  $[U]_d$  using each attribute.

$\underline{B}X$	The B-lower approximation
$a_1\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1, x_4\}, \{x_7\}, \{x_{12}\}\}$
$a_1\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_3\}, \{x_{10}\}, \{x_{11}\}\}$
$a_1\{x_5, x_6\}$	$\{\{x_5\}, \{x_6\}\}$
$a_2\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1\}, \{x_4\}, \{x_7\}, \{x_8\}, \{x_{12}\}\}$
$a_2\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_9\}, \{x_{11}\}\}$
$a_2\{x_5, x_6\}$	$\emptyset$
$a_3\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1, x_7, x_8\}, \{x_4\}, \{x_{12}\}\}$
$a_3\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_3\}, \{x_9, x_{11}\}, \{x_{10}\}\}$
$a_3\{x_5, x_6\}$	$\{\{x_5\}, \{x_6\}\}$
$a_4\{x_1, x_4, x_7, x_8, x_{12}\}$	$\{\{x_1, x_7, x_8, x_{12}\}, \{x_4\}\}$
$a_4\{x_2, x_3, x_9, x_{10}, x_{11}\}$	$\{\{x_2\}, \{x_3\}, \{x_9, x_{11}\}, \{x_{10}\}\}$
$a_4\{x_5, x_6\}$	$\{\{x_5\}, \{x_6\}\}$

 Table IV: The positive domain of  $[U]_d$  with respect to each attribute and weights of each attribute according to Eq.(8).

Attribute	Domain	$\zeta$
$POS_{a_1}(D)$	$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_{10}, x_{11}, x_{12}\}$	$\frac{10}{12}$
$POS_{a_2}(D)$	$\{x_1, x_2, x_4, x_7, x_8, x_9, x_{11}, x_{12}\}$	$\frac{10}{12}$
$POS_{a_3}(D)$	$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$	$\frac{10}{12}$
$POS_{a_4}(D)$	$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$	$\frac{10}{12}$

$\{x_8, x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}\}$ , for any elements  $Y$ , where  $Y \in U/a_1$ , if  $Y$  meets the condition:  $Y \subseteq X$ , where  $X \in U/D$ , then  $Y$  belongs to the domain  $POS_{a_1}(D)$ , for

example, when  $Y = \{x_1, x_4\}$  and  $X = \{x_1, x_4, x_7, x_8, x_{12}\}$ , it satisfies  $Y \subseteq X$ , so  $\{x_1, x_4\}$  belongs to  $POS_{a_1}(D)$ . However, if  $Y = \{x_8, x_9\}$ ,  $Y$  is not a subset of any elements in  $U/D$ , so  $\{x_8, x_9\}$  does not belong to  $POS_{a_1}(D)$ . Thus, according to Eq.(8), we can obtain the degree of support of  $a_i$  with respect to the decision attribute  $D$  in Table IV, which will be regarded as the weights of each attribute in the classification problem.

### B. Construction of BBA of Target Concept

As discussed in the introduction section, the traditional way of voting decision will cause uncertainty when using RSC, and directly affect the final classification accuracy. The evidence theory has a good ability to deal with the uncertainty problem, and evidence theory generally describes such concept of uncertainty through BBAs. However, the BBAs in evidence theory are always given by experts depending on their own experience, which cannot be obtained directly in practical problems. Thus, this requires that, when solving such problems, the corresponding BBAs are first constructed and calculated before using them to make decisions. Referring to the construction methods of BBAs in [21], [22], [23], we propose in this paper a new construction method for the BBA based on so-called attribute polygon in RST. Each polygon represents an attribute and each vertex in a polygon represents one category. That is to say, if it is a two-classification problem, the attribute polygon is the line segment; Similarly, if it is the three-classification problem, such polygon is the triangle, and so on. Figure 2 illustrates the corresponding four polygons which represent for two, three, four and five classification problems.

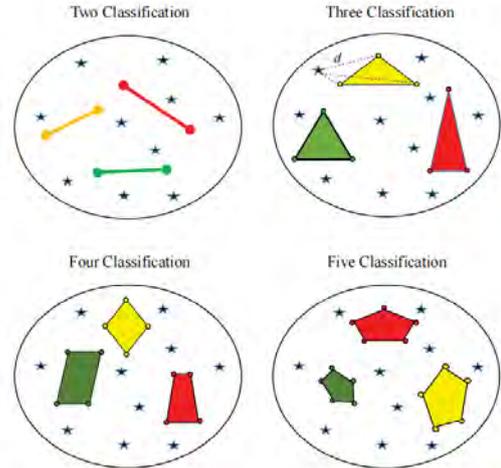


Figure 2: Attribute polygon. Pentagram represents the test example and for example, in three classification, the distances (dotted line) are calculated between the value of one attribute of pentagram and the vertices of one attribute triangle.

Besides, the coordinates of all vertices in all attribute polygons are calculated according to  $[U]_d = \{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$ .

Then, the Euclidean distance is used to calculate the distance between test example and each attribute polygon. Finally, we can get the belief mass value of this example belonging to each class with respect to one attribute by using Eq.(9) and Eq.(10).

$$m_{a_i}^{x^*}(\theta_s) = \alpha e^{\gamma_s d^\beta}, \quad (9)$$

$$m_{a_i}^{x^*}(\Theta) = 1 - \alpha e^{\gamma_s d^\beta}, \quad (10)$$

where  $\alpha$ ,  $\gamma_s$  and  $\beta$  are turning parameters and according to the recommendations given in [24], these parameters are set to  $\alpha = 0.95$ ,  $\gamma_s = -2$  and  $\beta = 1$ . Besides,  $d$  is the distance between the vertices of  $a_i$  attribute polygon and each attribute value of text example  $x^*$ . Next, we will show how to calculate BBAs through Example 1.

### Example 1 revisited:

According to the decision attribute  $d$  in Table I, we know that this simple example is a three-class problem because  $D = \{d_1, d_2, d_3\}$ , so we need to construct the triangles. Because the decision table has four condition attributes, we need to construct four triangles. In order to show how to calculate the coordinates of vertices in each attribute triangle, we give the calculation steps as follows: Based on the partitions of the decision attribute  $d : \{\{x_1, x_4, x_7, x_8, x_{12}\}, \{x_2, x_3, x_9, x_{10}, x_{11}\}, \{x_5, x_6\}\}$ , we can obtain the coordinates of each category with respect to attribute  $a_1$ :

- the coordinate of class one with respect to  $a_1$ :

$$\frac{1}{|X(a_1)|} \sum_{x \in X(a_1)} f(x, a_1) = 5.06,$$

where  $X(a_1) = \{x_1, x_4, x_7, x_8, x_{12}\}$  and  $|\cdot|$  denotes the cardinality;

- the coordinate of class two with respect to  $a_1$ :

$$\frac{1}{|X(a_1)|} \sum_{x \in X(a_1)} f(x, a_1) = 5.6800,$$

where  $X(a_1) = \{x_2, x_3, x_9, x_{10}, x_{11}\}$ ;

- the coordinate of class three with respect to  $a_1$ :

$$\frac{1}{|X(a_1)|} \sum_{x \in X(a_1)} f(x, a_1) = 6.6000,$$

where  $X(a_1) = \{x_5, x_6\}$ .

Here,  $f(x_i, a_j)$  is the value of the cell of the Table I corresponding to value  $x_i$  and attribute  $a_j$ .

Table V: All coordinates of three classes in each attribute.

Attribute	Class 1	Class 2	Class 3
$a_1$	5.0600	5.6800	6.600
$a_2$	3.6800	2.6400	2.8500
$a_3$	1.3400	3.9400	5.4000
$a_4$	0.2200	1.2400	2.0000

Similarly, we can calculate all the coordinates of three classes of four attributes, which is given in Table V as follows. Then, we randomly select a test example, which is denoted as  $x^* = \{5.1000, 3.5000, 1.4000, 0.2000\}$ . Based on the Euclidean distance<sup>4</sup>, the corresponding distances between  $x^*$  and each attribute polygon is given in Table VI.

Table VI: Distances between target  $x^*$  and all vertices of attribute polygons.

Distance	Class 1	Class 2	Class 3
$a_1 \leftrightarrow x^*$	0.0400	0.5800	1.5000
$a_2 \leftrightarrow x^*$	0.1800	0.8600	0.6500
$a_3 \leftrightarrow x^*$	0.0600	2.5400	4.0000
$a_4 \leftrightarrow x^*$	0.0200	1.0400	1.8000

Based on Eq.(9) and Eq.(10), we can transform these values of distances into belief mass so as to obtain the BBAs of each attribute, which is given in Table VII.

Table VII: BBAs of  $x^*$  with respect to each attribute.

$m(\cdot)$	Class 1	Class 2	Class 3	$\Theta$
$m_1(\cdot)$	0.7778	0.0523	0.0005	0.1694
$m_2(\cdot)$	0.3862	0.0129	0.0368	0.5641
$m_3(\cdot)$	0.7038	0.0000	0.0000	0.2962
$m_4(\cdot)$	0.8596	0.0052	0.0043	0.1309

Finally, we use PCR5 formula Eq.(2) to combine the weight of each attribute and the BBAs of each attribute so as to obtain the final BBA of  $x^*$ <sup>5</sup>. According to the fusion result, we can draw a conclusion that  $x^*$  belong to class 1 based on maximum of belief mass principle, which is consistent with the label of  $x^*$  in the original dataset.

$$m_{fusion}(\theta_1) = \mathbf{0.8827}; m_{fusion}(\theta_2) = 0.0009;$$

$$m_{fusion}(\theta_3) = 0.0007; m_{fusion}(\Theta) = 0.1157;$$

## IV. THE SUMMARY OF RSCD

On the next page, we give a brief pseudo-code of RSCD in Algorithm 1. Because RSCD in this paper is a data-driven model, so, first of all, we need to divide original dataset into training datasets and test samples (the experiments in this paper are using ten-fold cross validation). Afterwards, the training datasets are applied to construct attribute polygons and calculate the weights of attributes. Finally, we can obtain the corresponding BBAs of each test samples by calculating the distances between test examples and attribute polygons.

## V. SIMULATIONS

We have tested the different classifiers on real datasets given in the machine learning repository of the University of California Irvine (UCI) [25] and listed in Table VIII.

<sup>4</sup>The Euclidean distance  $d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}$  is used here.

<sup>5</sup>In the final BBA, for the sake of convenience,  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\Theta$  represent class 1, class 2, class 3 and unknown; And  $m_{fusion}(\cdot) = [(m_1(\cdot) \oplus m_2(\cdot)) \oplus m_3(\cdot)] \oplus m_4(\cdot)$ , where  $\oplus$  denotes PCR5 rule.

**Algorithm 1** Solving classification problem by RSCD

Dataset,  $\alpha = 0.95$ ,  $\gamma_s = -2$ , and  $\beta = 1$   
 The final BBA of test data:  $m_{fusion}(\cdot)$

- 1) Calculate the weights of attributes  $w_B$ , by

$$POS_B(D) = \bigcup_{X \in U/D} BX; \omega_B = \zeta_B^D = \frac{|POS_B(D)|}{|U|};$$

- 2) Calculate the BBA of each attribute, by

$$m_{a_i}^{x^*}(\theta_s) = \alpha e^{\gamma_s d^\beta}; m_{a_i}^{x^*}(\Theta) = 1 - \alpha e^{\gamma_s d^\beta};$$

- 3) Combine all BBAs of attributes sequentially, by  
 $m_{fusion}(\cdot) = 1 \quad i = 1$

- 4) **while**  $i \leq m$  **do**  $m_{fusion}(\cdot) = PCR5(m_{fusion}(\cdot), \omega_i \cdot m_i(\cdot))$  *Normalization*( $m_{fusion}(\cdot)$ ).

Table VIII: UCI datasets used in the experiments.

Datasets	Class Num.	Feature Dimention	Sample Num.
<i>Iris</i>	3	4	150
<i>Wine</i>	3	13	178
<i>Pima</i>	2	8	768
<i>Bupa</i>	2	6	345
<i>Ionosphere</i>	2	34	351

In our tests, we do not deal with the missing data problem, all the samples with missing values have been eliminated. Features of the samples are normalized by their means and standard deviations before their classification. As with the artificial datasets, we have evaluated the nearest neighbor (NN) classifier, the nearest class centroid (NC) classifier, two k-NN classifiers (one is with big  $k$  ( $k = 40$ ) and the other with a small  $k$  ( $k = 5$ )), and the ER-NN-NC classifier (both with DS+BetP option, and with PCR5+DSmP option) [26]. The results are listed in Table IX. As we can see in Table IX, RSCD performs better in three datasets (Iris, Pima and Bupa) and the classification results are close to ER-NN-NC on the other two datasets (Wine and Ionosphere).

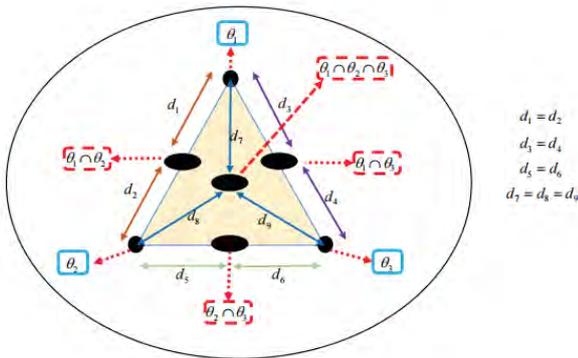


Figure 3: The principle of expanded attribute polygon.

## VI. DISCUSSIONS

In this paper, the Frame of Discernment (FoD) is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  where  $\theta_i$  represents the category and here we just consider singletons without compound focal elements<sup>6</sup>. Actually, some examples are difficult to be divided into a single class, and it may be possible to belong to two categories or several categories at the same time. On the basis of constructing attribute polygons in this paper, we can easily expand the mentioned principle above to more complex circumstances so as to ensure the particular target can belong to several classes simultaneously. The principle is illustrated in Figure 3: In this figure, we give a brief description of the expanded principle by using the three classification problem (triangle). In this triangle, three vertices (light blue and solid frame) represent single class, which is denoted by  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ . The difference is that, the centers of the three edges of such triangle and the center of gravity of this triangle are defined as compound focal elements, respectively. Specifically, the center of  $\theta_1$  and  $\theta_2$  is denoted as  $\theta_1 \cap \theta_2$ , in turn, we can define all the centers of all edges of this triangle. Besides, the center of gravity of this triangle is defined as  $\theta_1 \cap \theta_2 \cap \theta_3$ . Then, we can calculate all the coordinates of these centers and also the corresponding distances so as to obtain the BBAs of all attributes. To illustrate the principle of the expanded attribute polygon, we again revisit **Example 1** as follows: Since the extension method is mainly aimed at constructing BBAs, there is no impact on the calculation of attribute weights, so the following steps are only for BBAs calculation.

- Step 1: Calculate all relevant points in expanded polygon which are given in Table X. In Table X,  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  represent Class 1, Class 2 and Class 3.  $\theta_1 \cap \theta_2$  corresponds to the hypothesis for which the target belongs to two categories simultaneously, and so on. The coordinates of  $\theta_1 \cap \theta_2$  and  $\theta_1 \cap \theta_2 \cap \theta_3$  are calculated as for example by:

$$a_1(\theta_1 \cap \theta_2) = \frac{a_1(\theta_1) + a_1(\theta_2)}{2} = 5.37,$$

$$a_1(\theta_1 \cap \theta_2 \cap \theta_3) = \frac{a_1(\theta_1) + a_1(\theta_2) + a_1(\theta_3)}{3} = 5.78.$$

- Step 2: Based on Euclidean distance, we can obtain the corresponding distances between the target concept  $x^*$  and all relevant points in expanded polygon, which is given in Table XI.
- Step 3: According to Eq.(9) and Eq.(10), BBAs of  $x^*$  with respect to each attribute are shown in Table XII.
- Step 4: Sequentially combine all four BBAs with PCR5

<sup>6</sup>Here, we do not regard  $\Theta$  in Eq (10) as a compound focal element even though  $\Theta$  can be defined as  $\Theta = \theta_1 \cup \theta_2 \cup \dots \cup \theta_n$ . Because  $\Theta$  represents the ignorance or unknown of category of target concept, however, compound focal elements here mean that this target belongs to two categories or three categories at the same time.

Table IX: UCI datasets used in the experiments.

Classifiers	Iris(%)	Wine(%)	Pima(%)	Bupa(%)	Ionosphere(%)
<i>NN</i>	93.84	94.76	69.04	60.46	84.41
<i>NN(Center)</i>	92.09	95.68	72.70	56.54	79.25
<i>ER - NN - NC(DSmT + DSmP)</i>	95.15	<b>96.42</b>	73.38	60.96	<b>87.76</b>
<i>k - NN(k = 40)</i>	89.43	95.28	71.60	61.99	67.63
<i>k - NN(k = 5)</i>	95.65	95.28	72.10	59.57	82.41
<i>RSCD</i>	<b>98.00</b>	94.17	<b>74.50</b>	<b>62.87</b>	84.11

rule and then, we can get the final BBA as follows.

$$\begin{aligned}
m_{fusion}(\theta_1) &= \mathbf{0.8763}; m_{fusion}(\theta_2) = 0.0001; \\
m_{fusion}(\theta_3) &= 0.0000; m_{fusion}(\theta_1 \cap \theta_2) = 0.0240; \\
m_{fusion}(\theta_2 \cap \theta_3) &= 0.0000; m_{fusion}(\theta_1 \cap \theta_3) = 0.0006; \\
m_{fusion}(\theta_1 \cap \theta_2 \cap \theta_3) &= 0.0004; m_{fusion}(\Theta) = 0.0985.
\end{aligned}$$

Thus, we can also get the result that  $x^*$  belongs to Class 1 ( $\theta_1$ ). The biggest difference between the extension method and the RSCD is that the possible category of target is further divided so as to reduce the uncertainty in classification problem, which can be embodied in  $m_2(\cdot)$  in Table VII and Table XII. In RSCD, the assignment of  $x^*$  to  $\Theta$  with respect to  $a_2$  is 0.5640 (see the BBA  $m_2(\cdot)$  of Table VII), which means the class of  $x^*$  cannot be determined if the principle of maximum belief mass is applied. However, in expanded strategy,  $\Theta$  is further divided into  $\theta_1 \cap \theta_2$ ,  $\theta_2 \cap \theta_3$ ,  $\theta_1 \cap \theta_3$ ,  $\theta_1 \cap \theta_2 \cap \theta_3$ , which ensure the target can be labelled with the correct class.

## VII. CONCLUSION

In this paper, a new rough set classifier based on DSMT has been proposed to manage uncertainties using belief function theory. Our simulation results show clearly that RSCD performs well and its implementation is relatively simple since the attribute reduction in traditional rough set is avoided. In the implementation of RSCD, different types of combination rules can be used which give some flexibility to the users. In this paper, only one combination rule in DSMT (PCR5) has been tested. Of course many more could be implemented and tested, especially globally combing all BBAs in a single fusion step with PCR6 rule, which is left for future investigations. Also, The way of the attribute weights and BBAs' calculation used in RSCD is an open question and we plan to make investigations on this question, and evaluate the robustness of RSCD in future research works.

## VIII. ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61573097 and 91748106, in part by Key Laboratory of Integrated Automation of Process Industry (PAL-N201704), in part by the Fundamental Research Funds for the Central Universities (3208008401), in part by the Qing Lan Project and Six Major Top-talent Plan, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

## REFERENCES

- [1] Z. Pawlak, *Rough sets: Theoretical aspects of reasoning about data*, Springer Science & Business Media, Vol. 9, 2012.
- [2] P. Maji, S. Paul, *Rough set based maximum relevance-maximum significance criterion and gene selection from microarray data*, IJAR, Vol. 52(3), pp. 408–426, 2011.
- [3] S. Trabelsi, Z. Elouedi, P. Lingras, *Classification systems based on rough sets under the belief function framework*, IJAR, Vol. 52(9), pp. 1409–1432, 2011.
- [4] M. He, W. Ren, *Attribute reduction with rough set in context-aware collaborative filtering*, Chinese Journal of Electronics, Vol. 26(5), pp. 973–980, 2017.
- [5] D. Liang, Z. Xu, D. Liu, *A new aggregation method-based error analysis for decision-theoretic rough sets and its application in hesitant fuzzy information systems*, IEEE Trans. on Fuzzy Systems, Vol. 25(6), pp. 1685–1697, 2017.
- [6] U. Kumar, H. Inbarani, *PSO-based feature selection and neighbor-hood rough set-based classification for BCI multiclass motor imagery task*, Neural Computing and Applications, Vol. 28(11), pp.3239–3258, 2017.
- [7] Y.-S. Chen, C.-H. Cheng, *Assessing mathematics learning achievement using hybrid rough set classifiers and multiple regression analysis*, Applied Soft Computing, Vol. 13(2), pp. 1183–1192, 2013.
- [8] Y.-S. Chen, C.-H. Cheng, *Evaluating industry performance using extracted RGR rules based on feature selection and rough sets classifier*, Expert Systems with Applications, Vol. 36(5), pp. 9448–9456, 2009.
- [9] Y.-S. Chen, C.-H. Cheng, *Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry*, Knowledge-Based Systems, Vol. 39, pp. 224–239, 2013.
- [10] Y. Chen, Y. Xue, Y. Ma, F. Xu, *Measures of uncertainty for neighborhood rough sets*, Knowledge-Based Systems, Vol. 120(C), pp. 226–235, 2017.
- [11] T. Zheng, L. Zhu, *Uncertainty measures of neighborhood system-based rough sets*, Knowledge-Based Systems, Vol. 86(C), pp. 57–65, 2015.
- [12] D. Chen, Y. Yang, Z. Dong, *An incremental algorithm for attribute reduction with variable precision rough sets*, Applied Soft Computing, Vol. 45, pp. 129–149, 2016.
- [13] C. Gao, Z. Lai, J. Zhou, C. Zhao, D. Miao, *Maximum decision entropy-based attribute reduction in decision-theoretic rough set model*, Knowledge-Based Systems, 2017.
- [14] X. Jia, L. Shang, B. Zhou, Y. Yao, *Generalized attribute reduct in rough set theory*, Knowledge-Based Systems, Vol. 91, pp. 204–218, 2016.
- [15] J. Dezert, F. Smarandache, *An introduction to DSMT*, CoRR, Vol. abs/0903.0279, 2009. <http://arxiv.org/abs/0903.0279>
- [16] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of Fusion 2013 Int. Conference on Information Fusion, pp. 1119–1126, 2013.
- [17] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, Vols. 1–4, American Research hPress, Gallup, NM, USA, 2004–2015, <https://www.onera.fr/staff/jean-dezert/references>
- [18] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [19] P. Smets, *Analyzing the combination of conflicting belief functions*, Information fusion, Vol. 8(4), pp. 387–412, 2007.
- [20] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, in Advances and Applications of DSMT for Information Fusion: Collected Works (Volume 2), F. Smarandache & J. Dezert (Ed.), pp. 69–88, 2006.

Table X: All relevant points of three classes in each attribute with expanded polygon.

Attribute	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1 \cap \theta_2$	$\theta_2 \cap \theta_3$	$\theta_1 \cap \theta_3$	$\theta_1 \cap \theta_2 \cap \theta_3$
$a_1$	5.0600	5.6800	6.6000	5.3700	6.1400	5.8300	5.7800
$a_2$	3.6800	2.6400	2.8500	3.1600	2.7450	3.2650	3.0567
$a_3$	1.3400	3.9400	5.4000	2.6400	4.6700	3.3700	3.5600
$a_4$	0.2200	1.2400	2.0000	0.7300	1.6200	1.1100	1.1533

Table XI: Distances between target  $x^*$  and all points of expanded attribute polygons.

Distance	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1 \cap \theta_2$	$\theta_2 \cap \theta_3$	$\theta_1 \cap \theta_3$	$\theta_1 \cap \theta_2 \cap \theta_3$
$a_1 \leftrightarrow x^*$	0.0400	0.5800	1.5000	0.2700	1.0400	0.7300	0.6800
$a_2 \leftrightarrow x^*$	0.1800	0.8600	0.6500	0.3400	0.7550	0.2350	0.4433
$a_3 \leftrightarrow x^*$	0.0599	2.5400	4.0000	1.2400	3.2700	1.9700	2.1600
$a_4 \leftrightarrow x^*$	0.0200	1.0400	1.8000	0.5300	1.4200	0.9100	0.9533

Table XII: BBAs of  $x^*$  with respect to each attribute in expanded polygon.

$m(\cdot)$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_1 \cap \theta_2$	$\theta_2 \cap \theta_3$	$\theta_1 \cap \theta_3$	$\theta_1 \cap \theta_2 \cap \theta_3$	$\Theta$
$m_1(\cdot)$	0.7473	0.0293	0.0001	0.1880	0.0019	0.0119	0.0161	0.0054
$m_2(\cdot)$	0.3227	0.0055	0.0192	0.1235	0.0102	0.2319	0.0665	0.2205
$m_3(\cdot)$	0.6628	0.0000	0.0000	0.0006	0.0000	0.0000	0.0000	0.3366
$m_4(\cdot)$	0.8426	0.0019	0.0000	0.0395	0.0002	0.0040	0.0031	0.1087

- [21] Z.-G. Liu, Q. Pan, J. Dezert, *A new belief-based K-nearest neighbor classification method*, Pattern Recognition, Vol. 46(3), pp. 834–844, 2013.
- [22] L. Jiao, Q. Pan, X. Feng, F. Yang, *An evidential k-nearest neighbor classification method with weighted attributes*, in Proc. of Fusion 2013 Int. Conf. on Information Fusion, pp. 145–150, 2013.
- [23] Z.-G. Liu, Q. Pan, J. Dezert, *Classification of uncertain and imprecise data based on evidence theory*, Neurocomputing, Vol. 133, pp. 459–470, 2014.
- [24] T. Denceux, *A k -Nearest Neighbor Classification Rule Based on Dempster-Shafer Theory*, Springer Berlin Heidelberg, 2008.
- [25] M. Lichman, *UCI machine learning repository*, 2013. <http://archive.ics.uci.edu/ml>
- [26] D. Han, J. Dezert, Y. Yang, C. Han, *New neighborhood classifiers based on evidential reasoning*, in Proc. of Fusion 2013 Int. Conference on Information Fusion, pp. 158–165, 2013.

# Canonical Decomposition of Dichotomous Basic Belief Assignment

Jean Dezert<sup>a</sup>, Florentin Smarandache<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, F-91761 Palaiseau, France.

<sup>b</sup>Department of Mathematics, Univ. of New Mexico, Gallup, NM, USA.

Emails: jean.dezert@onera.fr, smarand@unm.edu

Originally published as: J. Dezert, F. Smarandache, *Canonical Decomposition of Dichotomous Basic Belief Assignment*, Int. J. of Intelligent Systems, Vol. 35, No. 7, pp. 1105–1125, July 2020, and reprinted with permission.

**Abstract**—In this paper, we prove that any dichotomous basic belief assignment (BBA)  $m$  can be expressed as the combination of two simple belief assignments  $m_p$  and  $m_c$  called respectively the pros and cons BBAs thanks to the proportional conflict redistribution rule no 5 (PCR5). This decomposition always exists and is unique and we call it the canonical decomposition of the BBA  $m$ . We also show that canonical decompositions do not exist in general if we use the conjunctive rule, the disjunctive rule, Dempster’s rule, Dubois and Prade’s or Yager’s rules, or even the averaging rule of combination. We give some numerical examples of canonical decompositions and discuss of the potential interest of this canonical decomposition for applications in information fusion.

**Keywords:** Belief Functions, Contra-evidence, Pro-evidence, PCR5, Canonical Decomposition.

## I. INTRODUCTION

The belief functions (BF) introduced by Shafer in the mid of 1970’s [1] from Dempster’s works are well known and used in the artificial intelligence community to model epistemic uncertainty and to reason with it for information fusion. In Dempster-Shafer theory, the combination of basic belief assignments (BBAs) provided by distinct sources of evidence is done with Dempster’s rule of combination which suffers of serious drawbacks in high conflict situation as discussed by Zadeh [16], [17], but also in very low conflict situations [4]. As a matter of fact many rules of combination have been proposed in the literature [2] (Vol. 2), among them the combination of two sources of evidence based on the proportional conflict redistribution principle no5 (PCR5 rule) [8] has been shown successful in applications, and well justified theoretically. However its complexity remains one of its limitations to prevent its use in large fusion problems.

In this study, we show how the fusion of dichotomous BBAs could be done thanks to their PCR5-based canonical decomposition which is always possible. Such decomposition of dogmatic or nondogmatic BBA has never been presented in the literature so far. Only a canonical decomposition based on conjunctive rule involving improper BBA has been proposed by Smets in 1995 [3] and extended later by Denœux [12] to develop the cautious rule of combination. Here the canonical decomposition we present is done differently, and we show that any dichotomous BBA is always the result of the PCR5 fusion of a simple proper pro-evidence BBA  $m_p$  with a

simple proper contra-evidence BBA  $m_c$ , and we show that this decomposition is unique.

This paper is organized as follows. After a brief recall of basics of belief functions in section II, we present the canonical decomposition problem (CDP) in section III and we show the impossibility to realize the CDP of a non dogmatic BBA with conjunctive rule, disjunctive rule, Yager’s and Dubois & Prade rules, and even with the averaging rule of combination. In section IV, we analyze the CDP based on Dempster’s rule of combination and we show that it cannot be done for a dogmatic BBA. In section V, we prove that the canonical decomposition based on PCR5 rule always exist for all the cases. In section VI, we present some particular decompositions of a dichotomous BBA (including dogmatic BBA). Some numerical examples are presented in section VII, and potential interests of this PCR5-based canonical decomposition are discussed in section VIII. The last section concludes this paper and opens a challenging question for application of this new approach.

## II. BASICS OF BELIEF FUNCTIONS

BF have been introduced by Shafer in [1] to model epistemic uncertainty. We assume that the answer<sup>1</sup> of the problem under concern belongs to a known (or given) finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are mutually exclusive<sup>2</sup>. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A proper Basic Belief Assignment (BBA) associated with a given source of evidence is defined [1] as a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . In some BF related frameworks, like in Smets Transferable Belief Model (TBM) [3],  $m(\emptyset)$  is allowed to take a positive value. In this case,  $m(\cdot)$  is said improper because it does not satisfy Shafer’s definition [1]. The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are respectively defined from a proper BBA  $m(\cdot)$  by

$$Bel(A) = \sum_{B \in 2^\Theta | B \subseteq A} m(B), \quad (1)$$

<sup>1</sup>That is, the solution, or the decision to take.

<sup>2</sup>This is so-called Shafer’s model of FoD [2].

and

$$Pl(A) = \sum_{B \in 2^\Theta | A \cap B \neq \emptyset} m(B) = 1 - Bel(\bar{A}), \quad (2)$$

where  $\bar{A}$  is the complement of  $A$  in  $\Theta$ .

$Bel(A)$  and  $Pl(A)$  are usually interpreted, respectively, as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$ .  $A$  is called a focal element (FE) of  $m(\cdot)$  if  $m(A) > 0$ . When all FEs are singletons then  $m(\cdot)$  is called a *Bayesian BBA* [1] and its corresponding  $Bel(\cdot)$  function is equal to  $Pl(\cdot)$ , and they are homogeneous to a (subjective) probability measure  $P(\cdot)$ . The vacuous BBA, or VBBA for short, representing a totally ignorant source is defined as<sup>3</sup>  $m_v(\Theta) = 1$ . A dogmatic BBA is a BBA such that  $m(\Theta) = 0$ . If  $m(\Theta) > 0$  the BBA  $m(\cdot)$  is nondogmatic. A simple BBA is a BBA that has at most two focal sets and one of them is  $\Theta$ . A dichotomous non dogmatic mass of belief is a BBA having three focal elements  $A$ ,  $\bar{A}$  and  $A \cup \bar{A}$  with  $A$  and  $\bar{A}$  subsets of  $\Theta$ .

In his Mathematical Theory of Evidence [1], Shafer proposed to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD  $\Theta$  with Dempster's rule (i.e. the normalized conjunctive rule). The justification and behavior of Dempster's rule have been disputed over the years from many counter-examples involving high and low conflicting sources (from both theoretical and practical standpoints) as reported in [4]–[7].

Many rules of combination exist in the literature<sup>4</sup>, among them we recommend the rule based on the proportional conflict redistribution principle no5 (PCR5 rule) [8] which has been shown successful in applications and well justified theoretically. That is why we analyze it in details for solving the BF canonical decomposition problem (BF-CDP). PCR5 transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is entirely preserved in this fusion process. (see [2], Vol. 2 and Vol. 3 for full justification and examples). The PCR5 combination of two BBAs  $m_1$  and  $m_2$  defined on the same FoD  $\Theta$ , denoted by  $m_{PCR5} = PCR5(m_1, m_2)$ , is mathematically defined as  $m_{PCR5}(\emptyset) = 0$ , and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{PCR5}(X) = \sum_{\substack{x_1, x_2 \in 2^\Theta \\ x_1 \cap x_2 = X}} m_1(x_1)m_2(x_2) + \sum_{\substack{X_2 \in 2^\Theta \\ X_2 \cap X = \emptyset}} \left[ \frac{m_1(X)^2 m_2(X_2)}{m_1(X) + m_2(X_2)} + \frac{m_2(X)^2 m_1(X_2)}{m_2(X) + m_1(X_2)} \right], \quad (3)$$

where all denominators in (3) are different from zero. If a denominator is zero, that fraction is discarded. The properties of PCR5 can be found in [9]. Extension of PCR5 for combining qualitative BBA's can be found in [2], Vol. 2 and 3. All propositions/sets are in a canonical form. A variant

<sup>3</sup>The complete ignorance is denoted  $\Theta$  in Shafer's book [1].

<sup>4</sup>see [2], Vol. 2 for a detailed list of fusion rules.

of PCR5, called PCR6 has been proposed by Martin and Osswald in [2], Vol. 2, for combining  $s > 2$  sources. The general formulas for PCR5 and PCR6 rules are also given in [2], Vol. 2. PCR6 coincides with PCR5 when one combines two sources. The difference between PCR5 and PCR6 lies in the way the proportional conflict redistribution is done as soon as three (or more) sources are involved in the fusion. From the implementation point of view, PCR6 is simpler to implement than PCR5. For convenience, very basic (not optimized) Matlab codes of PCR5 and PCR6 fusion rules can be found in [2], [10] and from the toolboxes repository on the web [11]. In the sequel we work with PCR5 rule because only two BBAs are involved in the canonical decomposition process we present.

### III. THE CANONICAL DECOMPOSITION PROBLEM

We consider a dichotomous (simplest) FoD  $\Theta$  made of only two exclusive elements  $A$  and  $\bar{A}$ , that is  $\Theta = \{A, \bar{A}\}$  and we consider a given proper<sup>5</sup> BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  of the form

$$m(A) = a, \quad m(\bar{A}) = b, \quad m(A \cup \bar{A}) = 1 - a - b, \quad (4)$$

with  $0 < a < 1$ ,  $0 < b < 1$ , and  $a + b < 1$ .

The conditions  $0 < a < 1$  and  $0 < b < 1$  mean that  $A$  and  $\bar{A}$  are FEs of the BBA. The restriction  $a + b < 1$  means that the BBA is nondogmatic. This assumption of nondogmaticity of the BBA  $m(\cdot)$  can be justified because most (if not all) states of belief, being based on imperfect and not entirely conclusive evidence, should be represented by nondogmatic BFs, even if the mass  $m(\Theta)$  is very small as argued by Denœux in [12] (p. 240). In fact, we can always slightly modify a dogmatic BBA  $m(\cdot)$  in a nondogmatic BBA by discounting it with some small discount rate  $\epsilon > 0$ , and letting  $\epsilon$  tend towards 0 [3]. The case of dogmatic belief, as well as degenerate cases with  $a = 0$  and  $b = 0$  will be discussed in Section VI. Note that his assumption of nondogmaticity of the BBA  $m(\cdot)$  is necessary for Smets canonical decomposition [3], but it is not essential for our PCR5-based canonical decomposition because it also works with a dogmatic BBA as discussed in section VI.

The belief function canonical decomposition problem can be expressed as follows:

Given a nondogmatic BBA  $m(\cdot)$  as in (4) and a chosen rule of combination, find the two following simple proper BBAs  $m_p$  and  $m_c$  of the form

$$m_p(A) = x, \quad m_p(A \cup \bar{A}) = 1 - x, \quad (5)$$

$$m_c(\bar{A}) = y, \quad m_c(A \cup \bar{A}) = 1 - y, \quad (6)$$

with  $(x, y) \in [0, 1] \times [0, 1]$ , such that  $m = Fusion(m_p, m_c)$ , for a chosen rule of combination denoted  $Fusion(\cdot, \cdot)$ .

$m_p(\cdot)$  is called the *pro-BBA* (or pro-evidence) of  $A$ , and  $m_c(\cdot)$  the *contra-BBA* (or contra-evidence) of  $A$ . In the section V we prove that this decomposition is always possible and unique and we call it the (PCR5-based) canonical decomposition of the BBA  $m(\cdot)$ . The BBA  $m_p(\cdot)$  is interpreted as

<sup>5</sup>which means that  $m(\emptyset) = 0$ .

a source of evidence providing uncertain evidence in favor of  $A$ , whereas  $m_c(\cdot)$  is interpreted as a source of evidence providing uncertain evidence against  $A$ . The BBA  $m(\cdot)$  can be interpreted as the result of the PCR5 fusion of these two (pros and cons) aspects of evidence about  $A$ .

It is worth noting that this BF-CDP must not be confused with canonical decomposition problem addressed by Smets in [3] in his TBM framework, which is based on conjunctive rule of combination and which involves, in general, improper BBAs, called generalized simple BBA (GSBBA) in Smets terminology.

#### A. Impossibility of decompositions by some well-known rules

Here we analyze briefly the impossibility of a canonical decomposition for some well-known rules of combination.

1) **Conjunctive rule:** We consider  $x > 1$  and  $y > 1$  so that the two BBAs are really informative (otherwise they become vacuous and useless from decision-making standpoint). In this case we always have a conflict between  $m_p(\cdot)$  and  $m_c(\cdot)$  resulting of the conjunctive rule of combination. That is

$$m_{conj}(\emptyset) = m_p(A)m_c(\bar{A}) = x \cdot y > 0 \quad (7)$$

Hence  $m_{conj}(\emptyset) \neq 0$  is incompatible with the constraint  $m(\emptyset) = 0$ . Therefore, the canonical decomposition of the BBA  $m(\cdot)$  expressed as the conjunctive fusion of pros and cons BBAs  $m_p(\cdot)$  and  $m_c(\cdot)$  is impossible to get in general<sup>6</sup>, but in the very degenerate cases where  $a = 0$ , or  $b = 0$ , or  $a = 0$  and  $b = 0$  which would involve vacuous BBAs in the decomposition and of course will be useless.

2) **Disjunctive rule:** If we consider the disjunctive rule of combination of  $m_p(\cdot)$  and  $m_c(\cdot)$  we will always obtain the vacuous BBA because  $m_p(A)m_c(\bar{A})$ ,  $m_p(A)m_c(A \cup \bar{A})$ ,  $m_p(A \cup \bar{A})m_c(\bar{A})$  and  $m_p(A \cup \bar{A})m_c(A \cup \bar{A})$  will all be committed to the uncertainty  $A \cup \bar{A}$ . Therefore the combination result is nothing but the vacuous belief assignment  $m_v$ , that is  $Disj(m_p, m_c) = m_v$ . In conclusion, we cannot make a decomposition of the BBA  $m(\cdot)$  based on the disjunctive rule in general because if  $m(\cdot)$  is informative (e.g. not vacuous) one always has  $a + b < 1$  so that  $m(A \cup \bar{A}) < 1$  whereas the disjunctive rule of  $m_p(\cdot)$  and  $m_c(\cdot)$  will always provide  $m(A \cup \bar{A}) = 1$ .

3) **Yager's and Dubois & Prade rules:** Due to the particular simple form of BBAs  $m_p(A)$  and  $m_c(\cdot)$ , Yager's rule [13] and Dubois-Prade rule [14] coincide. Based on these rules we are searching  $x$  and  $y$  in  $[0, 1]$  such that

$$m(A) = a = x(1 - y) \quad (8)$$

$$m(\bar{A}) = b = (1 - x)y \quad (9)$$

$$m(A \cup \bar{A}) = 1 - a - b = (1 - x)(1 - y) + xy \quad (10)$$

Because the third equation is dependent of the two first, we have only to solve the following system of equations

<sup>6</sup>that is for any  $a$  and  $b$  values of mass of FEs  $A$  and  $\bar{A}$  of the BBA  $m(\cdot)$ .

$x - xy = a$  and  $y - xy = b$ . Assuming<sup>7</sup>  $y < 1$ , one gets from the first equation  $x = \frac{a}{1-y}$ . By replacing  $x$  by its expression in the second equation  $y - xy = b$  we have to find  $y$  in  $[0, 1]$  such that (after basic algebraic simplifications)

$$y^2 + (a - b - 1)y + b = 0 \quad (11)$$

This second-order equation admits one or two real solutions  $y_1$  and  $y_2$  if and only if the discriminant is null or positive respectively, that is if  $(a - b - 1)^2 - 4b \geq 0$ . However this discriminant can become negative depending on the values of  $a$  and  $b$ . For instance, for  $a = 0.4$  and  $b = 0.5$ , we have  $(a - b - 1)^2 - 4b = -0.79$  which means that there is no real solution for the equation  $y^2 - 1.1 \cdot y + 0.5 = 0$ . Therefore, in general, the canonical decomposition of the BBA  $m(\cdot)$  cannot be accomplished from Yager's and Dubois & Prade rules of combination.

4) **Averaging rule:** Suppose we combine  $m_p(\cdot)$  and  $m_c(\cdot)$  with the averaging rule. Then we are searching  $x$  and  $y$  in  $[0, 1]$  such that

$$m(A) = a = (x + 0)/2 \quad (12)$$

$$m(\bar{A}) = b = (0 + y)/2 \quad (13)$$

$$m(A \cup \bar{A}) = 1 - a - b = ((1 - x) + (1 - y))/2 \quad (14)$$

This means that  $x = 2a$  and  $y = 2b$  with  $x$  and  $y$  in  $[0, 1]$ . So, if  $a > 0.5$  or  $b > 0.5$  the canonical decomposition is impossible to make with the averaging rule of combination. Therefore, in general, the averaging rule is not able to provide a canonical decomposition of the BBA  $m(\cdot)$ .

#### IV. DECOMPOSITION BASED ON DEMPSTER'S RULE

Let consider a nondogmatic BBA  $m(A) = a$ ,  $m(\bar{A}) = b$  and  $m(A \cup \bar{A}) = 1 - a - b$  with  $0 \leq a, b \leq 1$  and  $1 - a - b > 0$ , and let's see if a decomposition of  $(\cdot)$  is possible based on Dempster's rule of combination [1]. For this, we are searching  $x$  and  $y$  in  $[0, 1]$  such that  $xy \neq 1$  and

$$m(A) = a = \frac{x(1 - y)}{1 - xy}, \quad (15)$$

$$m(\bar{A}) = b = \frac{y(1 - x)}{1 - xy}, \quad (16)$$

$$m(A \cup \bar{A}) = 1 - a - b = \frac{(1 - x)(1 - y)}{1 - xy}. \quad (17)$$

Because the third equality is redundant with the two first, we just have to solve the system of two equations expressed as

$$(1 - xy)a = x(1 - y), \quad (18)$$

$$(1 - xy)b = y(1 - x). \quad (19)$$

That is, one should have

$$x - xy + axy = a, \quad (20)$$

$$y - xy + bxy = b, \quad (21)$$

<sup>7</sup>taking  $y = 1$  would mean that  $x(1 - y) = 0$  but  $m(A) = a$  with  $a \neq 0$  in general, so that the choice of  $y = 1$  is not possible.

with the constraints  $0 < x < 1$  and  $0 < y < 1$ . So one must have

$$x = \frac{a}{1-y+ay}, \quad y \neq \frac{1}{1-a}, \quad (22)$$

and solve the equation  $y - xy + bxy = b$  with  $x$  expressed as function of  $y$  as above. We get the equation for  $a \neq 1$

$$(a-1)y^2 + (1+b-a)y - b = 0, \quad (23)$$

whose solutions have the form

$$y = \frac{-(1+b-a) \pm \sqrt{\Delta}}{2(a-1)}, \quad (24)$$

where the discriminant  $\Delta$  is given by

$$\begin{aligned} \Delta &= (1+b-a)^2 - 4(1-a)b \\ &= 1 + b^2 + a^2 + 2b - 2a - 2ab + 4ab - 4b \\ &= a^2 + b^2 + 1 - 2b + 2ab - 2a \\ &= (a+b-1)^2 = (1-a-b)^2. \end{aligned}$$

One sees that  $\Delta$  is strictly positive because  $a + b < 1$  ( $m$  being a nondogmatic BBA). So, there exist two real solutions  $y_1$  and  $y_2$  of (23) of the form

$$y_1 = \frac{-(1+b-a) + \sqrt{\Delta}}{2(a-1)} = \frac{b}{1-a}, \quad (25)$$

$$y_2 = \frac{-(1+b-a) - \sqrt{\Delta}}{2(a-1)} = \frac{1-a}{1-a} = 1. \quad (26)$$

For the case  $a \neq 1$ , the second ‘‘solution’’  $y_2 = 1$  implies  $x = \frac{a}{1-y_2+ay_2} = \frac{a}{a} = 1$  which is not an acceptable solution<sup>8</sup> because one must have  $xy \neq 1$ . The solution  $(x, y)$  of the decomposition problem for  $a \neq 1$  is actually given by the first solution  $y_1$ , that is

$$y = y_1 = \frac{b}{1-a} \in [0, 1) \quad (27)$$

$$x = \frac{a}{1-y+ay} = \frac{a}{1-b} \in [0, 1) \quad (28)$$

The case  $a = 1$  corresponding to the dogmatic BBA given by  $m(A) = a = 1$ ,  $m(\bar{A}) = b = 0$ ,  $m(A \cup \bar{A}) = 1 - a - b = 0$  is analyzed in details in Section VI - See lemma right after Theorem 4.

In summary, the unique solution of decomposition of a nondogmatic BBA with  $0 < a < 1$ ,  $0 < b < 1$  and  $a + b < 1$  using Dempster’s rule is  $x = \frac{a}{1-b}$  and  $y = \frac{b}{1-a}$ .

**Example 1:** Consider  $m(A) = a = 0.6$ ,  $m(\bar{A}) = b = 0.2$  and  $m(A \cup \bar{A}) = 1 - a - b = 0.2$ . The solution  $(x, y)$  of the decomposition of  $m(\cdot)$  based on Dempster’s rule is  $x = \frac{a}{1-b} = \frac{0.6}{1-0.2} = 0.75$  and  $y = \frac{b}{1-a} = \frac{0.2}{1-0.6} = 0.5$ . Therefore,  $m_p(A) = x = 0.75$ ,  $m_p(A \cup \bar{A}) = 1 - x = 0.25$  and  $m_c(\bar{A}) = y = 0.5$ ,  $m_c(A \cup \bar{A}) = 1 - y = 0.5$ . It can be verified that  $m_p \oplus m_c = m$ , where  $\oplus$  represents symbolically Dempster’s rule of combination [1].

<sup>8</sup>otherwise the denominators of Eqs. (15)–(17) will be equal to zero.

## V. DECOMPOSITION BASED ON PCR5 RULE

In this section we prove that the decomposition of a dichotomous nondogmatic BBA  $m(\cdot)$  based on PCR5 rule of combination is always possible and unique. Suppose we combine  $m_p(\cdot)$  and  $m_c(\cdot)$  with the PCR5 rule of combination. Then we are searching  $(x, y) \in [0, 1]^2$  satisfying

$$m(A) = a = x(1-y) + \frac{x^2y}{x+y} = \frac{x^2 + xy - xy^2}{x+y}, \quad (29)$$

$$m(\bar{A}) = b = (1-x)y + \frac{xy^2}{x+y} = \frac{y^2 + xy - x^2y}{x+y}, \quad (30)$$

$$m(A \cup \bar{A}) = 1 - a - b = 1 - x - y + xy, \quad (31)$$

under the constraints  $(a, b) \in [0, 1]^2$ , and  $0 < a + b < 1$ .

The equations (29) and (30) can be rewritten as

$$x - \frac{xy^2}{x+y} = a, \quad (32)$$

$$y - \frac{x^2y}{x+y} = b, \quad (33)$$

from which (31) is redundant because (29) + (30) gives

$$x + y - xy = a + b. \quad (34)$$

Therefore  $(1-x)(1-y) = 1 - (a+b)$  and that is why the constraint  $a + b \leq 1$  is necessary<sup>9</sup> for the existence of the solution  $(x, y)$ .

With  $x$  and  $y$  in  $[0, 1]$  the solutions of (32) and (33) verify

$$x \geq a, \quad (35)$$

$$y \geq b. \quad (36)$$

Moreover, the equality (34) implies

$$x(1-y) = a + b - y \Rightarrow y \leq a + b, \quad (37)$$

$$y(1-x) = a + b - x \Rightarrow x \leq a + b. \quad (38)$$

For  $x \neq 1$ , from (34) one gets  $y = \frac{a+b-x}{1-x}$  and from (32) one has

$$x^2 + xy - xy^2 = ax + ay. \quad (39)$$

Putting this expression of  $y$  in (39), yields the equation

$$x^2 + (x-a)\frac{a+b-x}{1-x} - x\left(\frac{a+b-x}{1-x}\right)^2 - ax = 0, \quad (40)$$

which can be expressed after elementary algebraic calculation as

$$\begin{aligned} x^4 + (-a-2)x^3 + (2a+b)x^2 \\ + (a+b-ab-b^2)x + (-a^2-ab) = 0. \end{aligned} \quad (41)$$

This equation of degree 4 has at most four real solutions. We have to take only the solution  $x$  from the open interval  $(0, 1)$  and  $y = (a+b-x)/(1-x)$  with  $y \in [0, 1]$ .

The general expression of the solutions of this quartic equation [15] is very complicate to obtain analytically even

<sup>9</sup>In fact we use the constraint  $a + b < 1$  because in this section we consider only nondogmatic BBA. The canonical decomposition of a dichotomous dogmatic BBA will be analyzed in the section VI.

with modern symbolic computing systems like Maple™, or Mathematica™, but the solutions can be easily calculated numerically by these computing systems, and even with Matlab™system (thanks to the *fsolve* command) as soon as the numerical values are committed to  $a$  and to  $b$ . Another method to make the decomposition consists to solve numerically the system of equations  $\frac{x^2+xy-xy^2}{x+y} = a$  and  $\frac{y^2+xy-x^2y}{x+y} = b$  for numerical values committed to  $a$  and  $b$  thanks to Mathematica™, Maple™, or Matlab™computing systems for instance. Of course the solutions provided by the two methods are the same.

**Example 2:** Let consider  $m(A) = 0.6$ ,  $m(\bar{A}) = 0.3$  and  $m(A \cup \bar{A}) = 0.1$ , therefore  $a = 0.6$  and  $b = 0.3$ . The quartic equation (41) becomes

$$x^4 - 2.6x^3 + 1.5x^2 + 0.63x - 0.54 = 0. \quad (42)$$

The four solutions of this quartic equation provided by the computing system<sup>10</sup> are approximately

$$\begin{aligned} x_1 &\approx 0.7774780438, \\ x_2 &\approx 0.9297589637, \\ x_3 &\approx 1.419151582, \\ x_4 &\approx -0.5263885898, \end{aligned}$$

which are shown on the graph of figure 1 obtained easily from Desmos online tool<sup>11</sup>.

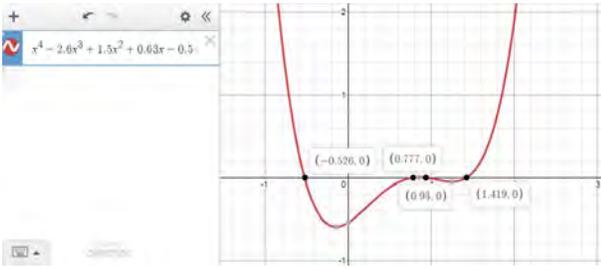


Figure 1. Plot of the quartic function.

Clearly  $x_3$  and  $x_4$  are not acceptable solutions because they do not belong to  $[0, 1]$ . If we take  $x_1 \approx 0.7774780438$  then will get  $y_1 = (a + b - x_1)/(1 - x_1) = (0.9 - x_1)/(1 - x_1) \approx 0.5506061437$ , so the pair  $(x_1, y_1) \in [0, 1]^2$  is a solution of the decomposition problem of the BBA  $m(\cdot)$ . If we take  $x_2 \approx 0.9297589637$  then will get  $y_2 = (a + b - x_2)/(1 - x_2) = (0.9 - x_2)/(1 - x_2) \approx -0.4236692006$ , so we see that  $y_2 \notin [0, 1]$  and therefore the pair  $(x_2, y_2)$  cannot be a solution of the decomposition problem of the BBA  $m(\cdot)$ . Therefore the canonical masses  $m_p(\cdot)$  and  $m_c(\cdot)$  are given by

$$\begin{aligned} m_p(A) &\approx 0.7774780438, & m_p(A \cup \bar{A}) &\approx 0.2225219562, \\ m_c(\bar{A}) &\approx 0.5506061437, & m_c(A \cup \bar{A}) &\approx 0.4493938563. \end{aligned}$$

<sup>10</sup>We did also obtain the same solutions with Maple™, and also with Matlab™.

<sup>11</sup><https://www.desmos.com/calculator>

It can be verified that the PCR5 combination of the BBAs  $m_p$  and  $m_c$ , denoted  $PCR5(m_p, m_c)$ , is equal to the BBA  $m(\cdot)$ . The following important theorem holds.

**Theorem 1:** Consider a dichotomous FoD  $\Theta = \{A, \bar{A}\}$  with  $A \neq \Theta$  and  $A \neq \emptyset$  and a nondogmatic BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  defined on  $\Theta$  by  $m(A) = a$ ,  $m(\bar{A}) = b$ , and  $m(A \cup \bar{A}) = 1 - a - b$ , where  $a, b \in [0, 1]$  and  $a + b < 1$ . Then the BBA  $m(\cdot)$  has a unique canonical decomposition using PCR5 rule of combination of the form  $m = PCR5(m_p, m_c)$  with pro-evidence  $m_p(A) = x$ ,  $m_p(A \cup \bar{A}) = 1 - x$  and contra-evidence  $m_c(\bar{A}) = y$ ,  $m_c(A \cup \bar{A}) = 1 - y$ , where  $x, y \in [0, 1]$ .

**Proof:** Based on (29)-(30), we have to prove that the following system  $S_{a,b}$  of equations always admits one and only one solution  $(x, y) \in [0, 1] \times [0, 1]$

$$S_{a,b} : \begin{cases} h(x, y) = a, \\ h(y, x) = b, \end{cases} \quad (43)$$

with  $h(x, y) = \frac{x^2+xy-xy^2}{x+y} = x - \frac{xy^2}{x+y}$ . The  $h$  function can be prolonged in  $(0, 0)$  by continuity by setting  $h(0, 0) = 0$ .

One has to prove the existence of a unique  $x \in [a, a + b] \subset [0, 1]$  and  $y \in [b, a + b] \subset [0, 1]$  solutions of  $S_{a,b}$ , or equivalently solutions of  $y = \frac{a+b-x}{1-x}$  and of (41)  $P(x) = 0$  with

$$\begin{aligned} P(x) \triangleq &x^4 + (-a - 2)x^3 + (2a + b)x^2 \\ &+ (a + b)(1 - b)x - a(a + b). \end{aligned} \quad (44)$$

Because<sup>12</sup>  $\lim_{x \rightarrow -\infty} P(x) = +\infty$  and<sup>13</sup>  $P(a) < 0$ , there exists  $x_1 \in (-\infty, a)$  such that  $P(x_1) = 0$ . The solution  $x_1$  is not acceptable because  $x_1 \notin [a, a + b]$ . Because<sup>14</sup>  $P(1) < 0$  and  $\lim_{x \rightarrow +\infty} P(x) = +\infty$ , there exists also  $x_4 \in (1, +\infty)$  such that  $P(x_4) = 0$ . The solution  $x_4$  is not acceptable because  $x_4 \notin [a, a + b]$ . For  $a + b \neq 1$ , one has<sup>15</sup>  $P(a + b) > 0$  and  $P(1) < 0$ . Therefore there exists  $x_3 \in (a + b, 1)$  such that  $P(x_3) = 0$  but this solution  $x_3$  is also not acceptable because  $x_3 \notin [a, a + b]$ . Because  $P(a) < 0$  and  $P(a + b) > 0$  there exists  $x_2 \in [a, a + b]$  such that  $P(x_2) = 0$  which is the only satisfactory solution. The value  $y_2$  is given by  $y_2 = \frac{a+b-x_2}{1-x_2}$ , and one has  $y_2 > 0$  because  $x_2 < a + b$  and  $y_2 < 1$  because  $a + b < 1$ . Moreover, from (33),  $y_2 - b = \frac{x_2 y_2}{x_2 + y_2}$  which is always positive, therefore  $y_2 > b$ , and from (34)  $y_2 - (a + b) = x_2(y_2 - 1)$  which is always negative, therefore  $y_2 < a + b$ . This completes the proof of Theorem 1.

<sup>12</sup> $P(x)$  being polynomial, it is continuous and if  $P(c)P(d) < 0$  there exist at least one solution between  $[c, d]$ . Therefore, we are not sure a priori there is only one solution between  $[c, d]$ . In our case, the signs of  $P(x)$  for  $x = -\infty, a, a + b, 1, +\infty$  are respectively  $+, -, +, -$  and  $+$ . But because one has four intervals, into each interval it is not possible to have more than one solution (because otherwise will get five or more solutions, while this equation has only up to four real solutions). Therefore in each interval there exists only one real solution.

<sup>13</sup>because  $P(a) = a^2b - ab(a + b) = -ab^2$ .

<sup>14</sup>because  $P(1) = -1 + a + b + (a + b)(1 - b - a) = -(a + b - 1)^2$ .

<sup>15</sup>because from (40),  $P(a + b) = (a + b)/(1 - a - b)^2 = (a + b)^2 - a(a + b) \Rightarrow P(a + b) = b(a + b)(1 - a - b)^2 > 0$ .

## VI. PARTICULAR CASES OF DECOMPOSITIONS

Here we examine the canonical decomposition of particular cases, including dogmatic BBA.

A. *Dogmatic BBA*:  $a + b = 1$

**Theorem 2:** Any dogmatic BBA defined by  $m(A) = a$  and  $m(\bar{A}) = b$ , where  $a, b \in [0, 1]$  and  $a + b = 1$ , has a canonical decomposition using PCR5 rule of combination of the form  $m = PCR5(m_p, m_c)$  with  $m_p(A) = x$ ,  $m_p(A \cup \bar{A}) = 1 - x$  and  $m_c(\bar{A}) = y$ ,  $m_c(A \cup \bar{A}) = 1 - y$  where  $x, y \in [0, 1]$ .

**Proof:** Any solution of  $S_{a,b}$  verifies

$$x - a = \frac{xy^2}{x + y}, \quad (45)$$

$$y - b = \frac{x^2y}{x + y}, \quad (46)$$

and therefore from (45)+(46) one has

$$(x - y) - (a - b) = \frac{xy(y - x)}{x + y}, \quad (47)$$

which can be rewritten as

$$(x - y)\left[1 + \frac{xy}{x + y}\right] = (a - b). \quad (48)$$

This means that differences  $(x - y)$  and  $(a - b)$  have the same sign. Moreover from (34) with  $a + b = 1$  one has  $x + y - xy = 1$ , or equivalently  $(1 - x)(1 - y) = 0$  which is satisfied if  $x = 1$ , or if  $y = 1$  or both equal one. We must distinguish three cases as follows:

- If  $a < b$  then<sup>16</sup>  $x < y$  therefore  $y = 1$  and  $h(x, 1) = a$ . Solving  $h(x, 1) = a$  is equivalent to solve  $x^2 - ax - a = 0$  which admits only one positive solution  $x \in [a, a + b = 1]$  given by  $x = \frac{a + \sqrt{a^2 + 4a}}{2}$ . Note if  $a + b = 1$  and  $a < b$ , then necessarily  $a < 0.5$ .
- If  $a > b$  then  $x > y$  therefore  $x = 1$  and  $h(1, y) = b$ . Solving  $h(1, y) = b$  is equivalent to solve  $y^2 - by - b = 0$  which admits only one positive solution  $y \in [b, a + b = 1]$  given by  $y = \frac{b + \sqrt{b^2 + 4b}}{2}$ . Note if  $a + b = 1$  and  $a > b$ , then necessarily  $b < 0.5$ .
- If  $a = b$  and  $a + b = 1$  then  $a = b = 0.5$  and  $x = y = 1$ .

So we have proved that a decomposition based on PCR5 always exists and it is unique also for any dogmatic dichotomous BBA. Therefore, this decomposition of dogmatic dichotomous BBA is canonical, which completes the proof of Theorem 2.

**Theorem 3:** Any dogmatic BBA  $m(A) = a$ ,  $m(\bar{A}) = b$  with  $a + b = 1$  and  $0 < a < 1$  is not decomposable from Yager's rule and Dubois-Prade rule of combination.

**Proof:** We have the following system of equations to solve

$$x - xy = a, \quad (49)$$

$$y - xy = b. \quad (50)$$

From (49) and (50), we get  $a - b = x - xy - (y - xy) = x - y$ , so  $y = x - a + b$ . After replacing this expression of  $y$  into (49) and algebraic manipulations, we have to solve

$$x^2 - 2ax + a = 0,$$

whose solutions are of the form

$$x = a \pm \sqrt{a(a - 1)}.$$

For  $0 < a < 1$  the system has no real solutions because  $a(a - 1) < 0$ , which completes the proof of Theorem 3.

**Theorem 4:** Any dogmatic BBA  $m(A) = a$ ,  $m(\bar{A}) = b$  with  $a + b = 1$  is not decomposable from Dempster's rule of combination for the case when  $(a, b) \neq (1, 0)$  and  $(a, b) \neq (0, 1)$ .

**Proof:** We have the following system of equations to solve with  $0 \leq x, y \leq 1$  and  $1 - xy \neq 0$

$$\frac{x - xy}{1 - xy} = a, \quad (51)$$

$$\frac{y - xy}{1 - xy} = b. \quad (52)$$

After adding the two equations (51) and (52) and because  $a + b = 1$ , we obtain  $\frac{x - xy + y - xy}{1 - xy} = a + b = 1$ , whence  $x + y - 2xy = 1 - xy$ , or  $x + y - xy = 1$ , or  $x + y(1 - x) = 1$ , or  $y(1 - x) = 1 - x$ , or  $\frac{1 - x}{1 - x} = 1$  when  $x \neq 1$ . From (52), one should have  $\frac{y - xy}{1 - xy} = b$  with  $y = 1$ , that is  $\frac{1 - x \cdot 1}{1 - x \cdot 1} = b$ , or  $1 = b$  which is false because if  $0 < a < 1$  then  $b = 1 - a \neq 1$ . This completes the proof of theorem 4.

**Lemma:** The dogmatic BBAs  $m(A) = 1$ ,  $m(\bar{A}) = 0$  (case  $(a, b) = (1, 0)$ ), or  $m(A) = 0$ ,  $m(\bar{A}) = 1$  (case  $(a, b) = (0, 1)$ ) have infinitely many decompositions based on Dempster's rule of combination.

**Proof:** For the case  $(a, b) = (1, 0)$  one has to solve with  $0 \leq x, y \leq 1$  and  $1 - xy \neq 0$  the system of equations

$$\frac{x - xy}{1 - xy} = 1, \quad \text{and} \quad \frac{y - xy}{1 - xy} = 0. \quad (53)$$

This system is satisfied for  $x = 1$  and  $y \in [0, 1)$ , that is any value in  $[0, 1)$  can be chosen for  $y$ .

For the case  $(a, b) = (0, 1)$  one has to solve with  $0 \leq x, y \leq 1$  and  $1 - xy \neq 0$  the system of equations

$$\frac{x - xy}{1 - xy} = 0, \quad \text{and} \quad \frac{y - xy}{1 - xy} = 1. \quad (54)$$

This system is satisfied for  $y = 1$  and  $x \in [0, 1)$ , that is any value in  $[0, 1)$  can be chosen for  $x$ . Therefore one sees that for the case  $(a, b) = (1, 0)$  and the case  $(a, b) = (0, 1)$  there is no unique decomposition of these BBAs from Dempster's rule of combination, which completes the proof of the lemma.

<sup>16</sup>because  $(x - y)$  and  $(a - b)$  have the same sign.

*B. Case when  $a = 0$  and  $b = 0$  (i.e.,  $m$  is the vacuous BBA)*

This is the most degenerate case where the BBA  $m(\cdot)$  corresponds to the vacuous BBA. For averaging rule, conjunctive rule, Yager's, Dubois-Prade's, Dempster's and PCR5 rules one has  $x = 0$  and  $y = 0$  (conflict between canonical masses is zero). In fact the vacuous BBA  $m(\cdot)$  can always be interpreted as the fusion of  $m_p$  and  $m_c$ , where  $m_p$  and  $m_c$  are also vacuous BBAs. This degenerate case has no particular interest in practice but to model the total ignorant state of knowledge.

*C. Case when  $a = 0$ , or  $b = 0$*

In the case  $a = 0$  and  $0 < b \leq 1$ , then for conjunctive rule, Yager's, Dubois-Prade's, Dempster's and PCR5 rules one has  $x = 0$  and  $y = b$  (conflict between canonical masses is zero) and  $m(\cdot)$  corresponds to the fusion of vacuous pro-evidence  $m_p = m_v$  with the contra-evidence  $m_c = m$ . In the case  $0 < a \leq 1$  and  $b = 0$ , then for conjunctive rule, Yager's, Dubois-Prade's, Dempster's and PCR5 rules one has  $x = a$  and  $y = 0$  (conflict between canonical masses is zero) and  $m(\cdot)$  corresponds to the fusion of the pro-evidence  $m_p = m$  with the vacuous contra-evidence  $m_c = m_v$ . These cases have no particular interest because they can be seen just as the combination of pros (or cons) BBA with the vacuous BBA

*D. Case when  $a = b \in (0, 0.5)$*

**Theorem 5:** In the case  $a = b \in (0, 0.5)$ , the BBA  $m(A) = m(\bar{A}) = a$  and  $m(A \cup \bar{A}) = 1 - 2a$  can be canonically decomposed from PCR5 rule with the BBAs  $m_p(A) = 1 - \sqrt{1 - 2a}$ ,  $m_p(A \cup \bar{A}) = \sqrt{1 - 2a}$  and  $m_c(\bar{A}) = 1 - \sqrt{1 - 2a}$ ,  $m_c(A \cup \bar{A}) = \sqrt{1 - 2a}$ .

**Proof:** From (29) and (30), one has  $\frac{x^2 + xy - xy^2}{x + y} = a$  and one has also in this case  $\frac{y^2 + xy - x^2y}{x + y} = b = a$ . Therefore  $x^2 + xy - xy^2 = y^2 + xy - x^2y$ , or  $x^2 - xy^2 - y^2 + x^2y = 0$ , or  $(x - y)(x + y + xy) = 0$ .  $x \geq 0$  and  $y \geq 0$  because they represent the masses. Therefore  $x + y + xy \geq 0$ . The sum  $x + y + xy = 0$  if and only if  $x = y = 0$ , but this produces the degenerate case, which is corresponding to  $a = b = 0$  (i.e. the vacuous BBA). Yet, in our theorem's hypothesis we assumed  $a, b \in (0, 0.5)$ , so  $a > 0$ , and  $b > 0$ . Therefore  $x + y + xy > 0$ . Hence  $x = y$ . Therefore the canonical BBAs must be of the form  $m_p(A) = x$ ,  $m_p(A \cup \bar{A}) = 1 - x$  and  $m_c(\bar{A}) = x$ ,  $m_c(A \cup \bar{A}) = 1 - x$ . So one must solve the equation<sup>17</sup>  $x - x^2 + \frac{x^2}{2} = m(A) = a$ , or equivalently  $\frac{1}{2}x^2 - x + a = 0$ , whose solutions are  $x_1 = 1 + \sqrt{1 - 2a}$ , and  $x_2 = 1 - \sqrt{1 - 2a}$ . For  $0 < a < 0.5$ , the solution  $x_1 > 1$  is not admissible because  $x_1 \notin [0, 1]$ . The solution  $x_2$  is acceptable because if  $0 < a < 0.5$ , then  $0 < 2a < 1$ , or  $-1 < -1 + 2a < 0$ , or (by multiplying by -1 the inequalities)  $1 > 1 - 2a > 0$ , or  $0 < 1 - 2a < 1$ , or  $\sqrt{0} < \sqrt{1 - 2a} < \sqrt{1}$ , or  $0 > -\sqrt{1 - 2a} > -1$ , or  $1 > 1 - \sqrt{1 - 2a} > 0$  hence  $x_2 \in (0, 1)$ . This completes the proof of Theorem 5.

<sup>17</sup>In fact, we have also the second equation  $x - x^2 + \frac{x^2}{2} = m(\bar{A}) = b = a$  to solve which is the same as the first one.

VII. EXAMPLES

We give in Tables I-IX some numerical examples of PCR5-based canonical decompositions of BBA  $m(\cdot)$  for different sampled values of  $a$  and  $b$  for convenience. These numerical examples may be useful for researchers working with belief functions and interested by this new type of decomposition in their own examples. The values have been approximated at the 10th digit.

$(a, b)$	$x$	$y$
(0.1,0.1)	0.1055728059	0.1055728059
(0.1,0.2)	0.1155063468	0.2085867463
(0.1,0.3)	0.1283308324	0.3116654549
(0.1,0.4)	0.1445620975	0.4155040377
(0.1,0.5)	0.1653570911	0.5207531320
(0.1,0.6)	0.1926613985	0.6284087006
(0.1,0.7)	0.2298437881	0.7403124237
(0.1,0.8)	0.2834628414	0.8604398965
(0.1,0.9)	0.3701562119	1

Table I  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.1$ .

$(a, b)$	$x$	$y$
(0.2,0.1)	0.2085867463	0.1155063468
(0.2,0.2)	0.2254033308	0.2254033308
(0.2,0.3)	0.2477759456	0.3353044255
(0.2,0.4)	0.2763932022	0.4472135955
(0.2,0.5)	0.3133633342	0.5630877072
(0.2,0.6)	0.3628331876	0.6861104563
(0.2,0.7)	0.4339764332	0.8233289109
(0.2,0.8)	0.5582575695	1

Table II  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.2$ .

$(a, b)$	$x$	$y$
(0.3,0.1)	0.3116654549	0.1283308324
(0.3,0.2)	0.3353044255	0.2477759456
(0.3,0.3)	0.3675444680	0.3675444680
(0.3,0.4)	0.4098895428	0.4916206002
(0.3,0.5)	0.4669657064	0.6247896197
(0.3,0.6)	0.5506061437	0.7774780438
(0.3,0.7)	0.7178908346	1

Table III  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.3$ .

$(a, b)$	$x$	$y$
(0.4,0.1)	0.4155040377	0.1445620975
(0.4,0.2)	0.4472135955	0.2763932022
(0.4,0.3)	0.4916206002	0.4098895428
(0.4,0.4)	0.5527864045	0.5527864045
(0.4,0.5)	0.6442577571	0.7188975951
(0.4,0.6)	0.8633249581	1

Table IV  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.4$ .

$(a, b)$	$x$	$y$
(0.5,0.1)	0.5207531320	0.1653570911
(0.5,0.2)	0.5630877072	0.3133633342
(0.5,0.3)	0.6247896197	0.4669657064
(0.5,0.4)	0.7188975951	0.6442577571
(0.5,0.5)	1	1

Table V  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.5$ .

$(a, b)$	$x$	$y$
(0.6,0.1)	0.6284087006	0.1926613985
(0.6,0.2)	0.6861104563	0.3628331876
(0.6,0.3)	0.7774780438	0.5506061437
(0.6,0.4)	1	0.8633249581

Table VI  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.6$ .

$(a, b)$	$x$	$y$
(0.7,0.1)	0.7403124237	0.2298437881
(0.7,0.2)	0.8233289109	0.4339764332
(0.7,0.3)	1	0.7178908346

Table VII  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.7$ .

$(a, b)$	$x$	$y$
(0.8,0.1)	0.8604398965	0.2834628414
(0.8,0.2)	1	0.5582575695

Table VIII  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.8$ .

$(a, b)$	$x$	$y$
(0.9,0.1)	1	0.3701562119

Table IX  
DECOMPOSITION OF BBA WHEN  $m(A) = 0.9$ .

Figures 2 and 3 show the shapes of the pro-evidence  $x = f(a, b)$  and the contra-evidence  $y = g(a, b)$  surfaces proving graphically the existence of canonical decomposition based on PCR5 at the sampling rate of 0.025. The values  $(a, b)$  for which  $a + b > 1$  are not acceptable and  $f(a, b)$  and  $g(a, b)$  have been set to zero in the figures.

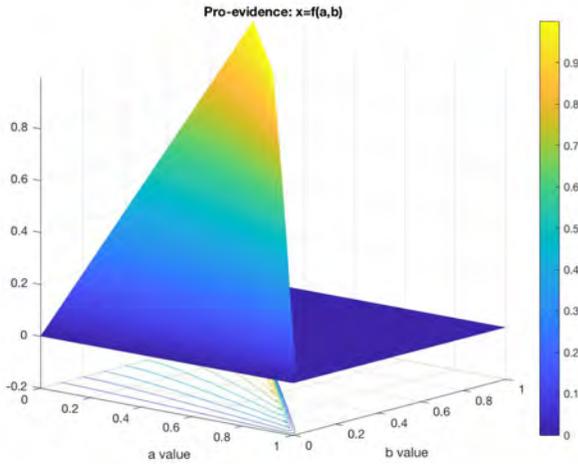


Figure 2. Plot of  $x = f(a, b)$  pro-evidence surface.

### VIII. INTEREST OF CANONICAL DECOMPOSITION

The canonical decomposition based on PCR5 offers several practical interests and advantages that are briefly listed here.

- 1) From the theoretical standpoint, one has proved that the canonical decomposition based on PCR5 rule always

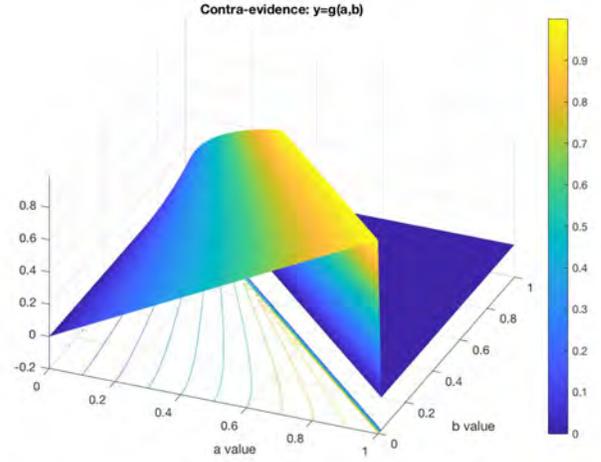


Figure 3. Plot of  $y = g(a, b)$  contra-evidence surface.

exists in all the cases for nondogmatic or dogmatic BBAs contrariwise to other rules of combination that only work in some restrictive cases. Therefore this decomposition is more general and mathematically well justified.

- 2) This canonical decomposition of any dichotomous BBA  $m(\cdot)$  into the pro-evidence  $m_p(\cdot)$  and the contra-evidence  $m_c(\cdot)$  allows to define now the notion of internal conflict of a (dichotomous) source of evidence, denoted  $K_{int}(m)$ , by

$$K_{int}(m) \triangleq m_p(A)m_c(\bar{A}), \quad (55)$$

where  $m_p(A) = x$  and  $m_c(\bar{A}) = y$  are the canonical factors of the BBA  $m(\cdot)$  based on PCR5 rule of combination. It is worth noting that the BBA  $m(\cdot)$  has no internal conflict, if and only if at least one of its factor is the vacuous belief mass, that is if  $x = 0$  or  $y = 0$ , or both, which makes sense. For instance the BBA  $m(A) = 0.3$  and  $m(A \cup \bar{A}) = 0.7$  does not carry internal conflict because  $m_p = m$  and  $m_c = m_v$  (the vacuous BBA) so that its internal conflict  $K_{int}(m) \triangleq m_p(A)m_c(\bar{A}) = 0.3 \cdot 0 = 0$ . In fact in this example the BBA  $m(\cdot)$  carries only uncertain pro-evidence, and vacuous contra-evidence. This internal conflict measure should contribute somehow in the definition of the information content carried by a (dichotomous) source of evidence. This aspect however is not detailed in this paper and is left for future research works. It is clear that the maximum of internal conflict  $K_{int}(m) = 1$  is obtained for the dogmatic BBA  $m(A) = m(\bar{A}) = 0.5$  whose canonical decomposition by PCR5 is  $m_p(A) = 1$  and  $m_c(\bar{A}) = 1$  which shows the full conflict between the pro-evidence  $m_p(\cdot)$  and the contra-evidence  $m_c(\cdot)$  of the source. Of course, there is no internal conflict for the vacuous BBA. More precisely,  $K_{int}(m_v) = 0$  because

if  $a = b = 0$  then one has  $x = y = 0$  calculated from PCR5-based decomposition. Figure 4 shows the internal conflict  $K_{int}(m)$  of a dichotomous BBA  $m$ .

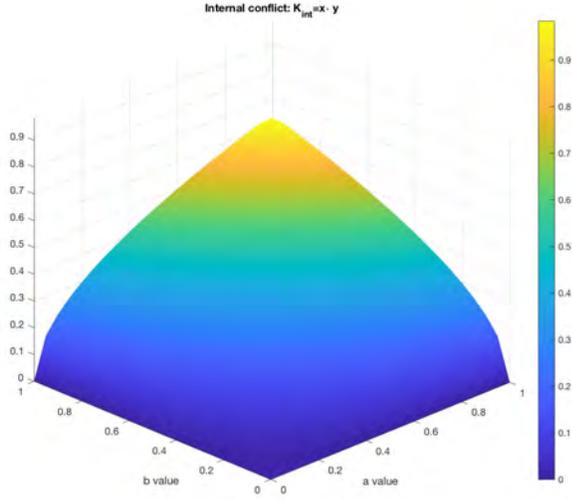


Figure 4. Internal conflict  $K_{int}(m)$ .

- 3) This canonical decomposition allows also to define the notion of level of uncertainty  $U(m)$  of a dichotomous source of evidence  $m(\cdot)$  as the conjunction of the uncertainties of pro and contra evidences, that is

$$\begin{aligned} U(m) &\triangleq m_p(A \cup \bar{A})m_c(A \cup \bar{A}) \\ &= (1-x)(1-y) = 1-x-y+xy \\ &= 1-x-y+K_{int}(m). \end{aligned} \quad (56)$$

Because of PCR5-based decomposition one gets (as already shown in (31))  $U(m) = 1 - a - b$  which always belongs to  $[0, 1]$ . The formula (56) is interesting because it clearly shows the link between the pro-evidence value  $x$ , the contra-evidence value  $y$  and the internal conflict  $K_{int}(m) = xy$ . Clearly, if  $x = 0$  and  $y = 0$ , then  $K_{int}(m) = 0$  and the uncertainty is maximal (i.e.  $U(m) = 1$ ) because the dichotomous BBA  $m$  is the vacuous BBA  $m(A \cup \bar{A}) = 1$ . It can be verified that a dichotomous BBA  $m$  has no uncertainty ( $U(m) = 0$ ) if and only if  $x = 1$ , or  $y = 1$ , or both which means that  $m(\cdot)$  is a Bayesian dichotomous BBA.

- 4) The canonical decomposition allows also to adjust/revise easily a dichotomous source of evidence (if needed) according the knowledge one has on it. For instance, suppose one knows that the source which provides the BBA  $m(\cdot)$  usually over estimates with a reinforcement factor of  $\beta_p = 20\%$  the belief mass committed to hypothesis  $A$  but is always fair (unbiased) when committing its mass to  $\bar{A}$ . Under this condition, we make the canonical decomposition of  $m(\cdot)$  to get  $m_p(\cdot)$

and  $m_c(\cdot)$  and we have to discount<sup>18</sup> the pro-evidence  $m_p(\cdot)$  with the discounting rate of  $\alpha_p = 1/(1 + \beta_p)$  to get the new unbiased BBA  $m'_p(\cdot)$  and keep the contra-evidence  $m_c(\cdot)$  unchanged, so that the corrected (unbiased) BBA  $m'(\cdot)$  will be obtained by the PCR5 combination of  $m'_p(\cdot)$  with  $m_c(\cdot)$ . Of course similar principles can be applied to discount (or reinforce)  $m_c(\cdot)$  as we prefer (and when necessary) by choosing the adequate discounting (or reinforcing) factors.

- 5) This canonical decomposition opens the door to new rules of combination for the fusion of  $S \geq 2$  (dichotomous) distinct<sup>19</sup> BBAs  $m_s(\cdot)$ ,  $s = 1, 2, \dots, S$ . After making their canonical decompositions to get  $S$  pro-evidences  $m_{p,s} = (m_{p,s}(A), m_{p,s}(\bar{A}), m_{p,s}(A \cup \bar{A}))$  equal to  $(x_s, 0, 1 - x_s)$ , and  $S$  contra-evidences  $m_{c,s} = (m_{c,s}(A), m_{c,s}(\bar{A}), m_{c,s}(A \cup \bar{A}))$  equal to  $(0, y_s, 1 - y_s)$  for  $s = 1, 2, \dots, S$ , one can for instance combine the  $S$  informative non-conflicting pro-evidences  $m_{p,s}$  altogether by the conjunctive rule (or any rule one prefers) to get the combined pro-evidence  $m_p(\cdot)$ , and do similarly to combine altogether the non conflicting contra-evidences  $m_{c,s}$  to get the combined contra-evidence  $m_c(\cdot)$ . Once  $m_p(\cdot)$  and  $m_c(\cdot)$  are calculated, we combine them with PCR5 to get the final resulting BBA. Processing this way will greatly simplify the combination of many dichotomous BBAs. Once the decomposition of each dichotomous BBA is done, we could also consider to apply some importance discounting [10] with rates  $\beta_s$  to combine separately the set of BBAs  $\{m_{p,s}, s = 1, \dots, S\}$  and the set of BBAs  $\{m_{c,s}, s = 1, \dots, S\}$  before making their PCR5 combination.

## IX. CONCLUSIONS

In this study, we have proved that any dichotomous basic belief assignment (nondogmatic, or dogmatic) can be decomposed into two simpler proper belief assignments called the pro-evidence and contra-evidence that can be combined with PCR5 rule to retrieve the original BBA. This canonical decomposition is unique and is always possible. No simple explicit form of the expression of the solution exists but the solution can be found quite easily with numerical solvers (Matlab, Maple, etc). We have also shown that the decomposition of any dichotomous basic belief assignment cannot be done in all the cases with other well-known rules of combination, which reinforce the interest of PCR5 principle for BF combination. This PCR5-based canonical decomposition allows also to establish the notion of internal conflict of a dichotomous source of evidence which could be helpful in some applications. It offers the possibility to combine several dichotomous sources of evidence based on the fusion of their canonical components. This will be presented in details in a forthcoming publication. The open challenging question is how to extend this notion of

<sup>18</sup>We use classical Shafer's discounting method [1].

<sup>19</sup>i.e., cognitively independent.

canonical decomposition for working with more general basic belief assignments to make their combination more effective (if possible), and how could we define a measure of (uncertain) information thanks to this canonical decomposition.

#### ACKNOWLEDGMENT

Jean Dezert is very grateful to Professor Christine Marois of the Faculté de Droit d'Économie et de Gestion of Orléans University, France, for her inspiration and valuable help for the establishment of the concise proof of Theorem 1.

#### REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [2] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, Vols. 1–4, American Research Press, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [3] P. Smets, *The canonical decomposition of a weighted belief*, in Proc. of Int. Joint Conf. on Artificial Intelligence, pp. 1896–1901, Morgan Kaufman, San Mateo, CA, USA, 1995.
- [4] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Fusion 2012 Int. Conf., Singapore, July 9–12, 2012.
- [5] A. Tchamova, J. Dezert, *On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory*, in Proc. of 6th IEEE Int. Conf. on Int. Syst., Sofia, Bulgaria, Sept. 6–8, 2012.
- [6] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intelligent Syst., Vol. 29(3), pp. 223–252, March 2014.
- [7] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of Fusion 2013 Int. Conf., Istanbul, Turkey, July 2013.
- [8] F. Smarandache, J. Dezert, *Information Fusion Based on New Proportional Conflict Redistribution Rules*, in Proc. of the 8th Int. Conf. on Information Fusion, Philadelphia, USA, 25-29 July, 2005.
- [9] J. Dezert, F. Smarandache, *Non Bayesian conditioning and deconditioning*, in Proc. of International Workshop on Belief Functions, Brest, France, April 2010.
- [10] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of Fusion 2010 Int. Conf., Edinburgh, UK, July 2010.
- [11] <https://bfasociety.org/>
- [12] T. Dencœur, *Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence*, Artificial Intelligence, Vol. 172, pp. 234–264, 2008.
- [13] R. Yager, *On the Dempster-Shafer framework and new combination rules*, Information Sciences, Vol. 41, pp. 93–138, 1987.
- [14] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Computational Intelligence, Vol. 4, pp. 244–264, 1988.
- [15] E. Weisstein, *Quartic Equation*, From MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/QuarticEquation.html>
- [16] L.A. Zadeh, *On the validity of Dempster's rule of combination*, ERL Memo M79/24, Department of EECS, Univ. of California, Berkeley, U.S.A., 1979.
- [17] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, Vol. 7(2), pp. 85–90, 1986.

# Canonical Decomposition of Basic Belief Assignment for Decision-Making Support

Jean Dezert<sup>a</sup>, Florentin Smarandache<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, F-91761 Palaiseau, France.

<sup>b</sup>Department of Mathematics, Univ. of New Mexico, Gallup, NM, USA.

Emails: jean.dezert@onera.fr, smarand@unm.edu

Originally published as: J. Dezert, F. Smarandache, *Canonical Decomposition of Basic Belief Assignment for Decision-Making Support*, in Proc. of 7th Int. Conf. on Modelling and Development of Intelligent Systems (MDIS 2020), Lucian Blaga Univ. of Sibiu, Sibiu, Romania, Oct. 22–24, 2020, and reprinted with permission.

**Abstract**—We present a new methodology for decision-making support based on belief functions thanks to a new theoretical canonical decomposition of dichotomous basic belief assignments (BBAs) that has been developed recently. This decomposition based on proportional conflict redistribution rule no 5 (PCR5) always exists and is unique. This new PCR5-based decomposition method circumvents the exponential complexity of the direct fusion of BBAs with PCR5 rule and it allows to fuse quickly many sources of evidences. The method we propose in this paper provides both a decision and an estimation of the quality of the decision made, which is appealing for decision-making support systems.

**Keywords:** Decision-Making, Belief Functions, PCR5.

## I. INTRODUCTION

This paper deals with the decision-making support problem from many sources of evidence characterized by belief functions (BF) defined over a same frame of discernment. Belief functions introduced by Shafer [1] are appealing to model epistemic uncertainty. They are well-known and used in the artificial intelligence community to fuse uncertain information and to make a decision. However, many debates in scientific community started with Zadeh’s criticism [2], [3] - see additional references in [4] - have bloomed on the validity of Dempster’s rule of combination and its counter-intuitive behavior (not only in high conflicting situations, but also in low conflicting situations as well). That is why many rules of combination have been developed by different researchers [5] (Vol. 2) over the last decades. In this work we consider only the rule based on the proportional conflict redistribution principle no 5 (PCR5 rule) to combine basic belief assignments (BBAs). This choice is motivated not only by its conflict redistribution principle, but also by its ability to generate a unique canonical decomposition of any dichotomous BBA that will be convenient for decision-making from many sources of evidence.

This paper is organized as follows. After a brief recall of basics of belief functions in Section II, we present succinctly the canonical decomposition of a (dichotomous) BBA in Section III based on [6]. Then we propose a new decision-making support methodology that exploits this canonical decomposition in Section IV for working in a general framework with many (non dichotomous) sources of evidences, with basic illustrative examples. Conclusions are given in Section V.

## II. BASICS OF BELIEF FUNCTIONS

### A. Definitions

The answer<sup>1</sup> of the problem under concern is supposed to belong to a given finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ . All elements of  $\Theta$  are mutually exclusive<sup>2</sup>. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A Basic Belief Assignment (BBA) given by a source of evidence is defined [1] as  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is the mass of belief of  $A$ . Belief and plausibility functions are respectively defined from  $m(\cdot)$  by

$$Bel(A) = \sum_{B \in 2^\Theta | B \subseteq A} m(B), \quad (1)$$

and

$$Pl(A) = \sum_{B \in 2^\Theta | A \cap B \neq \emptyset} m(B) = 1 - Bel(\bar{A}). \quad (2)$$

where  $\bar{A}$  is the complement of  $A$  in  $\Theta$ .

$Bel(A)$  and  $Pl(A)$  are usually interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$ .  $A$  is called a Focal Element (FE) of  $m(\cdot)$  if  $m(A) > 0$ . When all focal elements are singletons then  $m(\cdot)$  is called a *Bayesian BBA* [1] and its corresponding  $Bel(\cdot)$  function is equal to  $Pl(\cdot)$  and they are homogeneous to a (subjective) probability measure  $P(\cdot)$ . The vacuous BBA (VBBA for short) representing a totally ignorant source is defined as<sup>3</sup>  $m_v(\Theta) = 1$ . A dogmatic BBA is a BBA such that  $m(\Theta) = 0$ . If  $m(\Theta) > 0$  the BBA  $m(\cdot)$  is nondogmatic. A simple BBA is a BBA that has at most two focal sets and one of them is  $\Theta$ . A FoD is a dichotomous FoD if it has only two elements, say  $\Theta = \{A, \bar{A}\}$  with  $A \neq \emptyset$  and  $A \neq \Theta$ . A dichotomous BBA is a BBA defined over a dichotomous FoD.

### B. PCR5 Rule of Combination

The combination of distinct sources of evidence characterized by their BBAs is done by Dempster’s rule of combination in Shafer’s mathematical theory of evidence [1]. The

<sup>1</sup>I.e. the solution, or the decision to take.

<sup>2</sup>This is so-called Shafer’s model of FoD [5].

<sup>3</sup>The complete ignorance is denoted  $\Theta$  in Shafer’s book [1].

justification and behavior of Dempster's rule (corresponding to the normalized conjunctive rule) have been disputed from many counter-examples involving high and low conflicting sources (from both theoretical and practical standpoints) as reported in [4]. Many alternatives to Dempster's rule are now available [5], Vol. 2. Among them, we consider in the sequel the PCR5 rule which transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that a more sophisticated and precise distribution is done with the PCR5 fusion process. The PCR5 rule is presented in details (with justification and examples) in [5], Vol. 2 and Vol. 3. We only briefly recall for convenience its formula for the fusion of two BBAs, which is symbolically noted as  $m_{PCR5} = PCR5(m_1, m_2)$ , where  $PCR5(\cdot, \cdot)$  represents the PCR5 fusion rule for two BBAs. With this PCR5 rule, one has  $m_{PCR5}(\emptyset) = 0$ , and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{PCR5}(X) = m_{Conj}(X) + \sum_{\substack{X_2 \in 2^\Theta \\ X_2 \cap X = \emptyset}} \left[ \frac{m_1(X)^2 m_2(X_2)}{m_1(X) + m_2(X_2)} + \frac{m_2(X)^2 m_1(X_2)}{m_2(X) + m_1(X_2)} \right], \quad (3)$$

where  $m_{Conj}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1) m_2(X_2)$  is the conjunctive rule, and where all denominators in (3) are different from zero. If a denominator is zero, that fraction is discarded. Extension of PCR5 for combining qualitative BBA's can be found in [5], Vols. 2 & 3. All propositions/sets are in a canonical form. A variant of PCR5, called PCR6 has been proposed by Martin and Osswald in [5], Vol. 2, for combining  $s > 2$  sources. The general formulas for PCR5 and PCR6 rules are also given in [5], Vol. 2. PCR6 coincides with PCR5 when one combines two sources. The difference between PCR5 and PCR6 lies in the way the proportional conflict redistribution is done as soon as three (or more) sources are involved in the fusion.

### III. CANONICAL DECOMPOSITION OF A DICHOTOMOUS BASIC BELIEF ASSIGNMENT

Because the canonical decomposition of a dichotomous BBA has been presented in details in [6], we only make a succinct presentation here. A FoD is a dichotomous FoD if it is made of only two elements, say  $\Theta = \{A, \bar{A}\}$  with  $A \cup \bar{A} = \Theta$  and  $A \cap \bar{A} = \emptyset$ .  $A$  is different from  $\Theta$  and from Empty-Set because we want to work with informative FoD. A dichotomous BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  has the general form

$$m(A) = a, \quad m(\bar{A}) = b, \quad m(A \cup \bar{A}) = 1 - a - b, \quad (4)$$

with  $a, b \in [0, 1]$  and  $a + b \leq 1$ .

The canonical decomposition problem consists in finding the two following simpler BBAs  $m_p$  and  $m_c$  of the form

$$m_p(A) = x, \quad m_p(A \cup \bar{A}) = 1 - x, \quad (5)$$

$$m_c(\bar{A}) = y, \quad m_c(A \cup \bar{A}) = 1 - y, \quad (6)$$

with  $(x, y) \in [0, 1] \times [0, 1]$ , such that  $m = Fusion(m_p, m_c)$ , for a chosen rule of combination denoted by  $Fusion(\cdot, \cdot)$ . The simple BBA  $m_p(\cdot)$  is called the *pro-BBA* (or pro-evidence) of  $A$ , and the simple BBA  $m_c(\cdot)$  the *contra-BBA* (or contra-evidence) of  $A$ . The BBA  $m_p(\cdot)$  is interpreted as a source of evidence providing an uncertain evidence in favor of  $A$ , whereas  $m_c(\cdot)$  is interpreted as a source of evidence providing an uncertain contrary evidence about  $A$ . In [6], we proved that this decomposition always exists and is unique if we use the PCR5 fusion rule. In the vacuous BBA case when  $a = 0$  and  $b = 0$ , the BBA  $m(\cdot)$  can be interpreted as the PCR5 fusion of two degenerate pro- and contra-evidences BBAs  $m_p(\cdot)$  and  $m_c(\cdot)$  which coincide with the vacuous BBA with  $x = 0$  and  $y = 0$ . Hence any (Bayesian, or non Bayesian) dichotomous BBA  $m(\cdot)$  can be always interpreted as the result of the PCR5 fusion of these two (pros and cons) aspects of evidence about  $A$ . It is worth noting that this type of canonical decomposition is different of Smets' canonical decomposition problem [7] which needs to work with generalized simple BBA which are not *stricto sensu* valid BBAs as defined by Shafer [1].

For the case of dichotomous dogmatic BBA, the expression of solutions  $x$  and  $y$  of canonical decomposition are as follows [6]:

- if  $a = b$  and  $a + b = 1$  then  $a = b = 0.5$  and  $x = y = 1$ ;
- if  $a < b$  then  $x < y$ , and we have

$$\begin{cases} y = 1, \\ x = \frac{a + \sqrt{a^2 + 4a}}{2}; \end{cases}$$

- if  $a > b$  then  $x > y$ , and we have

$$\begin{cases} x = 1, \\ y = \frac{b + \sqrt{b^2 + 4b}}{2}. \end{cases}$$

For the case of dichotomous non-dogmatic BBA, the expression of solutions  $x$  and  $y$  of the canonical decomposition do not have simple analytical expression because one has to find  $x$  and  $y$  solutions of the system

$$a = x(1 - y) + \frac{x^2 y}{x + y} = \frac{x^2 + xy - xy^2}{x + y}, \quad (7)$$

$$b = (1 - x)y + \frac{xy^2}{x + y} = \frac{y^2 + xy - x^2 y}{x + y}, \quad (8)$$

under the constraints  $(a, b) \in [0, 1]^2$ , and  $0 < a + b < 1$ . In fact, we have proved in [6] that  $x \in [a, a + b] \subset [0, 1]$  and  $y \in [b, a + b] \subset [0, 1]$ , but the explicit expression of  $x$  and  $y$  are very complicated to obtain analytically (even with modern symbolic computing systems like Mathematica™, or Maple™) because after algebraic calculation, and for  $x \neq 1$ , one has to solve the following quartic equation which has at most four real solutions with only a valid one in  $[a, a + b]$

$$x^4 + (-a - 2)x^3 + (2a + b)x^2 + (a + b - ab - b^2)x + (-a^2 - ab) = 0, \quad (9)$$

and then compute  $y$  by  $y = (a + b - x)(1 - x)$ .

Once the numerical values are committed to  $a$  and to  $b$  the numerical (approximate) solutions  $x$  and then  $y$  can be easily obtained by a standard numerical solver. For instance, with Matlab™ we can use the `fsolve` command, and this is what we use to make the canonical decomposition of dichotomous non-dogmatic BBA.

### A. Canonical Decompositions From Other Well-Known Rules

In [6] we did prove that this type of canonical decomposition cannot be obtained by the conjunctive rule only, because if  $m_p$  and  $m_c$  exist and if  $x > 0$  and  $y > 0$  then  $m_{Conj}(\emptyset) = x \cdot y > 0$  which means that  $m = Conj(m_p, m_c)$  is not a proper BBA as defined by Shafer's. If we use the disjunctive rule of combination we will always obtain the vacuous BBA as the result<sup>4</sup> of  $Disj(m_p, m_c)$  because  $m_p(A)m_c(\bar{A})$ ,  $m_p(A)m_c(A \cup \bar{A})$ ,  $m_p(A \cup \bar{A})m_c(\bar{A})$  and  $m_p(A \cup \bar{A})m_c(A \cup \bar{A})$  will all be committed to the uncertainty  $A \cup \bar{A}$ . So for any choice of  $m_p$  and  $m_c$  we always get same result (the vacuous BBA) when using the disjunctive rule making the canonical decomposition of non vacuous dichotomous BBA  $m$  just impossible. Due to the particular simple form of BBAs  $m_p(\cdot)$  and  $m_c(\cdot)$ , Yager's rule [8] and Dubois-Prade rule [9] coincide, and we have to search  $x$  and  $y$  in  $[0, 1]$  such that  $m(A) = a = x(1 - y)$  and  $m(\bar{A}) = b = (1 - x)y$ . Assuming<sup>5</sup>  $y < 1$ , one gets from the first equation  $x = a/(1 - y)$ . By replacing  $x$  by its expression in the second equation  $y - xy = b$  we have to find  $y$  in  $[0, 1]$  such that (after basic algebraic simplifications)  $y^2 + (a - b - 1)y + b = 0$ . This 2nd order equation admits one or two real solutions  $y_1$  and  $y_2$  if and only if the discriminant is null or positive respectively, that is if  $(a - b - 1)^2 - 4b \geq 0$ . However this discriminant can become negative depending on the values of  $a$  and  $b$ . For instance, for  $a = 0.3$  and  $b = 0.6$ , we have  $(a - b - 1)^2 - 4b = -0.71$  which means that there is no real solution for the equation  $y^2 - 1.3 \cdot y + 0.6 = 0$ . Therefore, in general (that is for all possible values  $a$  and  $b$  of the BBA  $m$ ), the canonical decomposition of the BBA  $m(\cdot)$  cannot be obtained from Yager's and Dubois & Prade rules of combination. If we use the averaging rule, we are searching  $x$  and  $y$  in  $[0, 1]$  such that  $m(A) = a = (x + 0)/2$  and  $m(\bar{A}) = b = (0 + y)/2$ , which means that  $x = 2a$  and  $y = 2b$  with  $x$  and  $y$  in  $[0, 1]$ . So, if  $a > 0.5$  or  $b > 0.5$  the canonical decomposition is impossible to make with the averaging rule of combination. Therefore, in general, the averaging rule is not able to provide a canonical decomposition of the BBA  $m(\cdot)$ .

If we consider the canonical decomposition of a dichotomous non-dogmatic BBA ( $a + b < 1$ ) using Dempster's rule of combination [1], denoted  $DS(m_p, m_c)$ , we have to obtain

<sup>4</sup> $Disj(m_p, m_c)$  denotes the disjunctive fusion of  $m_p$  with  $m_c$ .

<sup>5</sup>Taking  $y = 1$  would mean that  $x(1 - y) = 0$  but  $m(A) = a$  with  $a \neq 0$  in general, so the choice of  $y = 1$  is not possible.

$x$  and  $y$  in  $[0, 1]$  such that<sup>6</sup>  $xy \neq 1$  and

$$m(A) = a = \frac{x(1 - y)}{1 - xy}, \quad (10)$$

$$m(\bar{A}) = b = \frac{y(1 - x)}{1 - xy}, \quad (11)$$

with the constraints  $0 < x < 1$  and  $0 < y < 1$ .

Therefore,

$$x = \frac{a}{1 - y + ay}, \quad y \neq \frac{1}{1 - a}, \quad (12)$$

and we solve the equation  $y - xy + bxy = b$  with  $x$  expressed as function of  $y$  as above. We get the equation for  $a \neq 1$

$$(a - 1)y^2 + (1 + b - a)y - b = 0, \quad (13)$$

whose two solutions are  $y_1 = b/(1 - a)$  and  $y_2 = 1$  - see [6] for details.

For the case  $a \neq 1$ , the second "solution"  $y_2 = 1$  implies  $x = \frac{a}{1 - y_2 + ay_2} = \frac{a}{a} = 1$  which is not an acceptable solution<sup>7</sup> because one must have  $xy \neq 1$ . The solution  $(x, y)$  of the decomposition problem for  $a \neq 1$  is actually given by the first solution  $y_1$ , that is

$$y = y_1 = \frac{b}{1 - a} \in [0, 1), \quad (14)$$

$$x = \frac{a}{1 - y + ay} = \frac{a}{1 - b} \in [0, 1). \quad (15)$$

The analysis of the case  $a = 1$  corresponding to the dogmatic BBA given by  $m(A) = a = 1$ ,  $m(\bar{A}) = b = 0$ ,  $m(A \cup \bar{A}) = 1 - a - b = 0$  shows that this BBA is not canonically decomposable by Dempster's rule. Why? Because one has to solve with  $0 \leq x, y \leq 1$  and  $1 - xy \neq 0$  the system of equations  $(x - xy)/(1 - xy) = 1$  and  $(y - xy)(1 - xy) = 0$  which is satisfied for  $x = 1$  and  $y \in [0, 1)$ , that is any value in  $[0, 1)$  can be chosen for  $y$ . Similarly, for the case  $(a, b) = (0, 1)$  one has to solve with  $0 \leq x, y \leq 1$  and  $1 - xy \neq 0$  the system of equations  $(x - xy)/(1 - xy) = 0$  and  $(y - xy)/(1 - xy) = 1$  which is satisfied for  $y = 1$  and  $x \in [0, 1)$ , that is any  $x$  value in  $[0, 1)$  can be chosen. Therefore one sees that for the case  $(a, b) = (1, 0)$  and the case  $(a, b) = (0, 1)$  there is no unique decomposition of these dogmatic BBAs from Dempster's rule of combination. More generally, any dogmatic BBA  $m(A) = a$ ,  $m(\bar{A}) = b$  with  $a + b = 1$  is not decomposable from Dempster's rule of combination for the case when  $(a, b) \neq (1, 0)$  and  $(a, b) \neq (0, 1)$  - See Theorem 4 with its proof in [6].

In summary, the canonical decomposition based on Dempster's rule of combination is possible only for nondogmatic BBA with  $0 < a < 1$ ,  $0 < b < 1$  and  $a + b < 1$  and we have  $x = \frac{a}{1 - b}$  and  $y = \frac{b}{1 - a}$ . Dempster's rule does not allow to obtain a canonical decomposition if the BBA is a Bayesian (dogmatic) dichotomous BBA.

<sup>6</sup>The third equality  $m(A \cup \bar{A}) = 1 - a - b = \frac{(1 - x)(1 - y)}{1 - xy}$  being redundant with (10) and (11) is useless.

<sup>7</sup>Otherwise the denominator of (10) and (11) will equal zero.

### Example where Dempster's canonical decomposition is possible

Consider  $m(A) = a = 0.6$ ,  $m(\bar{A}) = b = 0.2$  and  $m(A \cup \bar{A}) = 1 - a - b = 0.2$ . The solution  $(x, y)$  of the decomposition of  $m(\cdot)$  based on Dempster's rule is

$$x' = \frac{a}{1-b} = \frac{0.6}{1-0.2} = 0.75,$$

and

$$y' = \frac{b}{1-a} = \frac{0.2}{1-0.6} = 0.50.$$

Therefore, the pro- and contra- evidential BBAs  $m_p$  and  $m_c$  are given by

$$\begin{aligned} m_p(A) &= x = 0.75, & m_p(A \cup \bar{A}) &= 1 - x = 0.25, \\ m_c(\bar{A}) &= y = 0.50, & m_c(A \cup \bar{A}) &= 1 - y = 0.50. \end{aligned}$$

It can be verified that  $DS(m_p, m_c) = m$ .

If we make the PCR5-based canonical decomposition, we will obtain in this example  $x \approx 0.6861$  and  $y \approx 0.3628$ . Therefore, the pro- and contra- evidential BBAs  $m_p$  and  $m_c$  based on the PCR5-based canonical decomposition are

$$\begin{aligned} m_p(A) &= x = 0.6861, & m_p(A \cup \bar{A}) &= 1 - x = 0.3139, \\ m_c(\bar{A}) &= y = 0.3628, & m_c(A \cup \bar{A}) &= 1 - y = 0.6372. \end{aligned}$$

It can be verified that  $PCR5(m_p, m_c) = m$ .

In the case where Dempster's rule can be applied for making the canonical decomposition (that is when  $a + b < 1$ ) we see that the canonical values (parameters)  $x$  and  $y$  can be very different from those obtained with PCR5 rule as shown in the previous example. This is normal because the principles of conflicting information redistribution of Dempster's rule and PCR5 rule are very different, and there is no link between parameters  $x$  and  $y$  obtained with Dempster's rule versus those obtained from PCR5. In PCR5 rule the conflict is a refined conflict, i.e. the conflict is split into partial conflicts, so in PCR5 the total conflict is more accurately redistributed than in Dempster's rule because each partial conflict is redistributed only to the elements involved into it, while in Dempster's rule the total conflict is redistributed to all focal elements, therefore even the elements that were not involved in the conflict receive conflicting mass, which is inaccurate.

It is worth noting that the internal conflict of  $m$  based on Dempster's rule will be in this example  $xy = 0.75 \cdot 0.5 = 0.375$ , whereas the internal conflict of  $m$  based on PCR5 rule will be only  $xy \approx 0.6861 \cdot 0.3628 \approx 0.2489$ . In fact we can attest that the internal conflict obtained from PCR5-based canonical decomposition is always lesser (or equal) to the internal conflict obtained from Dempster-based canonical decomposition. Although such claim cannot be proved algebraically<sup>8</sup>, we can always make a fine sampling of  $(a, b)$  values in  $[0, 1)$  satisfying  $a + b < 1$  to evaluate numerically

<sup>8</sup>Because there is no simple analytical expressions for solutions  $x$  and  $y$  of PCR5-based canonical decomposition.

$x$  and  $y$  and compare the internal conflict  $xy$  to the internal conflict, denoted  $x'y' = \frac{a}{1-b} \cdot \frac{b}{1-a}$ , obtained with Dempster-based canonical decomposition. In doing this we see that the difference  $\Delta = x'y' - xy$  is always greater (or equal) to zero as clearly shown in Figure 1. This means that the PCR5-based canonical decomposition is more efficient than Dempster-based canonical decomposition because it always yield pro- and contra-evidences which are less conflicting when using PCR5 rule than when using Dempster's rule, which is normal.

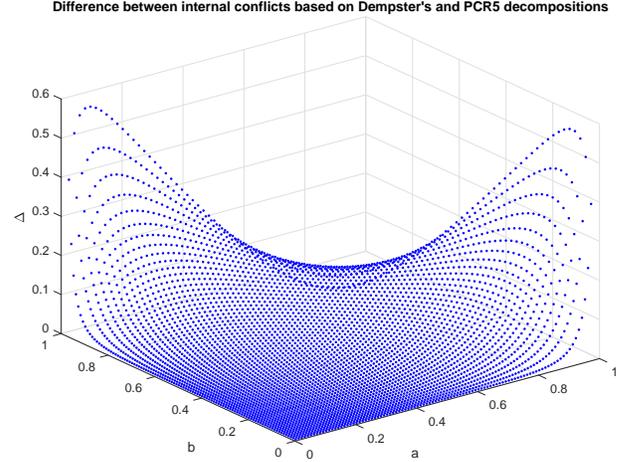


Figure 1. Plot of  $\Delta = x'y' - xy$  as function of  $a$  and  $b$ .

It is important to keep in mind that Dempster-based canonical decomposition is only possible for non-dogmatic BBAs (when  $a + b < 1$ ) but cannot be obtained with dogmatic BBAs, whereas PCR5-based canonical decomposition works for all types of dichotomous BBAs (dogmatic and non-dogmatic ones).

### B. Simple Example of PCR5-Based Canonical Decomposition

Let consider  $m(A) = 0.3$ ,  $m(\bar{A}) = 0.4$  and  $m(A \cup \bar{A}) = 1 - m(A) - m(\bar{A}) = 0.3$ , therefore  $a = 0.3$  and  $b = 0.4$ . The quartic equation (9) becomes

$$x^4 - 2.3x^3 + x^2 + 0.42x - 0.21 = 0. \quad (16)$$

The four solutions of this quartic equation are approximately<sup>9</sup>

$$\begin{aligned} x_1 &\approx 1.5203, \\ x_2 &\approx -0.4243, \\ x_3 &\approx 0.7942, \\ x_4 &\approx 0.4099. \end{aligned}$$

One sees that  $x_1$  and  $x_2$  are not acceptable solutions because they do not belong to  $[0, 1]$ . If we take  $x_3 \approx 0.7942$  then will get  $y_3 = (a + b - x_3)/(1 - x_3) = (0.7 - x_3)/(1 - x_3) \approx -0.4576$ . We see that  $y_3 \notin [0, 1]$  and therefore the pair  $(x_3, y_3)$  cannot be a solution of the PCR5-based canonical

<sup>9</sup>The solutions can be easily obtained with the `roots` command of Matlab™.

decomposition problem for the BBA  $m(\cdot)$  of this example. If we take  $x_4 \approx 0.4099$  then will get  $y_4 = (a+b-x_4)/(1-x_4) = (0.7-x_4)/(1-x_4) \approx 0.4916$  which belongs to  $[0, 1]$ . So the pair  $(x_4, y_4) \in [0, 1]^2$  is the unique solution of the canonical decomposition problem. Therefore the canonical masses  $m_p(\cdot)$  and  $m_c(\cdot)$  are given by

$$m_p(A) \approx 0.4099, \quad m_p(A \cup \bar{A}) \approx 0.5901,$$

and

$$m_c(\bar{A}) \approx 0.4916, \quad m_c(A \cup \bar{A}) \approx 0.5084.$$

It can be verified that  $PCR5(m_p, m_c) = m$ .

### C. Advantages and Limitation of PCR5-Based Decomposition

The PCR5-based canonical decomposition offers the following advantages:

- 1) It is well justified theoretically.
- 2) It gives us access to the simpler pro- and contra-evidences  $m_p(\cdot)$  and  $m_c(\cdot)$  which are unique and always exist for any possible (dogmatic, or non-dogmatic) dichotomous BBA  $m(\cdot)$ .
- 3) It allows to define clearly the notion of internal conflict of a dichotomous source of evidence simply as  $K_{int}(m) \triangleq m_p(A)m_c(\bar{A})$ .
- 4) It always provides less conflicting pro- and contra-evidences than what we would obtain with Dempster's rule when considering non-dogmatic dichotomous BBA  $m(\cdot)$ . This proves the superiority of PCR5-based canonical decomposition over Dempster's-based canonical decomposition in general.
- 5) It allows also to adjust or revise<sup>10</sup> quite easily a dichotomous source of evidence (if needed) according to the knowledge one has on it by reinforcing or discounting its pro- or contra-evidential BBA.
- 6) It can be easily achieved with classical numerical solvers on the shelf.
- 7) The decomposition can be done off-line for many sampled  $(a, b)$  values at any precision we want, and stored in computer memory for working directly with  $m_p(\cdot)$  and  $m_c(\cdot)$  instead of making the decomposition on the fly. This is of prime importance for real-time applications where this method could be used.
- 8) It allows to establish efficient fast<sup>11</sup> suboptimal PCR5 fusion scheme, see [10] for details, examples and evaluations.

The only important limitation of this PCR5-based canonical decomposition is that it applies only to dichotomous BBAs, and it seems very difficult (maybe impossible) to use or to extend it for making directly some new canonical decomposition of non dichotomous BBAs. Because of this limitation the use of PCR5-based canonical decomposition appears, at first glance, quite restrictive for being really useful in applications involving non dichotomous BBAs. Of course

in applications working with dichotomous BBAs (like those in robotics or for autonomous vehicle navigation using belief-based perception based on grid occupancy) this PCR5-based canonical decomposition may have a great interest. In fact we have already used it for belief-based inter-criteria analysis in [11] and that is why we do not present our results in this work. Nevertheless we will show in the next section how this PCR5-based canonical decomposition could be used for the decision-making support in a more general context involving many non-dichotomous BBAs. This is a problem which has not been addressed in [6].

## IV. DECISION-MAKING USING PCR5-BASED DECOMPOSITION

In this section we propose a new simple general decision-making scheme based on PCR5-based canonical decomposition of dichotomous BBA. We consider  $S > 2$  distinct sources of evidence characterized by their BBAs<sup>12</sup>  $m_s^\Theta(\cdot)$  defined over the same (possibly non dichotomous) FoD  $\Theta = \{\theta_1, \dots, \theta_n\}$ , with  $n > 1$ .

Can we exploit the PCR5-based canonical decomposition in this context to make a decision? How? We answer positively to the first question and explain in details how we can proceed. For this, we need to express the problem in the framework of dichotomous BBAs that has been presented in the previous section. More precisely, suppose one has a BBA  $m^\Theta(\cdot)$  defined on  $2^\Theta$  with  $|\Theta| \geq 2$ , then based on Bel and Pl formulas (1)-(2), it is always possible to calculate  $Bel^\Theta(X)$  and  $Pl^\Theta(X)$  for any  $X \in 2^\Theta$ . From  $Bel^\Theta(X)$  and  $Pl^\Theta(X)$  one can always build a simpler coarsened dichotomous BBA on the dichotomous (coarsened) FoD  $\Theta_X \triangleq \{X, \bar{X}\}$  if  $X \neq \emptyset$  and  $X \neq \Theta_X$  as follows

$$m^{\Theta_X}(X) = Bel^\Theta(X), \quad (17)$$

$$m^{\Theta_X}(\bar{X}) = 1 - Pl^\Theta(X), \quad (18)$$

$$m^{\Theta_X}(X \cup \bar{X}) = Pl^\Theta(X) - Bel^\Theta(X). \quad (19)$$

Hence,  $Bel^{\Theta_X}(X) = m^{\Theta_X}(X) = Bel^\Theta(X)$  and  $Pl^{\Theta_X}(X) = m^{\Theta_X}(X) + m^{\Theta_X}(X \cup \bar{X}) = Bel^\Theta(X) + Pl^\Theta(X) - Bel^\Theta(X) = Pl^\Theta(X)$ . This dichotomous BBA  $m^{\Theta_X}(\cdot)$  can always be decomposed canonically into its pro- and contra-evidences  $m_p^{\Theta_X}(\cdot)$  and  $m_c^{\Theta_X}(\cdot)$ .

Therefore, instead of combining  $S > 1$  non dichotomous BBAs  $m_s^\Theta(\cdot)$  for  $s = 1, 2, \dots, S$  altogether from which a decision is classically drawn, we propose to make the decision from the set of all combined coarsened BBAs relatively to each possible dichotomous frame of discernment  $\Theta_X$ . Of course this decision-scheme is only suboptimal because the whole information is not processed (combined) altogether, but separately using only the coarsened (less informative) BBAs  $m_s^{\Theta_X}(X)$ . However, this method allows to use fast suboptimal PCR5 fusion of  $m_s^{\Theta_X}(X)$  thanks to PCR5-based canonical decomposition as presented in [10] which can be applied with many (hundreds or even thousands) sources of dichotomous

<sup>10</sup>This point is not detailed here because is out of the scope of this paper.

<sup>11</sup>Where the complexity is linear with the number of dichotomous BBAs to fuse.

<sup>12</sup>For clarity, we need to introduce in the notations a superscript to indicate the FoD we are working on.

BBAs. With this simple suboptimal decision-scheme we can easily restrict the domain  $\mathcal{D}$  on which the decisions can be made, for instance  $\mathcal{D}$  can be chosen as the set of singletons of  $2^\Theta$ , or any other subset of  $2^\Theta$  depending on the application under concern as it will be shown in the next section. The generic steps of the method we propose are as follows:

- **Inputs:** BBAs  $m_s^\Theta(\cdot)$ ,  $s = 1, \dots, S$ , and the decision domain  $\mathcal{D} \subset 2^\Theta$ .
- **Step 1:** For  $s = 1, \dots, S$ , coarsening of  $m_s^\Theta(\cdot)$  into dichotomous BBA  $m_s^{\Theta_X}(\cdot)$ , for each  $X \in \mathcal{D}$  based on (17)-(19).
- **Step 2:** For  $s = 1, \dots, S$ , PCR5-based canonical decomposition of  $m_s^{\Theta_X}(\cdot)$  to get pro- and contra-evidences  $m_{p,s}^{\Theta_X}(\cdot)$  and  $m_{c,s}^{\Theta_X}(\cdot)$ .
- **Step 3:** Conjunctive fusion of all the pro-evidences  $m_{p,s}^{\Theta_X}(\cdot)$  to get  $m_p^{\Theta_X}(\cdot)$ .
- **Step 4:** Conjunctive fusion of all the contra-evidences  $m_{c,s}^{\Theta_X}(\cdot)$  to get  $m_c^{\Theta_X}(\cdot)$ .
- **Step 5:** PCR5 fusion of  $m_p^{\Theta_X}(\cdot)$  with  $m_c^{\Theta_X}(\cdot)$  to get  $m_{PCR5}^{\Theta_X}(\cdot)$  for  $X \in \mathcal{D}$ .
- **Step 6:** Decision-making from the set of the combined coarsened dichotomous BBAs  $\{m_{PCR5}^{\Theta_X}(\cdot), X \in \mathcal{D}\}$  to get the final decision  $\hat{X} \in \mathcal{D}$ .
- **Output:** the final decision  $\hat{X} \in \mathcal{D}$

In steps 3 and 4 we use the conjunctive fusion because there is no conflict between all pro-evidences  $m_{p,s}^{\Theta_X}(\cdot)$ , and there is also no conflict between all contra-evidences  $m_{c,s}^{\Theta_X}(\cdot)$ ,  $s = 1, \dots, S$ . The steps 1 to 5 do not require high computational burden and they can be done very quickly, specially if PCR5-based decompositions have been done off-line (as they should be) [10].

We must detail a bit more the principle of the decision-making for the step 6. Actually, the decision-making for step 6 can be interpreted as a decision-making problem from a set or coarsened BBAs  $m_{PCR5}^{\Theta_X}(\cdot)$  defined over different dichotomous FoD  $\Theta_X$  which are all the different coarsenings of the whole (refined original) FoD  $\Theta$ . In this paper we propose two methods to make the decision from the set of coarsened BBAs  $\{m_{PCR5}^{\Theta_X}(\cdot), X \in \mathcal{D}\}$ .

#### A. Method 1 for Step 6

This method is very simple. We take the decision  $\hat{X}$  corresponding to the largest value of  $m_{PCR5}^{\Theta_X}(X)$ , that is

$$\hat{X} = \arg \max_{X \in \mathcal{D}} (m_{PCR5}^{\Theta_X}(X)) \quad (20)$$

If there exist several arguments having the largest value (i.e. there is a tie), we select the one whose  $m_{PCR5}^{\Theta_X}(\bar{X})$  is smaller.

**Example 1 (without tie):** Suppose  $\Theta = \{A, B, C, D, E\}$  and we want to make a decision/choice only among the elements

of  $\mathcal{D} = \{A, B, C\}$ . Suppose after applying steps 1-5 we get the following 3 BBAs

$$\begin{aligned} m_{PCR5}^{\Theta_A}(A) &= 0.3, m_{PCR5}^{\Theta_A}(\bar{A}) = 0.2, m_{PCR5}^{\Theta_A}(A \cup \bar{A}) = 0.5, \\ m_{PCR5}^{\Theta_B}(B) &= 0.1, m_{PCR5}^{\Theta_B}(\bar{B}) = 0.5, m_{PCR5}^{\Theta_B}(B \cup \bar{B}) = 0.4, \\ m_{PCR5}^{\Theta_C}(C) &= 0.4, m_{PCR5}^{\Theta_C}(\bar{C}) = 0.3, m_{PCR5}^{\Theta_C}(C \cup \bar{C}) = 0.3. \end{aligned}$$

The decision will be  $\hat{X} = C$  because

$$m_{PCR5}^{\Theta_C}(C) > m_{PCR5}^{\Theta_A}(A) > m_{PCR5}^{\Theta_B}(B).$$

**Example 2 (with tie)** We consider same  $m_{PCR5}^{\Theta_B}(\cdot)$  and  $m_{PCR5}^{\Theta_C}(\cdot)$  as in example 1 but  $m_{PCR5}^{\Theta_A}(\cdot)$  is given by

$$m_{PCR5}^{\Theta_A}(A) = 0.4, m_{PCR5}^{\Theta_A}(\bar{A}) = 0.2, m_{PCR5}^{\Theta_A}(A \cup \bar{A}) = 0.4.$$

In this case, there is a tie between  $A$  and  $C$  because  $m_{PCR5}^{\Theta_A}(A) = m_{PCR5}^{\Theta_C}(C) = 0.4$ . But because  $m_{PCR5}^{\Theta_A}(\bar{A}) < m_{PCR5}^{\Theta_C}(\bar{C})$  we will take  $\hat{X} = A$  as the final decision.

The interest of this method is above all its simplicity, but it does not allow to quantify the quality (trustfulness) of the decision which is often useful and required in decision-making support systems, and that is why we propose a second method for the decision-making of step 6.

#### B. Method 2 for Step 6

This second method is a bit more sophisticated but it circumvents the exponential complexity of the direct PCR6 fusion of  $S \geq 2$  BBAs defined on non dichotomous FoD  $\Theta$ . Once the step 5 is accomplished we propose to fuse altogether the (coarsened) dichotomous  $m_{PCR5}^{\Theta_X}(\cdot)$  and to apply the decision-making method based on the distance between the belief intervals [12]. Because the fusion must operate on the same common frame, we need just to express each BBA  $m_{PCR5}^{\Theta_X}(\cdot)$  as a dichotomous BBA on  $\Theta$  which is denoted  $m_{PCR5}^{\Theta_X \uparrow \Theta}(\cdot)$ . This is done very easily by just expressing each  $\bar{X}$  as the disjunction of all elements of  $\Theta$  included in  $\bar{X}$ . The fusion of BBAs  $m_{PCR5}^{\Theta_X \uparrow \Theta}(\cdot)$  is done by the weighted averaging rule of combination, where each weighting factor depends on the decision-making easiness of the BBA  $m_{PCR5}^{\Theta_X}(\cdot)$  to fuse. The easier the decision-making, the higher the weighting factor. We summarize this method 2:

- 1) For each  $X \in \mathcal{D}$ , establish  $m_{PCR5}^{\Theta_X \uparrow \Theta}(\cdot)$  from  $m_{PCR5}^{\Theta_X}(\cdot)$
- 2) For each  $X \in \mathcal{D}$ , compute the weighting factor  $w(X)$  of  $m_{PCR5}^{\Theta_X \uparrow \Theta}(\cdot)$  by

$$w(X) = \frac{1}{C} (1 - h(m_{PCR5}^{\Theta_X \uparrow \Theta})), \quad (21)$$

where  $C$  is a normalization factor given by  $C = \sum_{X \in \mathcal{D}} (1 - h(m_{PCR5}^{\Theta_X \uparrow \Theta}))$ , and where  $h(m_{PCR5}^{\Theta_X \uparrow \Theta}) = H(m_{PCR5}^{\Theta_X \uparrow \Theta}) / H_{\max} \in [0, 1]$  is the normalized pignistic entropy of the BBA  $m_{PCR5}^{\Theta_X \uparrow \Theta}$  defined by  $H(m_{PCR5}^{\Theta_X \uparrow \Theta}) = - \sum_{X \in 2^\Theta} \text{BetP}(X) \log_2(\text{BetP}(X))$  and  $\text{BetP}(X)$  is the pignistic probability of  $X$  [13], and  $H_{\max} = \log_2 |\Theta|$ .

- 3) Make the weighting average of  $m_{PCR5}^{\Theta_X \uparrow \Theta}(\cdot)$  for all  $X \in \mathcal{D}$  to get the BBA

$$m^\Theta(\cdot) = \sum_{X \in \mathcal{D}} w(X) m_{PCR5}^{\Theta_X \uparrow \Theta}(\cdot). \quad (22)$$

- 4) From  $m^\Theta(\cdot)$  make the decision based on minimum of belief-interval distance [12], that is

$$\hat{X} = \arg \min_{X \in \mathcal{D}} d_{BI}(m^\Theta, m_X^\Theta), \quad (23)$$

where  $m_X^\Theta$  is the BBA focused on  $X$  that is  $m_X^\Theta(X) = 1$  and  $m_X^\Theta(Y) = 0$  if  $Y \neq X$ , and where  $d_{BI}(\cdot, \cdot)$  is the belief-interval distance defined by (see [12] for details, justification and examples)

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))}, \quad (24)$$

where  $N_c = 1/2^{|\Theta|-1}$  is a normalization factor to have  $d_{BI}(m_1, m_2) \in [0, 1]$ , and  $d_W(BI_1(X), BI_2(X))$  is the Wassertein's distance between belief intervals

$$BI_1(X) \triangleq [Bel_1(X), Pl_1(X)] = [a_1, b_1],$$

and

$$BI_2(X) \triangleq [Bel_2(X), Pl_2(X)] = [a_2, b_2],$$

given by

$$d_W([a_1, b_1], [a_2, b_2]) \triangleq \left[ \left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2 \right]^{\frac{1}{2}}.$$

- 5) The quality (or trustfulness) of the decision is given by

$$q(\hat{X}) \triangleq 1 - \frac{d_{BI}(m, m_{\hat{X}})}{\sum_{X \in \mathcal{D}} d_{BI}(m, m_X)}. \quad (25)$$

$q(\hat{X}) \in [0, 1]$  becomes maximum (equal to one) when  $d_{BI}(m^\Theta, m_{\hat{X}}^\Theta)$  is zero, which means that  $m^\Theta(\cdot)$  is focused only on  $\hat{X}$ . The higher  $q(\hat{X})$  is, the more confident in the decision  $\hat{X}$  we are. When there exists a tie between multiple decisions  $\{\hat{X}_j, j > 1\}$ , then the prudent decision corresponding to their disjunction  $\hat{X} = \cup_j \hat{X}_j$  should be preferred (if allowed), or we can apply the method 1 to resolve the tie, or in desperation select randomly  $\hat{X}$  among the elements  $\hat{X}_j$  involved in the tie.

Of course we could adopt a more complicate method where the averaging fusion could operate on all the possible dichotomous BBAs related with each element  $X \in 2^{\Theta \setminus \{\theta, \Theta\}}$  instead of  $X \in \mathcal{D}$ , but this would substantially increase the computational burden. Because the decision  $\hat{X}$  must be constrained to belong to  $\mathcal{D}$ , we restrict the fusion to be applied only for the dichotomous BBAs related to these elements only. By doing this we can reduce substantially the computational burden if  $|\mathcal{D}|$  is much lesser than  $2^{|\Theta|}$ .

For convenience, we show how works the method 2 in the previous Example 1 using the same  $\Theta$  and  $\mathcal{D} = \{A, B, C\}$ .

We have to make the weighted average of the three following BBAs

$$\begin{cases} m_{PCR5}^{\Theta_A \uparrow \Theta}(A) = 0.3, \\ m_{PCR5}^{\Theta_A \uparrow \Theta}(B \cup C \cup D \cup E) = 0.2, \\ m_{PCR5}^{\Theta_A \uparrow \Theta}(A \cup \bar{A} = \Theta) = 0.5, \end{cases}$$

$$\begin{cases} m_{PCR5}^{\Theta_B \uparrow \Theta}(B) = 0.1, \\ m_{PCR5}^{\Theta_B \uparrow \Theta}(A \cup C \cup D \cup E) = 0.5, \\ m_{PCR5}^{\Theta_B \uparrow \Theta}(B \cup \bar{B} = \Theta) = 0.4, \end{cases}$$

$$\begin{cases} m_{PCR5}^{\Theta_C \uparrow \Theta}(C) = 0.4, \\ m_{PCR5}^{\Theta_C \uparrow \Theta}(A \cup B \cup D \cup E) = 0.3, \\ m_{PCR5}^{\Theta_C \uparrow \Theta}(C \cup \bar{C} = \Theta) = 0.3, \end{cases}$$

with  $B \cup C \cup D \cup E = \bar{A}$ ,  $A \cup C \cup D \cup E = \bar{B}$  and  $A \cup B \cup D \cup E = \bar{C}$ . The pignistic entropies are respectively equal to  $H(m_{PCR5}^{\Theta_A \uparrow \Theta}) \approx 2.1710$ ,  $H(m_{PCR5}^{\Theta_B \uparrow \Theta}) \approx 2.3201$  and  $H(m_{PCR5}^{\Theta_C \uparrow \Theta}) \approx 2.0754$ , and their normalized values are  $h(A) \approx 2.1710/2.3219 = 0.9350$ ,  $h(B) \approx 2.3201/2.3219 = 0.9992$  and  $h(C) \approx 2.0754/2.3219 = 0.8938$ . From Eq. (21) we get the weighting factors  $w(A) \approx 0.37803$ ,  $w(B) \approx 0.00463$  and  $w(C) \approx 0.61734$ , and the weighted average BBA is

$$\begin{aligned} m^\Theta(A) &= w(A)m_{PCR5}^{\Theta_A \uparrow \Theta}(A) + w(B) \cdot 0 + w(C) \cdot 0 \approx 0.1134, \\ m^\Theta(B) &= w(A) \cdot 0 + w(B)m_{PCR5}^{\Theta_B \uparrow \Theta}(B) + w(C) \cdot 0 \approx 0.0005, \\ m^\Theta(C) &= w(A) \cdot 0 + w(B) \cdot 0 + w(C)m_{PCR5}^{\Theta_C \uparrow \Theta}(C) \approx 0.2469, \\ m^\Theta(B \cup C \cup D \cup E) &= w(A)m_{PCR5}^{\Theta_A \uparrow \Theta}(B \cup C \cup D \cup E) \\ &\quad + w(B) \cdot 0 + w(C) \cdot 0 \approx 0.0756, \\ m^\Theta(A \cup C \cup D \cup E) &= w(A) \cdot 0 \\ &\quad + w(B)m_{PCR5}^{\Theta_B \uparrow \Theta}(A \cup C \cup D \cup E) \\ &\quad + w(C) \cdot 0 \approx 0.0023, \\ m^\Theta(A \cup B \cup D \cup E) &= w(A) \cdot 0 + w(B) \cdot 0 \\ &\quad + w(C)m_{PCR5}^{\Theta_C \uparrow \Theta}(A \cup B \cup D \cup E) \\ &\quad \approx 0.1852, \\ m^\Theta(\Theta) &= w(A)m_{PCR5}^{\Theta_A \uparrow \Theta}(\Theta) + w(B)m_{PCR5}^{\Theta_B \uparrow \Theta}(\Theta) \\ &\quad + w(C)m_{PCR5}^{\Theta_C \uparrow \Theta}(\Theta) = 0.3761. \end{aligned}$$

From Eq. (24) we get

$$\begin{cases} d_{BI}(m^\Theta, m_A^\Theta) \approx 0.6818, \\ d_{BI}(m^\Theta, m_B^\Theta) \approx 0.7541, \\ d_{BI}(m^\Theta, m_C^\Theta) \approx 0.5874. \end{cases}$$

Because  $d_{BI}(m^\Theta, m_C^\Theta) < d_{BI}(m^\Theta, m_A^\Theta) < d_{BI}(m^\Theta, m_B^\Theta)$ , the final decision must be  $\hat{X} = C$  because it corresponds to the smallest  $d_{BI}$  distance value. This decision is the same as with method 1. Based on Eq. (25) one has  $q(\hat{X} = C) \approx 0.7096$  indicating a pretty good trustful decision because it is much greater than 0.5. If one have preferred  $\hat{X} = A$  (the second best choice) then  $q(\hat{X} = A) \approx 0.6630$  which is a bit worse, and for  $\hat{X} = B$  one gets the least trustful decision because  $q(\hat{X} = B) \approx 0.6273$ . Note that a

more optimistic attitude (if preferred) could be obtained by replacing the BetP probability by the DSMP probability [5] (Chap. 3 of Vol. 3) in the entropy derivation.

## V. CONCLUSIONS

In this work we have presented a very new methodology for decision-making under uncertainty in the framework of belief functions thanks to the unique PCR5-based canonical decomposition of any (dogmatic or non-dogmatic) dichotomous BBAs. We have shown that this new canonical decomposition provides less conflicting contra- and pro-evidences with respect to the decomposition based on Dempster's rule when the latter can be applied. Any BBAs defined on a general (non dichotomous) frame of discernment can be transformed into a set of coarsened dichotomous BBAs that can always be decomposed canonically and combined easily and quickly in one PCR5 fusion step to get a suboptimal fusion result for each element of the decision space under consideration. The final decision can be made in two ways: either by a simple comparative analysis of masses of elements of the decision space, or on the minimization of belief-interval distance which also offers the advantage of quantifying the quality of the decision. The evaluation of this new methodology for real applications is under progress and it will be reported in forthcoming publications.

## REFERENCES

- [1] G. Shafer, *A Mathematical theory of evidence*, Princeton University Press, 1976.
- [2] L.A. Zadeh, *On the validity of Dempster's rule of combination*, ERL Memo M79/24, Department of EECS, Univ. of California, Berkeley, U.S.A., 1979.
- [3] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, vol. 7(2), pp. 85–90, 1986.
- [4] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of bayesian fusion rule*, Int. J. of Intelligent Syst., Vol. 29(3), pp. 223–252, 2014.
- [5] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, Volumes 1–4, American Research Press, 2004–2015.
- [6] J. Dezert, F. Smarandache, *Canonical decomposition of dichotomous basic belief assignment*, International Journal of Intelligent Systems, pp. 1–21, 2020.
- [7] P. Smets, *The canonical decomposition of a weighted belief*, in Proc. of Int. Joint Conf. on Artif. Intell., pp. 1896–1901, San Mateo, CA, USA, 1995.
- [8] R. Yager, *On the Dempster-Shafer framework and new combination rules*, Information Sciences, Vol. 41, pp. 93–138, 1987.
- [9] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Computational Intelligence, Vol. 4, pp. 244–264, 1988.
- [10] J. Dezert, F. Smarandache, F. A. Tchamova, D. Han, *Fast fusion of basic belief assignments defined on a dichotomous frame of discernment*, in Proc. of Int. Conf. on Information Fusion (Fusion 2020), Pretoria, South Africa, 2020.
- [11] J. Dezert, S. Fidanova, A. Tchamova, *Fast BF-ICrA method for the evaluation of MO-ACO algorithm for WSN layout*, in Proc. of FedCSIS Int. Conference, Sofia, Bulgaria, 2020.
- [12] D. Han, J. Dezert, Y. Yang, *Belief interval based distances measures in the theory of belief functions*, IEEE Trans. on SMC, Vol. 48(6), pp. 833–850, 2018.
- [13] P. Smets, R. Kennes, *The transferable belief model*, Art. Intell., Vol. 66(2), pp. 191–234, 1994.

# Fast Fusion of Basic Belief Assignments Defined on a Dichotomous Frame of Discernment

Jean Dezert<sup>a</sup>, Florentin Smarandache<sup>b</sup>, Albena Tchamova<sup>c</sup>, Deqiang Han<sup>d</sup>

<sup>a</sup>The French Aerospace Lab, ONERA/DTIS, Palaiseau, France.

<sup>b</sup>Department of Mathematics, University of New Mexico, Gallup, NM, USA.

<sup>c</sup>Inst. of I&C Tech., Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>d</sup>Inst. of Integrated Automation, Xi'an Jiaotong Univ., Xi'an, China.

Emails: jean.dezert@onera.fr, smarand@unm.edu, tchamova@bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, F. Smarandache, A. Tchamova, D. Han, *Fast Fusion of Basic Belief Assignments Defined on a Dichotomous Frame of Discernment*, in Proc. of Int. Conf. on Information Fusion (Fusion 2020), Online Conference, July 6-9, 2020, and reprinted with permission.

**Abstract**—In this paper, we propose a new fusion approach to combine basic belief assignments (BBAs) defined on a dichotomous frame of discernment based on their canonical decomposition. In a companion paper, we have already proved that the canonical decomposition of this type of BBA (called dichotomous BBA) is always possible and unique thanks to the proportional conflict redistribution rule No 5 (PCR5). More precisely, any dichotomous BBA is always the PCR5 combination of two simpler basic belief assignments named respectively the pro-evidence, and the contra-evidence. From this interesting canonical decomposition, we present a new way of combining many dichotomous BBAs and we show that the computational time for fusing these dichotomous BBAs based on their canonical decomposition is quasi-linear with the number of sources to combine, contrary to the direct fusion of the dichotomous BBAs altogether.

**Keywords:** Information fusion, canonical decomposition, belief functions, PCR5 rule, PCR6 rule.

## I. INTRODUCTION

The belief functions (BF) introduced by Shafer in the mid of 1970's [1] from Dempster's works are well known and used in the artificial intelligence community to model epistemic uncertainty and to reason with it for information fusion and decision-making support. Dempster's rule to combine distinct sources of evidence characterized by their basic belief assignments (BBAs) defined on the same frame of discernment (FoD) is the historical and emblematic rule of combination in Dempster-Shafer Theory (DST). Unfortunately, Dempster's rule (denoted by DS rule for short) suffers of serious drawbacks in high conflict evidences as pointed out by Zadeh [2], [3], but more importantly also in some very low conflict situations [4] as well. That is why many rules have been proposed in the literature [5] (Vol.2), among them the combination of two sources of evidence based on the proportional conflict redistribution principle No. 5<sup>1</sup> (PCR5 rule) justified theoretically in [6], which has been shown successful in applications. However its complexity remains

<sup>1</sup>Actually PCR6 rule is preferentially used for the combination of more than two sources altogether. For two sources, PCR5 and PCR6 rules coincide and because canonical decomposition involved only two sources, we only need to work with PCR5 rule to combine the pro-evidence with its contra-evidence.

one of its limitations which prevents its use in fusion problems involving many sources of evidence to combine, and its non associativity property<sup>2</sup> which make it not so appealing because the fusion order matters when sequential PCR5 fusion is applied instead of global combination of the sources altogether.

In this work, we show how the fusion of many sources of evidences represented by BBAs defined on a same dichotomous frame of discernment can be easily done based on the PCR5-based canonical decomposition of the BBAs. Such decomposition of BBA has been proposed recently in [7].

We recall that another canonical decomposition based on conjunctive rule (but involving improper<sup>3</sup> BBA) had been proposed in 1995 by Smets [8], and extended later by Denœux [11] to develop the so-called cautious rule of combination. In this new approach we use our well justified canonical decomposition based on PCR5 which is strictly based on a proper (i.e. normal) BBAs as defined by Shafer himself. We have shown that any dichotomous BBA is always the result of the PCR5 fusion of a simple proper pro-evidence BBA  $m_p$  with a simple proper contra-evidence BBA  $m_c$ , and that this decomposition is unique. Based on this important result, we address in this work the problem of combination of many dichotomous BBAs based on their canonical decomposition.

This paper is organized as follows. After a brief recall of basics of belief functions in section II, we present briefly the canonical decomposition for any dichotomous BBA based on PCR5 rule of combination in section III which is explained in more details with proofs, and examples in [7]. The fusion of dichotomous BBAs based on the principle of canonical decompositions is detailed in section IV. Concluding remarks with perspectives are given in the last section.

## II. BASICS OF BELIEF FUNCTIONS

Belief functions (BF) have been introduced by Shafer in [1] to model epistemic uncertainty. We assume that the answer<sup>4</sup> of the problem under concern belongs to a known (or given) finite

<sup>2</sup>PCR5 is only quasi-associative.

<sup>3</sup>We call a BBA improper when it does not satisfy Shafer's original definition. Smets called it a generalized simple BBA (GSBBA).

<sup>4</sup>i.e. the solution, or the decision to take.

discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are mutually exclusive and exhaustive<sup>5</sup>. The FoD is said dichotomous when it involves only two elements (one subset and its complement), that is  $\Theta = \{A, \bar{A}\}$  where  $\bar{A}$  is the complement of  $A$  in  $\Theta$ . The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A proper Basic Belief Assignment (BBA) associated with a given source of evidence is defined [1] as a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . In some BF related frameworks, like in Smets Transferable Belief Model (TBM) [8],  $m(\emptyset)$  is allowed to take a positive value. In this case,  $m(\cdot)$  is said improper because it does not satisfy Shafer's definition [1]. The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are respectively defined from a proper BBA  $m(\cdot)$  by

$$Bel(A) = \sum_{B \in 2^\Theta | B \subseteq A} m(B), \quad (1)$$

and

$$Pl(A) = \sum_{B \in 2^\Theta | A \cap B \neq \emptyset} m(B) = 1 - Bel(\bar{A}), \quad (2)$$

where  $\bar{A}$  is the complement of  $A$  in  $\Theta$ .

$Bel(A)$  and  $Pl(A)$  are interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$  in original Dempster's works [9], [10]. The quantities  $m(\cdot)$  and  $Bel(\cdot)$  are one-to-one and the following Möbius inverse formula holds (see [1], p. 39)

$$m(A) = \sum_{B \subseteq A \subseteq \Theta} (-1)^{|A-B|} Bel(B). \quad (3)$$

$A$  is called a Focal Element (FE) of  $m(\cdot)$  if  $m(A) > 0$ . When all focal elements are singletons,  $m(\cdot)$  is called a *Bayesian BBA* [1] and its corresponding  $Bel(\cdot)$  function is equal to  $Pl(\cdot)$  and they are homogeneous to a (subjective) probability measure  $P(\cdot)$ . The vacuous BBA, or VBBA for short, representing a totally ignorant source is defined as<sup>6</sup>  $m_v(\Theta) = 1$ . A dichotomous BBA is a BBA defined on a dichotomous FoD. A dogmatic BBA is a BBA such that  $m(\Theta) = 0$ . If  $m(\Theta) > 0$  the BBA  $m(\cdot)$  is nondogmatic. A simple BBA is a BBA that has at most two focal sets and one of them is  $\Theta$ . A dichotomous non dogmatic mass of belief is a BBA having three focal elements  $A$ ,  $\bar{A}$  and  $A \cup \bar{A}$  with  $A$  and  $\bar{A}$  subsets of  $\Theta$ .

In his *Mathematical Theory of Evidence* [1], Shafer proposed to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD  $\Theta$  with Dempster's rule (i.e. the normalized conjunctive rule).

For the combination of two BBAs, Dempster's rule formula [1] is given by  $m_{DS}(\emptyset) = 0$  and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{DS}(X) = \frac{1}{K_{12}} \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2), \quad (4)$$

with  $K_{12} = 1 - \sum_{X_1, X_2 \in 2^\Theta | X_1 \cap X_2 = \emptyset} m_1(X_1)m_2(X_2)$ .

The justification and behavior of Dempster's rule have been disputed over the years from many counter-examples involving high and low conflicting sources (from both theoretical and practical standpoints) as reported in [4], [12]–[14]. Many rules of combination exist in the literature<sup>7</sup>, among them we recommend the rule based on the proportional conflict redistribution principle No. 5 (PCR5 rule) [6] which has been shown to be successful in applications and well justified theoretically. That is why we analyze it in details for solving the BF canonical decomposition problem (BF-CDP). PCR5 transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is entirely preserved in this fusion process (see [5], Vol. 2 and Vol. 3 for full justification and examples):  $m_{PCR5}(\emptyset) = 0$  and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{PCR5}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2) + \sum_{\substack{X_2 \in 2^\Theta \\ X_2 \cap X = \emptyset}} \left[ \frac{m_1(X)^2 m_2(X_2)}{m_1(X) + m_2(X_2)} + \frac{m_2(X)^2 m_1(X_2)}{m_2(X) + m_1(X_2)} \right], \quad (5)$$

where all denominators in (5) are different from zero. If a denominator is zero, that fraction is discarded. The properties of PCR5 can be found in [15]. Extension of PCR5 for combining qualitative BBA's can be found in [5], Vol. 2 & 3. A variant of PCR5, called PCR6 has been proposed by Martin and Osswald in [5], Vol. 2, for combining  $s > 2$  sources. The general formulas for PCR5 and PCR6 rules are also given in [5], Vol. 2. PCR6 coincides with PCR5 when one combines two sources. The difference between PCR5 and PCR6 lies in the way the proportional conflict redistribution is done as soon as three (or more) sources are involved in the fusion. From the implementation point of view, PCR6 is simpler to implement than PCR5. For convenience, very basic (not optimized) Matlab<sup>TM</sup> codes of PCR5 and PCR6 fusion rules can be found in [5], [16] and from the toolboxes repository on the web [17]. The main drawback of PCR5 and PCR6 rules is their combinatorial complexity when the number of source is big. Even for combining BBAs defined on a simple dichotomous frame of discernment, the computational time for combining more than 20 sources can take several hours<sup>8</sup>.

Our main motivation and contribution is to propose a faster fusion method to combine many dichotomous BBAs in order

<sup>7</sup>see [5], Vol. 2 for a detailed list of fusion rules.

<sup>8</sup>due to the exponential complexity of the PCR6 rule (as shown in Figure 4). For our simulations, we did use a MacBook Pro 2.8 GHz Intel Core i7 with 16 Go 1600 MHz DDR3 memory running Matlab<sup>TM</sup> R2018a.

<sup>5</sup>This is so-called Shafer's model of FoD [5].

<sup>6</sup>The complete ignorance is denoted  $\Theta$  in Shafer's book [1].

to overcome the combinatorial complexity problem by establishing a new effective (approximating) fusion method based on the new PCR5-based canonical decomposition principle. It is worth noting that our new method is very different of the method based on the clustering of non conflicting BBAs followed by a discounting step and the conjunctive rule presented in [18].

### III. CANONICAL DECOMPOSITION OF DICHOTOMOUS BBA

A FoD  $\Theta = \{A, \bar{A}\}$  is called dichotomous if it consists of only two elements  $A$  and  $\bar{A}$  with  $A \cup \bar{A} = \Theta$  and  $A \cap \bar{A} = \emptyset$ .  $A$  is different from  $\Theta$  and from Empty-Set because we want to work with informative FoD. Indeed, the very special frame  $\{\Theta, \emptyset\}$  does not bring any useful information since the only possible BBA with such frame is the vacuous BBA. So, we consider a given proper<sup>9</sup> BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  of the form

$$m(A) = a, \quad m(\bar{A}) = b, \quad m(A \cup \bar{A}) = 1 - a - b, \quad (6)$$

with  $0 < a < 1$ ,  $0 < b < 1$  and  $a + b < 1$ .

The conditions  $0 < a < 1$  and  $0 < b < 1$  mean that  $A$  and  $\bar{A}$  are focal elements of the BBA. The restriction  $a + b < 1$  means that the BBA is nondogmatic.

This assumption of nondogmaticity of the BBA  $m(\cdot)$  is necessary for Smets canonical decomposition [8], but it is not essential for our PCR5-based canonical decomposition (as we will show in the sequel) because our decomposition also works directly with a dogmatic BBA. Of course any dogmatic BBA can always be modified as a non-dogmatic BBA by using a very small discounting number ( $\epsilon > 0$ ) so that, in practice, Smets' decomposition can always be applied, but this is not sufficient to prove that Smets approach always provides relevant results. Why? just because we know (and we have proved) that Dempster's (normalized conjunctive rule) and even the conjunctive rule in Smets' TBM suffers of serious drawbacks - see justifications in our aforementioned references. That is why we explore in this work another way of making a canonical decomposition, which is, for now, limited to dichotomous BBA.

Our canonical decomposition problem consists in finding the two following simple proper BBAs  $m_p$  and  $m_c$  of the form

$$m_p(A) = x, \quad m_p(A \cup \bar{A}) = 1 - x, \quad (7)$$

$$m_c(\bar{A}) = y, \quad m_c(A \cup \bar{A}) = 1 - y, \quad (8)$$

with  $(x, y) \in [0, 1] \times [0, 1]$ , such that  $m = \text{Fusion}(m_p, m_c)$ , for a chosen rule of combination denoted by  $\text{Fusion}(\cdot, \cdot)$ . The simple BBA  $m_p(\cdot)$  is called the *pro-BBA* (or pro-evidence) of  $A$ , and the simple BBA  $m_c(\cdot)$  the *contra-BBA* (or contra-evidence) of  $A$ . The BBA  $m_p(\cdot)$  is interpreted as a source of evidence providing an uncertain evidence in favor of  $A$ , whereas  $m_c(\cdot)$  is interpreted as a source of evidence providing an uncertain contrary evidence about  $A$ .

This decomposition is possible with Dempster's rule only if  $0 < a < 1$ ,  $0 < b < 1$  and  $a + b < 1$ , and in this case we

have  $x = \frac{a}{1-b}$  and  $y = \frac{b}{1-a}$ . However, any dogmatic BBA  $m(A) = a$ ,  $m(\bar{A}) = b$  with  $a + b = 1$  is not decomposable from Dempster's rule for the case when  $(a, b) \neq (1, 0)$  and  $(a, b) \neq (0, 1)$  (see Theorem 4 in [7]), and the dogmatic BBAs  $m(A) = 1$ ,  $m(\bar{A}) = 0$  (case  $(a, b) = (1, 0)$ ), or  $m(A) = 0$ ,  $m(\bar{A}) = 1$  (case  $(a, b) = (0, 1)$ ) have infinitely many decompositions based on Dempster's rule of combination (see Lemma in [7]). In [7], we have shown that our canonical decomposition cannot be achieved from conjunctive, disjunctive, Yager's [19] or Dubois-Prade [20] rules of combination, neither from averaging rule. However, such type of decomposition is unique and is always possible in all cases of dichotomous BBA  $m(\cdot)$  using the PCR5 rule of combination. In [7], we did prove the following Theorem.

**Theorem 1:** Consider a dichotomous FoD  $\Theta = \{A, \bar{A}\}$  with  $A \neq \Theta$  and  $A \neq \emptyset$  and a nondogmatic BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  defined on  $\Theta$  by  $m(A) = a$ ,  $m(\bar{A}) = b$ , and  $m(A \cup \bar{A}) = 1 - a - b$ , where  $a, b \in [0, 1]$  and  $a + b < 1$ . Then the BBA  $m(\cdot)$  has a unique canonical decomposition using PCR5 rule of combination of the form  $m = \text{PCR5}(m_p, m_c)$  with pro-evidence  $m_p(A) = x$ ,  $m_p(A \cup \bar{A}) = 1 - x$  and contra-evidence  $m_c(\bar{A}) = y$ ,  $m_c(A \cup \bar{A}) = 1 - y$  where  $x, y \in [0, 1]$ .

Moreover, we also proved in [7] that the canonical decomposition also exists even if the dichotomous BBA is dogmatic (i.e. Bayesian) and the following theorem also holds.

**Theorem 2:** Any dogmatic BBA defined by  $m(A) = a$  and  $m(\bar{A}) = b$ , where  $a, b \in [0, 1]$  and  $a + b = 1$ , has a canonical decomposition using PCR5 rule of combination of the form  $m = \text{PCR5}(m_p, m_c)$  with  $m_p(A) = x$ ,  $m_p(A \cup \bar{A}) = 1 - x$  and  $m_c(\bar{A}) = y$ ,  $m_c(A \cup \bar{A}) = 1 - y$  where  $x, y \in [0, 1]$ .

Theorems 1 & 2 prove that the decomposition based on PCR5 always exists and it is unique for any dichotomous (nondogmatic, or dogmatic) BBA.

For the case of dichotomous dogmatic BBA considered in Theorem 2, the expression of solutions  $x$  and  $y$  can be established explicitly as follows - see [7] for details

- If  $a < b$  then  $x < y$ , and we have  $y = 1$  and  $x = \frac{a + \sqrt{a^2 + 4a}}{2}$ .
- If  $a > b$  then  $x > y$ , and we have  $x = 1$  and  $y = \frac{b + \sqrt{b^2 + 4b}}{2}$ .
- If  $a = b$  and  $a + b = 1$  then  $a = b = 0.5$  and  $x = y = 1$ .

For the case of dichotomous nondogmatic BBA considered in Theorem 1, one has to find  $x$  and  $y$  solutions of the system

$$a = x(1 - y) + \frac{x^2 y}{x + y} = \frac{x^2 + xy - xy^2}{x + y}, \quad (9)$$

$$b = (1 - x)y + \frac{xy^2}{x + y} = \frac{y^2 + xy - x^2 y}{x + y}. \quad (10)$$

under the constraints  $(a, b) \in [0, 1]^2$ , and  $0 < a + b < 1$ . In fact, it has been proved in [7] that  $x \in [a, a + b] \subset [0, 1]$  and  $y \in [b, a + b] \subset [0, 1]$ , but the explicit expression of  $x$  and  $y$  are very complicated to obtain analytically (even with modern symbolic computing systems like Mathematica<sup>TM</sup>, or

<sup>9</sup>which means that  $m(\emptyset) = 0$ .

Maple™) because after algebraic calculation, and for  $x \neq 1$ , one has to solve the following quartic equation which has at most four real solutions with only a valid one in  $[a, a + b]$

$$x^4 + (-a - 2)x^3 + (2a + b)x^2 + (a + b - ab - b^2)x + (-a^2 - ab) = 0, \quad (11)$$

and then compute  $y$  as  $y = \frac{a+b-x}{1-x}$ .

Fortunately, the solutions can be easily calculated numerically by these computing systems, and even with Matlab™system<sup>10</sup> as soon as the numerical values are committed to  $a$  and to  $b$ , and this is what we do in our simulations in the sequel.

**Example 1:** Let consider  $\Theta = \{A, \bar{A}\}$  and  $m(A) = 0.6$ ,  $m(\bar{A}) = 0.3$  and  $m(A \cup \bar{A}) = 1 - m(A) - m(\bar{A}) = 0.1$ . Hence,  $a = 0.6$  and  $b = 0.3$ . The quartic equation (11) becomes

$$x^4 - 2.6x^3 + 1.5x^2 + 0.63x - 0.54 = 0. \quad (12)$$

The four solutions of this quartic equation provided by the computing system<sup>11</sup> are approximately

$$\begin{aligned} x_1 &\approx 0.7774780438, & x_2 &\approx 0.9297589637, \\ x_3 &\approx 1.4191515820, & x_4 &\approx -0.5263885898. \end{aligned}$$

Clearly  $x_3$  and  $x_4$  are not acceptable solutions because they don't belong to  $[0, 1]$ . If we take  $x_1 \approx 0.7774780438$ , then we will get  $y_1 = (a + b - x_1)/(1 - x_1) = (0.9 - x_1)/(1 - x_1) \approx 0.5506061437$ . The pair  $(x_1, y_1) \in [0, 1]^2$  is a solution of the decomposition problem of the BBA  $m(\cdot)$ . If we take  $x_2 \approx 0.9297589637$ , then we will get  $y_2 = (a + b - x_2)/(1 - x_2) = (0.9 - x_2)/(1 - x_2) \approx -0.4236692006$ . We see that  $y_2 \notin [0, 1]$ , and therefore the pair  $(x_2, y_2)$  cannot be a solution of the decomposition problem of the BBA  $m(\cdot)$ . Therefore the canonical masses  $m_p(\cdot)$  and  $m_c(\cdot)$  are given by

$$\begin{aligned} m_p(A) &\approx 0.7774780438, & m_p(A \cup \bar{A}) &\approx 0.2225219562, \\ m_c(\bar{A}) &\approx 0.5506061437, & m_c(A \cup \bar{A}) &\approx 0.4493938563. \end{aligned}$$

It can be verified that the PCR5 combination of BBAs  $m_p$  and  $m_c$ , denoted by  $PCR5(m_p, m_c)$ , is equal to the BBA  $m(\cdot)$ .

Of course there are necessarily numerical approximations involved by the proposed decomposition because this decomposition is obtained by numerical solvers. This may have some little impact in the PCR5 fusion result but because PCR5 rule is numerically robust to small input changes (contrariwise to Dempster's rule) the PCR5 result will not change substantially with small changes (due to small numerical imprecisions) in the values of BBAs to combine.

#### A. Particular cases

1) *Case  $(a, b) = (0, 0)$  (i.e.  $m$  is the vacuous BBA):* This is the most degenerate case where the BBA  $m(\cdot)$  corresponds to the vacuous BBA. For averaging rule, conjunctive rule, Yager's, Dubois-Prade's, Dempster's and PCR5 rules one has

$x = 0$  and  $y = 0$  (conflict between canonical masses is zero). In fact the vacuous BBA  $m(\cdot)$  can always be interpreted as the fusion of  $m_p$  and  $m_c$ , where  $m_p$  and  $m_c$  are also vacuous BBAs. This degenerate case has no particular interest in practice but to model the total ignorant state of knowledge.

2) *Case when  $a = 0$ , or  $b = 0$ :* In the case  $a = 0$  and  $0 < b \leq 1$ , then for conjunctive rule, Yager's, Dubois-Prade's, Dempster's and PCR5 rules one has  $x = 0$  and  $y = b$  (conflict between canonical masses is zero) and  $m(\cdot)$  corresponds to the fusion of vacuous pro-evidence  $m_p = m_v$  with the contra-evidence  $m_c = m$ . In the case  $0 < a \leq 1$  and  $b = 0$ , then for conjunctive rule, Yager's, Dubois-Prade's, Dempster's and PCR5 rules one has  $x = a$  and  $y = 0$  (conflict between canonical masses is zero) and  $m(\cdot)$  corresponds to the fusion of the pro-evidence  $m_p = m$  with the vacuous contra-evidence  $m_c = m_v$ . These cases have no particular interest because they can be seen just as the combination of pro (or contra) BBA with the vacuous BBA.

3) *Case when  $a = b \in (0, 0.5)$ :* In this case, the BBA  $m(A) = m(\bar{A}) = a$  and  $m(A \cup \bar{A}) = 1 - 2a$  can be canonically decomposed from PCR5 rule with the BBAs  $m_p(A) = 1 - \sqrt{1 - 2a}$ ,  $m_p(A \cup \bar{A}) = \sqrt{1 - 2a}$  and  $m_c(\bar{A}) = 1 - \sqrt{1 - 2a}$ ,  $m_c(A \cup \bar{A}) = \sqrt{1 - 2a}$  - see details and proof in [7].

#### B. Benefits of canonical decomposition

The canonical decomposition based on PCR5 offers several interests and advantages that are briefly listed.

- 1) This canonical decomposition of  $m(\cdot)$  into the pro-evidence  $m_p(\cdot)$  and the contra-evidence  $m_c(\cdot)$  allows to define the notion of internal conflict of a dichotomous source of evidence, denoted by  $K_{int}(m)$ , as

$$K_{int}(m) \triangleq m_p(A)m_c(\bar{A}), \quad (13)$$

where  $m_p(A) = x$  and  $m_c(\bar{A}) = y$  are the canonical factors of the BBA  $m(\cdot)$  based on PCR5 rule of combination.

- 2) The canonical decomposition also allows to adjust/revise easily a dichotomous source of evidence (if needed) according to the knowledge one has on it. For instance, if one knows that a source over (or under) estimate the hypothesis  $A$ , then one could apply an adjustment (based on some discounting or reinforcing factors) on the pro (or contra) evidence to de-bias this source of evidence.
- 3) This canonical decomposition can help to develop new fast rules of combination for the fusion of  $S \geq 2$  (dichotomous) distinct<sup>12</sup> BBAs  $m_s(\cdot) = (m_s(A), m_s(\bar{A}), m_s(A \cup \bar{A})) = (a_s, b_s, 1 - a_s - b_s)$ ,  $s = 1, 2, \dots, S$ . This is presented next.

#### IV. FAST FUSION OF DICHOTOMOUS BBAS

In this section, we show how to combine many dichotomous BBAs defined on the same FoD  $\Theta$  thanks to their canonical decompositions.

<sup>10</sup>thanks to the *fsolve* Matlab™ command.

<sup>11</sup>We did get same solutions with Maple™, and with Matlab™.

<sup>12</sup>i.e. cognitively independent.

### A. Principle of the fast fusion of dichotomous BBAs

The main idea for making the fast fusion of dichotomous BBAs is, at first, to decompose canonically each dichotomous BBA  $m_s(\cdot)$ , for  $s = 1, 2, \dots, S$  into their pro and contra evidences  $m_{p,s} = (m_{p,s}(A), m_{p,s}(\bar{A}), m_{p,s}(A \cup \bar{A})) = (x_s, 0, 1 - x_s)$  and  $m_{c,s} = (m_{c,s}(A), m_{c,s}(\bar{A}), m_{c,s}(A \cup \bar{A})) = (0, y_s, 1 - y_s)$ , and then to combine the pro-evidences  $m_{p,s}$  for  $s = 1, 2, \dots, S$  altogether on one hand to get a global pro-evidence  $m_p$ , and to combine the contra-evidences  $m_{c,s}$  for  $s = 1, 2, \dots, S$  altogether on the other hand to get a global contra-evidence  $m_c$ . The fusion step of pro and contra evidences is discussed in section IV-D. Once  $m_p$  and  $m_c$  are calculated, then one combines them with PCR5 fusion rule to get the final result. This general principle of the new fusion method is represented by the diagram of figure 1 for convenience.

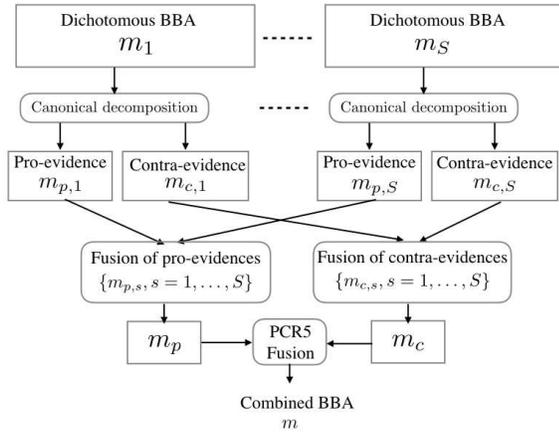


Figure 1. General principle of the fusion of dichotomous BBAs from their canonical decompositions.

This new fusion approach is interesting because the fusion of the pro-evidence  $m_{p,s}$  (resp. contra-evidences  $m_{c,s}$ ) is quite simple because there is no conflict between  $m_{p,s}$  (resp. between  $m_{c,s}$ ), so that their fusion can be done quite easily and a large number of sources can be combined without a high computational burden. In fact, with this fusion approach, only one PCR5 fusion step of simple (combined) canonical BBAs is needed at the very end of the fusion process. It is worth noting that in this work there is no link with the canonical decomposition proposed by Shafer and then extended by Smets because here we use another fusion rule based on the proportional conflict redistribution principle.

### B. Analysis of the effectiveness of this new fusion approach

Because the PCR5 rule<sup>13</sup> of combination is not associative, the fusion<sup>14</sup> of the canonical BBAs followed by their PCR5

<sup>13</sup>The same remark holds for PCR6 rule with more than two BBAs.

<sup>14</sup>We assume here that the fusion of all the pro-evidences (resp. contra-evidences) is done with PCR5 rule which coincides in this case with the conjunctive rule because there is no conflict between the pro-evidences (resp. the contra-evidences).

fusion will not provide in general the same result as the direct fusion of the dichotomous BBAs altogether but only an approximate result, which is normal.

The main question is to know how good is the approximation obtained by this new fusion method based on the fusion of pro-evidences and contra-evidences with respect to the direct fusion of the BBAs with PCR5 (or PCR6 when considering more than two sources to combine). To answer to this important question we make a statistical analysis of the quality of the combined result  $m$ , with respect to the direct PCR5, or PCR6 fusion of all BBAs altogether.

The measure of the goodness is obtained by the normalized (Euclidean) Belief Interval distance  $d_{BI}(m_{PCR5}, m)$  (for the case of two BBAs only), or by  $d_{BI}(m_{PCR6}, m)$  if more than two sources are considered in the fusion process, where  $m$  is the result of the fusion principle based on canonical decompositions, and  $m_{PCR5}$  (resp.  $m_{PCR6}$ ) is the result of the combination of original BBAs altogether with PCR5 (resp. PCR6) rule. The  $d_{BI}$  distance between two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined on the powerset of a given FoD  $\Theta = \{\theta_1, \dots, \theta_n\}$  has been proposed and justified in [21], [22]. It is defined by

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))}, \quad (14)$$

where  $N_c = 1/2^{n-1}$  is a normalization factor to make  $d_{BI}(m_1, m_2) \in [0, 1]$ , and  $d_W(BI_1(X), BI_2(X))$  is Wasserstein's distance [23] between belief intervals  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)] = [a_1, b_1]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)] = [a_2, b_2]$ . Here,  $d_W^2(BI_1(X), BI_2(X))$  entering in (14) is given by

$$d_W^2([a_1, b_1], [a_2, b_2]) \triangleq \left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2. \quad (15)$$

Figure 2 shows the normalized histogram (i.e. the empirical probability distribution) of the distance values  $d_{BI}^E(m_{PCR5}, m)$  based on 20000 random<sup>15</sup> generations of dichotomous BBAs  $m_1$  and  $m_2$ . One observes that the new fusion approach based on the canonical decompositions of BBAs (with the conjunctive fusion of pro-evidences, and the conjunctive fusion of contra-evidences) provides a solution which is very close to what we obtain from the direct application of PCR5 rule, with a mean of 0.0287 and a standard deviation of 0.0289. In 98.20% of cases, the final decision (based on the min of  $d_{BI}^E$  decision-making strategy explained in [22]) based on  $m_{PCR5}$ , or on  $m$  are in agreement. This means that the decision agreement (DA) rate is 98.20%.

Figures 3 show the normalized histograms of the  $d_{BI}^E(m_{PCR6}, m)$  values based also on 20000 random runs for the fusion of 6 dichotomous BBAs respectively. We use

<sup>15</sup>For this, we generate three random numbers uniformly distributed in  $[0, 1]$  and we normalize them to generate randomly a dichotomous BBA.

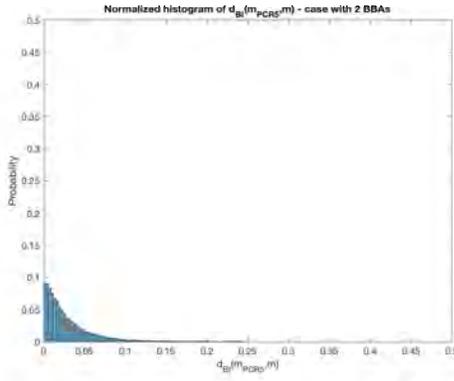


Figure 2. Normalized histogram of  $d_{BI}^E(m_{PCR5}, m)$  for 2 dichotomous BBAs case (20000 runs).

PCR6 rule instead of PCR5 rule to combine the 6 dichotomous BBAs altogether because PCR6 rule has been recognized to be more effective than PCR5 in applications [5] (Vol.2 - Chap. 2). As we can observe, the shape of the histograms is a bit different from the histogram of fig. 2, but what matters is that the mean value and the standard deviation of the  $d_{BI}^E$  distance are still low (0.1119 and 0.0392 respectively) indicating that the approximation obtained by this new fusion method is globally very good. Also the decision based on this new fusion approach is globally coherent with the decision taken by the direct PCR6 fusion of the BBAs (95.84% of decision coherence).

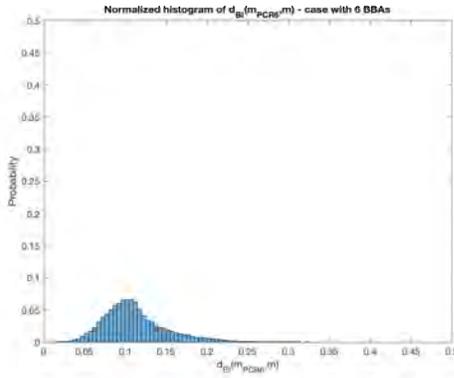


Figure 3. Normalized histogram of  $d_{BI}^E(m_{PCR6}, m)$  for 6 dichotomous BBAs case (20000 runs).

Several Monte Carlo simulations have been done with different numbers of dichotomous BBAs to combine. The results obtained based on 20000 runs Monte Carlo simulations are summarized in the Table I.

The second column of Table I indicates the mean value, denoted by  $\text{mean}(d_{BI}^E)$ , of the normalized Euclidean belief Interval distance between the direct fusion of the BBAs by the PCR5 (when combining 2 BBAs only), or PCR6 rule (when combining more than two BBAs) and the new fusion

# of BBAs	$\text{mean}(d_{BI}^E)$	$\text{std}(d_{BI}^E)$	Decision Agreement (%)
2	0.0287	0.0289	98.20
3	0.0578	0.0373	97.52
4	0.0838	0.0394	96.69
5	0.1008	0.0397	96.05
6	0.1119	0.0392	95.84
7	0.1169	0.0385	95.40
8	0.1200	0.0374	94.89
9	0.1211	0.0365	94.25
10	0.1204	0.0348	94.21

Table I  
COMPARATIVE EVALUATION OF CANONICAL DECOMPOSITION FUSION METHOD W.R.T. THE DIRECT PCR-BASED FUSION METHOD.

rule based on their canonical decomposition. The third column of the Table I shows the corresponding standard deviation values denoted by  $\text{std}(d_{BI}^E)$ . The last column indicates the decision agreement (DA) factor between the decision taken from the direct fusion method, and the indirect (canonical decomposition based) method. As we can see, the DA factors are very high which means that most of the time the decisions taken from the direct fusion method and from the indirect fusion method are the same.

After a deep analysis of our simulation results, one can attest that the decision-making disagreement occurs when the numerical values of the mass of  $A$  and the mass of  $\bar{A}$  are very close. This indicates a very high ambiguity in the decision to take in such situation which can be easily tracked in practice by evaluating the quality indicator of the decision-making - see [22] for details.

In this paper we did not investigate the quality of the approximation of the fusion result based on this canonical decomposition when replacing the PCR5 fusion step of  $m_p$  and  $m_c$  by other rules of combination because the core of the canonical decomposition is based on PCR5.

### C. Computational time of the new fusion method

Because of very high combinatorial complexity (and thus high computational time) required for applying the direct PCR6 fusion of many BBAs, we did only make the performance evaluation up to the fusion of ten BBAs only with PCR6. We conjecture that the performances of this new fusion method based on canonical decomposition will very slowly degrade with the increase of the number of BBAs involved in the fusion process. Of course the new fusion method based on this canonical decomposition does not suffer of combinatorial complexity limitation which is of great interest in some applications (like in multi-spectral imagery for detection and classification) because many (hundreds or even thousands) of dichotomous BBAs could be easily combined very quickly. Actually with this method what takes a bit time is only the canonical decomposition done by the numerical solver<sup>16</sup>.

Figure 4 shows the average (based on 50 random runs) computational time (in seconds) of the direct PCR6 fusion of the BBAs altogether (red plot), and the average computational time of the new fusion method based on canonical decomposition (blue plot). It is clear that the computational

<sup>16</sup>We did use Matlab™ *fsolve* function for this.

time of the direct PCR6 fusion method (the red curve) grows exponentially with the number of sources, whereas the computational time grows only slowly and quasi-linearly with the new method proposed in this work.

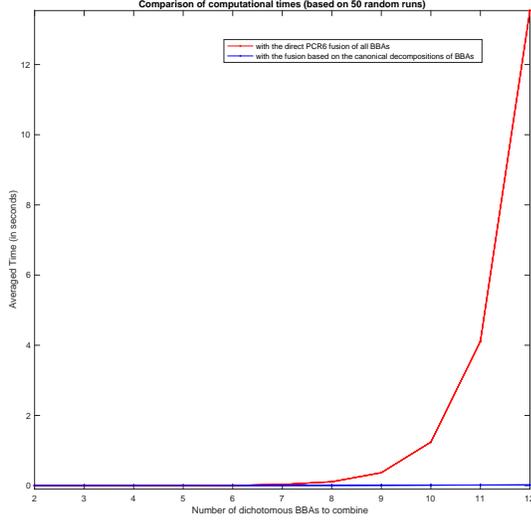


Figure 4. Computing time versus number of BBAs to combine.

Based on a set of 1000 random dichotomous BBAs, figure 5 shows that the computational time (in seconds) of the fusion based on the canonical decomposition is a quasi-linear<sup>17</sup> function of the number of BBAs to combine.

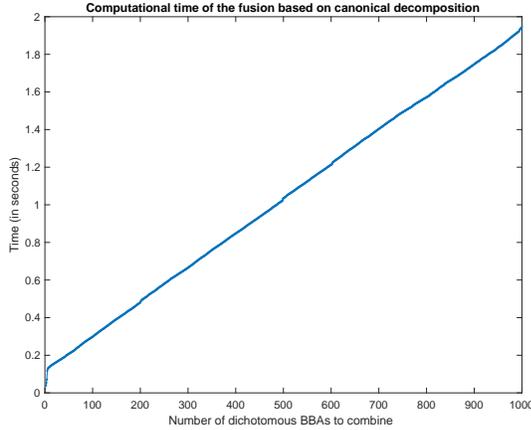


Figure 5. Computing time versus number of BBAs to combine.

Figures 4 and 5 show the computational times including the canonical decomposition itself done on the fly. Of course, the canonical decomposition could be done off-line once for all and stored in the computer memory (if necessary) - see for instance the  $(x, y)$  values given in [7] for convenience. If we have  $n$  dichotomous BBAs to combine, we have to make their canonical decomposition at first and because the  $n$

<sup>17</sup>It is not strictly linear because the time for the numerical *fsolve* search of pro-evidence and contra-evidence factors for making the canonical decomposition is not constant.

pro-evidence BBAs to combine (resp. contra evidence BBAs) have a very simple structure their conjunctive fusion  $m_p(A)$  is obtained very quickly by the direct product of  $n$  real numbers, that is  $m_p(A) = \prod_{i=1}^n m_{p,i}(A)$ , and we need also a subtraction because  $m_p(A \cup \bar{A}) = 1 - m_p(A)$ . The complexity of this fast suboptimal PCR fusion approach (once the canonical decomposition is available) is therefore  $2(n-1)$  multiplications and 2 subtractions for making the conjunctive fusion of  $m_{p,i}$  and the conjunctive fusion of  $m_{c,i}$ , and 7 additions and 5 multiplications for making the PCR5 fusion of  $m_p$  with  $m_c$ . There is no need to use the commonality function or the Smets canonical decomposition to make the fusion of these dichotomous BBAs. These figures show clearly the real advantage of the fusion of dichotomous BBAs based on their canonical decompositions in term of computational time, and that is why we can say that the new proposed method is really a fast fusion method with respect to the direct PCR5 or PCR6 rule of combination when working with a dichotomous frame of discernment.

#### D. On the fusion of pro and contra evidences

In the previous analysis, we did use the conjunctive rule for the intermediary fusion step of pro-evidences in one hand, and the intermediary fusion step of contra-evidences in the other hand. It is worth noting that the intermediary step of fusion of pro-evidences, and the intermediary step of fusion of contra-evidences can be done in parallel which offers a computational advantage with respect to the direct fusion method (if one has many sources to combine in a specific application). This parallelization cannot be achieved in general with the other existing rules of combination of evidences.

Because of the fusion principle depicted in Figure 1, this new fusion method offers also the possibility (if one prefers for some own specific reasons) of selecting other fusion rules for the intermediary fusion steps for combining the pro-evidences, and the contra-evidences. Of course the choice of the fusion rules used for the combination of pro-evidences and the combination of contra-evidences impacts the final result, but depending on the type of rules chosen we can obtain an associative rule, an idempotent rule, and even a new cautious rule. For example, let's consider the same type of fusion rule for combining the pro-evidences  $m_{p,s}$   $s = 1, \dots, S$ , and for combining contra-evidences  $m_{c,s}$   $s = 1, \dots, S$  and consider the following cases:

- 1) If we use the conjunctive rule [5] (Vol. 1), denoted by  $Conj(\dots)$  (as we did previously in our Monte Carlo simulations for histogram plots), then

$$m_p = Conj(m_{p,1}, \dots, m_{p,S}),$$

and one has  $m_p(A \cup \bar{A}) = \prod_{s=1}^S (1 - x_s)$  and  $m_p(A) = 1 - m_p(A \cup \bar{A})$ . Because the conjunctive rule is associative the fusion of pro-evidences can be done sequentially. Similarly, for the fusion of contra-evidences using the conjunctive rule one has  $m_c(A \cup \bar{A}) = \prod_{s=1}^S (1 - y_s)$  and  $m_p(\bar{A}) = 1 - m_c(A \cup \bar{A})$ . Because there is no conflict between the pro-evidences (resp. contra-evidences), the

fusion result of the pro-evidences (resp. the contra-evidences) by PCR5 (or PCR6) rules is equivalent to the conjunctive fusion result. The conjunctive rule however is not idempotent in general but in very specific cases where only one focal element gets all the mass of belief.

- 2) If we prefer to use the averaging rule, then we will have  $m_p(A) = \frac{1}{S} \sum_{s=1}^S x_s$  and  $m_p(A \cup \bar{A}) = 1 - m_p(A) = \frac{1}{S} \sum_{s=1}^S (1 - x_s)$ , and  $m_c(\bar{A}) = \frac{1}{S} \sum_{s=1}^S y_s$  and  $m_c(A \cup \bar{A}) = 1 - m_c(\bar{A}) = \frac{1}{S} \sum_{s=1}^S (1 - y_s)$ . Because the averaging rule is not associative, the sequential fusion of pro-evidences (and contra-evidences) is not recommended, however the averaging rule allows to get an idempotent fusion rule based on canonical decompositions if needed.
- 3) We could also prefer to use the min rule to build a new cautious rule of combination which will be associative and idempotent. For this, we just have to take  $m_p(A) = \min_{s=1, \dots, S}(x_s)$  and  $m_p(A \cup \bar{A}) = 1 - m_p(A)$ . Similarly,  $m_c(\bar{A}) = \min_{s=1, \dots, S}(y_s)$  and  $m_c(A \cup \bar{A}) = 1 - m_c(\bar{A})$ .

## V. CONCLUSIONS

In this research paper, we did propose a new fusion method to combine very quickly many BBAs defined on a dichotomous frame of discernment thanks to their unique canonical decompositions. This new interesting method can be parallelized and offers the advantage to have a quasi-linear computational time with the number of sources. For now, this method is limited to the fusion of many BBAs that are defined on a simple (dichotomous) frame of discernment. After some unsuccessful attempts, it appears that the development of a fast fusion method based on the canonical decomposition principle for working with non-dichotomous frames of discernment is actually a very difficult problem that we want to address to the scientific community working with belief functions as a future research challenge. This very new method brings already a significant benefit for real application involving inter-criteria analysis for the evaluation of multiple-objective ant colony optimization algorithm for wireless sensor networks deployment that should be reported in a forthcoming publication.

## REFERENCES

- [1] G. Shafer, *A mathematical theory of evidence*, Princeton Univ. Press, 1976.
- [2] L.A. Zadeh, *On the validity of Dempster's rule of combination*, ERL Memo M79/24, Department of EECS, Univ. of California, Berkeley, U.S.A., 1979.
- [3] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, Vol. 7(2), pp. 85–90, 1986.
- [4] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Proc. of Fusion 2012, Singapore, July 9–12, 2012.
- [5] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion – Collected works*, American Research Press, Vol. 1–4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [6] F. Smarandache, J. Dezert, *Information fusion based on new proportional conflict redistribution rules*, in Proc. of the 8th Int. Conf. on Information Fusion, Philadelphia, USA, 25–29 July, 2005.
- [7] J. Dezert, F. Smarandache, *Canonical decomposition of dichotomous basic belief assignment*, International Journal of Intelligent Systems, March 2020.

- [8] P. Smets, *The canonical decomposition of a weighted belief*, Int. Joint Conf. on Artificial Intelligence, Morgan Kaufman, San Mateo, CA, USA, pp. 1896–1901, 1995.
- [9] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The Annals of mathematical Statistics, Vol. 38, No. 2, April 1967.
- [10] A.P. Dempster, *The generalization of bayesian inference*, Technical report No. 20, Dept. of Statistics, Harvard University, November 15, 1967.
- [11] T. Deneux, *Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence*, Artificial Intelligence Vol. 172, pp. 234–264, 2008.
- [12] A. Tchamova, J. Dezert, *On the behavior of Dempster's rule of combination and the foundations of Dempster-Shafer theory*, in Proc. of 6th IEEE Int. Conf. on Int. Syst., Sofia, Bulgaria, Sept. 6–8, 2012.
- [13] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of bayesian fusion rule*, Int. J. of Intelligent Syst., Vol. 29, No. 3, pages 223–252, March 2014.
- [14] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of Int. Conf. on Information Fusion (Fusion 2013), Istanbul, Turkey, July 2013.
- [15] J. Dezert, F. Smarandache, *Non bayesian conditioning and deconditioning*, Int. Workshop on Belief Functions, Brest, France, April 2010.
- [16] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of Int. Conf. on Information Fusion (Fusion 2010), Edinburgh, UK, July 2010.
- [17] <https://bfasociety.org/>
- [18] K. Zhou, A. Martin, Q. Pan, *A belief combination rule for a large number of sources*, Infinite Study, 17 pages, 2018.
- [19] R. Yager, *On the Dempster-Shafer framework and new combination rules*, Information Sciences, Vol. 41, pp. 93–138, 1987.
- [20] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Computational Intelligence, Vol. 4, pp. 244–264, 1988.
- [21] D. Han, J. Dezert, Y. Yang, *Belief interval based distances measures in the theory of belief functions*, IEEE Trans on SMC: Systems, Vol. 48, No. 6, pp. 833–850, June 2018.
- [22] J. Dezert, D. Han, J.-M. Tacnet, S. Carlados, Y. Yang, *Decision-making with belief interval distance*, Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21–23, 2016.
- [23] A. Irpino, R. Verde, *Dynamic clustering of interval data using a Wasserstein-based distance*, Pattern Rec. Letters, Vol. 29, pp.1648–1658, 2008.

# Fast BF-ICrA Method for the Evaluation of MO-ACO Algorithm for WSN Layout

Jean Dezert<sup>a</sup>, Stefka Fidanova<sup>b</sup>, Albena Tchamova<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, F-91761 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

Emails: jean.dezert@onera.fr, stefka@parallel.bas.bg, tchamova@bas.bg

Originally published as: J. Dezert, S. Fidanova, A. Tchamova, *Fast BF-ICrA Method for the Evaluation of MO-ACO Algorithm for WSN Layout*, in Proc. of FedCSIS Int. Conf., Sofia, Bulgaria, Sept. 6–9, 2020, and reprinted with permission.

**Abstract**—In this paper, we present a fast Belief Function based Inter-Criteria Analysis (BF-ICrA) method based on the canonical decomposition of basic belief assignments defined on a dichotomous frame of discernment. This new method is then applied for evaluating the Multiple-Objective Ant Colony Optimization (MO-ACO) algorithm for Wireless Sensor Networks (WSN) deployment.

**Keywords:** Inter-Criteria Analysis, belief functions, information fusion, canonical decomposition, PCR5 rule.

## I. INTRODUCTION

In our previous work [1] we propose a new and improved version of classical Atanassov's InterCriteria Analysis (ICrA) [2]–[4] approach based on Belief Functions (BF-ICrA). This method proposes a better construction of Inter-Criteria Matrix that fully exploits all the information of the score matrix, and the closeness measure of agreement between criteria based on belief interval distance. In [5], we show how the fusion of many sources of evidences represented by Basic Belief Assignments (BBAs) defined on a same dichotomous frame of discernment can be fast and easily done thanks to the Proportional Conflict Redistribution rule no.5 (PCR5) and based on canonical decomposition of the BBAs, proposed recently in [6]. In the recent paper we consider BF-ICrA based on this promising technique. Then we show how to apply it for the evaluation of the Multiple-Objective Ant Colony Optimization (MO-ACO) algorithm for Wireless Sensor Networks (WSN) deployment. After a condensed presentation of basics of belief functions in Section II, including the short description of canonical decomposition of dichotomous BBAs approach, and the main steps of fast fusion method of dichotomous BBAs, in Section III the BF-ICrA method is described and analyzed. Section IV is devoted to the multi-objective ACO algorithm. In Section V the results of the fast BF-ICrA method with the MO-ACO algorithm for WSN layout deployment is presented and discussed. Conclusion is given in Section VI.

## II. BASICS OF BELIEF FUNCTIONS

### A. Basic definitions

Belief functions (BF) have been introduced by Shafer in [7] to model epistemic uncertainty and to combine distinct sources of evidence thanks to Dempster's rule of combination.

In Shafer's framework, we assume that the answer<sup>1</sup> of the problem under concern belongs to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are mutually exclusive and exhaustive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A proper Basic Belief Assignment (BBA) associated with a given source of evidence is defined [7] as a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are respectively defined from a proper BBA  $m(\cdot)$  by

$$Bel(A) = \sum_{B \in 2^\Theta | B \subseteq A} m(B), \quad (1)$$

and

$$Pl(A) = \sum_{B \in 2^\Theta | A \cap B \neq \emptyset} m(B) = 1 - Bel(\bar{A}), \quad (2)$$

where  $\bar{A}$  is the complement of  $A$  in  $\Theta$ .

$Bel(A)$  and  $Pl(A)$  are usually interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$ . The quantities  $m(\cdot)$  and  $Bel(\cdot)$  are one-to-one and linked by the Möbius inverse formula (see [7], p. 39).  $A$  is called a Focal Element (FE) of  $m(\cdot)$  if  $m(A) > 0$ . When all focal elements are singletons,  $m(\cdot)$  is called a *Bayesian BBA* [7] and its corresponding  $Bel(\cdot)$  function is equal to  $Pl(\cdot)$  and they are homogeneous to a (subjective) probability measure  $P(\cdot)$ . The vacuous BBA, representing a totally ignorant source, is defined as  $m_v(\Theta) = 1$ . A dichotomous BBA is a BBA defined on a FoD which has only two proper subsets, for instance  $\Theta = \{A, \bar{A}\}$  with  $A \neq \Theta$  and  $A \neq \emptyset$ . A dogmatic BBA is a BBA such that  $m(\Theta) = 0$ . If  $m(\Theta) > 0$  the BBA  $m(\cdot)$  is nondogmatic. A simple BBA is a BBA that has at most two focal sets and one of them is  $\Theta$ . A dichotomous non dogmatic mass of belief is a BBA having three focal elements  $A$ ,  $\bar{A}$  and  $A \cup \bar{A}$  with  $A$  and  $\bar{A}$  subsets of  $\Theta$ .

In his Mathematical Theory of Evidence [7], Shafer proposed to combine  $s \geq 2$  distinct sources of evidence represented by BBAs with Dempster's rule (i.e. the normalized conjunctive rule), which unfortunately behaves counterintuitively

<sup>1</sup>i.e. the solution, or the decision to take.

both in high and low conflicting situations as reported in [8]–[11]. In our previous works (see [12], Vol. 2 and Vol. 3 for full justification and examples) we did propose new rules of combination based on different Proportional Conflict Redistribution (PCR) principles, and we have shown the interest of the PCR rule No 5 (PCR5) for combining two BBAs, and PCR rule No 6 (PCR6) for combining more than two BBAs altogether [12], Vol. 2. PCR6 coincides with PCR5 when one combines two sources. The difference between PCR5 and PCR6 lies in the way the proportional conflict redistribution is done as soon as three (or more) sources are involved in the fusion. PCR5 transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is entirely preserved in this fusion process.

The general (complicate) formulas for PCR5 and PCR6 rules are given in [12], Vol. 2. The fusion of two BBAs based on PCR5 (or PCR6) rule which will be use for canonical decomposition of a dichotomous BBA is obtained by the formula

$$m_{PCR5}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2) + \sum_{\substack{X_2 \in 2^\Theta \\ X_2 \cap X = \emptyset}} \left[ \frac{m_1(X)^2 m_2(X_2)}{m_1(X) + m_2(X_2)} + \frac{m_2(X)^2 m_1(X_2)}{m_2(X) + m_1(X_2)} \right], \quad (3)$$

where all denominators in (3) are different from zero. If a denominator is zero, that fraction is discarded.

From the implementation point of view, PCR6 is simpler to implement than PCR5. For convenience, very basic (not optimized) Matlab™ codes of PCR5 and PCR6 fusion rules can be found in [12], [13] and from the toolboxes repository on the web [14]. The main drawback of PCR5 and PCR6 rules is their very high combinatorial complexity when the number of source is big, as well as the cardinality of the FoD. In this case, PCR5 or PCR6 rules cannot be used directly because of memory overflow. Even for combining BBAs defined on a simple dichotomous FoD as those involved in the Inter-Criteria Analysis (ICrA), the computational time for combining more than 10 sources can take several hours<sup>2</sup>. That is why a fast fusion method to combine dichotomous BBAs is necessary, and we present it in the next subsections.

### B. Canonical decomposition of dichotomous BBA

A FoD  $\Theta = \{A, \bar{A}\}$  is called dichotomous if it consists of only two proper subsets  $A$  and  $\bar{A}$  with  $A \cup \bar{A} = \Theta$  and  $A \cap \bar{A} = \emptyset$ , where  $\bar{A}$  is the complement of  $A$  in  $\Theta$  and  $A$  is different from  $\Theta$  and from Empty-Set. We consider a given proper BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  of the general form

$$m(A) = a, \quad m(\bar{A}) = b, \quad m(A \cup \bar{A}) = 1 - a - b. \quad (4)$$

<sup>2</sup>with a MacBook Pro 2.8 GHz Intel Core i7 with 16 Go 1600 MHz DDR3 memory running Matlab™ R2018a.

The canonical decomposition problem consists in finding the two following simple proper BBAs  $m_p$  and  $m_c$  of the form

$$m_p(A) = x, \quad m_p(A \cup \bar{A}) = 1 - x, \quad (5)$$

$$m_c(\bar{A}) = y, \quad m_c(A \cup \bar{A}) = 1 - y, \quad (6)$$

with  $(x, y) \in [0, 1] \times [0, 1]$ , such that  $m = Fusion(m_p, m_c)$ , for a chosen rule of combination denoted by  $Fusion(\cdot, \cdot)$ . The simple BBA  $m_p(\cdot)$  is called the *pro-BBA* (or pro-evidence) of  $A$ , and the simple BBA  $m_c(\cdot)$  the *contra-BBA* (or contra-evidence) of  $A$ . The BBA  $m_p(\cdot)$  is interpreted as a source of evidence providing an uncertain evidence in favor of  $A$ , whereas  $m_c(\cdot)$  is interpreted as a source of evidence providing an uncertain contrary evidence about  $A$ .

In [6], we have shown that this decomposition is possible with Dempster's rule only if  $0 < a < 1$ ,  $0 < b < 1$  and  $a + b < 1$ , and we have  $x = \frac{a}{1-b}$  and  $y = \frac{b}{1-a}$ . However, any dogmatic BBA  $m(A) = a$ ,  $m(\bar{A}) = b$  with  $a + b = 1$  is not decomposable from Dempster's rule for the case when  $(a, b) \neq (1, 0)$  and  $(a, b) \neq (0, 1)$ , and the dogmatic BBAs  $m(A) = 1$ ,  $m(\bar{A}) = 0$ , or  $m(A) = 0$ ,  $m(\bar{A}) = 1$  have infinitely many decompositions based on Dempster's rule of combination. We have also proved that this canonical decomposition cannot be done from conjunctive, disjunctive, Yager's [15] or Dubois-Prade [16] rules of combination, neither from the averaging rule. The main result of [6] is that this canonical decomposition is unique and is always possible in all cases using the PCR5 rule of combination. This is very useful to implement a fast efficient approximating fusion method of dichotomous BBAs as presented in details in [5]. We recall the following two important theorems proved in [6].

**Theorem 1:** Consider a dichotomous FoD  $\Theta = \{A, \bar{A}\}$  with  $A \neq \emptyset$  and  $A \neq \Theta$  and a nondogmatic BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  defined on  $\Theta$  by  $m(A) = a$ ,  $m(\bar{A}) = b$ , and  $m(A \cup \bar{A}) = 1 - a - b$ , where  $a, b \in [0, 1]$  and  $a + b < 1$ . Then the BBA  $m(\cdot)$  has a unique canonical decomposition using PCR5 rule of combination of the form  $m = PCR5(m_p, m_c)$  with pro-evidence  $m_p(A) = x$ ,  $m_p(A \cup \bar{A}) = 1 - x$  and contra-evidence  $m_c(\bar{A}) = y$ ,  $m_c(A \cup \bar{A}) = 1 - y$  where  $x, y \in [0, 1]$ .

**Theorem 2:** Any dogmatic BBA defined by  $m(A) = a$  and  $m(\bar{A}) = b$ , where  $a, b \in [0, 1]$  and  $a + b = 1$ , has a canonical decomposition using PCR5 rule of combination of the form  $m = PCR5(m_p, m_c)$  with  $m_p(A) = x$ ,  $m_p(A \cup \bar{A}) = 1 - x$  and  $m_c(\bar{A}) = y$ ,  $m_c(A \cup \bar{A}) = 1 - y$  where  $x, y \in [0, 1]$ .

Theorems 1 & 2 prove that the decomposition based on PCR5 always exists and it is unique for any dichotomous (nondogmatic, or dogmatic) BBA.

For the case of dichotomous nondogmatic BBA considered in Theorem 1, one has to find  $x$  and  $y$  solutions of the system

$$a = x(1 - y) + \frac{x^2 y}{x + y} = \frac{x^2 + xy - xy^2}{x + y}, \quad (7)$$

$$b = (1 - x)y + \frac{xy^2}{x + y} = \frac{y^2 + xy - x^2 y}{x + y}, \quad (8)$$

under the constraints  $(a, b) \in [0, 1]^2$ , and  $0 < a + b < 1$ . The explicit expression of  $x$  and  $y$  are difficult to obtain analytically (even with modern symbolic computing systems like Mathematica™, or Maple™) because one has a quartic equation to solve whose general analytical expression of its solutions is very complicate. Fortunately, the solutions can be easily calculated numerically by these computing systems, and even with Matlab™ system (thanks to the *fsolve* function) as soon as the numerical values are committed to  $a$  and to  $b$ , and this is what we use in our simulations.

### C. Fast Fusion of dichotomous BBAs

The main idea for making the fast fusion of dichotomous BBAs  $m_s(\cdot)$ , for  $s = 1, 2, \dots, S$  defined on the same FoD  $\Theta$  is based on the three following main steps:

- 1) In the first step, one decomposes canonically each dichotomous BBA  $m_s(\cdot)$  into its pro and contra evidences  $m_{p,s} = (m_{p,s}(A), m_{p,s}(\bar{A}), m_{p,s}(A \cup \bar{A})) = (x_s, 0, 1 - x_s)$  and  $m_{c,s} = (m_{c,s}(A), m_{c,s}(\bar{A}), m_{c,s}(A \cup \bar{A})) = (0, y_s, 1 - y_s)$ ;
- 2) In the second step, one combines the pro-evidences  $m_{p,s}$  for  $s = 1, 2, \dots, S$  altogether to get a global pro-evidence  $m_p$ , and in parallel one combines all the contra-evidences  $m_{c,s}$  for  $s = 1, 2, \dots, S$  altogether to get a global contra-evidence  $m_c$ . The fusion step of pro and contra evidences is based on conjunctive rule of combination;
- 3) Once  $m_p$  and  $m_c$  are calculated, then one combines them with PCR5 fusion rule to get the final result.

Because the PCR5 rule of combination is not associative, the fusion of the canonical BBAs followed by their PCR5 fusion will not provide in general the same result as the direct fusion of the dichotomous BBAs altogether but only an approximate result, which is normal. However, this new fusion approach is interesting because the fusion of the pro-evidence  $m_{p,s}$  (resp. contra-evidences  $m_{c,s}$ ) is very simple because there is no conflict between  $m_{p,s}$  (resp. between  $m_{c,s}$ ), so that their fusion can be done quite easily and a large number of sources can be combined without a high computational burden. In fact, with this fusion approach, only one PCR5 fusion step of simple (combined) canonical BBAs is needed at the very end of the fusion process. In [5], we have proved with a Monte-Carlo simulation analysis that the approximation obtained by this new fusion method based on the fusion of pro-evidences and contra-evidences with respect to the direct fusion of the BBAs with PCR5 (or PCR6 when considering more than two sources to combine) is effective because the agreement between the decision taken from the direct fusion method, and the indirect (canonical decomposition based) method is very good. This new fusion method based on this canonical decomposition does not suffer of combinatorial complexity limitation which is of great interest in some applications because many (hundreds or even thousands) of dichotomous BBAs could be easily combined very quickly. Actually with this method what takes a bit time is only the canonical decomposition done by the

numerical solver. Our analysis [5] has shown that complexity of this fast approach is quasi-linear with the number of sources to combine.

### III. THE BF-ICRA METHOD

In [1], we did present an improved version of Atanassov's Inter-Criteria Analysis (ICrA) method [2]–[4] based on belief functions. This new method has been named BF-ICrA (Belief Function based Inter-Criteria Analysis) for short. It has already been applied to GPS surveying problems in [17]. We present briefly in this section the principles of BF-ICrA.

BF-ICrA starts with the construction of an  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  from the score matrix  $\mathbf{S} = [S_{ij}]$ . The BBA matrix  $\mathbf{M}$  is obtained as follows - see [18] for details and justification.

$$m_{ij}(A_i) = Bel_{ij}(A_i), \quad (9)$$

$$m_{ij}(\bar{A}_i) = Bel_{ij}(\bar{A}_i) = 1 - Pl_{ij}(A_i), \quad (10)$$

$$m_{ij}(A_i \cup \bar{A}_i) = Pl_{ij}(A_i) - Bel_{ij}(A_i), \quad (11)$$

where<sup>3</sup>

$$Bel_{ij}(A_i) \triangleq Sup_j(A_i)/A_{\max}^j, \quad (12)$$

$$Bel_{ij}(\bar{A}_i) \triangleq Inf_j(A_i)/A_{\min}^j, \quad (13)$$

with

$$Sup_j(A_i) \triangleq \sum_{k \in \{1, \dots, M\} | S_{kj} \leq S_{ij}} |S_{ij} - S_{kj}|, \quad (14)$$

$$Inf_j(A_i) \triangleq - \sum_{k \in \{1, \dots, M\} | S_{kj} \geq S_{ij}} |S_{ij} - S_{kj}|, \quad (15)$$

and

$$A_{\max}^j \triangleq \max_i Sup_j(A_i), \quad (16)$$

$$A_{\min}^j \triangleq \min_i Inf_j(A_i). \quad (17)$$

For another criterion  $C_{j'}$  and the  $j'$ -th column of the score matrix we will obtain another set of BBA values  $m_{ij'}(\cdot)$ . Applying this method for each column of the score matrix we are able to compute the BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  whose each component is in fact a triplet  $(m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i))$  of BBA values in  $[0, 1]$  such that  $m_{ij}(A_i) + m_{ij}(\bar{A}_i) + m_{ij}(A_i \cup \bar{A}_i) = 1$  for all  $i = 1, \dots, M$  and  $j = 1, \dots, N$ .

The next step of BF-ICrA approach is the construction of the  $N \times N$  Inter-Criteria Matrix  $\mathbf{K} = [K_{jj'}]$  from  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  where elements  $K_{jj'}$  corresponds to the BBA  $(m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  about positive consonance  $\theta$ , negative consonance  $\bar{\theta}$  and uncertainty between criteria  $C_j$  and  $C_{j'}$  respectively. The construction of the triplet  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  is based on two steps:

- **Step 1 (BBA construction):** Getting  $m_{jj'}^i(\cdot)$ .

For each alternative  $A_i$  for  $i = 1, \dots, M$ , we first compute the BBA  $(m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  for

<sup>3</sup>assuming that  $A_{\max}^j \neq 0$  and  $A_{\min}^j \neq 0$ . If  $A_{\max}^j = 0$  then  $Bel_{ij}(A_i) = 0$ , and if  $A_{\min}^j = 0$  then  $Pl_{ij}(A_i) = 1$ .

any two criteria  $j, j' \in \{1, 2, \dots, N\}$ . For this, we consider two sources of evidences (SoE) indexed by  $j$  and  $j'$  providing the BBA  $m_{ij}$  and  $m_{ij'}$  defined on the simple FoD  $\{A_i, \bar{A}_i\}$  and denoted  $m_{ij} = [m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i)]$  and  $m_{ij'} = [m_{ij'}(A_i), m_{ij'}(\bar{A}_i), m_{ij'}(A_i \cup \bar{A}_i)]$ . We also denote  $\Theta = \{\theta, \bar{\theta}\}$  the FoD about the relative state of the two SoE, where  $\theta$  means that the two SoE agree,  $\bar{\theta}$  means that they disagree and  $\theta \cup \bar{\theta}$  means that we don't know. Hence, two SoE are in total agreement if both commit their maximum belief mass to the same element  $A_i$  or to the same element  $\bar{A}_i$ . Similarly, two SoE are in total disagreement if each one commits its maximum mass of belief to one element and the other to its opposite, that is if one has  $m_{ij}(A_i) = 1$  and  $m_{ij'}(\bar{A}_i) = 1$ , or if  $m_{ij}(\bar{A}_i) = 1$  and  $m_{ij'}(A_i) = 1$ . Based on this very simple and natural principle, one can now compute the belief masses as follows:

$$m_{jj'}^i(\theta) = m_{ij}(A_i)m_{ij'}(A_i) + m_{ij}(\bar{A}_i)m_{ij'}(\bar{A}_i), \quad (18)$$

$$m_{jj'}^i(\bar{\theta}) = m_{ij}(A_i)m_{ij'}(\bar{A}_i) + m_{ij}(\bar{A}_i)m_{ij'}(A_i), \quad (19)$$

$$m_{jj'}^i(\theta \cup \bar{\theta}) = 1 - m_{jj'}^i(\theta) - m_{jj'}^i(\bar{\theta}). \quad (20)$$

$m_{jj'}^i(\theta)$  represents the degree of agreement between the BBA  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$  for the alternative  $A_i$ ,  $m_{jj'}^i(\bar{\theta})$  represents the degree of disagreement of the two BBAs and  $m_{jj'}^i(\theta \cup \bar{\theta})$  the level of uncertainty (i.e. how much we don't know if they agree or disagree). By construction  $m_{jj'}^i(\cdot) = m_{jj'}^i(\cdot), m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}) \in [0, 1]$  and  $m_{jj'}^i(\theta) + m_{jj'}^i(\bar{\theta}) + m_{jj'}^i(\theta \cup \bar{\theta}) = 1$ . This BBA modeling permits to build a set of  $M$  symmetrical Inter-Criteria Belief Matrices (ICBM)  $\mathbf{K}^i = [K_{jj'}^i]$  of dimension  $N \times N$  relative to each alternative  $A_i$  whose components  $K_{jj'}^i$  correspond to the triplet of BBA values  $m_{jj'}^i = (m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  modeling the belief of agreement and of disagreement between  $C_j$  and  $C_{j'}$  based on  $A_i$ .

- **Step 2 (fusion):** Getting  $\mathbf{m}_{jj'}^i(\cdot)$ .

In this step, one needs to combine the BBAs  $\mathbf{m}_{jj'}^i(\cdot)$  for  $i = 1, \dots, M$  altogether to get the component  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  of the Inter-Criteria Belief matrix<sup>4</sup> (ICBM)  $\mathbf{K} = [K_{jj'}]$ . For this and from the theoretical standpoint, we recommend to use the PCR6 fusion rule [12] (Vol. 3) because of known deficiencies of Dempster's rule.

Once the global Inter-Criteria Belief Matrix (ICBM)  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$  is calculated, we can identify the criteria that are in strong agreement, in strong disagreement, and those on which we are uncertain. For identifying the criteria that are in strong agreement, we evaluate the distance of each component of  $K_{jj'}$  with the BBA

<sup>4</sup>For the presentation convenience, the ICBM  $\mathbf{K}$ , with  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$ , is decomposed into three matrices  $\mathbf{K}(\theta) = [K_{jj'}^\theta = m_{jj'}(\theta)]$ ,  $\mathbf{K}(\bar{\theta}) = [K_{jj'}^{\bar{\theta}} = m_{jj'}(\bar{\theta})]$  and  $\mathbf{K}(\theta \cup \bar{\theta}) = [K_{jj'}^{\theta \cup \bar{\theta}} = 1 - m_{jj'}(\theta) - m_{jj'}(\bar{\theta})]$ .

representing the best agreement state and characterized by the specific BBA<sup>5</sup>  $m_T(\theta) = 1$ . From a similar approach we can also identify, if we want, the criteria that are in very strong disagreement using the distance of  $m_{jj'}(\cdot)$  with respect to the BBA representing the best disagreement state characterized by the specific BBA  $m_F(\bar{\theta}) = 1$ . We use the belief interval distance  $d_{BI}(m_1, m_2)$  presented in [19] for measuring the distance between the two BBAs.

#### A. Fast BF-ICrA method

The computational complexity of BF-ICrA is of course higher than the complexity of ICrA because it makes a more precise evaluation of local and global inter-criteria belief matrices with respect to inter-criteria matrices calculated by Atanassov's ICrA. The overall reduction of the computational burden of the original MCDM problem thanks to BF-ICrA depends highly on the problem under concern, the complexity and cost to evaluate each criteria involved in it, as well as the number of redundant criteria identified by BF-ICrA method.

The main drawback of BF-ICrA method is the PCR6 combination required in its step 2 for combining altogether the dichotomous BBAs  $m_{jj'}^i(\cdot)$ . Because of combinatorial complexity of PCR6 rule, it cannot work in reasonable computational time as soon as the number of sources to combine altogether is greater than 10, which prevents its use for solving ICrA problems involving more than 10 alternatives (as in the examples 2 and 3 presented in section V). That is why it is necessary to adapt the original BF-ICrA method for working with a large number of alternatives and criteria. For this, we can in step 2 of BF-ICrA exploit the method for the fast fusion of dichotomous BBAs presented in section II-C. More precisely, each dichotomous BBA  $m_{jj'}^i(\cdot)$  will be canonically decomposed in its pro-evidence  $m_{jj',p}^i(\cdot)$  and its contra-evidence  $m_{jj',c}^i(\cdot)$  that will be combined separately to get the global pro-evidence  $m_{jj',p}(\cdot)$  and the global contra-evidence  $m_{jj',c}(\cdot)$ . Then, the BBAs  $m_{jj',p}(\cdot)$  and  $m_{jj',c}(\cdot)$  are combined with PCR5 rule to get the BBAs  $m_{jj'}(\cdot)$  and, finally, the global Inter-Criteria Belief Matrix  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$ . The principle of this modified step 2 of BF-ICrA is summarized in the Figure 1 for convenience.

Another simpler fusion method to combine the dichotomous BBAs  $m_{jj'}^i(\cdot)$  would just consist to average them. In section V, we will show how these two methods behave in the examples chosen for the evaluation of MO-ACO Algorithm for optimal WSN deployment.

## IV. MULTI-OBJECTIVE ACO ALGORITHM

Recently Wireless Sensor Networks (WSNs) have attracted the attention of the research scientists community, conditioned by a set of challenges: theoretical and practical. WSNs consists of distributed sensor nodes and their main purpose is to monitor the real-time environmental status, based on gathering available sensor information, processing and transmitting the

<sup>5</sup>We use the index  $T$  in the notation  $m_T(\cdot)$  to refer that the agreement is true, and  $F$  in  $m_F(\cdot)$  to specify that the agreement is false.

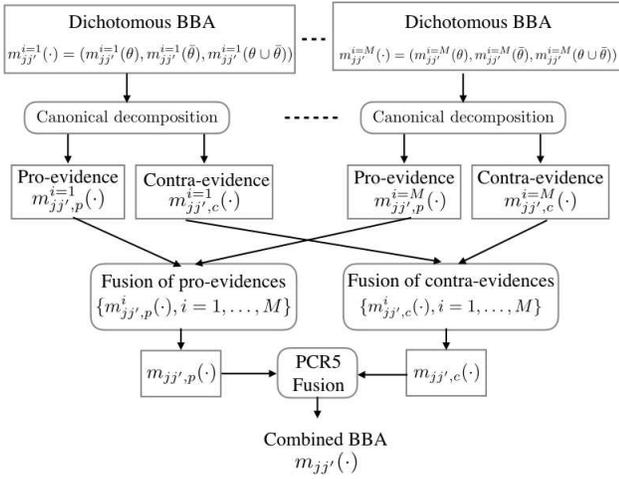


Figure 1. Principle of fast fusion of  $m_{jj'}^i(\cdot)$  of Step 2 of BF-ICrA.

collected data to the specified remote base station. It is a promising technology that is used in a coverage of application requiring minimum human contribution, ranging from civil and military to healthcare and environmental monitoring. One of the key mission of WSN is the full surveillance of the monitoring region with a minimal number of sensors and minimized energy consumption of the network. The lifetime of the sensors is strongly related to the amount of the power loaded in the battery, that is why the control of the energy consumption of sensors is an important active research problem. The small energy storage capacity of sensor nodes intrudes the possibility to gather the information directly to the main base. Because of this they transfer their data to the so called High Energy Communication Node (HECN), which is able to collect the information from across the network and to transmit it to the base computer for processing. The sensors transmit their data to the HECN, either directly or via hops, using closest sensors as communication relays. The WSN can have large numbers of nodes and the problem can be very complex.

In order to solve successfully the key mission of WSNs, in [20], we did apply multi-objective Ant Colony Optimization (ACO) to solve this hard, from the computational point of view, telecommunication problem. The number of ants is one of the key algorithm parameters in the ACO and it is important to find the optimal number of ants needed to achieve good solutions with minimal computational resources. In [20], the optimal solution was obtained by applying the classical Atanassov's ICRA method. In the next section we will present the results obtained by the fast BF-ICrA approach and compare their results.

The problem of designing a WSN is multi-objective, with two objective functions: 1) one wants to minimize the energy consumption of the nodes in the network, and 2) one wants to minimize the number of nodes. The full coverage of the network and connectivity are considered as constraints. For solving this problem, we have proposed to use a Multi-

Objective Ant Colony Optimization (MO-ACO) algorithm in [20] and we have studied the influence of the number of ants on the algorithm performance and quality of the achieved solutions. The computational resources, which the algorithm needs, are not negligible. The computational resources depends on the size of the solved problem and on the number of ants. The aim is to find a minimal number of ants which allow the algorithm to find good solution for WSN deployment.

The ACO algorithm uses a colony of artificial ants that behave as cooperating agents. With the help of the pheromone and the heuristic information they try to construct better solutions and to find the optimal ones. The pheromone corresponds to the global memory of the ants and the heuristic information is a some preliminary knowledge of the problem. The problem is represented by a graph and the solution is represented by a path in the graph or by tree in the graph. Ants start from random nodes and construct feasible solutions. When all ants construct their solution the pheromone is updated. The new, added, pheromone depends to the quality of the solution. The elements of the graph, which belong to better solutions will receive more pheromone and will be more desirable in the next iteration. In our implementation, we use the MAX-MIN Ant System (MMAS) which is one of the most successful ant approaches originally presented in [21]. In our case, the graph of the problem is represented by a square grid. The nodes of the graph are enumerated. The ants will deposit their pheromone on the nodes of the grid. We will deposit the sensors on the nodes of the grid too. The solution is represented by tree. An ant starts to create a solution starting from random node, which communicates with the HECN. Construction of the heuristic information is a crucial point in the ant algorithms. Our heuristic information represented by (21) is a product of three values.

$$\eta_{ij}(t) = s_{ij} l_{ij} (1 - b_{ij}) \quad (21)$$

where  $s_{ij}$  is the number of the new points (nodes of the graph) which the new sensor will cover, and which are not covered by other sensors, and

$$l_{ij} = \begin{cases} 1 & \text{if communication exists;} \\ 0 & \text{if there is no communication.} \end{cases} \quad (22)$$

and where  $b_{ij}$  is the solution matrix. The matrix element  $b_{ij}$  equals 1 when there is sensor on this position, otherwise  $b_{ij} = 0$ . With  $s_{ij}$ , we try to increase the number of points covered by one sensor and thus to decrease the number of sensors we need. With  $l_{ij}$ , we guarantee that all sensors will be connected. With  $b_{ij}$  we guarantee that maximum one sensor will be mapped on the same point. The search stops when transition probability  $p_{ij} = 0$  for all values of  $i$  and  $j$ . It means that there are no more free positions, or that all area is fully covered. At the end of every iteration the quantity of the pheromone is updated according to the rule:  $\tau_{ij} \leftarrow \rho \tau_{ij} + \Delta \tau_{ij}$ , with the increment  $\Delta \tau_{ij} = 1/F(k)$  if  $(i, j)$  belongs to the non-dominated solution constructed by ant  $k$ , or  $\Delta \tau_{ij} = 0$  otherwise. The parameter  $\rho$  is a pheromone

decreasing parameter chosen in  $[0, 1]$ . This parameter  $\rho$  models evaporation in the nature and decreases the influence of old information on the search process. After that, we add the new pheromone, which is proportional to the value of the fitness function constructed as  $F(k) = \frac{f_1(k)}{\max_i(f_1(i))} + \frac{f_2(k)}{\max_i(f_2(i))}$ , where  $f_1(k)$  is the number of sensors proposed by the  $k$ -th ant, and  $f_2(k)$  is the energy of the solution of the  $k$ -th ant. These are also the objective functions of the WSN layout problem. We normalize the values of two objective functions with their maximal achieved values from the first iteration.

V. APPLICATION OF THE FAST BF-ICRA METHOD

In this section we present the results of the fast BF-ICrA method with the MO-ACO algorithm for WSN layout deployment. Fidanova and Roeva have developed a software, which realizes the MO-ACO algorithm. This software can solve the problem at any rectangular area, the communication and the coverage radius can be different and can have any positive value. We can have regions in the area. The program was written in C language, and the tests were run on computer with an Intel Pentium 2.8GHz processor. In their tests, they use an example where the area is square. The coverage and communication radii cover 30 points. The HECN is fixed in the centre of the area. In the sequel we consider three examples of areas with three sizes:  $350 \times 350$  points,  $500 \times 500$  points, and  $700 \times 700$  points. The MO-ACO algorithm is based on 30 runs for each number of ants. We extract the Pareto front from the solutions of these 30 runs, and we show the achieved non dominated solutions (approximate Pareto fronts) for each case on which the BF-ICrA will be applied. The score matrices for each case is given in Tables I, II and III [20].

	ACO <sub>1</sub>	ACO <sub>2</sub>	ACO <sub>3</sub>	ACO <sub>4</sub>	ACO <sub>5</sub>	ACO <sub>6</sub>	ACO <sub>7</sub>	ACO <sub>8</sub>	ACO <sub>9</sub>	ACO <sub>10</sub>
111	30	36	30	30	30	30	30	30	30	30
112	30	36	30	30	30	30	30	30	30	30
113	28	35	28	30	30	30	28	28	28	28
114	26	26	26	26	26	26	26	26	26	26
115	26	26	26	26	26	26	26	26	26	26
116	26	26	26	26	26	25	25	26	25	25

Table I

THE 6 × 10 SCORE MATRIX S FOR 350 × 350 CASE (EXAMPLE 1).

	ACO <sub>1</sub>	ACO <sub>2</sub>	ACO <sub>3</sub>	ACO <sub>4</sub>	ACO <sub>5</sub>	ACO <sub>6</sub>	ACO <sub>7</sub>	ACO <sub>8</sub>	ACO <sub>9</sub>	ACO <sub>10</sub>
223	90	96	90	90	89	81	90	90	90	90
224	61	96	89	89	88	65	61	59	57	71
225	61	96	74	58	60	58	57	58	57	57
226	59	95	73	57	59	57	56	58	57	57
227	60	57	57	57	57	56	56	57	57	57
228	60	57	57	57	57	56	56	57	54	57
229	58	57	57	55	57	56	56	56	54	56
230	57	57	57	55	57	52	56	54	54	56
231	57	55	57	55	55	52	56	54	54	56
232	57	55	55	51	54	50	52	51	54	48
233	57	55	55	51	54	50	51	51	54	48
234	57	55	55	51	53	50	51	48	53	48
235	57	55	54	51	53	50	51	48	50	48
236	57	55	54	51	53	50	51	48	50	48
237	57	55	54	51	53	50	51	48	50	48
238	57	55	53	51	53	50	51	48	50	48
239	56	55	53	50	53	50	51	48	50	48
240	53	53	53	50	53	50	51	48	50	48
241	53	53	53	50	53	50	51	48	50	48
242	53	53	53	50	53	50	51	48	50	48
243	53	53	53	50	53	50	51	48	50	48
244	53	53	53	50	52	50	51	48	50	48

Table II

THE 22 × 10 SCORE MATRIX S FOR 500 × 500 CASE (EXAMPLE 2).

Each row of S corresponds to the number of sensors used in WSN to cover the area as indicated in the first column at the left side of the score matrix. Each column of S corresponds

	ACO <sub>1</sub>	ACO <sub>2</sub>	ACO <sub>3</sub>	ACO <sub>4</sub>	ACO <sub>5</sub>	ACO <sub>6</sub>	ACO <sub>7</sub>	ACO <sub>8</sub>	ACO <sub>9</sub>	ACO <sub>10</sub>
437	173	173	173	173	173	118	168	172	261	172
438	173	173	173	173	173	118	112	117	260	172
439	172	173	173	173	140	93	110	115	131	172
440	172	173	173	173	115	93	110	114	111	162
441	172	173	173	122	111	93	110	114	111	110
442	172	173	173	114	111	93	110	112	111	110
443	172	150	123	114	111	93	110	112	111	110
444	124	112	112	106	107	93	110	102	111	105
445	117	112	112	106	107	93	110	102	108	105
446	117	112	105	105	105	93	107	102	104	105
447	117	112	105	105	105	93	105	102	102	105
448	115	111	105	105	105	93	105	102	102	105
449	115	111	105	105	105	93	102	99	102	105
450	113	111	105	105	105	93	102	99	102	105
451	113	109	105	105	105	93	102	99	97	105
452	113	109	105	105	105	93	99	99	97	104
453	113	109	105	105	105	93	99	99	97	104
454	113	109	105	105	96	93	96	96	96	104
455	106	106	105	105	96	93	96	96	96	97

Table III

THE 19 × 10 SCORE MATRIX S FOR 700 × 700 CASE (EXAMPLE 3).

to ACO<sub>j</sub> algorithm used with  $j$  ants ( $j = 1, 2, \dots, 10$ ). Each element  $S_{ij}$  of S corresponds to the energy corresponding to this number of sensors and with the number of ants used for Multiple Objective ACO algorithm.

Application of BF-ICrA in example 1 (350 × 350 points)

In this example, one sees from the score matrix of the Table I that ACO<sub>1</sub>, ACO<sub>3</sub> and ACO<sub>9</sub> algorithms perform equally for all alternatives (i.e. all rows) and they define a first group/cluster of methods providing exactly the same performances. Similarly, ACO<sub>4</sub>, ACO<sub>5</sub> and ACO<sub>6</sub> constitute a second group of algorithms. The third group is made of ACO<sub>7</sub>, ACO<sub>8</sub> and ACO<sub>10</sub> algorithms. It is worth noting that these three groups {ACO<sub>1</sub>, ACO<sub>3</sub>, ACO<sub>9</sub>}, {ACO<sub>4</sub>, ACO<sub>5</sub>, ACO<sub>6</sub>}, and {ACO<sub>7</sub>, ACO<sub>8</sub>, ACO<sub>10</sub>} differ only very slightly, whereas the ACO<sub>2</sub> algorithm (i.e the 2nd column of the score matrix S) differs a bit more from all the three aforementioned groups.

**Example 1 with fast PCR6:** If we apply the fast BF-ICrA method using approximate PCR6 fusion rule based on the canonical decomposition of the  $M = 6$  dichotomous BBAs ( $m_{jj}^i(\theta), m_{jj'}^i(\theta), m_{jj'}^i(\theta \cup \theta')$ ), we get the matrix of mass of belief of agreement between criteria given in Table<sup>6</sup> IV.

0.865	0.821	0.865	0.790	0.790	0.790	0.806	0.806	0.865	0.806
0.821	0.928	0.821	0.950	0.950	0.950	0.805	0.805	0.821	0.805
0.865	0.821	0.865	0.790	0.790	0.790	0.806	0.806	0.865	0.806
0.790	0.950	0.790	1.000	1.000	1.000	0.795	0.795	0.790	0.795
0.790	0.950	0.790	1.000	1.000	1.000	0.795	0.795	0.790	0.795
0.806	0.805	0.806	0.795	0.795	0.795	0.843	0.843	0.806	0.843
0.806	0.805	0.806	0.795	0.795	0.795	0.843	0.843	0.806	0.843
0.865	0.821	0.865	0.790	0.790	0.790	0.806	0.806	0.865	0.806
0.806	0.805	0.806	0.795	0.795	0.795	0.843	0.843	0.806	0.843

Table IV

MATRIX  $K_{\approx PCR6}(\theta)$  FOR EXAMPLE 1.

The matrix of distances to full agreement based on fast BF-ICrA method, denoted by  $D_{\approx PCR6}(\theta)$ , is given in Table V.

In examining the table V, one sees that ACO<sub>1</sub>, ACO<sub>3</sub> and ACO<sub>9</sub> are at a small distance 0.134, with respect to other algorithms, so that they belong to the same group and behave similarly. Same remarks holds for the group {ACO<sub>4</sub>, ACO<sub>5</sub>, ACO<sub>6</sub>} because its inter-distance is zero,

<sup>6</sup>All the numerical values presented in the matrices have been truncated at their 3rd digit for typesetting convenience.

0.134	0.178	<b>0.134</b>	0.209	0.209	0.209	0.193	0.193	<b>0.134</b>	0.193
0.178	0.071	0.178	0.049	0.049	0.049	0.194	0.194	0.178	0.194
<b>0.134</b>	0.178	0.134	0.209	0.209	0.209	0.193	0.193	0.134	0.193
0.209	0.049	0.209	0	0	0	0.204	0.204	0.209	0.204
0.209	0.049	0.209	0	0	0	0.204	0.204	0.209	0.204
0.209	0.049	0.209	0	0	0	0.204	0.204	0.209	0.204
0.193	0.194	0.193	0.204	0.204	0.204	0.156	<b>0.156</b>	0.193	<b>0.156</b>
0.193	0.194	0.193	0.204	0.204	0.204	<b>0.156</b>	0.156	0.193	<b>0.156</b>
<b>0.134</b>	0.178	0.134	0.209	0.209	0.209	0.193	0.193	0.134	0.193
0.193	0.194	0.193	0.204	0.204	0.204	<b>0.156</b>	<b>0.156</b>	0.193	0.156

Table V

 MATRIX  $\mathbf{D}_{\approx PCR6}(\theta)$  WITH FAST BF-ICRA FOR EXAMPLE 1.

and for the group  $\{\text{ACO}_7, \text{ACO}_8, \text{ACO}_{10}\}$  because its inter-distance is 0.156. In a relative manner  $\text{ACO}_2$  appears closer to  $\{\text{ACO}_4, \text{ACO}_5, \text{ACO}_6\}$ , than  $\{\text{ACO}_1, \text{ACO}_3, \text{ACO}_9\}$  or  $\{\text{ACO}_7, \text{ACO}_8, \text{ACO}_{10}\}$ , which intuitively makes sense when comparing directly the columns of the matrix of Table I.

**Example 1 with averaging fusion:** The matrix of distances to full agreement based on BF-ICrA method using average fusion rule, denoted by  $\mathbf{D}_{\text{Aver.}}(\theta)$ , is given in Table VI.

0.084	0.082	0.084	0.081	0.081	0.081	0.156	0.156	0.084	0.156
0.082	0.030	0.082	0.016	0.016	0.016	0.142	0.142	0.082	0.142
0.084	0.082	0.084	0.081	0.081	0.081	0.156	0.156	0.084	0.156
0.081	0.016	0.081	0	0	0	0.138	0.138	0.081	0.138
0.081	0.016	0.081	0	0	0	0.138	0.138	0.081	0.138
0.081	0.016	0.081	0	0	0	0.138	0.138	0.081	0.138
0.156	0.142	0.156	0.138	0.138	0.138	0.198	0.198	0.156	0.198
0.156	0.142	0.156	0.138	0.138	0.138	0.198	0.198	0.156	0.198
0.084	0.082	0.084	0.081	0.081	0.081	0.156	0.156	0.084	0.156
0.156	0.142	0.156	0.138	0.138	0.138	0.198	0.198	0.156	0.198

Table VI

 MATRIX  $\mathbf{D}_{\text{Aver.}}(\theta)$  WITH BF-ICrA USING AVERAGING RULE FOR EXAMPLE 1.

One sees that only the group  $\{\text{ACO}_4, \text{ACO}_5, \text{ACO}_6\}$  can be clearly identified based on the averaging fusion rule. The other groups  $\text{ACO}_2$  appears also close to  $\{\text{ACO}_4, \text{ACO}_5, \text{ACO}_6\}$ . But  $\text{ACO}_1, \text{ACO}_3$  and  $\text{ACO}_9$  are closer to  $\{\text{ACO}_4, \text{ACO}_5, \text{ACO}_6\}$  also than in-between. Same remarks holds for  $\text{ACO}_7, \text{ACO}_8$ , and  $\text{ACO}_{10}$ . So one sees that the averaging fusion rule is not recommended for making the BF-ICrA in this example.

*Application of BF-ICrA in example 2 (500 × 500 points)*

**Example 2 with fast PCR6:** If we apply the fast BF-ICrA method using approximate PCR6 fusion rule based on the canonical decomposition of the  $M = 22$  dichotomous BBAs  $(m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$ , we get the following matrix of distances to full agreement, denoted by  $\mathbf{D}_{\approx PCR6}(\theta)$ , given in Table VII.

0.158	0.376	0.338	0.300	0.286	0.279	0.247	0.251	<b>0.225</b>	0.280
<b>0.376</b>	0.324	0.426	0.456	0.437	0.453	0.457	0.433	0.435	0.449
<b>0.338</b>	0.426	0.407	0.411	0.382	0.423	0.418	0.402	0.393	0.414
<b>0.300</b>	0.456	0.411	0.349	0.323	0.381	0.368	0.370	0.362	0.363
<b>0.286</b>	0.437	0.382	0.323	0.284	0.348	0.334	0.334	0.328	0.333
<b>0.279</b>	0.453	0.423	0.381	0.348	0.316	0.298	0.317	0.308	0.308
<b>0.247</b>	0.457	0.418	0.368	0.334	0.298	0.235	0.276	0.255	0.283
<b>0.251</b>	0.433	0.402	0.370	0.334	0.317	0.276	0.265	0.260	0.303
<b>0.225</b>	0.435	0.393	0.362	0.328	0.308	0.255	0.260	0.211	0.304
<b>0.280</b>	0.449	0.414	0.363	0.333	0.308	0.283	0.303	0.304	0.277

Table VII

 MATRIX  $\mathbf{D}_{\approx PCR6}(\theta)$  WITH FAST BF-ICrA FOR EXAMPLE 2.

Based on these results, one sees that no clear group can be identified but we emphasize in boldface in Table VII the minimal value for each row of the distance matrix  $\mathbf{D}_{\approx PCR6}(\theta)$

(diagonal elements excluded). We see that  $\text{ACO}_2$  is at the farthest distance of  $\text{ACO}_1$  because  $D_{12}(\theta) = 0.376$ , but in the mean time  $\text{ACO}_2$  is at closest distance to  $\text{ACO}_1$  because  $D_{2j}(\theta) > 0.376$  (for  $j > 2$ ) as shown in second line of Table VII. So we can conclude that  $\text{ACO}_2$  is not close to any other algorithm in fact. If we choose a ad-hoc distance threshold, say for instance 0.28, then we can identify the group  $\{\text{ACO}_1, \text{ACO}_7, \text{ACO}_8, \text{ACO}_9\}$ .

**Example 2 with averaging fusion:** The matrix of distances to full agreement based on BF-ICrA method using average fusion rule, denoted by  $\mathbf{D}_{\text{Aver.}}(\theta)$ , is given in Table VIII.

0.361	0.316	0.310	0.311	0.336	0.300	0.306	0.316	0.320	0.309
0.316	0.125	0.158	0.198	0.225	0.187	0.216	0.225	0.240	0.206
0.310	0.158	0.165	0.185	0.215	0.178	0.200	0.215	0.227	0.193
0.311	0.198	0.185	0.183	0.216	0.181	0.197	0.217	0.231	0.192
0.336	0.225	0.215	0.216	0.243	0.214	0.231	0.249	0.261	0.226
0.300	0.187	0.178	0.181	0.214	0.159	0.175	0.194	0.210	0.176
0.306	0.216	0.200	0.197	0.231	0.175	0.181	0.202	0.216	0.186
0.316	0.225	0.215	0.217	0.249	0.194	0.202	0.215	0.229	0.204
0.320	0.240	0.227	0.231	0.261	0.210	0.216	0.229	0.233	0.222
0.309	0.206	0.193	0.192	0.226	0.176	0.186	0.204	0.222	0.183

Table VIII

 MATRIX  $\mathbf{D}_{\text{Aver.}}(\theta)$  WITH BF-ICrA USING AVERAGING RULE FOR EXAMPLE 2.

Based on the average fusion rule there is no clear clustering of algorithms. However based on shortest inter-distance we could make the following distinct pairwise groupings  $\{\text{ACO}_2, \text{ACO}_3\}$ ,  $\{\text{ACO}_6, \text{ACO}_7\}$ ,  $\{\text{ACO}_4, \text{ACO}_{10}\}$ ,  $\{\text{ACO}_8, \text{ACO}_9\}$  and  $\{\text{ACO}_1, \text{ACO}_5\}$  if necessary, but remember that average fusion rule cannot provide the best result as shown in Example 1.

*Application of BF-ICrA in example 3 (700 × 700 points)*

**Example 3 with fast PCR6:** If we apply the fast BF-ICrA method using approximate PCR6 fusion rule based on the canonical decomposition of the  $M = 19$  dichotomous BBAs  $(m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$ , we get the matrix of distances to full agreement, denoted by  $\mathbf{D}_{\approx PCR6}(\theta)$ , given in Table IX.

0.313	0.388	0.465	0.498	0.469	0.500	0.426	0.451	0.498	0.477
0.388	0.339	0.403	0.496	0.461	0.500	0.421	0.440	0.497	0.464
0.465	0.403	0.348	0.493	0.456	0.500	0.416	0.437	0.495	0.457
0.498	0.496	0.493	0.362	0.385	0.500	0.376	0.391	0.470	0.303
0.469	0.461	0.456	0.385	0.230	0.380	0.256	0.288	0.300	0.324
0.500	0.500	0.500	0.500	0.380	0	0.312	0.356	0.308	0.500
0.426	0.421	0.416	0.376	0.256	0.312	0.137	0.185	0.272	0.330
0.451	0.440	0.437	0.391	0.288	0.356	0.185	0.205	0.314	0.351
0.498	0.497	0.495	0.470	0.300	0.308	0.272	0.314	0.283	0.438
0.477	0.464	0.457	0.303	0.324	0.500	0.330	0.351	0.438	0.228

Table IX

 MATRIX  $\mathbf{D}_{\approx PCR6}(\theta)$  WITH FAST BF-ICrA FOR EXAMPLE 3.

We observe that the average distance between ACO algorithms is much higher than in Tables V and VII of examples 1 and 2. This shows clearly the difficulty to precisely identify the clusters of similar algorithms because only few ACO algorithms perform actually very well for this third example. Eventually, and based on shortest inter-distance we could make the first pairwise group  $\{\text{ACO}_7, \text{ACO}_8\}$  because  $D_{78}(\theta) = 0.185$  is the minimal inter-distance we have between the ACO algorithms. Once the rows and columns of Table IX corresponding to  $\text{ACO}_7$  and  $\text{ACO}_8$  are eliminated, then the second best group will be  $\{\text{ACO}_5, \text{ACO}_9\}$  because  $D_{59}(\theta) = 0.300$ .

Similarly, we will get the group  $\{ACO_4, ACO_{10}\}$  because  $D_{4,10}(\theta) = 0.303$ , and then the group  $\{ACO_1, ACO_2\}$  because  $D_{12}(\theta) = 0.388$ . Finally we could also cluster  $ACO_3$  with  $ACO_6$  because  $D_{36}(\theta) = 0.500$ , although this distance of agreement is quite large to be considered as a trustable cluster.

**Example 3 with averaging fusion:** The matrix of distances to full agreement based on BF-ICrA method using average fusion rule, denoted by  $\mathbf{D}_{Aver.}(\theta)$ , is given in Table X.

0.170	0.154	0.142	0.221	0.351	0.350	0.392	0.345	0.332	0.298
0.154	0.120	0.092	0.167	0.321	0.295	0.369	0.313	0.290	0.261
0.142	0.092	0.042	0.114	0.289	0.237	0.342	0.279	0.242	0.224
0.221	0.167	0.114	0.054	0.255	0.139	0.327	0.260	0.184	0.177
0.351	0.321	0.289	0.255	0.339	0.245	0.391	0.355	0.287	0.324
0.350	0.295	0.237	0.139	0.245	0	0.304	0.242	0.115	0.247
0.392	0.369	0.342	0.327	0.391	0.304	0.390	0.368	0.336	0.387
0.345	0.313	0.279	0.260	0.355	0.242	0.368	0.328	0.288	0.341
0.332	0.290	0.242	0.184	0.287	0.115	0.336	0.288	0.190	0.279
0.298	0.261	0.224	0.177	0.324	0.247	0.387	0.341	0.279	0.261

Table X

MATRIX  $\mathbf{D}_{Aver.}(\theta)$  WITH BF-ICrA USING AVERAGING RULE FOR EXAMPLE 3.

Surprisingly, the use of averaging rule provides in this example lower distance values on average with respect to values given in Table IX. However no clear clustering of algorithms can be made because only few ACO algorithms perform actually very well for this third example. If we adopt the pairwise strategy to cluster algorithms, we will obtain now as first group  $\{ACO_2, ACO_3\}$  because  $D_{23}(\theta) = 0.092$ , as second group  $\{ACO_6, ACO_9\}$  because  $D_{69}(\theta) = 0.115$ , as third group  $\{ACO_4, ACO_{10}\}$  because  $D_{4,10}(\theta) = 0.177$ , as fourth group  $\{ACO_1, ACO_8\}$  because  $D_{18}(\theta) = 0.345$ , and finally we could also cluster  $ACO_5$  with  $ACO_7$  because  $D_{57}(\theta) = 0.391$ . one sees that there is no strong correlation between results obtained from BF-ICrA based on fast PCR6 and those based on averaging rule, which is not surprising because the rules are totally different. Nevertheless the group  $\{ACO_4, ACO_{10}\}$  is agreed by both methods here.

## VI. CONCLUSIONS

The fast Belief Function based Inter-Criteria Analysis method, using the canonical decomposition of basic belief assignments defined on a dichotomous frame of discernment was applied, tested and analysed in this paper. This new method was applied for evaluating the Multiple-Objective Ant Colony Optimization (MO-ACO) algorithm for Wireless Sensor Networks (WSN) deployment. Based on the BF-ICrA outcomes it was shown a very high correlation with fast PCR6 rule for the  $ACO_1, ACO_3$  and  $ACO_9$  group, for the  $ACO_4, ACO_5$  and  $ACO_6$  group, and for the  $ACO_7, ACO_8$  and  $ACO_{10}$  group of algorithms in example 1 (case of size  $350 \times 350$ ) as intuitively expected. This is because the considered ACO algorithms can solve the problem with good solution quality in example 1. These high correlations were not observed in the other two cases for example 2 (case of size  $500 \times 500$ ) and 3 (case of size  $700 \times 700$ ) because only few ACO algorithms perform actually very well for these examples. So, if we considered results in case of larger

problem sizes, the BF-ICrA results show that the number of ants has the significant influence on the obtained results, as already pointed out in [20].

## ACKNOWLEDGMENTS

This work is partially supported by the grant No BG05M20P001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014-2020) and co-financed by the European Union through the European structural and Investment funds and the Bulgarian scientific fund by the grant DN 12/5.

## REFERENCES

- [1] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Simplification of multi-criteria decision-making using inter-criteria analysis and belief functions*, in Proc. of Fusion 2019 Int. Conf., Canada, July 2–5, 2019.
- [2] K. Atanassov, D. Mavrov, V. Atanassova, *InterCriteria decision making: a new approach for multicriteria decision making, based on index matrices and intuitionistic fuzzy sets*, Issues IFSs GNs 11, pp. 1–8, 2014.
- [3] K. Atanassov, V. Atanassova, G. Gluhchev, *InterCriteria Analysis: Ideas and problems*, Notes on IFS, Vol. 21(1), pp. 81–88, 2015.
- [4] K. Atanassov et al., *An approach to a constructive simplification of multiagent multicriteria decision making problems via interCriteria analysis*, C.R. de l'Acad. Bulgare des Sci., Vol. 70(8), 2017.
- [5] J. Dezert, F. Smarandache, A. Tchamova, D. Han, *Fast Fusion of Basic Belief Assignments Defined on a Dichotomous Frame of Discernment*, In Proc. of Fusion 2020 (Online) conference, Pretoria, South Africa, July 2020.
- [6] J. Dezert, F. Smarandache, *Canonical Decomposition of Dichotomous Basic Belief Assignment*, Int. J. of Intelligent Systems, pp. 1–21, 2020.
- [7] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [8] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Proc. of Fusion 2012, Singapore, July 9–12, 2012.
- [9] A. Tchamova, J. Dezert, *On the Behavior of Dempster's Rule of Combination and the Foundations of Dempster-Shafer Theory*, IEEE IS-2012, Sofia, Bulgaria, Sept. 6–8, 2012.
- [10] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intelligent Syst., Vol. 29(3), pp. 223–252, March 2014.
- [11] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of Fusion 2013, Istanbul, Turkey, July 2013.
- [12] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, American Research Press, Vol. 1–4, 2004–2015, <http://www.onera.fr/staff/jean-dezert?page=2>
- [13] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proceedings of Fusion 2010 Int. Conf., Edinburgh, UK, July 2010.
- [14] <https://bfasociety.org/>
- [15] R. Yager, *On the Dempster-Shafer framework and new combination rules*, Information Sciences, Vol. 41, pp. 93–138, 1987.
- [16] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Comput. Intell., 4, 1988.
- [17] S. Fidanova, J. Dezert, A. Tchamova, *Inter-criteria analysis based on belief functions for GPS surveying problems*, in Proc. of IEEE Int. Symp. on INnovations in Intelligent SysTems and Applications (INISTA 2019), Sofia, Bulgaria, July 3–5, 2019.
- [18] J. Dezert, D. Han, H. Yin, *A New Belief Function Based Approach for Multi-Criteria Decision-Making Support*, in Proc. of Fusion 2016 Int. Conf, Heidelberg, July 2016
- [19] D. Han, J. Dezert, Y. Yang, *New Distance Measures of Evidence based on Belief Intervals*, in Proc. of Belief 2014 Int. Conf., Oxford, UK, Sept. 2014.
- [20] S. Fidanova, O. Roeva, *Multi-objective ACO Algorithm for WSN Layout: InterCriteria Analysis*, in Large-Scale Scientific Computing, Springer, 2020.
- [21] M. Dorigo, T. Stutzle, *Ant Colony Optimization*, MIT Press, Cambridge, 2004.

# Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments

Théo Dezert<sup>a</sup>, Jean Dezert<sup>b</sup>, Florentin Smarandache<sup>c</sup>

<sup>a</sup>Gustave Eiffel University, GERS-GeoEND, Bouguenais, France.

<sup>b</sup>The French Aerospace Lab, ONERA/DTIS, Palaiseau, France.

<sup>c</sup>Department of Mathematics, University of New Mexico, Gallup, NM, USA.

Emails: theo.dezert@univ-eiffel.fr, jean.dezert@onera.fr, smarand@unm.edu

Originally published as: T. Dezert, J. Dezert, F. Smarandache, *Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments*, Journal of Advances in Information Fusion (JAIF), Vol. 16, No. 1, June 2021, and reprinted with permission.

**Abstract**—This paper discusses and analyzes the behaviors of the proportional conflict redistribution rules no. 5 (PCR5) and no. 6 (PCR6) to combine several distinct sources of evidence characterized by their basic belief assignments defined over the same frame of discernment. After a brief review of these rules, the paper shows through simple examples why their behaviors can sometimes increase the uncertainty more than necessary which is detrimental to decision-making support drawn from the result of the combination. We present a theoretical improvement of these rules, and establish new PCR5<sup>+</sup> and PCR6<sup>+</sup> rules of combination. These new rules overcome the weakness of PCR5 and PCR6 rules by computing binary keeping-indexes that allow to keep only focal elements that play an effective role in the partial conflict redistribution. PCR5<sup>+</sup> and PCR6<sup>+</sup> rules are not associative but they preserve the neutrality of the vacuous belief assignment contrary to the PCR5 and PCR6 rules, and they make a more precise redistribution which does not increase improperly the mass of partial uncertainties.

**Keywords:** information fusion, belief functions, PCR5<sup>+</sup>, PCR6<sup>+</sup>, PCR5, PCR6 fusion rules..

## I. INTRODUCTION

There exist different theories based on distinct representations and modelings of uncertainty to deal with uncertain information to conduct information fusion [1]. The theory of probability [2], [3], the theory of fuzzy sets [4], [5], the possibility theory [6], [7], and the theory of belief functions [8]–[10] are the most well-known ones. This paper addresses the problem of information fusion in the mathematical framework of the belief functions introduced by Shafer from Dempster's works [11], [12]. The belief functions are often used in decision-making support applications because the experts are generally able to express only a belief in a hypothesis (or a set of hypotheses) from their partial knowledge, experience and from their own perception of the reality. To conduct information fusion, we need some efficient rules of combination that are able to manage the conflicting sources of evidence (if any), or expert opinions expressed in terms of belief functions. Readers interested in belief functions can found classical related papers in [13] and in the special issue [14] which includes also a list of good selected papers. It is worth to mention that the recent book of Cuzzolin [15] includes 2137 references, with many of them related to belief functions.

In this paper, we adopt the notion of conflict introduced by Shafer in [8] (p. 65). This notion of conflict is often adopted by researchers working with belief functions, as in [16] p. 17 for instance, because this notion is quite simple to understand. Different definitions and interpretations of conflict can be also found in [17]–[27] for readers interested in this topic. In this paper, two (or more) sources are said conflicting if they support incompatible (disjoint, or contradictory) hypotheses. We also work with distinct sources of evidences that are considered as (cognitively) independent and reliable. We do not consider, nor apply discounting techniques of belief assessments listed in [14] before combining them to keep the presentation and notations as simple as possible<sup>1</sup>.

While the conjunctive rule makes it possible to combine information between different sources of information by estimating the level of existing conflict, Dempster-Shafer (DS) rule [8], [16] proposes a distribution of this conflict on the hypotheses characterized by the sources of information. The normalization carried out by the DS rule may however be considered counter-intuitive especially when the level of conflict between the sources of information is high [28], [29], but also in some situations where the level of conflict between sources is low as shown in [30] showing a dictatorial behavior of DS rule. The Proportional Conflict Redistribution rules (PCR5 [31] and PCR6 [32], [33]) have been proposed to circumvent the problem of the DS rule to make a more judicious management of the conflict.

In this paper, we put forward a flawed behavior of these combination rules in some cases attributed to the non-neutrality of the vacuous BBA (Basic Belief Assignment), and we propose an improvement of these two combination rules (denoted by PCR5<sup>+</sup> and PCR6<sup>+</sup>) in order to ensure the neutrality property of the vacuous BBA. This is achieved by discarding specific elements implied in the partial conflict and which are not useful for making the conflict redistribution.

In the PCR rules [32]–[34] one redistributes the product of masses of belief of incompatible (i.e. conflicting) elements whose intersection is empty only to elements involved in

<sup>1</sup>Of course discounted belief assignments can also be combined by the rules presented in this paper.

this product and proportionally to their mass of belief. For instance, let's consider two elements  $A, B$  of a frame of discernment (FoD) with  $A \cap B = \emptyset$ , and three basic belief assignments  $m_1(\cdot), m_2(\cdot), m_3(\cdot)$  defined on this FoD with  $m_1(A) > 0, m_2(B) > 0$ , and  $m_3(A \cup B) > 0$ . The product  $m_1(A)m_2(B)m_3(A \cup B) > 0$  is called a *conflicting product hereafter because*  $A \cap B \cap (A \cup B) = \emptyset$ . Based on PCR5 (and PCR6) rule, we will redistribute the value of this product back to the focal<sup>2</sup> elements  $A, B$  and  $A \cup B$ , and proportionally to  $m_1(A), m_2(B)$  and  $m_3(A \cup B)$ . In the improved PCR rules developed in this paper we will redistribute this conflicting product only to the focal elements  $A$  and  $B$  since the focal element  $A \cup B$  is neither in conflict with  $A$ , nor with  $B$ . Such an improvement in the proportional conflict redistribution is made possible by defining a binary keeping-index for each focal element involved in the conflicting product. This index will allow the identification of elements of the conflicting product that will have an effective role in the proportional redistribution of conflicting product. All elements (if any) having a binary keeping-index equal to zero are discarded of the conflict redistribution process. This main idea is developed in this paper and illustrated with several examples. It allows to preserve the neutrality of the total ignorant source of evidence in the improved versions of PCR5 and PCR6 rules, which is often considered as a desirable property for a rule of combination of distinct and reliable sources of evidence.

For the reader not immersed in the belief mathematics notion, the comparative numerical examples of Example 1 of section III-B as compared with Example 1 revisited of section VII, provide a quick verification of the improvements.

This paper is organized as follows. We give the basics of belief functions in Section II. We present the PCR5 and PCR6 rules of combination in Section III with new general formulas in subsection III-C, and associated examples in Section IV. The flawed behavior of PCR5 and PCR6 rules are highlighted in Section V through specific examples. Then, Section VI proposes the mathematical expression of the new improved PCR5<sup>+</sup> and PCR6<sup>+</sup> rules of combination, as well as the very detailed procedure to select the focal elements for these new proportional redistributions. Finally, comparative results for relevant examples are shown in Section VII in order to compare the PCR5 and PCR6 results with the PCR5<sup>+</sup> and PCR6<sup>+</sup> results. Concluding remarks are given in section VIII. For convenience, two Matlab<sup>TM</sup> routines are also given in the appendix 3 of this paper for PCR5<sup>+</sup> and PCR6<sup>+</sup> rules of combination.

## II. BASICS OF BELIEF FUNCTIONS

We consider a given finite set  $\Theta$  of  $n > 1$  distinct elements  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  corresponding to the frame of discernment (FoD) of the fusion problem, or the decision-making problem, under concern. All elements of  $\Theta$  are mutually

<sup>2</sup>A focal element is an element (i.e. a subset) having a strictly positive mass of belief committed to it - see section II elements

exclusive<sup>3</sup> and each element is an elementary choice of the potential decision to take. The power set of  $\Theta$  is the set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) and it is usually denoted  $2^\Theta$  because its cardinality equals  $2^{|\Theta|}$ . We adopt Shafer's formalism whereby propositions are represented by subsets [8] (Chap.2, pp. 35–37). Hence, the propositions under concern are in one-to-one correspondance with subsets of  $\Theta$ . We also use classical notations of set theory [35], i.e.  $\emptyset$  for the empty set,  $A \cup B$  for the union<sup>4</sup> of sets  $A$  and  $B$  (which is the set of all objects that are a member of the set  $A$ , or the set  $B$ , or both),  $A \cap B$  for their intersection (which is the set of all objects that are members of both  $A$  and  $B$ ), etc. A Basic Belief Assignment (BBA) given by a source of evidence is defined by Shafer [8] in his Mathematical Theory of Evidence (known also as Dempster-Shafer Theory, or DST) as  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying

$$\begin{cases} m(\emptyset) = 0, \\ \sum_{A \in 2^\Theta} m(A) = 1, \end{cases} \quad (1)$$

where  $m(A)$  is the mass of belief exactly committed to  $A$ , what we usually call the mass of  $A$ . A BBA is said proper (or normal) if it satisfies Shafer's definition (1). The subset  $A \subseteq \Theta$  is called a Focal Element (FE) of the BBA  $m(\cdot)$  if and only if  $m(A) > 0$ . The empty set is not a focal element of a BBA because  $m(\emptyset) = 0$  according to definition (1). The set of all focal elements of a BBA  $m(\cdot)$  is denoted  $\mathcal{F}(m)$ . Its mathematical definition is  $\mathcal{F}(m) = \{X \in 2^\Theta | m(X) > 0\}$ . The cardinality  $|\mathcal{F}(m)|$  of the set  $\mathcal{F}(m)$  is denoted  $\mathcal{F}_m$ . The order of focal elements of  $\mathcal{F}(m)$  does not matter and all the focal elements are different. The set  $\mathcal{F}(m)$  of focal elements of  $m(\cdot)$  has at least one focal element, and at most  $2^{|\Theta|} - 1$  focal elements.

Belief and plausibility functions are respectively defined from  $m(\cdot)$  by [8]

$$Bel(A) = \sum_{X \in 2^\Theta | X \subseteq A} m(X), \quad (2)$$

and

$$Pl(A) = \sum_{X \in 2^\Theta | A \cap X \neq \emptyset} m(X) = 1 - Bel(\bar{A}). \quad (3)$$

where  $\bar{A}$  represents the complement of  $A$  in  $\Theta$ .

$Bel(A)$  and  $Pl(A)$  are usually interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$  [11], [12]. The functions  $m(\cdot), Bel(\cdot)$  and  $Pl(\cdot)$  are one-to-one. A belief function  $Bel(\cdot)$  is *Bayesian* if all Bel's focal elements are singletons [8] (Theorem 2.8 p. 45). In this case,  $m(X) = Bel(X)$  for any (singleton) focal element  $X$ , and  $m(\cdot)$  is called a *Bayesian BBA*. Corresponding  $Bel(\cdot)$  function is equal to  $Pl(\cdot)$  and these functions can be

<sup>3</sup>This standard assumption is called *Shafer's model of FoD* in DSmt (Dezert-Smarandache Theory) framework [34].

<sup>4</sup>We prefer the notation  $A \cup B$  for denoting the union of sets  $A$  and  $B$ , which is a formal mathematical notation for the union of two sets, instead of the notations  $AB$  or  $\{A, B\}$  used by some authors.

interpreted as a same (possibly subjective) probability measure  $P(\cdot)$ . The vacuous BBA (VBBA for short) representing a totally ignorant source is defined as  $m_v(\Theta) = 1$ .

### III. COMBINATION OF BBAS

This Section presents at first the conjunctive rule of combination which is one of the main rules to combine reliable sources of evidence and which allows to identify the conflicting information among the sources. Then we present the proportional conflict redistribution rules no. 5 (PCR5) [31] and no. 6 (PCR6) [32], [33] as alternatives of Dempster's rule of combination [8]. The development of these rules has been motivated by the counter-intuitive behavior of Dempster's rule [8] when combining high conflicting sources of evidences, but also when combining low conflicting sources of evidences as well<sup>5</sup>. The reader interested in this topic can refer to [13], [28]–[30] to see theoretical justifications and examples. In the following, and for simplicity, we restrain our presentation to the classical framework of belief functions, and we work with BBAs defined only on the power set  $2^\Theta$  of a FoD  $\Theta$ . PCR rules have been defined originally for working with Dedekind's lattice as well, see Chapter 1 of [34] (Volume 2). In this paper, we present simple general expressions of PCR5 and PCR6 fusion rules because they are more easy to understand than the original general formulas, and they afford expressions of the improved PCR5<sup>+</sup> and PCR6<sup>+</sup> rules in a direct and useful manner.

After a brief presentation of the main notations used in this paper, we will recall both PCR5 and PCR6 rules for historical and technical reasons. PCR5 has been developed at first, and then PCR6 has been proposed based on a modified redistribution principle inspired by PCR5. In this paper, we follow the logical and historical development of these PCR5 and PCR6 rules to make the presentation of their improved versions PCR5<sup>+</sup> and PCR6<sup>+</sup>. It seems easier to understand PCR6<sup>+</sup> fusion formula once the PCR5<sup>+</sup> formula will have been established. By presenting both rules, we offer to the readers a global deeper view on how these new rules work and their fundamental and mathematical differences in their conflict redistribution principles. In the sequel, all the introduced examples assume the model of Shafer's frame of discernment as in the classical DST framework.

#### A. Notations

When we make the combination of  $S \geq 2$  BBAs by the conjunctive rule, or by the PCR5 and PCR6 fusion rules, we have to compute the product of the masses of the focal elements composing any possible  $S$ -tuple of focal elements. Each possible  $S$ -tuple is noted by<sup>6</sup>

$$\mathbf{X}_j \triangleq (X_{j_1}, X_{j_2}, \dots, X_{j_S}) \in \mathcal{F}(m_1) \times \mathcal{F}(m_2) \times \dots \times \mathcal{F}(m_S),$$

where  $j_1 \in \{1, 2, \dots, \mathcal{F}_{m_1}\}$ ,  $j_2 \in \{1, 2, \dots, \mathcal{F}_{m_2}\}$ ,  $\dots$ ,  $j_S \in \{1, 2, \dots, \mathcal{F}_{m_S}\}$ . The element  $X_{j_i}$  is the focal element

of  $m_i(\cdot)$  that makes the  $i$ -th component of the  $j$ -th  $S$ -tuple  $\mathbf{X}_j$ .

For notation convenience also, the cartesian product  $\mathcal{F}(m_1) \times \mathcal{F}(m_2) \times \dots \times \mathcal{F}(m_S)$  is denoted by  $\mathcal{F}(m_1, \dots, m_S)$  in the sequel.

We have  $\mathcal{F} \triangleq |\mathcal{F}(m_1, \dots, m_S)| = \prod_{i=1}^S |\mathcal{F}(m_i)| = \prod_{i=1}^S \mathcal{F}_{m_i}$  products of masses of focal elements to consider and to calculate because we have  $\mathcal{F}_{m_1}$  focal elements in  $\mathcal{F}(m_1)$ ,  $\mathcal{F}_{m_2}$  focal elements in  $\mathcal{F}(m_2)$ ,  $\dots$ , and  $\mathcal{F}_{m_S}$  focal elements in  $\mathcal{F}(m_S)$ . Each product for  $j = 1$  to  $\mathcal{F}$  is of the form

$$\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) \triangleq \prod_{i=1}^S m_i(X_{j_i}). \quad (4)$$

There are two types of products:

- $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S})$  is called a *non-conflicting (mass) product* if

$$X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S} = X \neq \emptyset.$$

In this case,  $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S})$  is also noted by  $\pi_j(X)$  for short.

- $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S})$  is called a *conflicting (mass) product* if

$$X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S} = \emptyset.$$

In this case,  $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S})$  is also noted by  $\pi_j(\emptyset)$  for short.

It is worth noting that an element  $X \in 2^\Theta \setminus \{\emptyset\}$  may belong to sets of focal elements of the different BBAs to combine, and therefore a  $S$ -tuple  $\mathbf{X}_j$  can have duplicate components. Because all the BBAs are normalized, we always have

$$\sum_{j=1}^{\mathcal{F}} \pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) = 1. \quad (5)$$

As a simple example to illustrate our notations, let's consider two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined over the FoD  $\Theta = \{A, B, C\}$  with respectively two and three focal elements, say  $\mathcal{F}(m_1) = \{A, B \cup C\}$  and  $\mathcal{F}(m_2) = \{B, C, A \cup C\}$ . Here  $\mathcal{F}_{m_1} = |\mathcal{F}(m_1)| = 2$  and  $\mathcal{F}_{m_2} = |\mathcal{F}(m_2)| = 3$ . For  $j_1 = 1$  (the first focal element of  $m_1(\cdot)$ ) one has  $X_{j_1} = A$ , and for  $j_1 = 2$  (the second focal element of  $m_1(\cdot)$ ) one has  $X_{j_1} = B \cup C$ . Similarly, for  $j_2 = 1$  (the first focal element of  $m_2(\cdot)$ ) one has  $X_{j_2} = B$ , for  $j_2 = 2$  (the 2nd focal element of  $m_2(\cdot)$ ) one has  $X_{j_2} = C$ , and  $j_2 = 3$  (the 3rd focal element of  $m_2(\cdot)$ ) one has  $X_{j_2} = A \cup C$ . In this case we have  $\mathcal{F} = \mathcal{F}_{m_1} \cdot \mathcal{F}_{m_2} = 6$  products of masses to consider in the conjunctive fusion rule (see next sub-section) which are

$$\pi_1(A \cap B) = m_1(A)m_2(B),$$

$$\pi_2(A \cap C) = m_1(A)m_2(C),$$

$$\pi_3(A \cap (A \cup C)) = m_1(A)m_2(A \cup C),$$

$$\pi_4((B \cup C) \cap B) = m_1(B \cup C)m_2(B),$$

$$\pi_5((B \cup C) \cap C) = m_1(B \cup C)m_2(C),$$

$$\pi_6((B \cup C) \cap (A \cup C)) = m_1(B \cup C)m_2(A \cup C).$$

<sup>5</sup>Which is known as the dictatorial behavior of Dempster's rule [30].

<sup>6</sup>The symbol  $\triangleq$  means "equals by definition".

The products  $\pi_1$  and  $\pi_2$  are called *conflicting products* because

- for  $\pi_1$ , the focal elements  $A$  and  $B$  involved in  $\pi_1$  are incompatible (i.e. disjoint) because  $A \cap B = \emptyset$ .  $\pi_1(A \cap B)$  is of course equivalent to  $\pi_j(X_{j_1} \cap X_{j_2})$  with  $j = 1$  by taking  $X_{j_1} = A$  and  $X_{j_2} = B$ ;
- for  $\pi_2$ , one has  $A \cap C = \emptyset$ .  $\pi_2(A \cap C)$  is equivalent to  $\pi_j(X_{j_1} \cap X_{j_2})$  with  $j = 2$  by taking  $X_{j_1} = A$  and  $X_{j_2} = C$ , etc.

The products  $\pi_3, \dots$ , and  $\pi_6$  are not conflicting products because the focal elements involved in each product have non-empty intersection. Because  $m_1(A) + m_1(B \cup C) = 1$  and  $m_2(B) + m_2(C) + m_2(A \cup C) = 1$ , one has  $(m_1(A) + m_1(B \cup C))(m_2(B) + m_2(C) + m_2(A \cup C)) = 1$ , and therefore  $\sum_{j=1}^6 \pi_j = 1$ . This illustrates the formula (5).

In this paper,  $i \in \{1, \dots, S\}$  represents the index of the  $i$ -th source of evidence characterized by the BBA  $m_i(\cdot)$ , and  $j \in \{1, \dots, \mathcal{F}\}$  represents the index of the  $j$ -th product  $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S})$ .

### B. The conjunctive rule of combination

Let's consider  $S \geq 2$  distinct reliable sources of evidence characterized by their BBA  $m_s(\cdot)$  ( $s = 1, \dots, S$ ) defined on  $2^\Theta$ . Their conjunctive fusion<sup>7</sup> is defined for all  $A \in 2^\Theta$  by

$$\begin{aligned} m_{1,2,\dots,S}^{\text{Conj}}(A) &= \sum_{\substack{X_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = A}} \pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) \\ &= \sum_{\substack{X_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = A}} \prod_{i=1}^S m_i(X_{j_i}). \end{aligned} \quad (6)$$

The symbol  $\odot$  is also used in the literature, for instance in [36], to note the conjunctive fusion operator, i.e.  $m_{1,2,\dots,S}^{\text{Conj}}(A) = [m_1 \odot m_2 \odot \dots \odot m_S](A)$ .

The total conflicting mass between the  $S$  sources of evidence, denoted  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)$ , is nothing but the sum of all existing conflicting mass products, that is

$$\begin{aligned} m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) &= \sum_{\substack{X_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = \emptyset}} \pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) \\ &= 1 - \sum_{A \in 2^\Theta \setminus \{\emptyset\}} m_{1,2,\dots,S}^{\text{Conj}}(A). \end{aligned} \quad (7)$$

Note that the combined BBA  $m_{1,2,\dots,S}^{\text{Conj}}(\cdot)$  given in (6) is not a proper BBA because it does not satisfy Shafer's definition (1). In general the  $S$  sources of evidence to combine do not fully agree, and we have consequently  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) > 0$ .

Dempster's rule of combination (called also *orthogonal sum* by Shafer [8], p. 6) coincides with the normalized version of the conjunctive rule. It is defined by  $m_{1,2,\dots,S}^{\text{DS}}(A) =$

<sup>7</sup>The conjunctive fusion rule is also called Smets' rule of combination by some authors because it has been widely used by Philippe Smets in his works related to belief functions. But Smets himself call it conjunctive rule, see his last paper [20], p. 388.

$m_{1,2,\dots,S}^{\text{Conj}}(A)/(1 - m_{1,2,\dots,S}^{\text{Conj}}(\emptyset))$ , assuming  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) \neq 1$ . The DS upper notation refers to initials of Dempster and Shafer names because Dempster's rule has gained its popularity through Shafer's works on belief functions. Shafer uses the symbol  $\oplus$  to note Dempster's fusion operator, i.e.  $m_{1,2,\dots,S}^{\text{DS}}(A) = [m_1 \oplus m_2 \oplus \dots \oplus m_S](A)$  for  $A \neq \emptyset$ , and  $m_{1,2,\dots,S}^{\text{DS}}(\emptyset) = 0$ . A probabilistic analysis of Dempster's rule of combination can be found in [37], and the geometry of Dempster's rule is analyzed in [38].

**Example 1:** Consider  $\Theta = \{A, B\}$  and two following BBAs

$$\begin{aligned} m_1(A) &= 0.1, & m_1(B) &= 0.2, & m_1(A \cup B) &= 0.7, \\ m_2(A) &= 0.4, & m_2(B) &= 0.3, & m_2(A \cup B) &= 0.3. \end{aligned}$$

We have  $m_{1,2}^{\text{Conj}}(\emptyset) = 0.11$ , and

$$m_{1,2}^{\text{Conj}}(A) = 0.35, \quad m_{1,2}^{\text{Conj}}(B) = 0.33, \quad m_{1,2}^{\text{Conj}}(\Theta) = 0.21.$$

Symbolically we denote the conjunctive fusion of  $S$  sources as  $m_{1,2,\dots,S}^{\text{Conj}} = \text{Conj}(m_1, m_2, \dots, m_S)$ . This conjunctive rule is commutative and associative. This means that the sources can be combined altogether in one step, or sequentially in any order and it does not matter. Also, the total ignorant source represented by the vacuous (non-informative) BBA has no impact in the fusion result - see Lemma 1 below.

**Lemma 1:** The vacuous BBA  $m_v$  has a neutral impact in the conjunctive rule of combination, that is

$$\text{Conj}(m_1, m_2, \dots, m_S, m_v) = \text{Conj}(m_1, m_2, \dots, m_S). \quad (8)$$

**Proof:** see appendix 1.

The main drawback of this fusion rule is that it does not generate a proper BBA because  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) > 0$  in general, and also it can provide a fusion result  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)$  that quickly tends to one after only few steps of a sequential fusion processing of the sources which is not very useful for decision-making support. This is because the empty set  $\emptyset$  is the absorbing element for the conjunctive operation since  $\emptyset \cap A = \emptyset$  for all  $A \in 2^\Theta$  so that the mass committed to the empty set always increases through the repeated conjunctive fusion rule. The main interest of this rule is its ability to identify the partial conflicts and to provide a measure of the total level of conflict  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)$  between the sources which can be used to manage (select or discard) the sources in the fusion process if one prefers, see [39] for an application in geophysics for instance.

### C. PCR5 and PCR6 rules of combination

The Proportional Conflict Redistribution Rules (PCR) have been developed originally in the framework of DSMT (Dezert-Smarandache Theory) [31], [32], [34] but they can work also in the classical framework of Shafer's belief functions as well. Six rules have been proposed and they are referred as PCR1, ..., PCR6 rules of combination having different complexities, PCR1 being the most simplest (but less effective) one. All these rules share the same general principle which consists of three steps:

- apply the conjunctive rule (6);
- calculate the conflicting mass products  $\pi_j(\emptyset)$ ;
- redistribute the conflicting mass products  $\pi_j(\emptyset)$  proportionally on all non-empty sets involved in the conflict.

The way the conflicting mass product  $\pi_j(\emptyset)$  is redistributed yields to different versions of PCR combination rules that work for any degree of conflict. The sophistication/complexity and preciseness of PCR rules increases from the first PCR1 rule up to the last rule PCR6. The main disadvantage of these rules, aside their complexity, is their non-associativity properties which impose to combine all the BBAs altogether with PCR rules rather than sequentially to expect the best fusion result.

In this paper, we focus on the presentation of PCR5 and PCR6 only because they are the most well-known advanced fusion rules used so far in the belief functions community. A detailed presentation of other rules of combination encountered in the literature can be found in [40]. Symbolically, the PCR5 fusion and the PCR6 fusion of  $S \geq 2$  BBAs are respectively denoted  $m_{1,2,\dots,S}^{\text{PCR5}} = \text{PCR5}(m_1, m_2, \dots, m_S)$ , and  $m_{1,2,\dots,S}^{\text{PCR6}} = \text{PCR6}(m_1, m_2, \dots, m_S)$ .

Readers familiar with PCR rules could quickly read the example 1 given in section III-B, and the results obtained with classical and improved PCR5 and PCR6 rules in section VII to appreciate the discussion throughout the paper.

**The PCR5 rule of combination** [31]: This rule transfers the conflicting mass  $\pi_j(\emptyset)$  to all the elements involved in this conflict and proportionally to their individual masses, so that a more sophisticate and specific distribution is done with the PCR5 fusion process with respect to other existing rules (including Dempster's rule). The PCR5 rule is presented in details (with justification and examples) in [34], Vol. 2 and Vol. 3.

- The PCR5 fusion of two BBAs is obtained by  $m_{1,2}^{\text{PCR5}}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{1,2}^{\text{PCR5}}(A) = m_{1,2}^{\text{Conj}}(A) + \sum_{\substack{X \in 2^\Theta \\ X \cap A = \emptyset}} \left[ \frac{m_1(A)^2 m_2(X)}{m_1(A) + m_2(X)} + \frac{m_2(A)^2 m_1(X)}{m_2(A) + m_1(X)} \right], \quad (9)$$

where  $m_{1,2}^{\text{Conj}}(A)$  is the conjunctive rule formula (6) with  $S = 2$ , and where all denominators in (9) are different from zero. If a denominator is zero, that fraction is discarded. All propositions/sets are in a canonical form. We take the disjunctive normal form, which is a disjunction of conjunctions, and it is unique in Boolean algebra and simplest. For example,  $X = A \cap B \cap (A \cup B \cup C)$  it is not in a canonical form, but we simplify the formula and  $X = A \cap B$  is in a canonical form.

The PCR5 formula (9) for two BBAs can also be expressed by considering only the focal elements of  $m_1(\cdot)$  and  $m_2(\cdot)$  as follows

$$\begin{aligned} m_{1,2}^{\text{PCR5}}(A) &= m_{1,2}^{\text{Conj}}(A) \\ &+ \sum_{\substack{(X_{j_1}, X_{j_2}) \in \mathcal{F}(m_1) \times \mathcal{F}(m_2) \\ X_{j_1} \cap X_{j_2} = \emptyset \\ X_{j_1} = A}} m_1(X_{j_1}) \cdot \frac{m_1(X_{j_1}) m_2(X_{j_2})}{m_1(X_{j_1}) + m_2(X_{j_2})} \\ &+ \sum_{\substack{(X_{j_1}, X_{j_2}) \in \mathcal{F}(m_1) \times \mathcal{F}(m_2) \\ X_{j_1} \cap X_{j_2} = \emptyset \\ X_{j_2} = A}} m_2(X_{j_2}) \cdot \frac{m_1(X_{j_1}) m_2(X_{j_2})}{m_1(X_{j_1}) + m_2(X_{j_2})}, \end{aligned} \quad (10)$$

or equivalently, with shorthand  $\pi_j$  notations, as

$$\begin{aligned} m_{1,2}^{\text{PCR5}}(A) &= m_{1,2}^{\text{Conj}}(A) \\ &+ \sum_{\substack{j \in \{1, \dots, \mathcal{F}\} | \mathbf{X}_j \in \mathcal{F}(m_1, m_2) \\ X_{j_1} \cap X_{j_2} = \emptyset \\ A \in \mathbf{X}_j}} \left[ m_{i \in \{1,2\} | X_{j_i} = A} (X_{j_i}) \right. \\ &\quad \left. \cdot \frac{\pi_j(X_{j_1} \cap X_{j_2})}{m_1(X_{j_1}) + m_2(X_{j_2})} \right], \end{aligned} \quad (11)$$

where  $\mathcal{F} = |\mathcal{F}(m_1)| \cdot |\mathcal{F}(m_2)|$  is the total number of products  $\pi_j(X_{j_1} \cap X_{j_2}) = m_1(X_{j_1}) m_2(X_{j_2})$ , and  $A \in \mathbf{X}_j$  means that at least one component of  $\mathbf{X}_j$  equals  $A$ .

- The explicit formula of the PCR5 fusion of three BBAs is given in [41].
- A simple formulation of the general expression of the PCR5 fusion of  $S > 2$  basic belief assignments is obtained by redistributing each conflicting product  $\pi_j(\emptyset) = \pi_j(X_{j_1} \cap \dots \cap X_{j_S} = \emptyset) = \prod_{i=1}^S m_i(X_{j_i})$  to some elements of the power set of the FoD that are involved in the conflict. Each  $\pi_j(\emptyset)$  is redistributed proportionally to elements involved in this conflict based on the PCR5 redistribution principle. When an element  $A \in 2^\Theta$  is not involved in a conflicting product  $\pi_j(\emptyset)$ , i.e.  $A \notin \mathbf{X}_j$ , the conflicting product  $\pi_j(\emptyset)$  is not redistributed to  $A$ . If an element  $A$  is involved in the conflict  $X_{j_1} \cap \dots \cap X_{j_S} = \emptyset$ , i.e.  $A \in \mathbf{X}_j$  and  $\pi_j(\emptyset)$  occur, then the proportional redistribution of  $\pi_j(\emptyset)$  to  $A$  is given by

$$\begin{aligned} x_j(A) &\triangleq \left( \prod_{i \in \{1, \dots, S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \\ &\cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \prod_{i \in \{1, \dots, S\} | X_{j_i} = X} m_i(X_{j_i}) \right)}, \end{aligned} \quad (12)$$

where  $A \in \mathbf{X}_j$  means that at least one component of the  $S$ -tuple  $\mathbf{X}_j = (X_{j_1}, \dots, X_{j_S}) \in \mathcal{F}(m_1, \dots, m_S)$  equals  $A$ .

Finally the mass value of  $A$  obtained by the PCR5 rule is calculated by

$$m_{1,2,\dots,S}^{\text{PCR5}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1, \dots, \mathcal{F}\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} x_j(A), \quad (13)$$

where  $A \in \mathbf{X}_j \wedge \pi_j(\emptyset)$  is a shorthand notation meaning that at least one component of the  $S$ -tuple  $\mathbf{X}_j$  equals  $A$  and the components of  $\mathbf{X}_j$  are conflicting, i.e.  $X_{j_1} \cap \dots \cap X_{j_S} = \emptyset$ .

Therefore the general PCR5 formula can be expressed as  $m_{1,2,\dots,S}^{\text{PCR5}}(\emptyset) = 0$ , and for  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{1,2,\dots,S}^{\text{PCR5}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1,\dots,\mathcal{F}\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \prod_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \prod_{i \in \{1,\dots,S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right]. \quad (14)$$

It is worth noting that the formula (14) is a generalization of the formula (11), i.e. (14) coincides with (11) when  $S = 2$ .

This general PCR5 formula is equivalent to the original PCR5 formula given in [31] but it involves only the focal elements of the BBAs to combine which makes the derivation more efficient (less computationally demanding) than the original general PCR5 formula, specially when each BBA has only few focal elements. We use this new general PCR5 formula because it is relatively simple and easy to improve it into PCR5<sup>+</sup> formula - see section VI-B. The extension of PCR5 for combining qualitative<sup>8</sup> BBAs can be found in [34], Vol. 2 & 3, and in [33]. PCR5 rule is not associative and the best fusion result is obtained by combining the sources altogether at the same time when possible. A suboptimal fast fusion method using PCR5-based canonical decomposition [42] can be found in [43].

**The PCR6 rule of combination** [32]: A variant of PCR5 rule, called PCR6 rule, has been proposed by Martin and Osswald in [32], [33] for combining  $S > 2$  sources. Because PCR6 coincides with PCR5 when one combines two sources, we do not provide the PCR6 formula for two sources which is the same as (9). The difference between PCR5 and PCR6 lies in the way the proportional conflict redistribution is done as soon as three (or more) sources are involved in the fusion as it will be shown in the example 2 introduced in the next section. The explicit formula of the PCR6 fusion of three BBAs is given in [41] for convenience.

The PCR6 fusion of  $S > 2$  BBAs is obtained by  $m_{1,2,\dots,S}^{\text{PCR6}}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by<sup>9</sup>

$$m_{1,2,\dots,S}^{\text{PCR6}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1,\dots,\mathcal{F}\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right]. \quad (15)$$

The difference between the general PCR5 and PCR6 formulas is that the PCR5 proportional redistribution involves the products  $\prod_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i})$  of multiple same

<sup>8</sup>A qualitative BBA is a BBA whose values are labels (e.g. low, medium, high, etc) instead of real numbers.

<sup>9</sup>We wrote this PCR6 general formula in the style of PCR5 formula (14).

focal elements  $A$  (if any) in the conflict, whereas the PCR6 conflict redistribution principle works with their sum  $\sum_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i})$  instead. The next section presents some examples for PCR5 and PCR6 rules of combinations.

We use this general PCR6 formula instead of the original Martin-Osswald's PCR6 formula [32] because it is more easy to improve it into PCR6<sup>+</sup> formula - see section VI-B. From the implementation point of view, PCR6 is simpler to implement than PCR5. From the Decision-Making (DM) standpoint, PCR6 is better than PCR5 when  $S > 2$  as reported by Martin and Osswald in [32] (see also the Example 3 in the next section) in their applications. For convenience, some Matlab<sup>TM</sup> codes of PCR5 and PCR6 fusion rules can be found in the appendix of [44], also in Chap. 7 of [34] (Vol. 3), or from Arnaud Martin's web page [45]. PCR6 code (in R programming language) can be found also in iBelief package developed by Kuang Zhou and Arnaud Martin from the BFAS<sup>10</sup> repository [46], or directly from [47] as well. When we have only two BBAs to combine, PCR5 and PCR6 rules provide the same result because formulas (14) and (15) coincide for  $S = 2$ .

In this paper, we have voluntarily chosen to present the two rules PCR5, PCR6 and their improved versions mainly for historical reasons and because these two rules have strong theoretical links as we have shown. By doing this, we offer the possibility to readers (and potential users) to test each of these advanced fusion methods and evaluate their performances on their own applications. Even though PCR6 is posterior to PCR5, since some researchers have implemented and are using PCR5 fusion rule, it appears important to introduce the improved version of this rule. Furthermore, PCR5 goes back exactly on the tracks of the conjunctive rule, while PCR6 does not.

#### IV. EXAMPLES FOR PCR5 AND PCR6 FUSION RULES

Here we provide two simple examples showing the difference of the results between PCR5 and PCR6 rules. For convenience, all numerical values given in the examples of this paper have been rounded to six decimal places when necessary.

**Example 2:** We consider the simplest FoD  $\Theta = \{A, B\}$ , and the three following BBAs

$$\begin{aligned} m_1(A) &= 0.6, m_1(B) = 0.1, m_1(A \cup B) = 0.3, \\ m_2(A) &= 0.5, m_2(B) = 0.3, m_2(A \cup B) = 0.2, \\ m_3(A) &= 0.4, m_3(B) = 0.1, m_3(A \cup B) = 0.5. \end{aligned}$$

Because  $\mathcal{F}_{m_1} = |\mathcal{F}(m_1)| = 3$ ,  $\mathcal{F}_{m_2} = |\mathcal{F}(m_2)| = 3$  and  $\mathcal{F}_{m_3} = |\mathcal{F}(m_3)| = 3$ , we have  $\mathcal{F} = \mathcal{F}_{m_1} \cdot \mathcal{F}_{m_2} \cdot \mathcal{F}_{m_3} = 27$  products to consider. Fifteen products are non-conflicting and will enter in the calculation of  $m_{1,2,3}^{\text{Conj}}(A)$ ,  $m_{1,2,3}^{\text{Conj}}(B)$  and  $m_{1,2,3}^{\text{Conj}}(A \cup B)$ , and twelve products are conflicting products

<sup>10</sup>Belief Functions and Applications Society.

that will need to be proportionally redistributed. The conjunctive combination of these three BBAs is

$$\begin{aligned}
 m_{1,2,3}^{\text{Conj}}(A) &= m_1(A)m_2(A)m_3(A) \\
 &+ m_1(A)m_2(A)m_3(A \cup B) \\
 &+ m_1(A)m_2(A \cup B)m_3(A) \\
 &+ m_1(A \cup B)m_2(A)m_3(A) \\
 &+ m_1(A)m_2(A \cup B)m_3(A \cup B) \\
 &+ m_1(A \cup B)m_2(A)m_3(A \cup B) \\
 &+ m_1(A \cup B)m_2(A \cup B)m_3(A) \\
 &= 0.5370,
 \end{aligned}$$

$$\begin{aligned}
 m_{1,2,3}^{\text{Conj}}(B) &= m_1(B)m_2(B)m_3(B) \\
 &+ m_1(B)m_2(B)m_3(A \cup B) \\
 &+ m_1(B)m_2(A \cup B)m_3(B) \\
 &+ m_1(A \cup B)m_2(B)m_3(B) \\
 &+ m_1(B)m_2(A \cup B)m_3(A \cup B) \\
 &+ m_1(A \cup B)m_2(B)m_3(A \cup B) \\
 &+ m_1(A \cup B)m_2(A \cup B)m_3(B) \\
 &= 0.0900,
 \end{aligned}$$

$$\begin{aligned}
 m_{1,2,3}^{\text{Conj}}(A \cup B) &= m_1(A \cup B)m_2(A \cup B)m_3(A \cup B) \\
 &= 0.3 \cdot 0.2 \cdot 0.5 = 0.0300,
 \end{aligned}$$

and

$$\begin{aligned}
 m_{1,2,3}^{\text{Conj}}(\emptyset) &= 1 - m_{1,2,3}^{\text{Conj}}(A) - m_{1,2,3}^{\text{Conj}}(B) - m_{1,2,3}^{\text{Conj}}(A \cup B) \\
 &= 0.3430.
 \end{aligned}$$

In this example we have twelve partial conflicts, noted  $\pi_j(\emptyset)$  ( $j = 1, \dots, 12$ ), which are given by the following products

$$\begin{aligned}
 \pi_1(\emptyset) &= m_1(A)m_2(A)m_3(B) = 0.0300, \\
 \pi_2(\emptyset) &= m_1(A)m_2(B)m_3(A) = 0.0720, \\
 \pi_3(\emptyset) &= m_1(B)m_2(A)m_3(A) = 0.0200, \\
 \pi_4(\emptyset) &= m_1(B)m_2(B)m_3(A) = 0.0120, \\
 \pi_5(\emptyset) &= m_1(B)m_2(A)m_3(B) = 0.0050, \\
 \pi_6(\emptyset) &= m_1(A)m_2(B)m_3(B) = 0.0180, \\
 \pi_7(\emptyset) &= m_1(A \cup B)m_2(A)m_3(B) = 0.0150, \\
 \pi_8(\emptyset) &= m_1(A \cup B)m_2(B)m_3(A) = 0.0360, \\
 \pi_9(\emptyset) &= m_1(B)m_2(A)m_3(A \cup B) = 0.0250, \\
 \pi_{10}(\emptyset) &= m_1(A)m_2(B)m_3(A \cup B) = 0.0900, \\
 \pi_{11}(\emptyset) &= m_1(A)m_2(A \cup B)m_3(B) = 0.0120, \\
 \pi_{12}(\emptyset) &= m_1(B)m_2(A \cup B)m_3(A) = 0.0080.
 \end{aligned}$$

In applying the PCR5 formula (14), and the PCR6 formula (15) we obtain finally  $m_{1,2,3}^{\text{PCR5}}(\emptyset) = m_{1,2,3}^{\text{PCR6}}(\emptyset) = 0$ , and<sup>11</sup>

$$\begin{aligned}
 m_{1,2,3}^{\text{PCR5}}(A) &\approx 0.723281, \\
 m_{1,2,3}^{\text{PCR5}}(B) &\approx 0.182460, \\
 m_{1,2,3}^{\text{PCR5}}(A \cup B) &\approx 0.094259,
 \end{aligned}$$

$$\begin{aligned}
 m_{1,2,3}^{\text{PCR6}}(A) &\approx 0.743496, \\
 m_{1,2,3}^{\text{PCR6}}(B) &\approx 0.162245, \\
 m_{1,2,3}^{\text{PCR6}}(A \cup B) &\approx 0.094259.
 \end{aligned}$$

We see a difference between the BBAs  $m_{1,2,3}^{\text{PCR5}}$  and  $m_{1,2,3}^{\text{PCR6}}$  which is normal because the PCR principles are quite different. Using the PCR5 fusion rule the first partial conflicting mass  $\pi_1(\emptyset) = m_1(A)m_2(A)m_3(B) = 0.03$  will be redistributed back to  $A$  and  $B$  proportionally to  $m_1(A)m_2(A)$  and to  $m_3(B)$  as follows

$$\frac{x_1(A)}{m_1(A)m_2(A)} = \frac{x_1(B)}{m_3(B)} = \frac{\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)},$$

whence

$$x_1(A) = \frac{m_1(A)m_2(A)\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)} = 0.0225,$$

$$x_1(B) = \frac{m_3(B)\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)} = 0.0075.$$

We can verify  $\pi_1(\emptyset) = x_1(A) + x_1(B) = 0.03$ .

Using the PCR6 fusion rule the first partial conflicting mass  $\pi_1(\emptyset) = 0.03$  will be redistributed back to  $A$  and  $B$  proportionally to  $(m_1(A) + m_2(A))$  and to  $m_3(B)$ . So we will get the following redistributions  $x_1(A) = 0.0275$  for  $A$  and  $x_1(B) = 0.0025$  for  $B$  because

$$\frac{x_1(A)}{m_1(A) + m_2(A)} = \frac{x_1(B)}{m_3(B)} = \frac{\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)},$$

whence

$$x_1(A) = \frac{(m_1(A) + m_2(A))\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)} = 0.0275$$

$$x_1(B) = \frac{m_3(B)\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)} = 0.0025$$

We can verify  $\pi_1(\emptyset) = x_1(A) + x_1(B) = 0.03$ .

Note that for all the partial conflicts having no duplicate element involved in the conflicting product  $\pi_j(\emptyset)$  we make the same redistribution with PCR5 rule and with PCR6 rule. For instance, for  $\pi_7(\emptyset) = m_1(A \cup B)m_2(A)m_3(B) = 0.0150$  we get

$$\begin{aligned}
 \frac{x_7(A \cup B)}{m_1(A \cup B)} &= \frac{x_7(A)}{m_2(A)} = \frac{x_7(B)}{m_3(B)} \\
 &= \frac{\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)},
 \end{aligned}$$

whence  $\pi_7(\emptyset) = x_7(A \cup B) + x_7(A) + x_7(B) = 0.0150$  with

$$x_7(A \cup B) = \frac{m_1(A \cup B)\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)} = 0.0050,$$

$$x_7(A) = \frac{m_2(A)\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)} \approx 0.0083,$$

$$x_7(B) = \frac{m_3(B)\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)} \approx 0.0017.$$

The next example shows also the difference between PCR5 and PCR6 rules, and it justifies why PCR6 rule is usually preferred to PCR5 rule in applications.

<sup>11</sup>The symbol  $\approx$  means ‘‘approximately equal to’’.

**Example 3:** we consider the FoD  $\Theta = \{A, B, C\}$ , and the four very simple BBAs defined by

$$m_1(A \cup B) = 1, m_2(B) = 1, m_3(A \cup B) = 1, \text{ and } m_4(C) = 1.$$

These BBAs are in conflict because the intersection of their focal elements is  $(A \cup B) \cap A \cap (A \cup B) \cap C = \emptyset$ . In this example, one has only one product of masses to calculate, which is  $\pi_1((A \cup B) \cap A \cap (A \cup B) \cap C) = m_1(A \cup B)m_2(A)m_3(A \cup B)m_4(C) = 1$ . In fact this product is a conflicting product denoted  $\pi_1(\emptyset)$ . We can also denote it  $\pi(\emptyset)$  because the index  $j = 1$  is useless in this case. Moreover, these BBAs are also in total conflict because  $\pi(\emptyset) = m_1(A \cup B)m_2(A)m_3(A \cup B)m_4(C) = 1$ .

If one applies the PCR5 rule principle we get

$$\begin{aligned} \frac{x(A \cup B)}{m_1(A \cup B)m_3(A \cup B)} &= \frac{x(B)}{m_2(B)} = \frac{x(C)}{m_4(C)} \\ &= \frac{\pi(\emptyset)}{m_1(A \cup B)m_3(A \cup B) + m_2(B) + m_4(C)} \end{aligned}$$

whence  $x(A \cup B) = 1/3$ ,  $x(B) = 1/3$  and  $x(C) = 1/3$  so that

$$\begin{aligned} m_{1,2,3,4}^{\text{PCR5}}(A \cup B) &= x(A \cup B) = 1/3 \\ m_{1,2,3,4}^{\text{PCR5}}(B) &= x(B) = 1/3 \\ m_{1,2,3,4}^{\text{PCR5}}(C) &= x(C) = 1/3 \end{aligned}$$

This PCR5 result appears counter-intuitive because three sources among the four sources exclude definitely the hypothesis  $C$  because one has  $Pl_1(C) = Pl_2(C) = Pl_3(C) = 0$ , so it is intuitively expected that after the combination of all the four BBAs the mass committed to  $C$  should not be greater than  $1/4 = 0.25$ .

If one applies the PCR6 rule principle we get

$$\begin{aligned} \frac{x(A \cup B)}{m_1(A \cup B) + m_3(A \cup B)} &= \frac{x(B)}{m_2(B)} = \frac{x(C)}{m_4(C)} \\ &= \frac{\pi(\emptyset)}{m_1(A \cup B) + m_3(A \cup B) + m_2(B) + m_4(C)}, \end{aligned}$$

whence  $x(A \cup B) = 2/4$ ,  $x(B) = 1/4$  and  $x(C) = 1/4$  so that

$$\begin{aligned} m_{1,2,3,4}^{\text{PCR6}}(A \cup B) &= x(A \cup B) = 0.5, \\ m_{1,2,3,4}^{\text{PCR6}}(B) &= x(B) = 0.25, \\ m_{1,2,3,4}^{\text{PCR6}}(C) &= x(C) = 0.25, \end{aligned}$$

which is in better agreement with what we intuitively expect because  $m_{1,2,3,4}^{\text{PCR6}}(C)$  is not greater than than  $1/4$ . Of course in this example, Dempster's rule of combination cannot be simply applied because the conflict is total yielding a division by zero in Dempster's rule formula [8], but by using eventually some discounting methods to modify the BBAs to combine.

## V. FLAWED BEHAVIOR OF PCR5 AND PCR6 RULES

Formula (17) shows that in general PCR6 is not associative, and by combining two sources in a row each time and we continue doing that the results is different from the global combination of all sources using PCR6. The formula is true. Formula (18) says that in general PCR5 is different from PCR6, of course except the case when we combine only 2 sources. Formula (19) shows that in general PCR5 does not have the ignorance source as a neutral element.

The PCR5 and PCR6 rules of combination are not associative which means that the fusion of the BBAs must be done using general formulas (14) or (15) if one has more than two BBAs to combine, which is not very convenient. Therefore, the sequential PCR5 or PCR6 combination of  $S > 2$  BBAs are not in general equal to the global PCR5 or PCR6 fusion of the  $S$  BBAs altogether because the order of the combination of the sources does matter in the sequential combination. In general (i.e. when conflicts exist between the sources of evidence to combine) one has for  $S > 2$

$$\begin{aligned} \text{PCR5}(m_1, m_2, \dots, m_S) &\neq \\ \text{PCR5}(\text{PCR5}(\text{PCR5}(m_1, m_2), m_3), \dots, m_S), \end{aligned} \quad (16)$$

and

$$\begin{aligned} \text{PCR6}(m_1, m_2, \dots, m_S) &\neq \\ \text{PCR6}(\text{PCR6}(\text{PCR6}(m_1, m_2), m_3), \dots, m_S), \end{aligned} \quad (17)$$

and also for  $S > 2$  PCR5 fusion result is generally different of PCR6 fusion result that is

$$\text{PCR5}(m_1, m_2, \dots, m_S) \neq \text{PCR6}(m_1, m_2, \dots, m_S). \quad (18)$$

PCR5 and PCR6 rules can become computationally intractable for combining a large number of sources and for working with large FoD. This is a well-known limitation of these rules, but this is the price to pay to get better results than with classical rules.

Aside the complexity of these rules, it is worth to mention that the neutral impact property of the vacuous BBA  $m_v$  is lost in general when considering the PCR5 or PCR6 combination of  $S > 2$  BBAs altogether, that is

$$\text{PCR5}(m_1, \dots, m_{S-1}, m_v) \neq \text{PCR5}(m_1, \dots, m_{S-1}), \quad (19)$$

and

$$\text{PCR6}(m_1, \dots, m_{S-1}, m_v) \neq \text{PCR6}(m_1, \dots, m_{S-1}). \quad (20)$$

This is due to the redistribution principles used in PCR5 and in PCR6 rules. Example 4 shows the non-neutral impact of the vacuous BBA in PCR5 and PCR6 rules for convenience. Note that the vacuous BBA has a neutral impact in the fusion result if and only if one has only two BBAs to combine with PCR5, or PCR6, and one of them is the vacuous BBA because in this case there is no possible (partial) conflict to redistribute between any BBA  $m(\cdot)$  defined over the FoD  $\Theta$

and the vacuous BBA  $m_v(\cdot)$ . That is, for any BBA  $m_1(\cdot)$  one always has

$$\text{PCR5}(m_1, m_v) = \text{PCR6}(m_1, m_v) = m_1. \quad (21)$$

**Example 4:** we consider the FoD  $\Theta = \{A, B\}$  having only two elements, and the following four BBAs as follows:

$$\begin{aligned} m_1(A) &= 0.6, m_1(B) = 0.1, m_1(A \cup B) = 0.3, \\ m_2(A) &= 0.5, m_2(B) = 0.3, m_2(A \cup B) = 0.2, \\ m_3(A) &= 0.4, m_3(B) = 0.1, m_3(A \cup B) = 0.5, \\ m_4(A \cup B) &= 1 \end{aligned}$$

BBAs  $m_1$ ,  $m_2$  and  $m_3$  are as in example 2, and the BBA  $m_4$  is nothing but the vacuous BBA  $m_v$  defined over this FoD  $\Theta$ .

In example 2, we did obtain with  $\text{PCR5}(m_1, m_2, m_3)$  and with  $\text{PCR5}(m_1, m_2, m_3, m_4)$  the following resulting BBAs

$$\begin{aligned} m_{1,2,3}^{\text{PCR5}}(A) &\approx 0.723281, \\ m_{1,2,3}^{\text{PCR5}}(B) &\approx 0.182460, \\ m_{1,2,3}^{\text{PCR5}}(A \cup B) &\approx 0.094259, \end{aligned}$$

and

$$\begin{aligned} m_{1,2,3,4}^{\text{PCR5}}(A) &\approx 0.654604, \\ m_{1,2,3,4}^{\text{PCR5}}(B) &\approx 0.144825, \\ m_{1,2,3,4}^{\text{PCR5}}(A \cup B) &\approx 0.200571. \end{aligned}$$

Clearly,  $\text{PCR5}(m_1, m_2, m_3) \neq \text{PCR5}(m_1, m_2, m_3, m_4)$  even if  $m_4$  is the vacuous BBA.

Analogously, we did obtain with  $\text{PCR6}(m_1, m_2, m_3)$  and with  $\text{PCR6}(m_1, m_2, m_3, m_4)$

$$\begin{aligned} m_{1,2,3}^{\text{PCR6}}(A) &\approx 0.743496, \\ m_{1,2,3}^{\text{PCR6}}(B) &\approx 0.162245, \\ m_{1,2,3}^{\text{PCR6}}(A \cup B) &\approx 0.094259, \end{aligned}$$

and

$$\begin{aligned} m_{1,2,3,4}^{\text{PCR6}}(A) &\approx 0.647113, \\ m_{1,2,3,4}^{\text{PCR6}}(B) &\approx 0.128342, \\ m_{1,2,3,4}^{\text{PCR6}}(A \cup B) &\approx 0.224545. \end{aligned}$$

Therefore,  $\text{PCR6}(m_1, m_2, m_3) \neq \text{PCR6}(m_1, m_2, m_3, m_4)$ , even if  $m_4$  is the vacuous BBA.

This example 4 shows clearly that the vacuous BBA does not have a neutral impact in the PCR5 and PCR6 rules of combination. In fact, adding more vacuous BBAs  $m_v$  in the PCR5 or PCR6 fusion will increase more and more the mass of  $A \cup B$  while decreasing more and more the masses of  $A$  and of  $B$  with PCR5, and PCR6. When the number of vacuous BBAs  $m_v$  increases, we will have<sup>12</sup>  $m_{1,2,3,m_v,\dots,m_v}^{\text{PCR5/6}}(A \cup B) \rightarrow 1$ ,  $m_{1,2,3,m_v,\dots,m_v}^{\text{PCR5/6}}(A) \rightarrow 0$ , and  $m_{1,2,3,m_v,\dots,m_v}^{\text{PCR5/6}}(B) \rightarrow 0$ .

This is unsatisfactory because the vacuous BBA brings no useful information to exploit, and it is naturally expected that it

must not impact the fusion result in the combination of BBAs. This can be seen as a flaw of the behavior of PCR5 and PCR6 rules of combination.

To emphasize this flaw, we give in the example 5 a case where the mass committed to some partial uncertainties can increase more than necessary with PCR5 and with PCR6 rules of combination. This is detrimental for the quality of the fusion result and for decision-making because the result is more uncertain than it should be, and consequently the decision is more difficult to make.

**Example 5:** we consider the FoD  $\Theta = \{A, B, C, D, E\}$ , and the following three BBAs

$$\begin{cases} m_1(A \cup B) = 0.70, \\ m_1(C \cup D) = 0.06, \\ m_1(A \cup B \cup C \cup D) = 0.15, \\ m_1(E) = 0.09, \end{cases}$$

and

$$\begin{cases} m_2(A \cup B) = 0.06, \\ m_2(C \cup D) = 0.50, \\ m_2(A \cup B \cup C \cup D) = 0.04, \\ m_2(E) = 0.40, \end{cases}$$

and

$$\begin{cases} m_3(B) = 0.01 \\ m_3(A \cup B \cup C \cup D \cup E) = 0.99. \end{cases}$$

Note that the BBA  $m_3$  is not equal to the vacuous BBA but it is very close to the vacuous BBA because  $m_3(\Theta)$  is close to one.

If we make the  $\text{PCR6}(m_1, m_2)$  fusion of only the two BBAs  $m_1$  and  $m_2$  altogether, which is also equal to  $\text{PCR5}(m_1, m_2)$ , we obtain

$$\begin{cases} m_{1,2}^{\text{PCR6}}(A \cup B) \approx 0.465309, \\ m_{1,2}^{\text{PCR6}}(C \cup D) \approx 0.296299, \\ m_{1,2}^{\text{PCR6}}(A \cup B \cup C \cup D) \approx 0.023471, \\ m_{1,2}^{\text{PCR6}}(E) \approx 0.214921. \end{cases}$$

If we make the  $\text{PCR6}(m_1, m_2, m_3)$  fusion of all these three BBAs altogether we obtain

$$\begin{cases} m_{1,2,3}^{\text{PCR6}}(B) \approx 0.000962, \\ m_{1,2,3}^{\text{PCR6}}(A \cup B) \approx 0.286107, \\ m_{1,2,3}^{\text{PCR6}}(C \cup D) \approx 0.203454, \\ m_{1,2,3}^{\text{PCR6}}(A \cup B \cup C \cup D) \approx 0.012203, \\ m_{1,2,3}^{\text{PCR6}}(E) \approx 0.116038, \\ m_{1,2,3}^{\text{PCR6}}(A \cup B \cup C \cup D \cup E) \approx 0.381236. \end{cases}$$

One sees that combining the BBAs  $m_1$ ,  $m_2$  with the BBA  $m_3$  (where  $m_3$  is close to vacuous BBA, and therefore  $m_3$  is almost non-informative) generates a big increase of the belief of the uncertainty in the resulting BBA. This behaviour is clearly counter-intuitive because if the source is almost vacuous, only a small degradation of the uncertainty is expected and in the limit case when  $m_3$  is the vacuous BBA no impact of  $m_3$  on the fusion result should occur. Note that this behavior

<sup>12</sup>The notation  $m^{\text{PCR5/6}}$  indicates “ $m^{\text{PCR5}}$  or  $m^{\text{PCR6}}$ ” for convenience.

also occurs with PCR5( $m_1, m_2, m_3$ ) because one has for this example

$$\left\{ \begin{array}{l} m_{1,2,3}^{\text{PCR5}}(B) \approx 0.001103, \\ m_{1,2,3}^{\text{PCR5}}(A \cup B) \approx 0.286107, \\ m_{1,2,3}^{\text{PCR5}}(C \cup D) \approx 0.203384, \\ m_{1,2,3}^{\text{PCR5}}(A \cup B \cup C \cup D) \approx 0.012203, \\ m_{1,2,3}^{\text{PCR5}}(E) \approx 0.115967, \\ m_{1,2,3}^{\text{PCR5}}(A \cup B \cup C \cup D \cup E) \approx 0.381236. \end{array} \right.$$

The deep analysis of the partial conflict redistributions done in this interesting example reveals clearly the flaw of the principles of PCR5 and PCR6 rules of combination. Indeed, for this example one has  $\mathcal{F}_{m_1} \cdot \mathcal{F}_{m_2} \cdot \mathcal{F}_{m_3} = 4 \cdot 4 \cdot 2 = 32$  products  $\pi_j(X_{j_1} \cap X_{j_2} \cap X_{j_3}) = m_1(X_{j_1})m_2(X_{j_2})m_3(X_{j_3})$  to calculate, where  $X_{j_1} \in \mathcal{F}(m_1) = \{A \cup B, C \cup D, A \cup B \cup C \cup D, E\}$ ,  $X_{j_2} \in \mathcal{F}(m_2) = \{A \cup B, C \cup D, A \cup B \cup C \cup D, E\}$ , and  $X_{j_3} \in \mathcal{F}(m_3) = \{B, A \cup B \cup C \cup D \cup E\}$ . Among these 32 possible conjunctions of focal elements, twenty products corresponds to partial conflicts when  $X_{j_1} \cap X_{j_2} \cap X_{j_3} = \emptyset$ , which need to be redistributed properly to some elements of  $2^\Theta \setminus \{\emptyset\}$  according to the PCR5, or the PCR6 redistribution principles.

More precisely, we have to consider all the following products  $\pi_j$  for calculating the result

$$\begin{aligned} \pi_1(B) &= m_1(A \cup B)m_2(A \cup B)m_3(B) = 0.00042, \\ \pi_2(A \cup B) &= m_1(A \cup B)m_2(A \cup B)m_3(\Theta) = 0.04158, \\ \pi_3(\emptyset) &= m_1(A \cup B)m_2(C \cup D)m_3(B) = 0.0035, \\ \pi_4(\emptyset) &= m_1(A \cup B)m_2(C \cup D)m_3(\Theta) = 0.3465, \\ \pi_5(B) &= m_1(A \cup B)m_2(A \cup B \cup C \cup D)m_3(B) = 0.00028, \\ \pi_6(A \cup B) &= m_1(A \cup B)m_2(A \cup B \cup C \cup D)m_3(\Theta) = 0.02772, \\ \pi_7(\emptyset) &= m_1(A \cup B)m_2(E)m_3(B) = 0.0028, \\ \pi_8(\emptyset) &= m_1(A \cup B)m_2(E)m_3(\Theta) = 0.2772, \\ \pi_9(\emptyset) &= m_1(C \cup D)m_2(A \cup B)m_3(B) = 0.000036, \\ \pi_{10}(\emptyset) &= m_1(C \cup D)m_2(A \cup B)m_3(\Theta) = 0.003564, \\ \pi_{11}(\emptyset) &= m_1(C \cup D)m_2(C \cup D)m_3(B) = 0.0003, \\ \pi_{12}(C \cup D) &= m_1(C \cup D)m_2(C \cup D)m_3(\Theta) = 0.0297, \\ \pi_{13}(\emptyset) &= m_1(C \cup D)m_2(A \cup B \cup C \cup D)m_3(B) = 0.000024, \\ \pi_{14}(C \cup D) &= m_1(C \cup D)m_2(A \cup B \cup C \cup D)m_3(\Theta) \\ &= 0.002376, \\ \pi_{15}(\emptyset) &= m_1(C \cup D)m_2(E)m_3(B) = 0.00024, \\ \pi_{16}(\emptyset) &= m_1(C \cup D)m_2(E)m_3(\Theta) = 0.02376, \\ \pi_{17}(B) &= m_1(A \cup B \cup C \cup D)m_2(A \cup B)m_3(B) = 0.00009, \\ \pi_{18}(A \cup B) &= m_1(A \cup B \cup C \cup D)m_2(A \cup B)m_3(\Theta) = 0.00891, \\ \pi_{19}(\emptyset) &= m_1(A \cup B \cup C \cup D)m_2(C \cup D)m_3(B) = 0.00075, \\ \pi_{20}(C \cup D) &= m_1(A \cup B \cup C \cup D)m_2(C \cup D)m_3(\Theta) \\ &= 0.07425, \\ \pi_{21}(B) &= m_1(A \cup B \cup C \cup D)m_2(A \cup B \cup C \cup D)m_3(B) \\ &= 0.00006, \\ \pi_{22}(A \cup B \cup C \cup D) &= m_1(A \cup B \cup C \cup D)m_2(A \cup B \cup C \cup D) \\ &\quad \cdot m_3(\Theta) = 0.00594, \\ \pi_{23}(\emptyset) &= m_1(A \cup B \cup C \cup D)m_2(E)m_3(B) = 0.0006, \\ \pi_{24}(\emptyset) &= m_1(A \cup B \cup C \cup D)m_2(E)m_3(\Theta) = 0.0594, \\ \pi_{25}(\emptyset) &= m_1(E)m_2(A \cup B)m_3(B) = 0.000054, \end{aligned}$$

$$\begin{aligned} \pi_{26}(\emptyset) &= m_1(E)m_2(A \cup B)m_3(\Theta) = 0.005346, \\ \pi_{27}(\emptyset) &= m_1(E)m_2(C \cup D)m_3(B) = 0.00045, \\ \pi_{28}(\emptyset) &= m_1(E)m_2(C \cup D)m_3(\Theta) = 0.04455, \\ \pi_{29}(\emptyset) &= m_1(E)m_2(A \cup B \cup C \cup D)m_3(B) = 0.000036, \\ \pi_{30}(\emptyset) &= m_1(E)m_2(A \cup B \cup C \cup D)m_3(\Theta) = 0.003564, \\ \pi_{31}(\emptyset) &= m_1(E)m_2(E)m_3(B) = 0.00036, \\ \pi_{32}(E) &= m_1(E)m_2(E)m_3(\Theta) = 0.03564. \end{aligned}$$

The conjunctive rule gives

$$\begin{aligned} m_{1,2,3}^{\text{Conj}}(B) &= \pi_1(B) + \pi_5(B) + \pi_{17}(B) + \pi_{21}(B) = 0.00085, \\ m_{1,2,3}^{\text{Conj}}(A \cup B) &= \pi_2(A \cup B) + \pi_6(A \cup B) + \pi_{18}(A \cup B) \\ &= 0.07821, \\ m_{1,2,3}^{\text{Conj}}(C \cup D) &= \pi_{12}(C \cup D) + \pi_{14}(C \cup D) + \pi_{20}(C \cup D) \\ &= 0.106326, \\ m_{1,2,3}^{\text{Conj}}(A \cup B \cup C \cup D) &= \pi_{22}(A \cup B \cup C \cup D) = 0.00594, \\ m_{1,2,3}^{\text{Conj}}(E) &= \pi_{32}(E) = 0.03564. \end{aligned}$$

The total conflicting mass between these three BBAs is

$$\begin{aligned} m_{1,2,3}^{\text{Conj}}(\emptyset) &= \sum_{j=3,4,7,\dots,11,13,15,16,19,23,\dots,31} \pi_j(\emptyset) \\ &= 1 - m_{1,2,3}^{\text{Conj}}(B) - m_{1,2,3}^{\text{Conj}}(A \cup B) - m_{1,2,3}^{\text{Conj}}(C \cup D) \\ &\quad - m_{1,2,3}^{\text{Conj}}(A \cup B \cup C \cup D) - m_{1,2,3}^{\text{Conj}}(E) = 0.773034. \end{aligned}$$

Let's examine how the  $m_{1,2,3}^{\text{PCR5}}(\Theta) \approx 0.381236$  value is obtained based on the PCR5 redistribution principle. Based on the structures of  $\pi_j(\emptyset)$  products, we have to consider only products involving a proportional redistribution to  $\Theta$ . So we get a proportional redistribution to  $\Theta$  only from the following products

$$\begin{aligned} \pi_4(\emptyset) &= m_1(A \cup B)m_2(C \cup D)m_3(\Theta) = 0.3465, \\ \pi_8(\emptyset) &= m_1(A \cup B)m_2(E)m_3(\Theta) = 0.2772, \\ \pi_{10}(\emptyset) &= m_1(C \cup D)m_2(A \cup B)m_3(\Theta) = 0.003564, \\ \pi_{16}(\emptyset) &= m_1(C \cup D)m_2(E)m_3(\Theta) = 0.02376, \\ \pi_{24}(\emptyset) &= m_1(A \cup B \cup C \cup D)m_2(E)m_3(\Theta) = 0.0594, \\ \pi_{26}(\emptyset) &= m_1(E)m_2(A \cup B)m_3(\Theta) = 0.005346, \\ \pi_{28}(\emptyset) &= m_1(E)m_2(C \cup D)m_3(\Theta) = 0.04455, \\ \pi_{30}(\emptyset) &= m_1(E)m_2(A \cup B \cup C \cup D)m_3(\Theta) = 0.003564. \end{aligned}$$

Because there is no duplicate focal elements in each of these products, the PCR5 and PCR6 redistributions to  $\Theta$  will be the same in this example.

The proportional redistribution of  $\pi_4(\emptyset)$  to  $\Theta$  is

$$x_4(\Theta) = \frac{m_3(\Theta)\pi_4(\emptyset)}{m_1(A \cup B) + m_2(C \cup D) + m_3(\Theta)} \approx 0.156637.$$

The proportional redistribution of  $\pi_8(\emptyset)$  to  $\Theta$  is

$$x_8(\Theta) = \frac{m_3(\Theta)\pi_8(\emptyset)}{m_1(A \cup B) + m_2(E) + m_3(\Theta)} \approx 0.131305.$$

The proportional redistribution of  $\pi_{10}(\emptyset)$  to  $\Theta$  is

$$x_{10}(\Theta) = \frac{m_3(\Theta)\pi_{10}(\emptyset)}{m_1(C \cup D) + m_2(A \cup B) + m_3(\Theta)} \approx 0.003179.$$

The proportional redistribution of  $\pi_{16}(\emptyset)$  to  $\Theta$  is

$$x_{16}(\Theta) = \frac{m_3(\Theta)\pi_{16}(\emptyset)}{m_1(C \cup D) + m_2(E) + m_3(\Theta)} \approx 0.016222.$$

The proportional redistribution of  $\pi_{24}(\emptyset)$  to  $\Theta$  is

$$x_{24}(\Theta) = \frac{m_3(\Theta)\pi_{24}(\emptyset)}{m_1(A \cup B \cup C \cup D) + m_2(E) + m_3(\Theta)} \approx 0.038186.$$

The proportional redistribution of  $\pi_{26}(\emptyset)$  to  $\Theta$  is

$$x_{26}(\Theta) = \frac{m_3(\Theta)\pi_{26}(\emptyset)}{m_1(E) + m_2(A \cup B) + m_3(\Theta)} \approx 0.004643.$$

The proportional redistribution of  $\pi_{28}(\emptyset)$  to  $\Theta$  is

$$x_{28}(\Theta) = \frac{m_3(\Theta)\pi_{28}(\emptyset)}{m_1(E) + m_2(C \cup D) + m_3(\Theta)} \approx 0.027914.$$

The proportional redistribution of  $\pi_{30}(\emptyset)$  to  $\Theta$  is

$$x_{30}(\Theta) = \frac{m_3(\Theta)\pi_{30}(\emptyset)}{m_1(E) + m_2(A \cup B \cup C \cup D) + m_3(\Theta)} \approx 0.003150.$$

Therefore we finally obtain the quite big value for the mass committed to  $\Theta$

$$\begin{aligned} m_{1,2,3}^{\text{PCR5}}(\Theta) &= x_4(\Theta) + x_8(\Theta) + x_{10}(\Theta) + x_{16}(\Theta) + x_{24}(\Theta) \\ &\quad + x_{26}(\Theta) + x_{28}(\Theta) + x_{30}(\Theta) \\ &\approx 0.381236. \end{aligned}$$

We see clearly why PCR5 (and PCR6) redistributes some mass to uncertainty  $\Theta$  although the focal element  $\Theta$  is not in conflict with other focal elements involved in each product  $\pi_4(\emptyset)$ ,  $\pi_8(\emptyset)$ ,  $\pi_{10}(\emptyset)$ ,  $\pi_{16}(\emptyset)$ ,  $\pi_{24}(\emptyset)$ ,  $\pi_{26}(\emptyset)$ ,  $\pi_{28}(\emptyset)$  and  $\pi_{30}(\emptyset)$ , which is an undesirable behavior that we want to avoid. That is why we propose in the next section some improvement of PCR5 and PCR6 rules of combination.

## VI. IMPROVEMENT OF PCR5 AND PCR6 RULES

To circumvent the weakness of the original PCR5 and PCR6 redistribution principles, we propose an improvement of these rules that will be denoted as PCR5<sup>+</sup> and PCR6<sup>+</sup> in the sequel. These new rules are not redundant with PCR5 nor with PCR6 when combining more than two BBAs altogether.

The very simple and basic idea to improve PCR5 and PCR6 redistribution principles is to discard the elements that contain all the other elements implied in the partial conflict  $\pi_j(\emptyset)$  calculation. Indeed, the elements discarded are regarded as non-informative and not useful for making the conflict redistribution.

For instance, if we consider the previous Example 5, the conflicting mass with PCR5<sup>+</sup> and PCR6<sup>+</sup> for the conflicting product  $\pi_4(\emptyset) = m_1(A \cup B)m_2(C \cup D)m_3(\Theta)$  will be proportionally redistributed back only to  $A \cup B$  and to  $C \cup D$  but not to  $\Theta$  because  $A \cup B \subseteq \Theta$  and  $C \cup D \subseteq \Theta$ . Thus

with PCR5<sup>+</sup> and PCR6<sup>+</sup> rules we will make the following redistribution:

$$\frac{x_4(A \cup B)}{m_1(A \cup B)} = \frac{x_4(C \cup D)}{m_2(C \cup D)} = \frac{\pi_4(\emptyset)}{m_1(A \cup B) + m_2(C \cup D)}.$$

Here,  $x_4(\Theta)$  is set to 0 with PCR5<sup>+</sup> and PCR6<sup>+</sup> principles because no proportion of  $\pi_4(\emptyset)$  must be redistributed to  $\Theta$ .

However, with PCR5 and PCR6 rule we make the redistributions according to

$$\begin{aligned} \frac{x_4(A \cup B)}{m_1(A \cup B)} &= \frac{x_4(C \cup D)}{m_2(C \cup D)} = \frac{x_4(\Theta)}{m_3(\Theta)} \\ &= \frac{\pi_4(\emptyset)}{m_1(A \cup B) + m_2(C \cup D) + m_3(\Theta)}. \end{aligned}$$

### A. Selection of focal elements for proportional redistribution

The main issue to improve PCR5 and PCR6 rules of combination is how to identify in each conflicting product  $\pi_j(\emptyset)$  the set of elements to keep for making the improved proportional redistribution.

In this section we propose a solution of this problem that can be easily implemented. For convenience, we give also the basic Matlab™ codes of PCR5<sup>+</sup> and PCR6<sup>+</sup> in appendix 3.

Let's consider  $\pi_j(\emptyset) = m_1(X_{j_1})m_2(X_{j_2}) \dots m_S(X_{j_S})$  a conflicting product<sup>13</sup> where  $X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S} = \emptyset$ . We denote by  $\mathcal{X}_j = \{X_1, \dots, X_{s_j}, s_j \leq S\}$  the set of all distinct components of the  $S$ -tuple  $\mathbf{X}_j$  related with the conflicting product  $\pi_j(\emptyset)$ . The order of the elements in  $\mathcal{X}_j$  does not matter. The number  $s_j$  of elements in  $\mathcal{X}_j$  can be less than  $S$  because it is possible to have duplicate focal elements in  $\pi_j(\emptyset)$ . We consider in  $\mathcal{X}_j$  only the distinct focal elements involved in  $\pi_j(\emptyset)$  (see the next example) and we will define their binary *keeping-index indicator* which will allow to know if each element of  $\mathcal{X}_j$  needs to be kept in the proportional conflict redistribution, or not, in the improved PCR5 and PCR6 rules of combination.

For each element  $X_l \in \mathcal{X}_j$  we first define its binary *containing indicator*  $\delta_j(X_{l'}, X_l)$  with respect to  $X_{l'} \in \mathcal{X}_j$  to characterize if  $X_l$  contains (includes)  $X_{l'}$  in wide sense, or not. Therefore, we take  $\delta_j(X_{l'}, X_l) = 1$  if  $X_{l'} \cap X_l = X_{l'}$ , or equivalently if  $X_{l'} \subseteq X_l$ , and  $\delta_j(X_{l'}, X_l) = 0$  otherwise. The definition of this binary *containing indicator* is summarized by the formula

$$\delta_j(X_{l'}, X_l) \triangleq \begin{cases} 1 & \text{if } X_{l'} \subseteq X_l, \\ 0 & \text{if } X_{l'} \not\subseteq X_l. \end{cases} \quad (22)$$

Of course  $\delta_j(X_l, X_l) = 1$  because  $X_l \cap X_l = X_l$ , and we have  $\delta_j(X_{l'}, X_l) = 0$  as soon as  $|X_{l'}| > |X_l|$ , where  $|X_{l'}|$  and  $|X_l|$  are the cardinalities of  $X_{l'}$  and  $X_l$  respectively. We have

<sup>13</sup>We consider  $S > 2$  BBAs because for  $S = 2$  BBAs no improper increasing of uncertainty occurs with PCR5 or PCR6.

also  $\delta_j(X_{l'}, X_l) = 0$  when  $X_{l'} \cap X_l \neq X_{l'}$ . For  $X_l = \Theta$ , we have  $\delta_j(X_{l'}, X_l) = \delta_j(X_{l'}, \Theta) = 1$  for any  $X_{l'} \in \mathcal{X}_j$ .

To know if a focal element  $X_{j_i} \in \mathbf{X}_j$  must be kept, or not, in the proportional redistribution of the  $j$ -th conflicting mass  $\pi_j(\emptyset)$  with PCR5<sup>+</sup> and PCR6<sup>+</sup> rules, we have to determinate its binary *keeping-index*  $\kappa_j(X_{j_i})$ . For this, we define  $\kappa_j(X_{j_i}) \in \{0, 1\}$  as follows

$$\kappa_j(X_{j_i}) \triangleq 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_i}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l). \quad (23)$$

The value  $\kappa_j(X_{j_i}) = 1$  stipulates that the focal element  $X_{j_i} \in \mathbf{X}_j$  must receive some proportional redistribution from the conflicting mass  $\pi_j(\emptyset)$ . The value  $\kappa_j(X_{j_i}) = 0$  indicates that the focal element  $X_{j_i}$  will not be involved in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

The binary keeping-index can also be defined equivalently as

$$\kappa_j(X_{j_i}) = \begin{cases} 1 & \text{if } c(X_{j_i}) \text{ is true,} \\ 1 - \prod_{\substack{X_{l'} \in \mathcal{X}_j \\ X_{l'} \neq X_{j_i} \\ |X_{l'}| \leq |X_{j_i}|}} \delta_j(X_{l'}, X_{j_i}) & \text{if } c(X_{j_i}) \text{ is false,} \end{cases} \quad (24)$$

where the condition  $c(X_{j_i})$  is defined as

$$c(X_{j_i}) \triangleq \exists X_l \in \mathcal{X}_j \text{ such } |X_l| > |X_{j_i}| \text{ and } \kappa_j(X_l) = 1.$$

Because this second definition of  $\kappa_j(X_{j_i})$  is self-referencing, we need to calculate the binary keeping-indexes iteratively starting by the element of  $\mathcal{X}_j$  of highest cardinality (say  $X$ ), then for elements of  $\mathcal{X}_j$  of cardinality  $|X| - 1$  (if any), then for elements of  $\mathcal{X}_j$  of cardinality  $|X| - 2$  (if any), etc. From the implementation standpoint the definition (24) is more efficient than the direct definition (23).

**Remark 1:** We always have  $\kappa_j(\Theta) = 0$  if  $\Theta \in \mathcal{X}_j$  because  $\Theta$  always includes all other focal elements of  $\mathcal{X}_j$  and  $\Theta$  has the highest cardinality, so  $\delta_j(X_{l'}, \Theta) = 1$  for all  $X_{l'} \in \mathcal{X}_j$ . Therefore the binary keeping-index formula (23) reduces to

$$\kappa_j(\Theta) = 1 - \prod_{X_{l'} \in \mathcal{X}_j} \delta_j(X_{l'}, \Theta) = 1 - \underbrace{1 \cdot 1 \cdot \dots \cdot 1}_{|\mathcal{X}_j| \text{ terms}} = 0.$$

**Remark 2:** For a given FoD and a given number of BBAs to combine, it is always possible to calculate off-line the values of the binary keeping-indexes of focal elements of all possible combinations of focal elements involved in conflicting products  $\pi_j(\emptyset) > 0$  because the binary keeping-index depends only on the structure of the focal elements, and not on the numerical mass values of the focal elements. This remark is important, especially in applications where we have thousands or millions of fusion steps to make because we will not have to recalculate in each fusion step the binary keeping-indexes for each  $\pi_j(\emptyset)$  even if the input BBAs values to combine change.

**Remark 3:** It is worth to recall that PCR5<sup>+</sup> and PCR6<sup>+</sup> have interest if and only if we have more than two ( $S > 2$ ) BBAs to combine. If we have only two BBAs to combine ( $S = 2$ ) we always get  $m_{\text{PCR5}} = m_{\text{PCR5}^+} = m_{\text{PCR6}} = m_{\text{PCR6}^+}$  because in this case the PCR5, PCR5<sup>+</sup>, PCR6, PCR6<sup>+</sup> rules coincide.

For convenience, we illustrate the calculation of these binary keeping-indexes based on the direct calculation (23) for different examples.

**Example 6:** We consider the FoD  $\Theta = \{A, B, C, D\}$ , six BBAs, and the  $j$ -th conflicting (assumed strictly positive) product whose structure is as follows

$$\pi_j(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C)m_4(B \cup C) \\ \cdot m_5(A \cup B \cup C)m_6(A \cup B \cup C \cup D)$$

In this product  $\pi_j(\emptyset)$  we have the duplicate focal element  $B \cup C$  because it appears both in  $m_2(B \cup C)$  and in  $m_4(B \cup C)$ . The focal elements entering in each BBA of  $\pi_j(\emptyset)$  are respectively  $X_{j_1} = A$ ,  $X_{j_2} = B \cup C$ ,  $X_{j_3} = A \cup C$ ,  $X_{j_4} = B \cup C$ ,  $X_{j_5} = A \cup B \cup C$ , and  $X_{j_6} = A \cup B \cup C \cup D = \Theta$ . So we have to consider only the following set of distinct focal elements for this  $\pi_j(\emptyset)$  product

$$\mathcal{X}_j = \{X_1 = A, X_2 = B \cup C, X_3 = A \cup C, \\ X_4 = A \cup B \cup C, X_5 = A \cup B \cup C \cup D\}$$

Therefore, considering only  $X_{l'} \neq X_l$  and  $|X_{l'}| \leq |X_l|$  that are conditions entering in formula (23), we have the following binary containing indicator  $\delta_j(X_{l'}, X_l)$  values:

$$\begin{aligned} \delta_j(X_1, X_2) &= 0 \text{ because } (X_1 = A) \not\subseteq (X_2 = B \cup C), \\ \delta_j(X_1, X_3) &= 1 \text{ because } (X_1 = A) \subseteq (X_3 = A \cup C), \\ \delta_j(X_1, X_4) &= 1 \text{ because } (X_1 = A) \subseteq (X_4 = A \cup B \cup C), \\ \delta_j(X_1, X_5) &= 1 \text{ because } (X_1 = A) \subseteq (X_5 = \Theta), \\ \delta_j(X_2, X_3) &= 0 \text{ because } (X_2 = B \cup C) \not\subseteq (X_3 = A \cup C), \\ \delta_j(X_2, X_4) &= 1 \text{ because } (X_2 = B \cup C) \subseteq (X_4 = A \cup B \cup C), \\ \delta_j(X_2, X_5) &= 1 \text{ because } (X_2 = B \cup C) \subseteq (X_5 = \Theta), \\ \delta_j(X_3, X_2) &= 0 \text{ because } (X_3 = A \cup C) \not\subseteq (X_2 = B \cup C), \\ \delta_j(X_3, X_4) &= 1 \text{ because } (X_3 = A \cup C) \subseteq (X_4 = A \cup B \cup C), \\ \delta_j(X_3, X_5) &= 1 \text{ because } (X_3 = A \cup C) \subseteq (X_5 = \Theta), \\ \delta_j(X_4, X_5) &= 1 \text{ because } (X_4 = A \cup B \cup C) \subseteq (X_5 = \Theta). \end{aligned}$$

The binary keeping-indexes  $\kappa_j(X_{j_i})$  for  $i = 1, 2, \dots, 6$  are calculated based on the formula (23) as follows:

- For the focal element  $X_{j_1} = A = X_1$  of  $\mathcal{X}_j$  having  $|X_{j_1}| = 1$ , we get

$$\begin{aligned} \kappa_j(A) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_1}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\ &= 1 - [\delta_j(X_1, X_2)\delta_j(X_1, X_3)\delta_j(X_1, X_4)\delta_j(X_1, X_5) \\ &\quad \cdot \delta_j(X_2, X_3)\delta_j(X_2, X_4)\delta_j(X_2, X_5)\delta_j(X_3, X_2) \\ &\quad \cdot \delta_j(X_3, X_4)\delta_j(X_3, X_5)\delta_j(X_4, X_5)] \\ &= 1 - 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 = 1. \end{aligned}$$

Hence the focal element  $X_{j_1} = A$  will be kept in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

- For the focal element  $X_{j_2} = B \cup C = X_2$  of  $\mathcal{X}_j$  having  $|X_{j_2}| = 2$ , we get

$$\begin{aligned} \kappa_j(B \cup C) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_2}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\ &= 1 - [\delta_j(X_1, X_2)\delta_j(X_1, X_3)\delta_j(X_1, X_4) \\ &\quad \cdot \delta_j(X_1, X_5)\delta_j(X_2, X_3)\delta_j(X_2, X_4) \\ &\quad \cdot \delta_j(X_2, X_5)\delta_j(X_3, X_2)\delta_j(X_3, X_4) \\ &\quad \cdot \delta_j(X_3, X_5)\delta_j(X_4, X_5)] \\ &= 1 - 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 = 1. \end{aligned}$$

Hence the focal element  $X_{j_2} = B \cup C$  will be kept in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

- For the focal element  $X_{j_3} = A \cup C = X_3$  of  $\mathcal{X}_j$  having  $|X_{j_3}| = 2$ , we get

$$\begin{aligned} \kappa_j(A \cup C) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_3}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\ &= 1 - [\delta_j(X_1, X_2)\delta_j(X_1, X_3)\delta_j(X_1, X_4) \\ &\quad \cdot \delta_j(X_1, X_5)\delta_j(X_2, X_3)\delta_j(X_2, X_4) \\ &\quad \cdot \delta_j(X_2, X_5)\delta_j(X_3, X_2)\delta_j(X_3, X_4) \\ &\quad \cdot \delta_j(X_3, X_5)\delta_j(X_4, X_5)] \\ &= 1 - 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 = 1. \end{aligned}$$

Hence the focal element  $X_{j_3} = A \cup C$  will be kept in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

- For the duplicate focal element  $X_{j_4} = B \cup C$  of  $\mathcal{X}_j$  having  $|X_{j_4}| = 2$ , we have  $\kappa_j(X_{j_4}) = 1$  because  $X_{j_4} = X_{j_2}$  and  $\kappa_j(X_{j_2}) = 1$ .
- For the focal element  $X_{j_5} = A \cup B \cup C = X_4$  of  $\mathcal{X}_j$  having  $|X_{j_5}| = 3$ , we get

$$\begin{aligned} \kappa_j(A \cup B \cup C) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_5}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\ &= 1 - [\delta_j(X_1, X_4)\delta_j(X_1, X_5) \\ &\quad \cdot \delta_j(X_2, X_4)\delta_j(X_2, X_5)\delta_j(X_3, X_4) \\ &\quad \cdot \delta_j(X_3, X_5)\delta_j(X_4, X_5)] \\ &= 1 - 1 \cdot 1 = 0. \end{aligned}$$

Hence the focal element  $X_{j_5} = A \cup B \cup C$  will be discarded in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

- For the focal element  $X_{j_6} = A \cup B \cup C \cup D = \Theta = X_5$  of  $\mathcal{X}_j$  having  $|X_{j_6}| = 4$ , we get

$$\begin{aligned} \kappa_j(\Theta) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_6}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\ &= 1 - \delta_j(X_1, X_5)\delta_j(X_2, X_5)\delta_j(X_3, X_5)\delta_j(X_4, X_5) \\ &= 1 - 1 \cdot 1 \cdot 1 \cdot 1 = 0. \end{aligned}$$

This result illustrates the validity of the aforementioned remark 1. Hence the focal element  $X_{j_5} = A \cup B \cup C \cup D = \Theta$  will be discarded in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

In summary, the conflicting product  $\pi_j(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C)m_4(B \cup C)m_5(A \cup B \cup C)m_6(\Theta)$  will be redistributed only to the three focal elements  $A$ ,  $B \cup C$  and  $A \cup C$  with the improved rules PCR5<sup>+</sup> and PCR6<sup>+</sup>, whereas it would have been redistributed to all five focal elements  $A$ ,  $B \cup C$ ,  $A \cup C$ ,  $A \cup B \cup C$  and  $\Theta$  with the classical PCR5 and PCR6 rules. Thus, two focal elements were discarded.

**Example 7:** This example is somehow an extension of example 6 by including a new element  $E$  in the FoD. So, the FoD is  $\Theta = \{A, B, C, D, E\}$ , seven BBAs, and the  $j$ -th conflicting (assumed strictly positive) product whose structure is as follows

$$\begin{aligned} \pi_j(\emptyset) &= m_1(A \cup E)m_2(B \cup C \cup E)m_3(A \cup C \cup E)m_4(B \cup C \cup E) \\ &\quad \cdot m_5(A \cup B \cup C \cup E)m_6(A \cup B \cup C \cup D \cup E)m_7(A). \end{aligned}$$

In this product  $\pi_j(\emptyset)$  we have the duplicate focal element  $B \cup C \cup E$  because it appears both in  $m_2(B \cup C \cup E)$  and in  $m_4(B \cup C \cup E)$ . The focal elements entering in each BBA of  $\pi_j(\emptyset)$  are respectively  $X_{j_1} = A \cup E$ ,  $X_{j_2} = B \cup C \cup E$ ,  $X_{j_3} = A \cup C \cup E$ ,  $X_{j_4} = B \cup C \cup E$ ,  $X_{j_5} = A \cup B \cup C \cup E$ ,  $X_{j_6} = A \cup B \cup C \cup D \cup E = \Theta$  and  $X_{j_7} = A$ . So we have to consider only the following set of distinct focal elements for this  $\pi_j(\emptyset)$  product

$$\mathcal{X}_j = \{X_1 = A \cup E, X_2 = B \cup C \cup E, X_3 = A \cup C \cup E, X_4 = A \cup B \cup C \cup E, X_5 = A \cup B \cup C \cup D \cup E, X_6 = A\}.$$

Therefore, considering only  $X_{l'} \neq X_l$  and  $|X_{l'}| \leq |X_l|$  that are conditions entering in formula (23), we have the following binary containing indicator  $\delta_j(X_{l'}, X_l)$  values:

$$\begin{aligned} \delta_j(X_6, X_1) &= 1 \text{ because } (X_6 = A) \subseteq (X_1 = A \cup E), \\ \delta_j(X_6, X_2) &= 0 \text{ because } (X_6 = A) \not\subseteq (X_2 = B \cup C \cup E), \\ \delta_j(X_6, X_3) &= 1 \text{ because } (X_6 = A) \subseteq (X_3 = A \cup C \cup E), \\ \delta_j(X_6, X_4) &= 1 \text{ because } (X_6 = A) \subseteq (X_4 = A \cup B \cup C \cup E), \\ \delta_j(X_6, X_5) &= 1 \text{ because } (X_6 = A) \subseteq (X_5 = \Theta), \\ \delta_j(X_1, X_2) &= 0 \text{ because } (X_1 = A \cup E) \not\subseteq (X_2 = B \cup C \cup E), \\ \delta_j(X_1, X_3) &= 1 \text{ because } (X_1 = A \cup E) \subseteq (X_3 = A \cup C \cup E), \\ \delta_j(X_1, X_4) &= 1 \text{ because } (X_1 = A \cup E) \subseteq (X_4 = A \cup B \cup C \cup E), \\ \delta_j(X_1, X_5) &= 1 \text{ because } (X_1 = A \cup E) \subseteq (X_5 = \Theta), \\ \delta_j(X_2, X_3) &= 0 \text{ because } (X_2 = B \cup C \cup E) \not\subseteq (X_3 = A \cup C \cup E), \\ \delta_j(X_2, X_4) &= 1 \text{ because } (X_2 = B \cup C \cup E) \subseteq (X_4 = A \cup B \cup C \cup E), \end{aligned}$$

$$\begin{aligned}
 \delta_j(X_2, X_5) &= 1 \text{ because } (X_2 = B \cup C \cup E) \subseteq (X_5 = \Theta), \\
 \delta_j(X_3, X_2) &= 0 \text{ because } (X_3 = A \cup C \cup E) \not\subseteq (X_2 = B \cup C \cup E), \\
 \delta_j(X_3, X_4) &= 1 \text{ because } (X_3 = A \cup C \cup E) \subseteq (X_4 = A \cup B \cup C \cup E), \\
 \delta_j(X_3, X_5) &= 1 \text{ because } (X_3 = A \cup C \cup E) \subseteq (X_5 = \Theta), \\
 \delta_j(X_4, X_5) &= 1 \text{ because } (X_4 = A \cup B \cup C \cup E) \subseteq (X_5 = \Theta).
 \end{aligned}$$

The binary keeping-indexes  $\kappa_j(X_{j_i})$  for  $i = 1, 2, \dots, 7$  are calculated based on the formula (23) as follows

- For the focal element  $X_{j_1} = A \cup E = X_1$  of  $\mathcal{X}_j$  having  $|X_{j_1}| = 2$ , we get

$$\begin{aligned}
 \kappa_j(X_{j_1}) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_1}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_i(X_{l'}, X_l) \\
 &= 1 - [\delta_j(X_1, X_2)\delta_j(X_1, X_3)\delta_j(X_1, X_4)\delta_j(X_1, X_5) \\
 &\quad \cdot \delta_j(X_2, X_3)\delta_j(X_2, X_4)\delta_j(X_2, X_5)\delta_j(X_3, X_2) \\
 &\quad \cdot \delta_j(X_3, X_4)\delta_j(X_3, X_5)\delta_j(X_4, X_5)\delta_j(X_6, X_1) \\
 &\quad \cdot \delta_j(X_6, X_2)\delta_j(X_6, X_3)\delta_j(X_6, X_4)\delta_j(X_6, X_5)] \\
 &= 1 - 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 = 1.
 \end{aligned}$$

Hence the focal element  $X_{j_1} = A \cup E$  will be kept in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

- For the focal element  $X_{j_2} = B \cup C \cup E = X_2$  of  $\mathcal{X}_j$  having  $|X_{j_2}| = 3$ , we get

$$\begin{aligned}
 \kappa_j(X_{j_2}) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_2}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\
 &= 1 - [\delta_j(X_1, X_2)\delta_j(X_1, X_3)\delta_j(X_1, X_4)\delta_j(X_1, X_5) \\
 &\quad \cdot \delta_j(X_2, X_3)\delta_j(X_2, X_4)\delta_j(X_2, X_5)\delta_j(X_3, X_2) \\
 &\quad \cdot \delta_j(X_3, X_4)\delta_j(X_3, X_5)\delta_j(X_4, X_5)\delta_j(X_6, X_2) \\
 &\quad \cdot \delta_j(X_6, X_3)\delta_j(X_6, X_4)\delta_j(X_6, X_5)] \\
 &= 1 - 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 \\
 &= 1.
 \end{aligned}$$

Hence the focal element  $X_{j_2} = B \cup C \cup E$  will also be kept in the proportional redistribution of the conflicting mass  $\pi_j(\emptyset)$ .

- For the focal element  $X_{j_3} = A \cup C \cup E = X_3$  of  $\mathcal{X}_j$  having  $|X_{j_3}| = 3$ , we get

$$\begin{aligned}
 \kappa_j(X_{j_3}) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_3}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\
 &= 1 - [\delta_j(X_1, X_2)\delta_j(X_1, X_3)\delta_j(X_1, X_4)\delta_j(X_1, X_5) \\
 &\quad \cdot \delta_j(X_2, X_3)\delta_j(X_2, X_4)\delta_j(X_2, X_5)\delta_j(X_3, X_2) \\
 &\quad \cdot \delta_j(X_3, X_4)\delta_j(X_3, X_5)\delta_j(X_4, X_5)\delta_j(X_6, X_2) \\
 &\quad \cdot \delta_j(X_6, X_3)\delta_j(X_6, X_4)\delta_j(X_6, X_5)] \\
 &= 1 - 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 \\
 &= 1.
 \end{aligned}$$

Hence the focal element  $X_{j_3} = A \cup C \cup E$  is also kept in the redistribution.

- For the duplicate focal element  $X_{j_4} = B \cup C \cup E$  having  $|X_{j_4}| = 3$ , we have  $\kappa_j(X_{j_4}) = 1$  because  $X_{j_4} = X_{j_2}$  and  $\kappa_j(X_{j_2}) = 1$ .

- For the focal element  $X_{j_5} = A \cup B \cup C \cup E = X_4$  having  $|X_{j_5}| = 4$ , we get

$$\begin{aligned}
 \kappa_j(X_{j_5}) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_5}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\
 &= 1 - [\delta_j(X_1, X_4)\delta_j(X_1, X_5)\delta_j(X_2, X_4)\delta_j(X_2, X_5) \\
 &\quad \cdot \delta_j(X_3, X_4)\delta_j(X_3, X_5)\delta_j(X_4, X_5)\delta_j(X_6, X_4) \\
 &\quad \cdot \delta_j(X_6, X_5)] \\
 &= 1 - 1 \cdot 1 = 0.
 \end{aligned}$$

Hence the focal element  $X_{j_5} = A \cup B \cup C \cup E$  must be ignored in the proportional redistribution.

- For  $X_{j_6} = A \cup B \cup C \cup D \cup E = \Theta = X_5$  having  $|X_{j_6}| = 5$ , we get

$$\begin{aligned}
 \kappa_j(X_{j_6}) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_6}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\
 &= 1 - [\delta_j(X_1, X_5)\delta_j(X_2, X_5)\delta_j(X_3, X_5)\delta_j(X_4, X_5) \\
 &\quad \cdot \delta_j(X_6, X_5)] \\
 &= 1 - 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 = 0.
 \end{aligned}$$

This result illustrates the validity of the aforementioned remark 1. Hence the focal element  $X_{j_6} = A \cup B \cup C \cup D \cup E$  must be ignored in the proportional redistribution.

- For the focal element  $X_{j_7} = A = X_6$  having  $|X_{j_7}| = 1$ , we get naturally (see our previous remark 1)

$$\begin{aligned}
 \kappa_j(X_{j_7}) &= 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j \\ X_{l'} \neq X_l \\ |X_{j_7}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l) \\
 &= 1 - [\delta_j(X_1, X_2)\delta_j(X_1, X_3)\delta_j(X_1, X_4)\delta_j(X_1, X_5) \\
 &\quad \cdot \delta_j(X_2, X_3)\delta_j(X_2, X_4)\delta_j(X_2, X_5)\delta_j(X_3, X_2) \\
 &\quad \cdot \delta_j(X_3, X_4)\delta_j(X_3, X_5)\delta_j(X_4, X_5)\delta_j(X_6, X_2) \\
 &\quad \cdot \delta_j(X_6, X_3)\delta_j(X_6, X_4)\delta_j(X_6, X_5)] \\
 &= 1 - 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 \cdot 0 \cdot 1 \cdot 1 \cdot 1 \\
 &= 1.
 \end{aligned}$$

Hence the focal element  $X_{j_7} = A$  must be kept in the proportional redistribution.

In summary, the conflicting product  $\pi_j(\emptyset) = m_1(A \cup E)m_2(B \cup C \cup E)m_3(A \cup C \cup E)m_4(B \cup C \cup E)m_5(A \cup B \cup C \cup E)m_6(\Theta)m_7(A)$  will be redistributed only to focal elements  $A \cup E$ ,  $B \cup C \cup E$ ,  $A \cup C \cup E$  and  $A$  with the improved rules PCR5<sup>+</sup> and PCR6<sup>+</sup>, whereas it would have been redistributed to all focal elements  $A \cup E$ ,  $B \cup C \cup E$ ,  $A \cup C \cup E$ ,  $A \cup B \cup C \cup E$ ,  $\Theta$  and  $A$  with the classical PCR5 and PCR6 rules.

**Example 8:** This is a somehow simplified version of example 6. We consider the FoD  $\Theta = \{A, B, C, D\}$ , only five BBAs, and suppose that the  $j$ -th conflicting (assumed strictly positive) product is as follows

$$\pi_j(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C)m_4(B \cup C) \cdot m_5(A \cup B \cup C \cup D).$$

Based on (23), it can be verified<sup>14</sup> that the binary keeping-indexes of focal elements involved in conflicting products are

$$\begin{aligned} \kappa_j(A) &= 1, \\ \kappa_j(B \cup C) &= 1, \\ \kappa_j(A \cup C) &= 1, \\ \kappa_j(A \cup B \cup C \cup D) &= 0. \end{aligned}$$

**Example 9:** We consider the FoD  $\Theta = \{A, B, C, D\}$ , seven BBAs, and suppose that the  $j$ -th conflicting (assumed strictly positive) product is as follows

$$\pi_j(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C)m_4(B \cup C) \cdot m_5(A \cup B \cup C \cup D)m_6(A \cup B \cup C)m_7(A \cup B \cup C).$$

Based on (23), it can be verified that the binary keeping-indexes of focal elements involved in conflicting products are

$$\begin{aligned} \kappa_j(A) &= 1, \\ \kappa_j(B \cup C) &= 1, \\ \kappa_j(A \cup C) &= 1, \\ \kappa_j(A \cup B \cup C \cup D) &= 0, \\ \kappa_j(A \cup B \cup C) &= 0. \end{aligned}$$

**Example 10:** We consider the FoD  $\Theta = \{A, B, C\}$ , three BBAs, and suppose that the  $j$ -th conflicting (assumed strictly positive) product is as follows

$$\pi_j(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C).$$

Based on (23), it can be verified that the binary keeping-indexes of focal elements involved in conflicting products are

$$\begin{aligned} \kappa_j(A) &= 1, \\ \kappa_j(B \cup C) &= 1, \\ \kappa_j(A \cup C) &= 1. \end{aligned}$$

**Example 11:** We consider the FoD  $\Theta = \{A, B, C\}$ , four BBAs, and suppose that the  $j$ -th conflicting (assumed strictly positive) product is as follows

$$\pi_j(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C)m_4(A \cup B).$$

Based on (23), it can be verified that the binary keeping-indexes of focal elements involved in conflicting products are

$$\begin{aligned} \kappa_j(A) &= 1, \\ \kappa_j(B \cup C) &= 1, \\ \kappa_j(A \cup C) &= 1, \\ \kappa_j(A \cup B) &= 1. \end{aligned}$$

**Example 12:** We consider the FoD  $\Theta = \{A, B, C\}$ , three BBAs, and suppose that the  $j$ -th conflicting (assumed strictly positive) product is as follows

$$\pi_j(\emptyset) = m_1(A \cup B \cup C)m_2(A)m_3(B \cup C).$$

Based on (23), it can be verified that the binary keeping-indexes of focal elements involved in conflicting products are

$$\begin{aligned} \kappa_j(A \cup B \cup C) &= 0, \\ \kappa_j(A) &= 1, \\ \kappa_j(B \cup C) &= 1. \end{aligned}$$

**Example 13:** We consider the FoD  $\Theta = \{A, B, C, D\}$ , and the three following BBAs

$$\begin{aligned} m_1(A \cup B) &= 0.8, m_1(C \cup D) = 0.2, \\ m_2(A \cup B) &= 0.4, m_2(C \cup D) = 0.6, \\ m_3(B) &= 0.1, m_3(A \cup B \cup C \cup D) = 0.9. \end{aligned}$$

We have  $\mathcal{F} = |\mathcal{F}(m_1)| \cdot |\mathcal{F}(m_2)| \cdot |\mathcal{F}(m_3)| = 2 \cdot 2 \cdot 2 = 8$  products  $\pi_j$  ( $j = 1, \dots, \mathcal{F}$ ) entering in the fusion process as follows

$$\begin{aligned} \pi_1(B) &= m_1(A \cup B)m_2(A \cup B)m_3(B) = 0.032, \\ \pi_2(A \cup B) &= m_1(A \cup B)m_2(A \cup B)m_3(\Theta) = 0.288, \\ \pi_3(\emptyset) &= m_1(A \cup B)m_2(C \cup D)m_3(B) = 0.048, \\ \pi_4(\emptyset) &= m_1(A \cup B)m_2(C \cup D)m_3(\Theta) = 0.432, \\ \pi_5(\emptyset) &= m_1(C \cup D)m_2(A \cup B)m_3(B) = 0.008, \\ \pi_6(\emptyset) &= m_1(C \cup D)m_2(A \cup B)m_3(\Theta) = 0.072, \\ \pi_7(\emptyset) &= m_1(C \cup D)m_2(C \cup D)m_3(B) = 0.012, \\ \pi_8(C \cup D) &= m_1(C \cup D)m_2(C \cup D)m_3(\Theta) = 0.108. \end{aligned}$$

Based on (23), it can be verified<sup>15</sup> that the binary keeping-indexes of focal elements involved in conflicting products  $\pi_3(\emptyset)$  to  $\pi_7(\emptyset)$  are

$$\begin{aligned} \kappa_3(A \cup B) &= 1, \kappa_3(C \cup D) = 1, \kappa_3(B) = 1, \\ \kappa_4(A \cup B) &= 1, \kappa_4(C \cup D) = 1, \kappa_4(\Theta) = 0, \\ \kappa_5(C \cup D) &= 1, \kappa_5(A \cup B) = 1, \kappa_5(B) = 1, \\ \kappa_6(C \cup D) &= 1, \kappa_6(A \cup B) = 1, \kappa_6(\Theta) = 0, \\ \kappa_7(C \cup D) &= 1, \kappa_7(B) = 1. \end{aligned}$$

In summary, once the binary keeping-index of  $\kappa_j(X_{j_i})$  of all focal elements  $X_{j_i}$  involved in a conflicting product  $\pi_j(\emptyset)$  are calculated, we can apply PCR5, or PCR6 redistribution principle only with the focal elements for which  $\kappa_j(X_{j_i}) = 1$ .

<sup>14</sup>The verification is left to the reader.

<sup>15</sup>The verification is left to the reader.

With this new improved method of proportional redistribution PCR5<sup>+</sup> and PCR6<sup>+</sup> rules will never increase the mass of non conflicting elements involved in each  $\pi_j(\emptyset)$  (if any), and in doing this way we will preserve the neutrality of the vacuous belief assignment in the PCR5<sup>+</sup> and PCR6<sup>+</sup> fusion rules, which is a very desirable behavior.

### B. Expressions of PCR5<sup>+</sup> and PCR6<sup>+</sup> fusion rules

The expressions of PCR5<sup>+</sup> and PCR6<sup>+</sup> fusion rules are proper modifications of PCR5 and PCR6 formulas (14) and (15) taking into account the selection of focal elements on which the proportional redistribution must apply thanks to the value of their binary keeping-index.

The PCR5<sup>+</sup> fusion of  $S > 2$  BBAs is obtained by  $m_{1,2,\dots,S}^{\text{PCR5}^+}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{1,2,\dots,S}^{\text{PCR5}^+}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1,\dots,S\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \kappa_j(A) \prod_{i \in \{1,\dots,S\} | X_{j_i}=A} m_i(X_{j_i}) \right) \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \kappa_j(X) \prod_{i \in \{1,\dots,S\} | X_{j_i}=X} m_i(X_{j_i}) \right)} \right]. \quad (25)$$

The PCR6<sup>+</sup> fusion of  $S > 2$  BBAs is obtained by  $m_{1,2,\dots,S}^{\text{PCR6}^+}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{1,2,\dots,S}^{\text{PCR6}^+}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1,\dots,S\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \kappa_j(A) \sum_{i \in \{1,\dots,S\} | X_{j_i}=A} m_i(X_{j_i}) \right) \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \kappa_j(X) \sum_{i \in \{1,\dots,S\} | X_{j_i}=X} m_i(X_{j_i}) \right)} \right], \quad (26)$$

where  $\kappa_j(A)$  and  $\kappa_j(X)$  are respectively the binary keeping-indexes of elements  $A$  and  $X$  involved in the conflicting product  $\pi_j(\emptyset)$ , that are calculated by the formula (23) or (24).

**Remark 4:** It is worth mentioning that PCR5<sup>+</sup> formula (25) is totally consistent with PCR5 formula (14) when all binary keeping-indexes are equal to one. Similarly, the PCR6<sup>+</sup> formula (26) reduces to PCR6 formula (15) if all binary keeping-indexes equal one.

**Theorem:** The vacuous BBA  $m_v$  has a neutral impact in PCR5<sup>+</sup> and PCR6<sup>+</sup> rules of combination.

**Proof:** see appendix 2.

### C. On the complexity of PCR5<sup>+</sup> and PCR6<sup>+</sup> fusion rules

The complexity of PCR5 and PCR6 rules is difficult to establish precisely because the number of computations highly depends on the structure of focal elements of the BBAs to combine, but definitely it is higher than Dempster's rule of combination. What about the complexity of PCR5<sup>+</sup> and PCR6<sup>+</sup> fusion rules? On the one hand, PCR5<sup>+</sup> and PCR6<sup>+</sup> seem more complex than PCR5 and PCR6 rules because one needs extra computational burden with respect to PCR5 and

PCR6 rules to calculate the binary keeping-indexes. But in fact, the calculation of binary keeping-indexes do not depend on the mass values of focal elements but only on their structure. Hence, the binary keeping-indexes can be calculated off-line once for all for many possible structures of focal elements of BBAs to combine. On the other hand, if the binary keeping-index calculation is done off-line, then PCR5<sup>+</sup> and PCR6<sup>+</sup> become less complex than PCR5 and PCR6 rule because some elements are discarded with PCR5<sup>+</sup> and PCR6<sup>+</sup> making the redistribution simpler and more effective than with PCR5 and PCR6 rules. It is not possible to say for sure if globally PCR5<sup>+</sup> and PCR6<sup>+</sup> are more (or less) complex than PCR5 and PCR6 because it really depends on the fusion problem under consideration and the structure of focal elements of BBAs to combine. If the sources of evidence to combine generate many partial conflicts to redistribute including many elements to discard, then PCR5<sup>+</sup> and PCR6<sup>+</sup> are more advantageous than PCR5 and PCR6 in terms of reduction of complexity.

## VII. EXAMPLES FOR PCR5<sup>+</sup> AND PCR6<sup>+</sup> FUSION RULES

Here we compare the results obtained with PCR5<sup>+</sup> and PCR6<sup>+</sup> with respect to those drawn from PCR5 and PCR6 rules on the examples from 1 to 13 in the previous sections. Since these following examples, for PCR5<sup>+</sup> and PCR6<sup>+</sup> fusion rules, respectively consider the same FoD and BBAs as those presented, they will be denoted as "revisited examples".

**Example 1 (revisited):** Consider  $\Theta = \{A, B\}$  and two following BBAs

$$m_1(A) = 0.1, \quad m_1(B) = 0.2, \quad m_1(A \cup B) = 0.7, \\ m_2(A) = 0.4, \quad m_2(B) = 0.3, \quad m_2(A \cup B) = 0.3.$$

Because there is only two BBAs to combine, we have

$$\text{PCR5}(m_1, m_2) = \text{PCR6}(m_1, m_2) \\ \text{PCR5}^+(m_1, m_2) = \text{PCR6}^+(m_1, m_2).$$

We have  $m_{1,2}^{\text{Conj}}(A) = 0.35$ ,  $m_{1,2}^{\text{Conj}}(B) = 0.33$ , and  $m_{1,2}^{\text{Conj}}(\Theta) = 0.21$ , and we have the two conflicting products  $\pi_1(\emptyset) = m_1(A)m_2(B) = 0.03$  and  $\pi_2(\emptyset) = m_2(A)m_1(B) = 0.08$  to redistribute.

Applying PCR5 principle for  $\pi_1(\emptyset) = 0.03$  we get

$$\frac{x_1(A)}{m_1(A)} = \frac{x_1(B)}{m_2(B)} = \frac{\pi_1(\emptyset)}{m_1(A) + m_2(B)},$$

whence  $x_1(A) = 0.1 \cdot \frac{0.03}{0.1+0.3} = 0.0075$  and  $x_1(B) = 0.3 \cdot \frac{0.03}{0.1+0.3} = 0.0225$ .

Applying PCR5 principle for  $\pi_2(\emptyset) = 0.08$  we get

$$\frac{x_2(A)}{m_2(A)} = \frac{x_2(B)}{m_1(B)} = \frac{\pi_2(\emptyset)}{m_2(A) + m_1(B)},$$

whence  $x_2(A) = 0.4 \cdot \frac{0.08}{0.4+0.2} \approx 0.0533$  and  $x_2(B) = 0.2 \cdot \frac{0.08}{0.4+0.2} \approx 0.0267$ .

Therefore we get

$$\begin{aligned} m_{1,2}^{\text{PCR5}}(A) &= m_{1,2}^{\text{PCR6}}(A) = m_{1,2}^{\text{Conj}}(A) + x_1(A) + x_2(A) \\ &= 0.35 + 0.0075 + 0.0533 = 0.4108, \\ m_{1,2}^{\text{PCR5}}(B) &= m_{1,2}^{\text{PCR6}}(B) = m_{1,2}^{\text{Conj}}(B) + x_1(B) + x_2(B) \\ &= 0.33 + 0.0225 + 0.0267 = 0.3792, \\ m_{1,2}^{\text{PCR5}}(A \cup B) &= m_{1,2}^{\text{PCR6}}(A \cup B) = m_{1,2}^{\text{Conj}}(A \cup B) = 0.21. \end{aligned}$$

If we want to apply  $\text{PCR5}^+$ , or  $\text{PCR6}^+$ , rule we need to compute the binary keeping-indexes of each focal element entering in the conflicting products  $\pi_1(\emptyset)$  and  $\pi_2(\emptyset)$ . In this example for  $\pi_1(\emptyset) = m_1(A)m_2(B)$  we have  $\mathcal{X}_1 = \{A, B\}$ , and for  $\pi_2(\emptyset) = m_2(A)m_1(B)$  we have  $\mathcal{X}_2 = \{A, B\}$ . Applying formula (22), we get  $\delta_1(A, B) = 0$  because  $A \not\subseteq B$ , and  $\delta_1(B, A) = 0$  because  $B \not\subseteq A$  (and also  $\delta_2(A, B) = 0$  and  $\delta_2(B, A) = 0$ ). Applying formula (23) we get the binary keeping-indexes  $\kappa_1(A) = 1$ ,  $\kappa_1(B) = 1$ ,  $\kappa_2(A) = 1$ , and  $\kappa_2(B) = 1$  indicating that the redistribution of  $\pi_1(\emptyset)$  must operate on all elements of  $\mathcal{X}_1 = \{A, B\}$ , and the redistribution of  $\pi_2(\emptyset)$  must also operate on all elements of  $\mathcal{X}_2 = \{A, B\}$ , so there is no element that must be discarded for making the improved redistribution in this example. Therefore  $\text{PCR5}^+$ , or  $\text{PCR6}^+$  results coincide with  $\text{PCR5}$  and  $\text{PCR6}$  results, that is  $m^{\text{PCR5}}(\cdot) = m^{\text{PCR6}}(\cdot) = m^{\text{PCR5}^+}(\cdot) = m^{\text{PCR6}^+}(\cdot)$  which is normal.

**Example 2** (revisited): Consider  $\Theta = \{A, B\}$  and the three following BBAs

$$\begin{aligned} m_1(A) &= 0.6, m_1(B) = 0.1, m_1(A \cup B) = 0.3, \\ m_2(A) &= 0.5, m_2(B) = 0.3, m_2(A \cup B) = 0.2, \\ m_3(A) &= 0.4, m_3(B) = 0.1, m_3(A \cup B) = 0.5. \end{aligned}$$

As shown in Section IV, for this example one has the following twelve conflicting products to redistribute when applying  $\text{PCR5}$ , or  $\text{PCR6}$  fusion formulas.

$$\begin{aligned} \pi_1(\emptyset) &= m_1(A)m_2(A)m_3(B) = 0.0300, \\ \pi_2(\emptyset) &= m_1(A)m_2(B)m_3(A) = 0.0720, \\ \pi_3(\emptyset) &= m_1(B)m_2(A)m_3(A) = 0.0200, \\ \pi_4(\emptyset) &= m_1(B)m_2(B)m_3(A) = 0.0120, \\ \pi_5(\emptyset) &= m_1(B)m_2(A)m_3(B) = 0.0050, \\ \pi_6(\emptyset) &= m_1(A)m_2(B)m_3(B) = 0.0180, \\ \pi_7(\emptyset) &= m_1(A \cup B)m_2(A)m_3(B) = 0.0150, \\ \pi_8(\emptyset) &= m_1(A \cup B)m_2(B)m_3(A) = 0.0360, \\ \pi_9(\emptyset) &= m_1(B)m_2(A)m_3(A \cup B) = 0.0250, \\ \pi_{10}(\emptyset) &= m_1(A)m_2(B)m_3(A \cup B) = 0.0900, \\ \pi_{11}(\emptyset) &= m_1(A)m_2(A \cup B)m_3(B) = 0.0120, \\ \pi_{12}(\emptyset) &= m_1(B)m_2(A \cup B)m_3(A) = 0.0080. \end{aligned}$$

With  $\text{PCR5}$  and  $\text{PCR6}$  the products  $\pi_1(\emptyset)$  to  $\pi_6(\emptyset)$  are redistributed to  $A$  and  $B$  only, whereas the products  $\pi_7(\emptyset)$  to  $\pi_{12}(\emptyset)$  are redistributed to  $A$ ,  $B$  and  $A \cup B$ . Applying  $\text{PCR5}$  formula (14), and  $\text{PCR6}$  formula (15) we obtain  $m_{1,2,3}^{\text{PCR5}}(\emptyset) = m_{1,2,3}^{\text{PCR6}}(\emptyset) = 0$  and

$$\begin{cases} m_{1,2,3}^{\text{PCR5}}(A) \approx 0.723281, \\ m_{1,2,3}^{\text{PCR5}}(B) \approx 0.182460, \\ m_{1,2,3}^{\text{PCR5}}(A \cup B) \approx 0.094259, \end{cases} \quad \text{and} \quad \begin{cases} m_{1,2,3}^{\text{PCR6}}(A) \approx 0.743496, \\ m_{1,2,3}^{\text{PCR6}}(B) \approx 0.162245, \\ m_{1,2,3}^{\text{PCR6}}(A \cup B) \approx 0.094259. \end{cases}$$

The calculation of the binary keeping-indexes by the formula (23) gives in this example

$$\begin{cases} \kappa_j(A) = 1, \kappa_j(B) = 1, & \text{for } j = 1, \dots, 6 \\ \kappa_j(A) = 1, \kappa_j(B) = 1, \kappa_j(A \cup B) = 0, & \text{for } j = 7, \dots, 12 \end{cases}$$

Therefore, if we apply the  $\text{PCR5}^+$  and  $\text{PCR6}^+$  improved rules of combination, we redistribute the products  $\pi_1(\emptyset)$  to  $\pi_6(\emptyset)$  to  $A$  and  $B$  (as for  $\text{PCR5}$  and  $\text{PCR6}$  rule), but the products  $\pi_7(\emptyset)$  to  $\pi_{12}(\emptyset)$  will be redistributed to  $A$ ,  $B$  only, and not to  $A \cup B$  because  $\kappa_j(A \cup B) = 0$  for  $j = 7, \dots, 12$ . So finally, we obtain  $m_{1,2,3}^{\text{PCR5}^+}(\emptyset) = m_{1,2,3}^{\text{PCR6}^+}(\emptyset) = 0$  and

$$\begin{cases} m_{1,2,3}^{\text{PCR5}^+}(A) \approx 0.768631, \\ m_{1,2,3}^{\text{PCR5}^+}(B) \approx 0.201369, \\ m_{1,2,3}^{\text{PCR5}^+}(A \cup B) = 0.03, \end{cases} \quad \text{and} \quad \begin{cases} m_{1,2,3}^{\text{PCR6}^+}(A) \approx 0.788847, \\ m_{1,2,3}^{\text{PCR6}^+}(B) \approx 0.181153, \\ m_{1,2,3}^{\text{PCR6}^+}(A \cup B) = 0.03. \end{cases}$$

We can verify that we obtain a more precise redistribution with  $\text{PCR5}^+$  (resp.  $\text{PCR6}^+$ ) rule with respect to  $\text{PCR5}$  (resp.  $\text{PCR6}$ ) rule because  $m_{1,2,3}^{\text{PCR5}^+}(A \cup B) < m_{1,2,3}^{\text{PCR5}}(A \cup B)$  and also  $m_{1,2,3}^{\text{PCR6}^+}(A \cup B) < m_{1,2,3}^{\text{PCR6}}(A \cup B)$ .

**Example 3** (revisited): we consider  $\Theta = \{A, B, C\}$ , and the four very simple BBAs defined by

$$m_1(A \cup B) = 1, m_2(B) = 1, m_3(A \cup B) = 1, \text{ and } m_4(C) = 1$$

These four basic belief assignments are in total conflict because  $(A \cup B) \cap A \cap (A \cup B) \cap C = \emptyset$ , and one has only one product  $\pi(\emptyset) = m_1(A \cup B)m_2(A)m_3(A \cup B)m_4(C) = 1$  to consider, so  $j = 1$  in this case and it can be omitted in the notations of the binary keeping-indexes.

As shown previously, one has

$$\begin{cases} m_{1,2,3,4}^{\text{PCR5}}(A \cup B) = 1/3, \\ m_{1,2,3,4}^{\text{PCR5}}(B) = 1/3, \\ m_{1,2,3,4}^{\text{PCR5}}(C) = 1/3, \end{cases} \quad \text{and} \quad \begin{cases} m_{1,2,3,4}^{\text{PCR6}}(A \cup B) = 0.5, \\ m_{1,2,3,4}^{\text{PCR6}}(B) = 0.25, \\ m_{1,2,3,4}^{\text{PCR6}}(C) = 0.25. \end{cases}$$

Because all focal elements  $A \cup B$ ,  $A$  and  $C$  entering in  $\pi(\emptyset)$  are conflicting then one has the binary keeping-indexes  $\kappa(A \cup B) = 1$ ,  $\kappa(A) = 1$  and  $\kappa(C) = 1$  i.e. all these elements will receive a redistribution of the conflicting mass  $\pi(\emptyset)$ . Therefore there is no restriction for making the redistribution. Consequently,  $\text{PCR5}^+$  result coincides with  $\text{PCR5}$  result, and  $\text{PCR6}^+$  result coincides with  $\text{PCR6}$  result.

**Example 4** (revisited): we consider  $\Theta = \{A, B\}$ , and the following four BBAs

$$\begin{aligned} m_1(A) &= 0.6, m_1(B) = 0.1, m_1(A \cup B) = 0.3, \\ m_2(A) &= 0.5, m_2(B) = 0.3, m_2(A \cup B) = 0.2, \\ m_3(A) &= 0.4, m_3(B) = 0.1, m_3(A \cup B) = 0.5, \\ m_4(A \cup B) &= 1 \quad (m_4 \text{ is the vacuous BBA}). \end{aligned}$$

The BBAs  $m_1$ ,  $m_2$  and  $m_3$  are the same as in Example 2, and the BBA  $m_4$  is the vacuous BBA. We have already shown that  $\text{PCR5}(m_1, m_2, m_3) \neq \text{PCR5}(m_1, m_2, m_3, m_4)$  even if  $m_4$  is the vacuous BBA, and

$$\begin{cases} m_{1,2,3,4}^{\text{PCR5}}(A) \approx 0.654604, \\ m_{1,2,3,4}^{\text{PCR5}}(B) \approx 0.144825, \\ m_{1,2,3,4}^{\text{PCR5}}(A \cup B) \approx 0.200571. \end{cases}$$

Similarly,  $\text{PCR6}(m_1, m_2, m_3) \neq \text{PCR6}(m_1, m_2, m_3, m_4)$ , and

$$\begin{cases} m_{1,2,3,4}^{\text{PCR6}}(A) \approx 0.647113, \\ m_{1,2,3,4}^{\text{PCR6}}(B) \approx 0.128342, \\ m_{1,2,3,4}^{\text{PCR6}}(A \cup B) \approx 0.224545. \end{cases}$$

Applying the  $\text{PCR5}^+$  formula (25), and the  $\text{PCR6}^+$  formula (26) we will obtain  $m_{1,2,3}^{\text{PCR5}^+}(\emptyset) = m_{1,2,3,4}^{\text{PCR6}^+}(\emptyset) = 0$  and

$$\begin{cases} m_{1,2,3,4}^{\text{PCR5}^+}(A) \approx 0.768631, \\ m_{1,2,3,4}^{\text{PCR5}^+}(B) \approx 0.201369, \\ m_{1,2,3,4}^{\text{PCR5}^+}(A \cup B) = 0.03, \end{cases} \quad \text{and} \quad \begin{cases} m_{1,2,3,4}^{\text{PCR6}^+}(A) \approx 0.788847, \\ m_{1,2,3,4}^{\text{PCR6}^+}(B) \approx 0.181153, \\ m_{1,2,3,4}^{\text{PCR6}^+}(A \cup B) = 0.03. \end{cases}$$

One has  $\text{PCR5}^+(m_1, m_2, m_3, m_4) = \text{PCR5}^+(m_1, m_2, m_3)$  and also  $\text{PCR6}^+(m_1, m_2, m_3, m_4) = \text{PCR6}^+(m_1, m_2, m_3)$  because with the improved proportional redistribution of  $\text{PCR5}^+$  and  $\text{PCR6}^+$  rules the vacuous BBA has always a neutral impact in the fusion result, which is what we intuitively expect.

**Example 5** (revisited): we consider  $\Theta = \{A, B, C, D, E\}$ , and the following three BBAs

$$\begin{cases} m_1(A \cup B) = 0.70, \\ m_1(C \cup D) = 0.06, \\ m_1(A \cup B \cup C \cup D) = 0.15, \\ m_1(E) = 0.09, \end{cases}$$

and

$$\begin{cases} m_2(A \cup B) = 0.06, \\ m_2(C \cup D) = 0.50, \\ m_2(A \cup B \cup C \cup D) = 0.04, \\ m_2(E) = 0.40, \end{cases}$$

and

$$\begin{cases} m_3(B) = 0.01, \\ m_3(A \cup B \cup C \cup D \cup E) = 0.99. \end{cases}$$

Note that the BBA  $m_3$  is not equal to the vacuous BBA but it is very close to the vacuous BBA because  $m_3(\Theta)$  is close to one.

If we consider the fusion of only the two first BBAs  $m_1$  and  $m_2$ , we have  $\text{PCR6}(m_1, m_2) = \text{PCR6}^+(m_1, m_2) = \text{PCR5}(m_1, m_2) = \text{PCR5}^+(m_1, m_2)$  because all these rules coincide when combining two BBAs.

$$\begin{cases} m_{1,2}^{\text{PCR6}}(A \cup B) \approx 0.465309, \\ m_{1,2}^{\text{PCR6}}(C \cup D) \approx 0.296299, \\ m_{1,2}^{\text{PCR6}}(A \cup B \cup C \cup D) \approx 0.023471, \\ m_{1,2}^{\text{PCR6}}(E) \approx 0.214921. \end{cases}$$

If we make the  $\text{PCR5}$ ,  $\text{PCR5}^+$ ,  $\text{PCR6}$  and  $\text{PCR6}^+$  fusion of these three BBAs altogether we obtain now different results which is normal, because for  $S > 2$  one has  $\text{PCR5}^+(m_1, \dots, m_S) \neq \text{PCR5}(m_1, \dots, m_S)$  and  $\text{PCR6}^+(m_1, \dots, m_S) \neq \text{PCR6}(m_1, \dots, m_S)$  in general. So, in this example 5 we get results shown in Tables I and II.

Focal Elements	$m_{1,2,3}^{\text{PCR5}}(\cdot)$	$m_{1,2,3}^{\text{PCR5}^+}(\cdot)$
$B$	0.001103	0.001107
$A \cup B$	0.286107	0.464483
$C \cup D$	0.203385	0.296186
$A \cup B \cup C \cup D$	0.012203	0.023408
$E$	0.115966	0.214816
$A \cup B \cup C \cup D \cup E$	0.381236	0

Table I  
EXAMPLE 5: RESULTS OF  $\text{PCR5}^+$  VERSUS  $\text{PCR5}$ .

Focal Elements	$m_{1,2,3}^{\text{PCR6}}(\cdot)$	$m_{1,2,3}^{\text{PCR6}^+}(\cdot)$
$B$	0.000962	0.000967
$A \cup B$	0.286107	0.464483
$C \cup D$	0.203454	0.296255
$A \cup B \cup C \cup D$	0.012203	0.023408
$E$	0.116038	0.214887
$A \cup B \cup C \cup D \cup E$	0.381236	0

Table II  
EXAMPLE 5: RESULTS OF  $\text{PCR6}^+$  VERSUS  $\text{PCR6}$ .

These values highlight the great ignorance of the results proposed by  $\text{PCR5}$  and  $\text{PCR6}$  when the third (almost vacuous) source of information is taken into account. Indeed,  $m_{1,2,3}^{\text{PCR5}}(\Theta) = m_{1,2,3}^{\text{PCR6}}(\Theta)$  is the greatest mass among the set of hypotheses, whereas the results proposed with  $\text{PCR5}^+$  and  $\text{PCR6}^+$  combination rules discard the ignorant information and propose results closer to those obtained by merging two sources. Indeed, the largest mass is allocated to  $A \cup B$ .

The next examples 6 to 12 are very simple examples involving only categorical BBAs so that only one conflicting product (equals to one) needs to be redistributed based on  $\text{PCR5}$ ,  $\text{PCR6}$ ,  $\text{PCR5}^+$  and  $\text{PCR6}^+$  rules. These examples offer the possibility to the reader to do the derivations manually for making a verification of our results.

**Example 6** (revisited): we consider  $\Theta = \{A, B, C, D\}$ , and the following categorical BBAs  $m_1(A) = 1$ ,  $m_2(B \cup C) = 1$ ,  $m_3(A \cup C) = 1$ ,  $m_4(B \cup C) = 1$ ,  $m_5(A \cup B \cup C) = 1$  and  $m_6(A \cup B \cup C \cup D) = 1$ . If we make the  $\text{PCR5}$ ,  $\text{PCR5}^+$ ,  $\text{PCR6}$  and  $\text{PCR6}^+$  fusion of these six BBAs altogether we obtain results given in Tables III and IV.

In this example, we have only one conflicting product  $\pi_1(\emptyset)$  to redistribute which is given by

$$\pi_1(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C)m_4(B \cup C) \cdot m_5(A \cup B \cup C)m_6(A \cup B \cup C \cup D).$$

Because  $\kappa_1(A \cup B \cup C) = 0$  and  $\kappa_1(A \cup B \cup C \cup D) = 0$ , these two disjunctions are discarded and more mass is committed

Focal Elements	$m_{1,2,3,4,5,6}^{\text{PCR5}}(\cdot)$	$m_{1,2,3,4,5,6}^{\text{PCR5}^+}(\cdot)$
$A$	1/5	1/3
$A \cup C$	1/5	1/3
$B \cup C$	1/5	1/3
$A \cup B \cup C$	1/5	0
$A \cup B \cup C \cup D$	1/5	0

Table III  
EXAMPLE 6: RESULTS OF PCR5<sup>+</sup> VERSUS PCR5.

Focal Elements	$m_{1,2,3,4,5,6}^{\text{PCR6}}(\cdot)$	$m_{1,2,3,4,5,6}^{\text{PCR6}^+}(\cdot)$
$A$	1/6	1/4
$A \cup C$	1/6	1/4
$B \cup C$	1/3	1/2
$A \cup B \cup C$	1/6	0
$A \cup B \cup C \cup D$	1/6	0

Table IV  
EXAMPLE 6: RESULTS OF PCR6<sup>+</sup> VERSUS PCR6.

to  $A$ ,  $A \cup C$  and  $B \cup C$  with PCR5<sup>+</sup> and PCR6<sup>+</sup> rules. There is more mass allocated to  $B \cup C$  with PCR6<sup>+</sup> and PCR6 than with PCR5<sup>+</sup> and PCR5 because two sources of information support this hypothesis.

**Example 7 (revisited):** we consider  $\Theta = \{A, B, C, D, E\}$ , and the following seven categorical BBAs  $m_1(A \cup E) = 1$ ,  $m_2(B \cup C \cup E) = 1$ ,  $m_3(A \cup C \cup E) = 1$ ,  $m_4(B \cup C \cup E) = 1$ ,  $m_5(A \cup B \cup C \cup E) = 1$ ,  $m_6(A \cup B \cup C \cup D \cup E) = 1$ , and  $m_7(A) = 1$ . If we make the PCR5, PCR5<sup>+</sup>, PCR6 and PCR6<sup>+</sup> fusion of these seven BBAs altogether we obtain results given in Tables V and VI.

Focal Elements	$m_{1,2,3,4,5,6,7}^{\text{PCR5}}(\cdot)$	$m_{1,2,3,4,5,6,7}^{\text{PCR5}^+}(\cdot)$
$A$	1/6	1/4
$A \cup E$	1/6	1/4
$A \cup C \cup E$	1/6	1/4
$B \cup C \cup E$	1/6	1/4
$A \cup B \cup C \cup E$	1/6	0
$A \cup B \cup C \cup D \cup E$	1/6	0

Table V  
EXAMPLE 7: RESULTS OF PCR5<sup>+</sup> VERSUS PCR5.

Focal Elements	$m_{1,2,3,4,5,6,7}^{\text{PCR6}}(\cdot)$	$m_{1,2,3,4,5,6,7}^{\text{PCR6}^+}(\cdot)$
$A$	1/7	1/5
$A \cup E$	1/7	1/5
$A \cup C \cup E$	1/7	1/5
$B \cup C \cup E$	2/7	2/5
$A \cup B \cup C \cup E$	1/7	0
$A \cup B \cup C \cup D \cup E$	1/7	0

Table VI  
EXAMPLE 7: RESULTS OF PCR6<sup>+</sup> VERSUS PCR6.

In this example 7, we have only one conflicting product  $\pi_1(\emptyset)$  to redistribute which is given by

$$\begin{aligned} \pi_1(\emptyset) = & m_1(A \cup E)m_2(B \cup C \cup E)m_3(A \cup C \cup E) \\ & \cdot m_4(B \cup C \cup E)m_5(A \cup B \cup C \cup E) \\ & \cdot m_6(A \cup B \cup C \cup D \cup E)m_7(A). \end{aligned}$$

Because  $\kappa_1(A \cup B \cup C \cup E) = 0$  and  $\kappa_1(A \cup B \cup C \cup D \cup E) = 0$ , these two disjunctions are discarded and more mass is committed to  $A$ ,  $A \cup E$ ,  $A \cup C \cup E$  and  $B \cup C \cup E$  with PCR5<sup>+</sup> and PCR6<sup>+</sup> rules. There is more mass allocated to  $B \cup C \cup E$  with PCR6<sup>+</sup> and PCR6 than with PCR5<sup>+</sup> and PCR5 because two sources of information support this hypothesis.

**Example 8 (revisited):** we consider  $\Theta = \{A, B, C, D\}$ , and the following categorical BBAs  $m_1(A) = 1$ ,  $m_2(B \cup C) = 1$ ,  $m_3(A \cup C) = 1$ ,  $m_4(B \cup C) = 1$  and  $m_5(A \cup B \cup C \cup D) = 1$ . If we make the PCR5, PCR5<sup>+</sup>, PCR6 and PCR6<sup>+</sup> fusion of these seven BBAs altogether we obtain results given in Tables VII and VIII.

Focal Elements	$m_{1,2,3,4,5}^{\text{PCR5}}(\cdot)$	$m_{1,2,3,4,5}^{\text{PCR5}^+}(\cdot)$
$A$	1/4	1/3
$A \cup C$	1/4	1/3
$B \cup C$	1/4	1/3
$A \cup B \cup C \cup D$	1/4	0

Table VII  
EXAMPLE 8: RESULTS OF PCR5<sup>+</sup> VERSUS PCR5.

Focal Elements	$m_{1,2,3,4,5}^{\text{PCR6}}(\cdot)$	$m_{1,2,3,4,5}^{\text{PCR6}^+}(\cdot)$
$A$	1/5	1/4
$A \cup C$	1/5	1/4
$B \cup C$	2/5	1/2
$A \cup B \cup C \cup D$	1/5	0

Table VIII  
EXAMPLE 8: RESULTS OF PCR6<sup>+</sup> VERSUS PCR6.

Because  $\kappa_1(A \cup B \cup C \cup D) = 0$ , this disjunction is discarded and more mass is committed to  $A$ ,  $A \cup C$  and  $B \cup C$  with PCR5<sup>+</sup> and PCR6<sup>+</sup> rules. There is more mass allocated to  $B \cup C$  with PCR6<sup>+</sup> and PCR6 than with PCR5<sup>+</sup> and PCR5 because two sources of information support this hypothesis.

**Example 9 (revisited):** we consider  $\Theta = \{A, B, C, D\}$ , and the following seven categorical BBAs  $m_1(A) = 1$ ,  $m_2(B \cup C) = 1$ ,  $m_3(A \cup C) = 1$ ,  $m_4(B \cup C) = 1$ ,  $m_5(A \cup B \cup C \cup D) = 1$ ,  $m_6(A \cup B \cup C) = 1$ , and  $m_7(A \cup B \cup C) = 1$ . If we make the PCR5, PCR5<sup>+</sup>, PCR6 and PCR6<sup>+</sup> fusion of these seven BBAs altogether we obtain results given in Tables IX and X.

Focal Elements	$m_{1,2,3,4,5,6,7}^{\text{PCR5}}(\cdot)$	$m_{1,2,3,4,5,6,7}^{\text{PCR5}^+}(\cdot)$
$A$	1/5	1/3
$A \cup C$	1/5	1/3
$B \cup C$	1/5	1/3
$A \cup B \cup C$	1/5	0
$A \cup B \cup C \cup D$	1/5	0

Table IX  
EXAMPLE 9: RESULTS OF PCR5<sup>+</sup> VERSUS PCR5.

Because  $\kappa_1(A \cup B \cup C \cup D) = 0$  and  $\kappa_1(A \cup B \cup C) = 0$ , these disjunctions are discarded and more mass is committed

Focal Elements	$m_{1,2,3,4,5,6,7}^{\text{PCR6}}(\cdot)$	$m_{1,2,3,4,5,6,7}^{\text{PCR6}^+}(\cdot)$
$A$	1/7	1/4
$A \cup C$	1/7	1/4
$B \cup C$	2/7	1/2
$A \cup B \cup C$	2/7	0
$A \cup B \cup C \cup D$	1/7	0

Table X

 EXAMPLE 9: RESULTS OF PCR6<sup>+</sup> VERSUS PCR6.

to  $A$ ,  $A \cup C$  and  $B \cup C$  with PCR5<sup>+</sup> and PCR6<sup>+</sup> rules. There is more mass allocated to  $B \cup C$  with PCR6<sup>+</sup> and PCR6 than with PCR5<sup>+</sup> and PCR5 because two sources of information support this hypothesis. Similarly, more mass is allocated to  $(A \cup B \cup C)$  with PCR6 than PCR5 since two sources of information support this hypothesis.

**Example 10** (revisited): we consider  $\Theta = \{A, B, C\}$ , and the following three categorical BBAs  $m_1(A) = 1$ ,  $m_2(B \cup C) = 1$ , and  $m_3(A \cup C) = 1$ . We have only one conflicting product  $\pi_1(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C) = 1$  to redistribute, and for this example we have  $\kappa_1(A) = 1$ ,  $\kappa_1(A \cup C) = 1$  and  $\kappa_1(B \cup C) = 1$  which means that all focal elements  $A$ ,  $A \cup C$  and  $B \cup C$  must be kept, and they must receive a mass through the proportional redistribution principle. Hence in this example we have  $m_{1,2,3}^{\text{PCR5}} = m_{1,2,3}^{\text{PCR6}} = m_{1,2,3}^{\text{PCR5}^+} = m_{1,2,3}^{\text{PCR6}^+}$ , and the combined masses are evenly distributed as shown in the Table XI.

Focal Elements	$m_{1,2,3}^{\text{PCR5}}(\cdot)$	$m_{1,2,3}^{\text{PCR5}^+}(\cdot)$	$m_{1,2,3}^{\text{PCR6}}(\cdot)$	$m_{1,2,3}^{\text{PCR6}^+}(\cdot)$
$A$	1/3	1/3	1/3	1/3
$A \cup C$	1/3	1/3	1/3	1/3
$B \cup C$	1/3	1/3	1/3	1/3

Table XI

 EXAMPLE 10: RESULTS OF PCR5, PCR5<sup>+</sup>, PCR6, PCR6<sup>+</sup>.

**Example 11** (revisited): we consider  $\Theta = \{A, B, C\}$ , and the following four categorical BBAs  $m_1(A) = 1$ ,  $m_2(B \cup C) = 1$ ,  $m_3(A \cup C) = 1$ , and  $m_4(A \cup B) = 1$ . Because we have only one conflicting product  $\pi_1(\emptyset) = m_1(A)m_2(B \cup C)m_3(A \cup C)m_4(A \cup B) = 1$  and  $\kappa_1(A) = 1$ ,  $\kappa_1(A \cup B) = 1$ ,  $\kappa_1(A \cup C) = 1$  and  $\kappa_1(B \cup C) = 1$  no hypothesis is discarded in the proportional conflict redistribution, and we get  $m_{1,2,3,4}^{\text{PCR5}} = m_{1,2,3,4}^{\text{PCR6}} = m_{1,2,3,4}^{\text{PCR5}^+} = m_{1,2,3,4}^{\text{PCR6}^+}$  with the merged masses being evenly distributed, that is  $m_{1,2,3,4}^{\text{PCR5}}(A) = 1/4$ ,  $m_{1,2,3,4}^{\text{PCR5}}(A \cup B) = 1/4$ ,  $m_{1,2,3,4}^{\text{PCR5}}(A \cup C) = 1/4$ , and  $m_{1,2,3,4}^{\text{PCR5}}(B \cup C) = 1/4$ .

**Example 12** (revisited): we consider  $\Theta = \{A, B, C\}$ , and the following three categorical BBAs  $m_1(A \cup B \cup C) = 1$ ,  $m_2(A) = 1$ ,  $m_3(B \cup C) = 1$ . If we make the PCR5 fusion, and the PCR5<sup>+</sup> fusion, of these three BBAs altogether we obtain results given in Table XII. Because  $\pi_1(\emptyset) = m_1(A \cup B \cup C)m_2(A)m_3(B \cup C)$ , we get  $\kappa_1(A \cup B \cup C) = 0$ ,  $\kappa_1(A) = 1$  and  $\kappa_1(B \cup C) = 1$  based on (23). Therefore, using the PCR5<sup>+</sup> combination rule, we get a

Focal Elements	$m_{1,2,3}^{\text{PCR5}}(\cdot)$	$m_{1,2,3}^{\text{PCR5}^+}(\cdot)$
$A$	1/3	1/2
$B \cup C$	1/3	1/2
$A \cup B \cup C$	1/3	0

Table XII

 EXAMPLE 12: RESULTS OF PCR5, PCR5<sup>+</sup>.

redistribution of the conflicting mass  $\pi_1(\emptyset) = 1$  only between  $A$  and  $B \cup C$ . In this example we have  $m_{1,2,3}^{\text{PCR5}} = m_{1,2,3}^{\text{PCR6}}$ , and  $m_{1,2,3}^{\text{PCR5}^+} = m_{1,2,3,4}^{\text{PCR6}^+}$  because no mass is allocated on the same hypothesis by two different sources.

**Example 13** (revisited): we consider  $\Theta = \{A, B, C, D\}$ , and the three following BBAs

$$\begin{aligned} m_1(A \cup B) &= 0.8, m_1(C \cup D) = 0.2, \\ m_2(A \cup B) &= 0.4, m_2(C \cup D) = 0.6, \\ m_3(B) &= 0.1, m_3(A \cup B \cup C \cup D) = 0.9. \end{aligned}$$

If we make the PCR5, PCR5<sup>+</sup>, PCR6 and PCR6<sup>+</sup> fusion of these seven BBAs altogether we obtain results given in Tables XIII and XIV.

Focal Elements	$m_{1,2,3}^{\text{PCR5}}(\cdot)$	$m_{1,2,3}^{\text{PCR5}^+}(\cdot)$
$B$	0.041797	0.041797
$A \cup B$	0.487632	0.613029
$C \cup D$	0.258327	0.345174
$A \cup B \cup C \cup D$	0.212244	0

Table XIII

 EXAMPLE 13: RESULTS OF PCR5<sup>+</sup> VERSUS PCR5.

Focal Elements	$m_{1,2,3}^{\text{PCR6}}(\cdot)$	$m_{1,2,3}^{\text{PCR6}^+}(\cdot)$
$B$	0.037676	0.037676
$A \cup B$	0.487632	0.613029
$C \cup D$	0.262448	0.349295
$A \cup B \cup C \cup D$	0.212244	0

Table XIV

 EXAMPLE 13: RESULTS OF PCR6<sup>+</sup> VERSUS PCR6.

Because  $\kappa_j(\Theta) = 0$  for any conflicting product  $\pi_j(\emptyset)$  involving  $\Theta$ , this hypothesis is discarded in the redistribution of  $\pi_4(\emptyset)$  and of  $\pi_6(\emptyset)$  (see Example 13 in subsection VI-A for details), and therefore more mass is redistributed to  $A \cup B$  and  $C \cup D$  with PCR5<sup>+</sup> and PCR6<sup>+</sup> rules. No more mass is committed to  $B$  with PCR5<sup>+</sup> and PCR6<sup>+</sup> respectively in comparison with PCR5 and PCR6. This is because  $B$  is not implied in any partial conflict with  $\Theta$  (cf. subsection VI-A for details).

## VIII. CONCLUSION

In this paper, after having demonstrated the flawed behavior of PCR5 and PCR6 rules of combination for  $S > 2$  BBAs (including possibly vacuous BBAs), we proposed improvements to correct these behaviors. A computation of a binary keeping-index has been detailed which makes it possible to discard

ignorant information sources for the calculation of each partial conflict. This binary keeping-index has been integrated into the original formulations of PCR5 and PCR6 in order to ensure the neutrality property of the vacuous BBA and to propose two new combination rules for a number of sources greater than 2: PCR5<sup>+</sup> and PCR6<sup>+</sup> rules. The interest of such combination rules could prove to be particularly important in an application case identifying many ignorant sources of information. In such a scenario, the preponderant ignorance of a certain number of sources will no longer obscure a more precise characterization provided by other sources. These new rules of combination have been already applied to risk analysis issues for geophysical and geotechnical data fusion in order to reinforce the levee protection characterizations [48].

#### APPENDIX 1: PROOF OF THE LEMMA 1

We prove that:  $m_{1,2,\dots,S,S+1}^{\text{Conj}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A)$ , for any  $A \in 2^\Theta \setminus \{\emptyset\}$ , where  $m_{S+1}(\Theta) = 1$  is the vacuous BBA  $m_v$ . The set of focal elements of  $m_{S+1}(\cdot)$  is  $\mathcal{F}(m_{S+1}) = \{\Theta\}$ , therefore  $\mathcal{F}_{m_{S+1}} = 1$  and  $X_{j_{S+1}} = \Theta$ . Based on the formula (6) written for  $S+1$  BBAs, we have

$$\begin{aligned} m_{1,2,\dots,S,S+1}^{\text{Conj}}(A) &= \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S, m_{S+1}) \\ X_{j_1} \cap \dots \cap X_{j_S} \cap X_{j_{S+1}} = A}} \pi_j(X_{j_1} \cap \dots \cap X_{j_S} \cap X_{j_{S+1}}) \\ &= \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S, m_{S+1}) \\ X_{j_1} \cap \dots \cap X_{j_S} \cap \Theta = A}} \prod_{i=1}^{S+1} m_i(X_{j_i}) \quad (27) \end{aligned}$$

Because  $X_{j_{S+1}} = \Theta$  is constant and  $m_{S+1}(X_{j_{S+1}}) = m_{S+1}(\Theta) = 1$ , one has

$$\prod_{i=1}^{S+1} m_i(X_{j_i}) = \left( \prod_{i=1}^S m_i(X_{j_i}) \right) \cdot m_{S+1}(\Theta) = \prod_{i=1}^S m_i(X_{j_i})$$

and  $X_{j_1} \cap \dots \cap X_{j_S} \cap X_{j_{S+1}} = X_{j_1} \cap \dots \cap X_{j_S} \cap \Theta = X_{j_1} \cap \dots \cap X_{j_S}$ . Therefore the formula (27) becomes

$$\begin{aligned} m_{1,2,\dots,S,S+1}^{\text{Conj}}(A) &= \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S, m_{S+1}) \\ X_{j_1} \cap \dots \cap X_{j_S} \cap \Theta = A}} \prod_{i=1}^{S+1} m_i(X_{j_i}) \\ &= \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = A}} \prod_{i=1}^S m_i(X_{j_i}) \\ &= m_{1,2,\dots,S}^{\text{Conj}}(A) \end{aligned}$$

which completes the proof of the Lemma 1.

#### APPENDIX 2: PROOF OF THE THEOREM

We prove that  $\text{PCR5}^+(m_1, \dots, m_S, m_{S+1}) = \text{PCR5}^+(m_1, \dots, m_S)$ , or equivalently that  $m_{1,2,\dots,S,S+1}^{\text{PCR5}^+}(A) = m_{1,2,\dots,S}^{\text{PCR5}^+}(A)$  for any  $A \in 2^\Theta \setminus \{\emptyset\}$ , where  $m_{S+1}(X_{j_{S+1}}) = m_{S+1}(\Theta) = 1$  is the vacuous BBA. It is worth noting that  $m_{1,2,\dots,S,S+1}^{\text{Conj}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A)$  for any  $A \in 2^\Theta \setminus \{\emptyset\}$  because the vacuous BBA  $m_{S+1}(\cdot)$  is the

neutral element of the conjunctive rule (see Lemma 1). It is important to note that when considering  $A = \Theta$ , we have always  $m_{1,2,\dots,S,S+1}^{\text{PCR5}^+}(\Theta) = m_{1,2,\dots,S,S+1}^{\text{Conj}}(\Theta) = m_{1,2,\dots,S}^{\text{Conj}}(\Theta) = m_{1,2,\dots,S}^{\text{PCR5}^+}(\Theta)$  because the binary keeping-index of  $\Theta$  is always equal to zero (see remark 1), i.e.  $\kappa_j(\Theta) = 0$ . Therefore all the redistribution terms to  $\Theta$  in PCR5<sup>+</sup> (and in PCR6<sup>+</sup>) formula are equal to zero when  $A = \Theta$ . So, we just have to consider  $A \neq \Theta$  to make the proof.

Because  $m_{S+1}(\cdot)$  is the vacuous BBA, its set of focal elements is  $\mathcal{F}(m_{S+1}) = \{\Theta\}$  and it contains only one focal element, i.e.  $|\mathcal{F}(m_{S+1})| = 1$ . Therefore

$$\begin{aligned} \mathcal{F} &= |\mathcal{F}(m_1)| \cdot |\mathcal{F}(m_2)| \cdot \dots \cdot |\mathcal{F}(m_S)| \cdot |\mathcal{F}(m_{S+1})| \quad (28) \\ &= |\mathcal{F}(m_1)| \cdot |\mathcal{F}(m_2)| \cdot \dots \cdot |\mathcal{F}(m_S)| \quad (29) \end{aligned}$$

This means that the number of conflicting products  $\pi_j(\emptyset)$  associated to the  $S+1$ -tuple  $\mathbf{X}_j = (X_{j_1}, \dots, X_{j_S}, \Theta) \in \mathcal{F}(m_1, \dots, m_S, m_{S+1})$  is equal to the number of conflicting products  $\pi_j(\emptyset)$  associated to  $S$ -tuple  $\mathbf{X}_j = (X_{j_1}, \dots, X_{j_S}) \in \mathcal{F}(m_1, \dots, m_S)$ . Moreover, we always have

$$\prod_{i=1}^{S+1} m_i(X_{j_i}) = \left( \prod_{i=1}^S m_i(X_{j_i}) \right) \cdot m_{S+1}(\Theta) = \prod_{i=1}^S m_i(X_{j_i})$$

Hence, we always have

$$\pi_j(X_{j_1} \cap \dots \cap X_{j_S} \cap \Theta = \emptyset) = \pi_j(X_{j_1} \cap \dots \cap X_{j_S} = \emptyset)$$

because  $X_{j_1} \cap \dots \cap X_{j_S} \cap \Theta = X_{j_1} \cap \dots \cap X_{j_S}$ .

Based on the formula (25) written for  $S+1$  BBAs, we have

$$\begin{aligned} m_{1,2,\dots,S,S+1}^{\text{PCR5}^+}(A) &= m_{1,2,\dots,S,S+1}^{\text{Conj}}(A) \\ &+ \sum_{j \in \{1, \dots, \mathcal{F}\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \kappa_j(A) \prod_{i \in \{1, \dots, S+1\} | X_{j_i} = A} m_i(X_{j_i}) \right) \right. \\ &\left. \cdot \frac{\pi_j(X_{j_1} \cap \dots \cap X_{j_S} \cap \Theta = \emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \kappa_j(X) \prod_{i \in \{1, \dots, S+1\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right] \quad (30) \end{aligned}$$

where  $\mathcal{F}$  is given by (28).

Because  $X_{j_{S+1}} = \Theta$  and because we consider  $A \neq \Theta$ , we have always

$$\prod_{i \in \{1, \dots, S+1\} | X_{j_i} = A} m_i(X_{j_i}) = \prod_{i \in \{1, \dots, S\} | X_{j_i} = A} m_i(X_{j_i})$$

Whether  $X \in \mathbf{X}_j = (X_{j_1}, \dots, X_{j_S})$  or  $X \in \mathbf{X}_j = (X_{j_1}, \dots, X_{j_S}, \Theta)$  the value of  $\kappa_j(X)$  is the same since the additional binary *containing indicator*  $\delta_j(X, \Theta)$  entering in the product of the computation of the binary *keeping-index* is always equal to 1 and does not modify  $\kappa_j(X)$  value, and of course when  $X = A$ . Because the binary *keeping-index* entering in the numerator and denominator of formula (30) removes the factor  $m_{S+1}(\Theta)$  from all products it belongs to

(since  $\Theta$  includes all elements of the product it belongs to), the formula (30) reduces to the following formula

$$\begin{aligned}
 m_{1,2,\dots,S,S+1}^{\text{PCR5}^+}(A) &= m_{1,2,\dots,S}^{\text{Conj}}(A) \\
 + \sum_{j \in \{1,\dots,S\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} & \left[ \left( \kappa_j(A) \prod_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \right. \\
 \cdot \frac{\pi_j(X_{j_1} \cap \dots \cap X_{j_S} = \emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \kappa_j(X) \prod_{i \in \{1,\dots,S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} & \left. \right] \\
 &= m_{1,2,\dots,S}^{\text{PCR5}^+}(A) \quad (31)
 \end{aligned}$$

where  $\mathbf{X}_j$  represents now the  $S$ -tuple  $(X_{j_1}, \dots, X_{j_S})$ , and  $\pi_j(\emptyset) = \pi_j(X_{j_1} \cap \dots \cap X_{j_S} = \emptyset)$ .

So, we have proved  $\text{PCR5}^+(m_1, \dots, m_S, m_{S+1}) = \text{PCR5}^+(m_1, \dots, m_S)$  when  $m_{S+1}$  is the vacuous BBA. Similarly, we can prove that  $\text{PCR6}^+(m_1, \dots, m_S, m_{S+1}) = \text{PCR6}^+(m_1, \dots, m_S)$  when  $m_{S+1}$  is the vacuous BBA. This completes the proof of the theorem.

### APPENDIX 3: CODES OF $\text{PCR5}^+$ AND $\text{PCR6}^+$ RULES

For convenience, we provide two basic Matlab<sup>TM</sup> codes for  $\text{PCR5}^+$  and  $\text{PCR6}^+$  for the fusion of  $S \geq 2$  BBAs for working with  $2^\Theta$ , i.e. working with Shafer's model. No input verification of input is done in the routines. It is assumed that the input matrix BBA is correct, both in dimension and in content. The derivation of all possible combinations is done with `combvec(Combinations,vec)` instruction which is included in the Matlab<sup>TM</sup> neural networks toolbox. This `combvec` call can be a very time-consuming task when the size of the problem increases. A standalone version of these codes is also available upon request to the authors. The  $j$ -th column of the BBA input matrix corresponds to the (vertical) BBA vector  $m_j(\cdot)$  associated with the  $j$ -th source  $s_j$ . Each element of a BBA matrix is in  $[0, 1]$  and the sum of each column must be one. If  $N$  is the cardinality of the frame  $\Theta$  and if  $S$  is the number of sources, then the size of the BBA input matrix is  $((2^N) - 1) \times S$ . Each column of the BBA matrix must use the classical binary encoding of elements. For example, if  $\Theta = \{A, B, C\}$ , then we encode the elements of  $2^\Theta \setminus \{\emptyset\}$  by the binary sequence  $001 \equiv A$ ,  $010 \equiv B$ ,  $011 \equiv A \cup B$ ,  $\dots$ ,  $111 \equiv A \cup B \cup C$ . The mass of empty set is not included in the BBA vector because it is always set to zero.

These codes can be used and shared for free for research purposes only. Commercial uses of these codes, or adaptation of them in any programming language, is not allowed without written agreement of the authors. These codes are provided by the copyright holders "as is" and any express or implied warranties are disclaimed. The copyright holder will not be liable for any direct, or indirect damages of the use of these codes. The authors would appreciate any feedback in the use of these codes, and publication using these codes should cite this paper in agreement for their use.

```

=====
function [mPCR5plus]=PCR5plusfusion(BBA)
=====
% Authors and copyrights: Theo Dezert & Jean Dezert
% Input: BBA=[m1 m2 ... mS]= Matrix of BBAs to combine with PCR5+
% Output: mPCR5plus is PCR5+(m1,m2,...,mS) fusion result
=====
NbrSources=size(BBA,2);CardTheta=log2(size(BBA,1)+1);
if(NbrSources==1), mPCR5plus=BBA(:,1);return, end
mPCR5plus=zeros(size(BBA,1),1);FocalElem= cell(NbrSources,1);
for i=1:NbrSources, FocalElem[i]=find(BBA(:,i)> 0)';end
Combinations=combvec(FocalElem(1:NbrSources))';
for c=1:size(Combinations,1)
    PC=Combinations(c,:);masseConj=diag(BBA(PC,:))';
    masseConj=prod(diag(BBA(PC,:))',2);Intersections=PC(1);
    for s=2:NbrSources, Intersections=bitand(Intersections,PC(s)); end
    if(Intersections==0)
        mPCR5plus(Intersections)=mPCR5plus(Intersections)+masseConj;
    else
        Binary=[];CardPC=[];KeepIndex=[];
        for i=1:NbrSources
            Binary(i,:)=bitget(PC(i),CardTheta:-1:1,'int8');
            CardPC(i,:)=sum(Binary(i,:)==1);
        end
        for j=1:NbrSources
            delta=[];
            for js=1:NbrSources
                if CardPC(js)>=CardPC(j)
                    for jp=1:NbrSources
                        if PC(jp)~=PC(js) && CardPC(jp)<=CardPC(js)
                            if sum(Binary(jp,:)==Binary(js,:))==CardTheta
                                delta=[delta 1]; else, delta=[delta 0];
                            end
                        end
                    end
                end
            end
            end
            if isempty(delta)==1
                KeepIndex(j,1)=1;
            else
                KeepIndex(j,1)=1-prod(delta);
            end
        end
        KeepIndex=KeepIndex';
        for i=1:NbrSources
            if KeepIndex(i)==1, KeepIndex(i)=masseConj(i); end
        end
        UQ=unique(PC);Proportions=0*UQ;DenPCR5=0;
        for u=1:size(UQ,2)
            SamePropositions=find(PC==UQ(u));
            MassProd=prod(KeepIndex(SamePropositions));
            Proportions(u)= MassProd*masseConj;DenPCR5=DenPCR5+MassProd;
        end
        Proportions=Proportions/DenPCR5;
        for u=1:size(UQ,2),mPCR5plus(UQ(u))=mPCR5plus(UQ(u))+Proportions(u); end
    end
end

=====
function [mPCR6plus]=PCR6plusfusion(BBA)
=====
% Authors and copyrights: Theo Dezert & Jean Dezert
% Input: BBA=[m1 m2 ... mS]= Matrix of BBAs to combine with PCR6+
% Output: mPCR6plus is PCR6+(m1,m2,...,mS) fusion result
=====
NbrSources=size(BBA,2);CardTheta=log2(size(BBA,1)+1);
if(NbrSources==1), mPCR6plus=BBA(:,1);return, end
mPCR6plus=zeros(size(BBA,1),1);FocalElem= cell(NbrSources,1);
for i=1:NbrSources, FocalElem[i]=find(BBA(:,i)> 0)';end
Combinations=combvec(FocalElem(1:NbrSources))';
for c=1:size(Combinations,1)
    PC=Combinations(c,:);masseConj=diag(BBA(PC,:))';
    masseConj=prod(diag(BBA(PC,:))',2);Intersections=PC(1);
    for s=2:NbrSources, Intersections=bitand(Intersections,PC(s));end
    if(Intersections==0)
        mPCR6plus(Intersections)=mPCR6plus(Intersections)+masseConj;
    else
        Binary=[];CardPC=[];KeepIndex=[];
        for i=1:NbrSources
            Binary(i,:)=bitget(PC(i),CardTheta:-1:1,'int8');
            CardPC(i,:)=sum(Binary(i,:)==1);
        end
        for j=1:NbrSources
            delta=[];
            for js=1:NbrSources
                if CardPC(js)>=CardPC(j)
                    for jp=1:NbrSources
                        if PC(jp)~=PC(js) && CardPC(jp)<=CardPC(js)
                            if sum(Binary(jp,:)==Binary(js,:))<=CardTheta
                                delta=[delta 1]; else, delta=[delta 0];
                            end
                        end
                    end
                end
            end
            end
            if isempty(delta)==1
                KeepIndex(j,1)=1;
            else
                KeepIndex(j,1)=1-prod(delta);
            end
        end
        KeepIndex=KeepIndex';IgnoringSetOffE=find(KeepIndex==0);
        masseConj(IgnoringSetOffE)=[];PC(IgnoringSetOffE)=[];
        for s=1:numel(masseConj)
            Proportion= masseConj(s) * (masseConj / (sum(masseConj,2)));
            mPCR6plus(PC(s))=mPCR6plus(PC(s))+Proportion;
        end
    end
end
end
    
```

## REFERENCES

- [1] J.Y. Halpern, *Reasoning about uncertainty*, The MIT Press, Cambridge, MA, USA, 2003.
- [2] A.N. Kolmogorov, *Foundations of the theory of probability*, Chelsea Publishing Company, New York, NY, USA, 1956.
- [3] A.N. Kolmogorov, *Three approaches to the quantitative definition of information*, Int. Journal of Computer Mathematics, Vol. 2, pp. 157–168, 1968.
- [4] L.A. Zadeh, *Fuzzy Sets*, Information and Control, Vol. 8, pp. 338–353, 1965.
- [5] D. Dubois, H. Prade (Editors), *Fundamentals of Fuzzy Sets*, Springer Sciences+Business Media, New York, NY, USA, 2000.
- [6] D. Dubois, H. Prade *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, NY, USA, 1980.
- [7] D. Dubois, H. Prade, *Possibility Theory and Its Applications: Where Do We Stand?*, Springer Handbook of Computational Intelligence, pp. 1–31, 2015.
- [8] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [9] R.R. Yager, J.M. Kacprzyk, M. Fedrizzi (Editors), *Advances in the Dempster-Shafer Theory of Evidence*, John Wiley & Sons, New York, NY, USA, 1994.
- [10] R.R. Yager, L. Liu (Editors), *Classical works of the Dempster-Shafer Theory of belief functions*, Studies in Fuzziness and Soft Computing, Vol. 219, Springer, 2008.
- [11] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. Math. Statist., Vol. 38, pp. 325–339, 1967.
- [12] A.P. Dempster, *A generalization of Bayesian inference*, J. of the Royal Statistical Society, Ser. B, Vol. 30, pp. 205–245, 1968.
- [13] R. Yager, *On the Dempster-Shafer framework and new combination rules*, Information Sciences, Vol. 41, pp. 93–138, 1987.
- [14] T. Denœux (Editor), *40 years of Research on Dempster-Shafer theory: a Retrospective*, Int. Journal of Approximate Reasoning, Special Issue, Vol. 79, 2016.
- [15] F. Cuzzolin, *The Geometry of Uncertainty*, Springer Nature, Switzerland, 2021.
- [16] K. Sentz, S. Ferson, *Combination of Evidence in Dempster-Shafer Theory*, SANDIA Tech. Report, SAND2002-0835, 96 pages, April 2002.
- [17] T. George, N.R. Pal, *Quantification of conflict in Dempster-Shafer framework: a new approach*, Int. J. Gen. Syst., Vol. 24(4), pp. 407–423, 1996.
- [18] M.J. Wierman, *Measuring conflict in evidence theory*, in Proc. of IFSA World Congress and 20th NAFIPS Int. Conf., Vol. 3(21), pp. 1741–1745, 2001.
- [19] W. Liu, *Analyzing the degree of conflict among belief functions*, Artif. Intell., Vol. 170, pp. 909–924, 2006.
- [20] P. Smets, *Analyzing the combination of conflicting belief functions*, Information Fusion, Vol. 8, pp. 387–412, 2007.
- [21] A. Martin, A.-L. Jousselme, C. Osswald, *Conflict measure for the discounting operation on belief functions*, in Proc. of Int. Conf. on Information Fusion (Fusion 2008), Cologne, Germany, 2008.
- [22] A. Martin, *About Conflict in the Theory of Belief Functions*, in Belief Functions: Theory and Applications. Advances in Intelligent and Soft Computing (T. Denœux & M.H. Masson Editors), Vol 164. Springer, Berlin, Heidelberg, 2012.
- [23] J. Schubert, *The internal conflict of a belief function*, in Proc. of the Second International Conference on Belief Functions, Compiègne, France, May 2012.
- [24] S. Destercke, T. Burger, *Toward an Axiomatic Definition of Conflict Between Belief Functions*, IEEE Trans. on Cybernetics, Vol. 43(2), pp. 585–596, April 2013.
- [25] T. Burger, *Geometric views on conflicting mass functions: From distances to angles*, International Journal of Approximate Reasoning, Vol. 70, pp. 36–50, 2016.
- [26] M. Daniel, V. Kratochvíl, *On Hidden Conflicts of Belief Functions*, in Proc. of 11th Conf. of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019), Prague, Czech Republic, Sept. 2019.
- [27] A. Martin, *Conflict management in information fusion with belief functions*, in Information quality in information fusion and decision making (Eloi Bossé & Galina L. Rogova Editors), Springer Verlag, pp. 79–97, April 2019.
- [28] L.A. Zadeh, *On the validity of Dempster's rule of combination of evidence*, ERL Memo M79/24, Department of EECS, Univ. of California, Berkeley, CA, U.S.A., 1979.
- [29] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, Vol. 7(2), pp. 85–90, 1986.
- [30] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of bayesian fusion rule*, Int. J. of Intelligent Syst., Vol. 29(3), pp. 223–252, 2014.
- [31] F. Smarandache, J. Dezert, *Proportional conflict redistribution rules for information fusion*, Chapter 1 of [34], Vol. 2, 2006.
- [32] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, Chapter 2 of [34], Vol. 2, 2006.
- [33] A. Martin, C. Osswald, J. Dezert, F. Smarandache, *General combination rules for qualitative and quantitative beliefs*, Journal of Advances in Information Fusion (JAIF), Vol. 3(2), pp. 67–89, Dec. 2008.
- [34] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion (collected works)*, American Research Press, Volumes. 1–4, 2004–2015. <https://www.onera.fr/fr/staff/jean-dezert/references>
- [35] P. Halmos, *Naive Set Theory*, D. Van Nostrand Company Inc., New York, NY, USA, 1960.
- [36] T. Denœux, *The cautious rule of combination for belief functions and some extensions*, in Proc. of Int. Information Fusion Conference (Fusion 2006), Firenze, Italy, July 2006.
- [37] I. Kramosil, *Probabilistic Analysis of Dempster Combination Rule*, Chap. 6 of Probabilistic Analysis of Belief Functions, Springer-Science+Business Media, New York, 2001.
- [38] F. Cuzzolin, *Geometry of Dempster's Rule of Combination*, IEEE Trans. on SMC, Part B: Cybernetics, Vol. 34(2), pp. 961–977, April 2004.
- [39] T. Dezert, Y. Fargier, S. Palma-Lopes, P. Côte, *Levee Characterization by Means of Data Fusion of In-Situ Geophysical and Geotechnical Information*, in Proc. of 25th European Meeting of Environmental and Engineering Geophysics, Vol. 2019, No. 1, pp. 1-5, European Association of Geoscientists & Engineers, Sept. 2012.
- [40] F. Smarandache, *An in-depth look at quantitative information fusion rules*, Chapter 8 of [34], Vol. 2, 2006.
- [41] F. Smarandache, J. Dezert, *Importance of sources using repeated fusion with the proportional conflict redistribution rules #5 and #6*, in Multispace & Multistructure Neutrosophic transdisciplinarity (100 collected papers of sciences), Vol. IV, North-European Scientific Publishers, Finland, 2010.
- [42] J. Dezert, F. Smarandache, *Canonical decomposition of dichotomous basic belief assignment*, Int. Journal of Intelligent Systems, pp. 1–21, 2020.
- [43] J. Dezert, F. Smarandache, A. Tchamova, D. Han, *Fast fusion of basic belief assignments defined on a dichotomous frame of discernment*, in Proc. of Int. Conf. on Information Fusion (Fusion 2020), Pretoria, South Africa, July 2020.
- [44] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of Int. Conf. on Information Fusion (Fusion 2010), Edinburgh, Scotland, UK, 26-29 July 2010.
- [45] Arnaud Martin's toolbox <http://people.irisa.fr/Arnaud.Martin/toolboxes/>
- [46] The Belief Functions and Applications Society <https://www.bfasociety.org/#software>
- [47] The Comprehensive R Archive Network <https://CRAN.R-project.org/package=ibelief>
- [48] T. Dezert, J. Dezert, *Improvement of Proportional Conflict Redistribution Fusion Rules for Levee Characterization*, in Proc. of ESREL 2021 Int. Conf., Angers, France, Sept. 2021.



# Explicit Formulas of PCR5 and PCR6 Fusion Rules for Three Bayesian Basic Belief Assignments

Jean Dezert<sup>a</sup>, Florentin Smarandache<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA/DTIS, Palaiseau, France.

<sup>b</sup>Department of Mathematics, University of New Mexico, Gallup, NM, USA.

Emails: jean.dezert@onera.fr, smarand@unm.edu

**Abstract**—This short paper presents the explicit formulas of the PCR5 and PCR6 rules of combination for three Bayesian basic belief assignments. We give a simple example to show how to apply them.

**Keywords:** information fusion, belief functions, PCR5, PCR6, PCR5<sup>+</sup>, PCR6<sup>+</sup> fusion rules..

## I. INTRODUCTION

Among many existing rules of combination of Basic Belief Assignments (BBAs), the conjunctive rule, Dempster-Shafer (DS) rule [1], and the Proportional Conflict Redistribution rules no 5 (PCR5) and no 6 (PCR6) are the most used rules of combination. While the conjunctive rule makes it possible to combine information between different sources of information represented by belief functions by estimating the level of existing conflict, DS rule [1], [2] proposes a distribution of this conflict on the hypotheses characterized by the sources of information. The normalization carried out by the DS rule may however be considered counter-intuitive especially when the level of conflict between the sources of information is high [3], [4], but also in some situations where the level of conflict between sources is low as shown in [5] showing a dictatorial behavior of DS rule. The Proportional Conflict Redistribution rules (PCR5 [6] and PCR6 [7], [8]) have been proposed to circumvent the problem of the DS rule to make a more judicious management of the conflict. Moreover, improved versions of PCR5 and PCR6 rules preserving the neutrality of a vacuous (i.e. a totally ignorant) source of evidence in the PCR process have been recently proposed in [9]. They are denoted by PCR5<sup>+</sup> and PCR6<sup>+</sup> fusion rules. We will not present in detail these improved rules here because we address the problem of fusing only Bayesian BBAs and for these particular type of BBAs PCR5<sup>+</sup> coincides with PCR5, and PCR6<sup>+</sup> coincides with PCR6 because there is no mass committed to partial and to total ignorances (i.e. to all possible disjunctions) involved in partial conflict to redistribute thanks to PCR5 and PCR6 principles.

After a brief recall of basics of belief functions in section II, we present the general formulas for PCR5 and PCR6 fusion rules in section III based on [9], with a simple example in section IV. In section V we present the direct formulas for PCR5 and PCR6 rules for three general (i.e. non-Bayesian)

BBAs, and the direct formulas for PCR5 and PCR6 rules for three Bayesian BBAs in section VI. A simple example of application of these formulas for the fusion of three Bayesian BBAs defined on the simple frame of discernment with two elements is given in section VII with complete calculation for convenience. Section VIII concludes this paper.

## II. BASICS OF BELIEF FUNCTIONS

We consider a given finite set  $\Theta$  of  $n > 1$  distinct elements  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  corresponding to the frame of discernment (FoD) of the fusion problem, or the decision-making problem, under concern. All elements of  $\Theta$  are mutually exclusive<sup>1</sup> and each element is an elementary choice of the potential decision to take. The power set of  $\Theta$  is the set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) and it is usually denoted  $2^\Theta$  because its cardinality equals  $2^{|\Theta|}$ . A Basic Belief Assignment (BBA) given by a source of evidence is defined by Shafer [1] in his Mathematical Theory of Evidence (known also as Dempster-Shafer Theory, or DST) as  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying

$$\begin{cases} m(\emptyset) = 0, \\ \sum_{A \in 2^\Theta} m(A) = 1, \end{cases} \quad (1)$$

where  $m(A)$  is the mass of belief exactly committed to  $A$ , what we usually call the mass of  $A$ . A BBA is said proper (or normal) if it satisfies Shafer's definition (1). The subset  $A \subseteq \Theta$  is called a Focal Element (FE) of the BBA  $m(\cdot)$  if and only if  $m(A) > 0$ . The empty set is not a focal element of a BBA because  $m(\emptyset) = 0$  according to definition (1). The set of all focal elements of a BBA  $m(\cdot)$  is denoted  $\mathcal{F}(m)$ . Its mathematical definition is  $\mathcal{F}(m) = \{X \in 2^\Theta | m(X) > 0\}$ . The cardinality  $|\mathcal{F}(m)|$  of the set  $\mathcal{F}(m)$  is denoted  $\mathcal{F}_m$ . The order of focal elements of  $\mathcal{F}(m)$  does not matter and all the focal elements are different. The set  $\mathcal{F}(m)$  of focal elements of  $m(\cdot)$  has at least one focal element, and at most  $2^{|\Theta|} - 1$  focal elements.

<sup>1</sup>This standard assumption is called *Shafer's model of FoD* in DSMT (Dezert-Smarandache Theory) framework [10].

Belief and plausibility functions are respectively defined from  $m(\cdot)$  by [1]

$$Bel(A) = \sum_{X \in 2^\Theta | X \subseteq A} m(X), \quad (2)$$

and

$$Pl(A) = \sum_{X \in 2^\Theta | A \cap X \neq \emptyset} m(X) = 1 - Bel(\bar{A}). \quad (3)$$

where  $\bar{A}$  represents the complement of  $A$  in  $\Theta$ .

$Bel(A)$  and  $Pl(A)$  are usually interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$  [11], [12]. The functions  $m(\cdot)$ ,  $Bel(\cdot)$  and  $Pl(\cdot)$  are one-to-one. A belief function  $Bel(\cdot)$  is *Bayesian* if all Bel's focal elements are singletons [1] (Theorem 2.8 p. 45). In this case,  $m(X) = Bel(X)$  for any (singleton) focal element  $X$ , and  $m(\cdot)$  is called a *Bayesian BBA*. Corresponding  $Bel(\cdot)$  function is equal to  $Pl(\cdot)$  and these functions can be interpreted as a same (possibly subjective) probability measure  $P(\cdot)$ . The vacuous BBA (VBBA for short) representing a totally ignorant source is defined as  $m_v(\Theta) = 1$ .

### III. PCR5 AND PCR6 RULES OF COMBINATION

#### A. The PCR5 rule of combination

The PCR5 rule [6] transfers the conflicting mass to all the elements involved in this conflict and proportionally to their individual masses, so that a more sophisticate and specific distribution is done with the PCR5 fusion process with respect to other existing rules (including Dempster's rule). The PCR5 rule is presented in details (with justification and examples) in [10], Vol. 2 and Vol. 3.

A simple formulation of the general expression of the PCR5 fusion of  $S > 2$  basic belief assignments is obtained by redistributing each conflicting product defined by

$$\pi_j(\emptyset) = \pi_j(X_{j_1} \cap \dots \cap X_{j_S} = \emptyset) = \prod_{i=1}^S m_i(X_{j_i}), \quad (4)$$

to some elements of the power set of the FoD that are involved in the conflict  $X_{j_1} \cap \dots \cap X_{j_S} = \emptyset$ . Each  $\pi_j(\emptyset)$  is redistributed proportionally to elements involved in this conflict based on the PCR5 redistribution principle. When an element  $A \in 2^\Theta$  is not involved in a conflicting product  $\pi_j(\emptyset)$ , i.e.  $A \notin \mathbf{X}_j$ , the conflicting product  $\pi_j(\emptyset)$  is not redistributed to  $A$ . If an element  $A$  is involved in the conflict  $X_{j_1} \cap \dots \cap X_{j_S} = \emptyset$ , i.e.  $A \in \mathbf{X}_j$  and  $\pi_j(\emptyset)$  occur, then the proportional redistribution of  $\pi_j(\emptyset)$  to  $A$  is given by

$$x_j(A) \triangleq \left( \prod_{i \in \{1, \dots, S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \prod_{i \in \{1, \dots, S\} | X_{j_i} = X} m_i(X_{j_i}) \right)}, \quad (5)$$

where  $A \in \mathbf{X}_j$  means that at least one component of the  $S$ -tuple  $\mathbf{X}_j$  equals  $A$ , with

$$\mathbf{X}_j \triangleq (X_{j_1}, X_{j_2}, \dots, X_{j_S}) \in \mathcal{F}(m_1, \dots, m_S),$$

where

- $j_1 \in \{1, 2, \dots, \mathcal{F}_{m_1}\}$ ,
- $j_2 \in \{1, 2, \dots, \mathcal{F}_{m_2}\}$ ,
- $\dots$ ,
- $j_S \in \{1, 2, \dots, \mathcal{F}_{m_S}\}$ ,
- and

$$\mathcal{F}(m_1, \dots, m_S) \triangleq \mathcal{F}(m_1) \times \mathcal{F}(m_2) \times \dots \times \mathcal{F}(m_S),$$

$$\mathcal{F} \triangleq |\mathcal{F}(m_1, \dots, m_S)| = \prod_{i=1}^S |\mathcal{F}(m_i)| = \prod_{i=1}^S \mathcal{F}_{m_i}.$$

The element  $X_{j_i}$  is the focal element of  $m_i(\cdot)$  that makes the  $i$ -th component of the  $j$ -th  $S$ -tuple  $\mathbf{X}_j$ .

The mass of  $A$  obtained by the PCR5 rule is

$$m_{1,2,\dots,S}^{\text{PCR5}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1, \dots, \mathcal{F}\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} x_j(A), \quad (6)$$

where  $A \in \mathbf{X}_j \wedge \pi_j(\emptyset)$  is a shorthand notation meaning that at least one component of the  $S$ -tuple  $\mathbf{X}_j$  equals  $A$  and the components of  $\mathbf{X}_j$  are conflicting, i.e.  $X_{j_1} \cap \dots \cap X_{j_S} = \emptyset$ .

The general PCR5 formula can be expressed as  $m_{1,2,\dots,S}^{\text{PCR5}}(\emptyset) = 0$ , and for  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{1,2,\dots,S}^{\text{PCR5}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1, \dots, \mathcal{F}\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \prod_{i \in \{1, \dots, S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \prod_{i \in \{1, \dots, S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right]. \quad (7)$$

where  $m_{1,2,\dots,S}^{\text{Conj}}(A)$  is the mass of  $A$  obtained by the conjunctive rule, that is

$$\begin{aligned} m_{1,2,\dots,S}^{\text{Conj}}(A) &= \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = A}} \pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) \\ &= \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = A}} \prod_{i=1}^S m_i(X_{j_i}). \end{aligned} \quad (8)$$

The total conflicting mass between the  $S$  sources of evidence, denoted  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)$ , is nothing but the sum of all existing conflicting mass products, that is

$$\begin{aligned} m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) &= \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = \emptyset}} \pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) \\ &= 1 - \sum_{A \in 2^\Theta \setminus \{\emptyset\}} m_{1,2,\dots,S}^{\text{Conj}}(A). \end{aligned} \quad (9)$$

Note that the combined BBA  $m_{1,2,\dots,S}^{\text{Conj}}(\cdot)$  given in (8) is not a proper BBA because it does not satisfy Shafer's definition (1). In general the  $S$  sources of evidence to combine do not fully agree, and we have consequently  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) > 0$ .

### B. The PCR6 rule of combination

A variant of PCR5 rule, called PCR6 rule, has been proposed by Martin and Osswald in [7], [8] for combining  $S > 2$  sources. The difference between PCR5 and PCR6 lies in the way the proportional conflict redistribution is done as soon as three (or more) sources are involved in the fusion. The PCR6 fusion of  $S > 2$  BBAs is obtained by  $m_{1,2,\dots,S}^{\text{PCR6}}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by<sup>2</sup>

$$m_{1,2,\dots,S}^{\text{PCR6}}(A) = m_{1,2,\dots,S}^{\text{Conj}}(A) + \sum_{j \in \{1,\dots,S\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right]. \quad (10)$$

The difference between the general PCR5 and PCR6 formulas is that the PCR5 proportional redistribution involves the products  $\prod_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i})$  of multiple same focal elements  $A$  (if any) in the conflict, whereas the PCR6 conflict redistribution principle works with their sum  $\sum_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i})$  instead.

PCR6 coincides with PCR5 when one combines two sources of evidence.

### IV. SIMPLE EXAMPLE OF PCR5 AND PCR6 FUSION RULES

Here we provide a simple example showing the difference of the results between PCR5 and PCR6 rules. This example has been already presented in [9]. For convenience, all numerical values have been rounded to six decimal places when necessary.

**Example 1:** We consider the simplest FoD  $\Theta = \{A, B\}$ , and the three following BBAs

$$\begin{aligned} m_1(A) &= 0.6, m_1(B) = 0.1, m_1(A \cup B) = 0.3, \\ m_2(A) &= 0.5, m_2(B) = 0.3, m_2(A \cup B) = 0.2, \\ m_3(A) &= 0.4, m_3(B) = 0.1, m_3(A \cup B) = 0.5. \end{aligned}$$

Because  $\mathcal{F}_{m_1} = |\mathcal{F}(m_1)| = 3$ ,  $\mathcal{F}_{m_2} = |\mathcal{F}(m_2)| = 3$  and  $\mathcal{F}_{m_3} = |\mathcal{F}(m_3)| = 3$ , we have  $\mathcal{F} = \mathcal{F}_{m_1} \cdot \mathcal{F}_{m_2} \cdot \mathcal{F}_{m_3} = 27$  non-zero products to consider. Fifteen products are non-conflicting and enter in the calculation of  $m_{1,2,3}^{\text{Conj}}(A)$ ,  $m_{1,2,3}^{\text{Conj}}(B)$  and  $m_{1,2,3}^{\text{Conj}}(A \cup B)$ , and twelve products are conflicting products

that need to be proportionally redistributed. The conjunctive combination of these three BBAs is

$$\begin{aligned} m_{1,2,3}^{\text{Conj}}(A) &= m_1(A)m_2(A)m_3(A) \\ &\quad + m_1(A)m_2(A)m_3(A \cup B) \\ &\quad + m_1(A)m_2(A \cup B)m_3(A) \\ &\quad + m_1(A \cup B)m_2(A)m_3(A) \\ &\quad + m_1(A)m_2(A \cup B)m_3(A \cup B) \\ &\quad + m_1(A \cup B)m_2(A)m_3(A \cup B) \\ &\quad + m_1(A \cup B)m_2(A \cup B)m_3(A) = 0.5370, \\ m_{1,2,3}^{\text{Conj}}(B) &= m_1(B)m_2(B)m_3(B) \\ &\quad + m_1(B)m_2(B)m_3(A \cup B) \\ &\quad + m_1(B)m_2(A \cup B)m_3(B) \\ &\quad + m_1(A \cup B)m_2(B)m_3(B) \\ &\quad + m_1(B)m_2(A \cup B)m_3(A \cup B) \\ &\quad + m_1(A \cup B)m_2(B)m_3(A \cup B) \\ &\quad + m_1(A \cup B)m_2(A \cup B)m_3(B) = 0.0900, \\ m_{1,2,3}^{\text{Conj}}(A \cup B) &= m_1(A \cup B)m_2(A \cup B)m_3(A \cup B) \\ &= 0.3 \cdot 0.2 \cdot 0.5 = 0.0300, \end{aligned}$$

and the total conflict between these three BBAs is given by

$$m_{1,2,3}^{\text{Conj}}(\emptyset) = 1 - m_{1,2,3}^{\text{Conj}}(A) - m_{1,2,3}^{\text{Conj}}(B) - m_{1,2,3}^{\text{Conj}}(A \cup B) = 0.3430.$$

In this example we have twelve partial conflicts, noted  $\pi_j(\emptyset)$  ( $j = 1, \dots, 12$ ), which correspond to the following products

$$\begin{aligned} \pi_1(\emptyset) &= m_1(A)m_2(A)m_3(B) = 0.0300, \\ \pi_2(\emptyset) &= m_1(A)m_2(B)m_3(A) = 0.0720, \\ \pi_3(\emptyset) &= m_1(B)m_2(A)m_3(A) = 0.0200, \\ \pi_4(\emptyset) &= m_1(B)m_2(B)m_3(A) = 0.0120, \\ \pi_5(\emptyset) &= m_1(B)m_2(A)m_3(B) = 0.0050, \\ \pi_6(\emptyset) &= m_1(A)m_2(B)m_3(B) = 0.0180, \\ \pi_7(\emptyset) &= m_1(A \cup B)m_2(A)m_3(B) = 0.0150, \\ \pi_8(\emptyset) &= m_1(A \cup B)m_2(B)m_3(A) = 0.0360, \\ \pi_9(\emptyset) &= m_1(B)m_2(A)m_3(A \cup B) = 0.0250, \\ \pi_{10}(\emptyset) &= m_1(A)m_2(B)m_3(A \cup B) = 0.0900, \\ \pi_{11}(\emptyset) &= m_1(A)m_2(A \cup B)m_3(B) = 0.0120, \\ \pi_{12}(\emptyset) &= m_1(B)m_2(A \cup B)m_3(A) = 0.0080. \end{aligned}$$

In applying the PCR5 formula (7), and the PCR6 formula (10) we obtain finally  $m_{1,2,3}^{\text{PCR5}}(\emptyset) = m_{1,2,3}^{\text{PCR6}}(\emptyset) = 0$ , and<sup>3</sup>

$$\begin{aligned} m_{1,2,3}^{\text{PCR5}}(A) &\approx 0.723281, \\ m_{1,2,3}^{\text{PCR5}}(B) &\approx 0.182460, \\ m_{1,2,3}^{\text{PCR5}}(A \cup B) &\approx 0.094259, \end{aligned}$$

and

$$\begin{aligned} m_{1,2,3}^{\text{PCR6}}(A) &\approx 0.743496, \\ m_{1,2,3}^{\text{PCR6}}(B) &\approx 0.162245, \\ m_{1,2,3}^{\text{PCR6}}(A \cup B) &\approx 0.094259. \end{aligned}$$

We see a difference between the BBAs  $m_{1,2,3}^{\text{PCR5}}$  and  $m_{1,2,3}^{\text{PCR6}}$  which is normal because the PCR principles are quite different.

<sup>2</sup>We wrote this PCR6 general formula in the style of PCR5 formula (7).

<sup>3</sup>The symbol  $\approx$  means "approximately equal to".

Using the PCR5 fusion rule the first partial conflicting mass  $\pi_1(\emptyset) = m_1(A)m_2(A)m_3(B) = 0.03$  is redistributed back to  $A$  and  $B$  proportionally to  $m_1(A)m_2(A)$  and to  $m_3(B)$  as follows

$$\frac{x_1(A)}{m_1(A)m_2(A)} = \frac{x_1(B)}{m_3(B)} = \frac{\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)},$$

whence

$$x_1(A) = \frac{m_1(A)m_2(A)\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)} = 0.0225,$$

$$x_1(B) = \frac{m_3(B)\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)} = 0.0075.$$

We can verify  $\pi_1(\emptyset) = x_1(A) + x_1(B) = 0.03$ .

Using the PCR6 fusion rule the first partial conflicting mass  $\pi_1(\emptyset) = 0.03$  is redistributed back to  $A$  and  $B$  proportionally to  $(m_1(A) + m_2(A))$  and to  $m_3(B)$ . So we get the following redistributions  $x_1(A) = 0.0275$  for  $A$  and  $x_1(B) = 0.0025$  for  $B$  because

$$\frac{x_1(A)}{m_1(A) + m_2(A)} = \frac{x_1(B)}{m_3(B)} = \frac{\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)},$$

whence

$$x_1(A) = \frac{(m_1(A) + m_2(A))\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)} = 0.0275,$$

$$x_1(B) = \frac{m_3(B)\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)} = 0.0025.$$

We can verify  $\pi_1(\emptyset) = x_1(A) + x_1(B) = 0.03$ .

Note that for all the partial conflicts having no duplicate element involved in the conflicting product  $\pi_j(\emptyset)$  we make the same redistribution with PCR5 rule and with PCR6 rule. For instance, for  $\pi_7(\emptyset) = m_1(A \cup B)m_2(A)m_3(B) = 0.0150$  we get

$$\frac{x_7(A \cup B)}{m_1(A \cup B)} = \frac{x_7(A)}{m_2(A)} = \frac{x_7(B)}{m_3(B)}$$

$$= \frac{\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)},$$

whence  $\pi_7(\emptyset) = x_7(A \cup B) + x_7(A) + x_7(B) = 0.0150$ , with

$$x_7(A \cup B) = \frac{m_1(A \cup B)\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)} = 0.0050,$$

$$x_7(A) = \frac{m_2(A)\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)} \approx 0.0083,$$

$$x_7(B) = \frac{m_3(B)\pi_7(\emptyset)}{m_1(A \cup B) + m_2(A) + m_3(B)} \approx 0.0017.$$

## V. PCR5 AND PCR6 RULES FOR THREE BBAS

The previous general formulas of PCR5 (7) and PCR6 (10) can be written more explicitly for the fusion of three BBAs

as follows (see [13] for details<sup>4</sup>) when working with a FoD  $\Theta$  with Shafer's model.

$$m_{1,2,3}^{\text{PCR5}}(A) = m_{1,2,3}^{\text{Conj}}(A)$$

$$+ \sum_{\substack{X, Y \in 2^\Theta \\ A \neq X, A \neq Y, X \neq Y \\ A \cap X \cap Y = \emptyset}} \left[ \frac{m_1(A)^2 m_2(X) m_3(Y)}{m_1(A) + m_2(X) + m_3(Y)} \right.$$

$$+ \frac{m_1(Y) m_2(A)^2 m_3(X)}{m_1(Y) + m_2(A) + m_3(X)}$$

$$+ \left. \frac{m_1(X) m_2(Y) m_3(A)^2}{m_1(X) + m_2(Y) + m_3(A)} \right]$$

$$+ \sum_{\substack{X \in 2^\Theta \\ A \cap X = \emptyset}} \left[ \frac{m_1(A)^2 m_2(X) m_3(X)}{m_1(A) + m_2(X) m_3(X)} \right.$$

$$+ \frac{m_1(X) m_2(A)^2 m_3(X)}{m_1(X) m_3(X) + m_2(A)}$$

$$+ \frac{m_1(X) m_2(X) m_3(A)^2}{m_1(X) m_2(X) + m_3(A)}$$

$$+ \sum_{\substack{X \in 2^\Theta \\ A \cap X = \emptyset}} \left[ \frac{m_1(A)^2 m_2(A)^2 m_3(X)}{m_1(A) m_2(A) + m_3(X)} \right.$$

$$+ \frac{m_1(X) m_2(A)^2 m_3(A)^2}{m_1(X) + m_2(A) m_3(A)}$$

$$+ \left. \frac{m_1(A)^2 m_2(X) m_3(A)^2}{m_1(A) m_3(A) + m_2(X)} \right]$$
(11)

and

$$m_{1,2,3}^{\text{PCR6}}(A) = m_{1,2,3}^{\text{Conj}}(A)$$

$$+ \sum_{\substack{X, Y \in 2^\Theta \\ A \neq X, A \neq Y, X \neq Y \\ A \cap X \cap Y = \emptyset}} \left[ \frac{m_1(A)^2 m_2(X) m_3(Y)}{m_1(A) + m_2(X) + m_3(Y)} \right.$$

$$+ \frac{m_1(Y) m_2(A)^2 m_3(X)}{m_1(Y) + m_2(A) + m_3(X)}$$

$$+ \left. \frac{m_1(X) m_2(Y) m_3(A)^2}{m_1(X) + m_2(Y) + m_3(A)} \right]$$

$$+ \sum_{\substack{X \in 2^\Theta \\ A \cap X = \emptyset}} \left[ \frac{m_1(A)^2 m_2(X) m_3(X)}{m_1(A) + m_2(X) + m_3(X)} \right.$$

$$+ \frac{m_1(X) m_2(A)^2 m_3(X)}{m_1(X) + m_2(A) + m_3(X)}$$

$$+ \frac{m_1(X) m_2(X) m_3(A)^2}{m_1(X) + m_2(X) + m_3(A)}$$

$$+ \sum_{\substack{X \in 2^\Theta \\ A \cap X = \emptyset}} \left[ \frac{(m_1(A) + m_2(A)) m_1(A) m_2(A) m_3(X)}{m_1(A) + m_2(A) + m_3(X)} \right.$$

$$+ \frac{(m_2(A) + m_3(A)) m_1(X) m_2(A) m_3(A)}{m_1(X) + m_2(A) + m_3(A)}$$

$$+ \left. \frac{(m_1(A) + m_3(A)) m_1(A) m_2(X) m_3(A)}{m_1(A) + m_2(X) + m_3(A)} \right]$$
(12)

It is worth mentioning that if some fractions involved in the formulas (11) and (12) have their denominators equal to zero,

<sup>4</sup>It is worth mentioning that PCR5 for three BBAs given in the section 2 of [13] is incorrect, and it must be replaced by formula (11) of this paper.

these fractions are just discarded. It can be easily verified on example 1 that PCR5 formula (11) gives the same result as with the formula (7), and that the PCR6 formula (12) gives the same result as with the formula (10).

## VI. PCR5 AND PCR6 RULES FOR THREE BAYESIAN BBAS

If we want to work with three Bayesian BBAs only, the focal elements of BBAs to combine are only singletons of the power set  $2^\Theta$ . In this particular case, the previous PCR5 and PCR6 formulas (11) and (12) can be simplified as

$$\begin{aligned}
 m_{1,2,3}^{\text{PCR5}}(A) &= m_1(A)m_2(A)m_3(A) \\
 &+ \sum_{\substack{X,Y \in \Theta \setminus \{A\} \\ X \neq Y}} \left[ \frac{m_1(A)^2 m_2(X) m_3(Y)}{m_1(A) + m_2(X) + m_3(Y)} \right. \\
 &\quad + \frac{m_1(Y) m_2(A)^2 m_3(X)}{m_1(Y) + m_2(A) + m_3(X)} \\
 &\quad \left. + \frac{m_1(X) m_2(Y) m_3(A)^2}{m_1(X) + m_2(Y) + m_3(A)} \right] \\
 &+ \sum_{X \in \Theta \setminus \{A\}} \left[ \frac{m_1(A)^2 m_2(X) m_3(X)}{m_1(A) + m_2(X) m_3(X)} \right. \\
 &\quad + \frac{m_1(X) m_2(A)^2 m_3(X)}{m_1(X) m_3(X) + m_2(A)} \\
 &\quad \left. + \frac{m_1(X) m_2(X) m_3(A)^2}{m_1(X) m_2(X) + m_3(A)} \right] \\
 &+ \sum_{X \in \Theta \setminus \{A\}} \left[ \frac{m_1(A)^2 m_2(A)^2 m_3(X)}{m_1(A) m_2(A) + m_3(X)} \right. \\
 &\quad + \frac{m_1(X) m_2(A)^2 m_3(A)^2}{m_1(X) + m_2(A) m_3(A)} \\
 &\quad \left. + \frac{m_1(A)^2 m_2(X) m_3(A)^2}{m_1(A) m_3(A) + m_2(X)} \right]
 \end{aligned} \tag{13}$$

$$\begin{aligned}
 m_{1,2,3}^{\text{PCR6}}(A) &= m_{1,2,3}^{\text{Conj}}(A) \\
 &+ \sum_{\substack{X,Y \in \Theta \setminus \{A\} \\ X \neq Y}} \left[ \frac{m_1(A)^2 m_2(X) m_3(Y)}{m_1(A) + m_2(X) + m_3(Y)} \right. \\
 &\quad + \frac{m_1(Y) m_2(A)^2 m_3(X)}{m_1(Y) + m_2(A) + m_3(X)} \\
 &\quad \left. + \frac{m_1(X) m_2(Y) m_3(A)^2}{m_1(X) + m_2(Y) + m_3(A)} \right] \\
 &+ \sum_{X \in \Theta \setminus \{A\}} \left[ \frac{m_1(A)^2 m_2(X) m_3(X)}{m_1(A) + m_2(X) + m_3(X)} \right. \\
 &\quad + \frac{m_1(X) m_2(A)^2 m_3(X)}{m_1(X) + m_3(X) + m_2(A)} \\
 &\quad \left. + \frac{m_1(X) m_2(X) m_3(A)^2}{m_1(X) + m_2(X) + m_3(A)} \right] \\
 &+ \sum_{X \in \Theta \setminus \{A\}} \left[ \frac{(m_1(A) + m_2(A)) m_1(A) m_2(A) m_3(X)}{m_1(A) + m_2(A) + m_3(X)} \right. \\
 &\quad + \frac{(m_1(A) + m_3(A)) m_1(A) m_2(X) m_3(A)}{m_1(A) + m_3(A) + m_2(X)} \\
 &\quad \left. + \frac{(m_2(A) + m_3(A)) m_1(X) m_2(A) m_3(A)}{m_1(X) + m_2(A) + m_3(A)} \right]
 \end{aligned} \tag{14}$$

In the formulas (13) and (14) the subset  $A$  is any singleton of  $2^\Theta$  (i.e. any element of  $\Theta$ ). For any non-singleton  $A$  of  $2^\Theta$  we have  $m_{1,2,3}^{\text{PCR5}}(A) = 0$  because the fusion of Bayesian BBAs by PCR5 and PCR6 always produces a Bayesian BBA. It is worth mentioning that if some fractions involved in the formulas (13) and (14) have their denominators equal to zero, these fractions are just discarded.

## VII. EXAMPLE OF PCR5 AND PCR6 FUSION OF THREE BAYESIAN BBAS

Here we provide a simple example showing the difference of the results between PCR5 and PCR6 rules for three Bayesian BBAs. For convenience, all numerical values have been rounded to six decimal places when necessary.

**Example 2:** We consider the simplest FoD  $\Theta = \{A, B\}$ , and the three following BBAs

$$\begin{aligned}
 m_1(A) &= 0.2, m_1(B) = 0.8, m_1(A \cup B) = 0, \\
 m_2(A) &= 0.1, m_2(B) = 0.9, m_2(A \cup B) = 0, \\
 m_3(A) &= 0.6, m_3(B) = 0.4, m_3(A \cup B) = 0.
 \end{aligned}$$

Because  $\mathcal{F}_{m_1} = |\mathcal{F}(m_1)| = 2$ ,  $\mathcal{F}_{m_2} = |\mathcal{F}(m_2)| = 2$  and  $\mathcal{F}_{m_3} = |\mathcal{F}(m_3)| = 2$ , we have  $\mathcal{F} = \mathcal{F}_{m_1} \cdot \mathcal{F}_{m_2} \cdot \mathcal{F}_{m_3} = 8$  non-zero products to consider. Two non-zero products are non-conflicting and enter in the calculation of  $m_{1,2,3}^{\text{Conj}}(A)$  and  $m_{1,2,3}^{\text{Conj}}(B)$ , and six non-zero products are conflicting products that need to be proportionally redistributed. The conjunctive combination of these three Bayesian BBAs is

$$\begin{aligned}
 m_{1,2,3}^{\text{Conj}}(A) &= m_1(A)m_2(A)m_3(A) = 0.012, \\
 m_{1,2,3}^{\text{Conj}}(B) &= m_1(B)m_2(B)m_3(B) = 0.288, \\
 m_{1,2,3}^{\text{Conj}}(A \cup B) &= m_1(A \cup B)m_2(A \cup B)m_3(A \cup B) = 0,
 \end{aligned}$$

and the total conflict between these three BBAs is

$$\begin{aligned}
 m_{1,2,3}^{\text{Conj}}(\emptyset) &= 1 - m_{1,2,3}^{\text{Conj}}(A) - m_{1,2,3}^{\text{Conj}}(B) - m_{1,2,3}^{\text{Conj}}(A \cup B) \\
 &= 0.70.
 \end{aligned}$$

In this example we have six partial conflicts, noted  $\pi_j(\emptyset)$  ( $j = 1, \dots, 6$ ), which are given by the following products

$$\begin{aligned}
 \pi_1(\emptyset) &= m_1(A)m_2(A)m_3(B) = 0.008, \\
 \pi_2(\emptyset) &= m_1(A)m_2(B)m_3(A) = 0.108, \\
 \pi_3(\emptyset) &= m_1(B)m_2(A)m_3(A) = 0.048, \\
 \pi_4(\emptyset) &= m_1(B)m_2(B)m_3(A) = 0.432, \\
 \pi_5(\emptyset) &= m_1(B)m_2(A)m_3(B) = 0.032, \\
 \pi_6(\emptyset) &= m_1(A)m_2(B)m_3(B) = 0.072.
 \end{aligned}$$

### A. PCR5 fusion of the three Bayesian BBAs of example 2

Using the general PCR5 fusion rule (7) with  $S = 3$  (i.e. 3 BBAs) we manage the the conflicting mass products as follows:

- Conflicting mass  $\pi_1(\emptyset) = m_1(A)m_2(A)m_3(B) = 0.008$  is redistributed back to  $A$  and  $B$  proportionally to  $m_1(A)m_2(A)$  and to  $m_3(B)$  as follows

$$\frac{x_1(A)}{m_1(A)m_2(A)} = \frac{x_1(B)}{m_3(B)} = \frac{\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)},$$

whence

$$x_1(A) = \frac{m_1(A)m_2(A)\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)} \approx 0.000381,$$

$$x_1(B) = \frac{m_3(B)\pi_1(\emptyset)}{m_1(A)m_2(A) + m_3(B)} \approx 0.007619.$$

We can verify  $\pi_1(\emptyset) = x_1(A) + x_1(B) = 0.008$ .

• Conflicting mass  $\pi_2(\emptyset) = m_1(A)m_2(B)m_3(A) = 0.108$  is redistributed back to  $A$  and  $B$  proportionally to  $m_1(A)m_3(A)$  and to  $m_2(B)$  as follows

$$\frac{x_2(A)}{m_1(A)m_3(A)} = \frac{x_2(B)}{m_2(B)} = \frac{\pi_2(\emptyset)}{m_1(A)m_3(A) + m_2(B)},$$

whence

$$x_2(A) = \frac{m_1(A)m_3(A)\pi_2(\emptyset)}{m_1(A)m_3(A) + m_2(B)} \approx 0.012706,$$

$$x_2(B) = \frac{m_2(B)\pi_2(\emptyset)}{m_1(A)m_3(A) + m_2(B)} \approx 0.095294.$$

• Conflicting mass  $\pi_3(\emptyset) = m_1(B)m_2(A)m_3(A) = 0.048$  is redistributed back to  $A$  and  $B$  proportionally to  $m_2(A)m_3(A)$  and to  $m_1(B)$  as follows

$$\frac{x_3(A)}{m_2(A)m_3(A)} = \frac{x_3(B)}{m_1(B)} = \frac{\pi_3(\emptyset)}{m_2(A)m_3(A) + m_1(B)},$$

whence

$$x_3(A) = \frac{m_2(A)m_3(A)\pi_3(\emptyset)}{m_2(A)m_3(A) + m_1(B)} \approx 0.003349,$$

$$x_3(B) = \frac{m_1(B)\pi_3(\emptyset)}{m_2(A)m_3(A) + m_1(B)} \approx 0.044651.$$

• Conflicting mass  $\pi_4(\emptyset) = m_1(B)m_2(B)m_3(A) = 0.432$  is redistributed back to  $A$  and  $B$  proportionally to  $m_3(A)$  and to  $m_1(B)m_2(B)$  as follows

$$\frac{x_4(A)}{m_3(A)} = \frac{x_4(B)}{m_1(B)m_2(B)} = \frac{\pi_4(\emptyset)}{m_3(A) + m_1(B)m_2(B)},$$

whence

$$x_4(A) = \frac{m_3(A)\pi_4(\emptyset)}{m_3(A) + m_1(B)m_2(B)} \approx 0.196364,$$

$$x_4(B) = \frac{m_1(B)m_2(B)\pi_4(\emptyset)}{m_3(A) + m_1(B)m_2(B)} \approx 0.235636.$$

• Conflicting mass  $\pi_5(\emptyset) = m_1(B)m_2(A)m_3(B) = 0.032$  is redistributed back to  $A$  and  $B$  proportionally to  $m_2(A)$  and to  $m_1(B)m_3(B)$  as follows

$$\frac{x_5(A)}{m_2(A)} = \frac{x_5(B)}{m_1(B)m_3(B)} = \frac{\pi_5(\emptyset)}{m_2(A) + m_1(B)m_3(B)},$$

whence

$$x_5(A) = \frac{m_2(A)\pi_5(\emptyset)}{m_2(A) + m_1(B)m_3(B)} \approx 0.007620,$$

$$x_5(B) = \frac{m_1(B)m_3(B)\pi_5(\emptyset)}{m_2(A) + m_1(B)m_3(B)} \approx 0.024380.$$

• Conflicting mass  $\pi_6(\emptyset) = m_1(A)m_2(B)m_3(B) = 0.072$  is redistributed back to  $A$  and  $B$  proportionally to  $m_1(A)$  and to  $m_2(B)m_3(B)$  as follows

$$\frac{x_6(A)}{m_1(A)} = \frac{x_6(B)}{m_2(B)m_3(B)} = \frac{\pi_6(\emptyset)}{m_1(A) + m_2(B)m_3(B)},$$

whence

$$x_6(A) = \frac{m_1(A)\pi_6(\emptyset)}{m_1(A) + m_2(B)m_3(B)} \approx 0.025714,$$

$$x_6(B) = \frac{m_2(B)m_3(B)\pi_6(\emptyset)}{m_1(A) + m_2(B)m_3(B)} \approx 0.046286.$$

Therefore in applying PCR5 formula we get

$$m_{1,2,3}^{\text{PCR5}}(A) = m_{1,2,3}^{\text{Conj}}(A) + x_1(A) + x_2(A) + x_3(A) \\ + x_4(A) + x_5(A) + x_6(A) \approx 0.258134,$$

$$m_{1,2,3}^{\text{PCR5}}(B) = m_{1,2,3}^{\text{Conj}}(B) + x_1(B) + x_2(B) + x_3(B) \\ + x_4(B) + x_5(B) + x_6(B) \approx 0.741866,$$

and because the result is a Bayesian BBA we have also

$$m_{1,2,3}^{\text{PCR5}}(A \cup B) = 0.$$

Now if we apply the PCR5 combination of the three Bayesian BBAs of example 2 using the direct formula (13), we have to work with  $\Theta = \{A, B\}$ . So, for the focal element  $A$  we must consider all  $X \in \Theta \setminus \{A\}$  in the second and third summations but  $\Theta \setminus \{A\} = \{B\}$ , hence  $X = B$  only. In the first summation there is no  $X, Y \in \Theta \setminus \{A\}$  such that  $X \neq Y$ , so the first summation does not exist for this example. For the focal element  $A$ , the formula (13) reduces to the simple expression

$$m_{1,2,3}^{\text{PCR5}}(A) = \underbrace{m_1(A)m_2(A)m_3(A)}_{m_{1,2,3}^{\text{Conj}}(A) = 0.012} \\ + \left[ \underbrace{\frac{m_1(A)^2 m_2(B)m_3(B)}{m_1(A) + m_2(B)m_3(B)}}_{x_6(A) \approx 0.025714} \right. \\ + \underbrace{\frac{m_1(B)m_2(A)^2 m_3(B)}{m_1(B)m_3(B) + m_2(A)}}_{x_5(A) \approx 0.007620} \\ + \left. \underbrace{\frac{m_1(B)m_2(B)m_3(A)^2}{m_1(B)m_2(B) + m_3(A)}}_{x_4(A) \approx 0.196364} \right] \\ + \left[ \underbrace{\frac{m_1(A)^2 m_2(A)^2 m_3(B)}{m_1(A)m_2(A) + m_3(B)}}_{x_1(A) \approx 0.000381} \right. \\ + \underbrace{\frac{m_1(A)^2 m_2(B)m_3(A)^2}{m_1(A)m_3(A) + m_2(B)}}_{x_2(A) \approx 0.012706} \\ + \left. \underbrace{\frac{m_1(B)m_2(A)^2 m_3(A)^2}{m_1(B) + m_2(A)m_3(A)}}_{x_3(A) \approx 0.003349} \right] \\ \approx 0.258134 \quad (15)$$

Similarly, for the example 2 and using the direct PCR5 formula (13) for three bayesian BBAs we have for the focal element  $B$

$$\begin{aligned}
 m_{1,2,3}^{\text{PCR5}}(B) &= \underbrace{m_1(B)m_2(B)m_3(B)}_{m_{1,2,3}^{\text{Conj}}(B)=0.288} \\
 &+ \left[ \underbrace{\frac{m_1(B)^2 m_2(A) m_3(A)}{m_1(B) + m_2(A) m_3(A)}}_{x_3(B) \approx 0.044651} \right. \\
 &\quad + \underbrace{\frac{m_1(A) m_2(B)^2 m_3(A)}{m_1(A) m_3(A) + m_2(B)}}_{x_2(B) \approx 0.095294} \\
 &\quad + \left. \underbrace{\frac{m_1(A) m_2(A) m_3(B)^2}{m_1(A) m_2(A) + m_3(B)}}_{x_1(B) \approx 0.007619} \right] \\
 &+ \left[ \underbrace{\frac{m_1(B)^2 m_2(B)^2 m_3(A)}{m_1(B) m_2(B) + m_3(A)}}_{x_4(B) \approx 0.235636} \right. \\
 &\quad + \underbrace{\frac{m_1(B)^2 m_2(A) m_3(B)^2}{m_1(B) m_3(B) + m_2(A)}}_{x_5(B) \approx 0.024380} \\
 &\quad + \left. \underbrace{\frac{m_1(A) m_2(B)^2 m_3(B)^2}{m_1(A) + m_2(B) m_3(B)}}_{x_6(B) \approx 0.046286} \right] \\
 &\approx 0.741866
 \end{aligned} \tag{16}$$

It is clear that the results obtained with the direct formula (13) are in agreement with those obtained by the general PCR5 formula (7) when  $S = 3$ .

### B. PCR6 fusion of the three Bayesian BBAs of example 2

Using the general PCR6 fusion rule (10) with  $S = 3$  (i.e. 3 BBAs) we manage the the conflicting mass products as follows:

- Conflicting mass  $\pi_1(\emptyset) = m_1(A)m_2(A)m_3(B) = 0.008$  is redistributed back to  $A$  and  $B$  proportionally to  $m_1(A) + m_2(A)$  and to  $m_3(B)$  as follows

$$\frac{x_1(A)}{m_1(A) + m_2(A)} = \frac{x_1(B)}{m_3(B)} = \frac{\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)},$$

whence

$$\begin{aligned}
 x_1(A) &= \frac{(m_1(A) + m_2(A))\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)} \approx 0.003429, \\
 x_1(B) &= \frac{m_3(B)\pi_1(\emptyset)}{m_1(A) + m_2(A) + m_3(B)} \approx 0.004571.
 \end{aligned}$$

We can verify  $\pi_1(\emptyset) = x_1(A) + x_1(B) = 0.008$ .

- Conflicting mass  $\pi_2(\emptyset) = m_1(A)m_2(B)m_3(A) = 0.108$  is redistributed back to  $A$  and  $B$  proportionally to  $m_1(A) + m_3(A)$  and to  $m_2(B)$  as follows

$$\frac{x_2(A)}{m_1(A) + m_3(A)} = \frac{x_2(B)}{m_2(B)} = \frac{\pi_2(\emptyset)}{m_1(A) + m_3(A) + m_2(B)},$$

whence

$$\begin{aligned}
 x_2(A) &= \frac{(m_1(A) + m_3(A))\pi_2(\emptyset)}{m_1(A) + m_3(A) + m_2(B)} \approx 0.050824, \\
 x_2(B) &= \frac{m_2(B)\pi_2(\emptyset)}{m_1(A) + m_3(A) + m_2(B)} \approx 0.057176.
 \end{aligned}$$

- Conflicting mass  $\pi_3(\emptyset) = m_1(B)m_2(A)m_3(A) = 0.048$  is redistributed back to  $A$  and  $B$  proportionally to  $m_2(A) + m_3(A)$  and to  $m_1(B)$  as follows

$$\frac{x_3(A)}{m_2(A) + m_3(A)} = \frac{x_3(B)}{m_1(B)} = \frac{\pi_3(\emptyset)}{m_2(A) + m_3(A) + m_1(B)},$$

whence

$$\begin{aligned}
 x_3(A) &= \frac{(m_2(A) + m_3(A))\pi_3(\emptyset)}{m_2(A) + m_3(A) + m_1(B)} \approx 0.022400, \\
 x_3(B) &= \frac{m_1(B)\pi_3(\emptyset)}{m_2(A) + m_3(A) + m_1(B)} \approx 0.025600.
 \end{aligned}$$

- Conflicting mass  $\pi_4(\emptyset) = m_1(B)m_2(B)m_3(A) = 0.432$  is redistributed back to  $A$  and  $B$  proportionally to  $m_3(A)$  and to  $m_1(B) + m_2(B)$  as follows

$$\frac{x_4(A)}{m_3(A)} = \frac{x_4(B)}{m_1(B) + m_2(B)} = \frac{\pi_4(\emptyset)}{m_3(A) + m_1(B) + m_2(B)},$$

whence

$$\begin{aligned}
 x_4(A) &= \frac{m_3(A)\pi_4(\emptyset)}{m_3(A) + m_1(B) + m_2(B)} \approx 0.112696, \\
 x_4(B) &= \frac{(m_1(B) + m_2(B))\pi_4(\emptyset)}{m_3(A) + m_1(B) + m_2(B)} \approx 0.319304.
 \end{aligned}$$

- Conflicting mass  $\pi_5(\emptyset) = m_1(B)m_2(A)m_3(B) = 0.032$  is redistributed back to  $A$  and  $B$  proportionally to  $m_2(A)$  and to  $m_1(B) + m_3(B)$  as follows

$$\frac{x_5(A)}{m_2(A)} = \frac{x_5(B)}{m_1(B) + m_3(B)} = \frac{\pi_5(\emptyset)}{m_2(A) + m_1(B) + m_3(B)},$$

whence

$$\begin{aligned}
 x_5(A) &= \frac{m_2(A)\pi_5(\emptyset)}{m_2(A) + m_1(B) + m_3(B)} \approx 0.002462, \\
 x_5(B) &= \frac{(m_1(B) + m_3(B))\pi_5(\emptyset)}{m_2(A) + m_1(B) + m_3(B)} \approx 0.029538.
 \end{aligned}$$

- Conflicting mass  $\pi_6(\emptyset) = m_1(A)m_2(B)m_3(B) = 0.072$  is redistributed back to  $A$  and  $B$  proportionally to  $m_1(A)$  and to  $m_2(B) + m_3(B)$  as follows

$$\frac{x_6(A)}{m_1(A)} = \frac{x_6(B)}{m_2(B) + m_3(B)} = \frac{\pi_6(\emptyset)}{m_1(A) + m_2(B) + m_3(B)},$$

whence

$$\begin{aligned}
 x_6(A) &= \frac{m_1(A)\pi_6(\emptyset)}{m_1(A) + m_2(B) + m_3(B)} \approx 0.009600, \\
 x_6(B) &= \frac{(m_2(B) + m_3(B))\pi_6(\emptyset)}{m_1(A) + m_2(B) + m_3(B)} \approx 0.062400.
 \end{aligned}$$

Therefore in applying PCR6 formula (10) with  $S = 3$  we get finally

$$\begin{aligned} m_{1,2,3}^{\text{PCR6}}(A) &= m_{1,2,3}^{\text{Conj}}(A) + x_1(A) + x_2(A) + x_3(A) \\ &\quad + x_4(A) + x_5(A) + x_6(A) \approx 0.213411, \\ m_{1,2,3}^{\text{PCR6}}(B) &= m_{1,2,3}^{\text{Conj}}(B) + x_1(B) + x_2(B) + x_3(B) \\ &\quad + x_4(B) + x_5(B) + x_6(B) \approx 0.786589, \end{aligned}$$

and because the result is a Bayesian BBA we have also

$$m_{1,2,3}^{\text{PCR6}}(A \cup B) = 0.$$

When using the direct formula (14) of PCR6 rule for the three bayesian BBAs of example 2 we obtain for the focal element  $A$

$$\begin{aligned} m_{1,2,3}^{\text{PCR6}}(A) &= \underbrace{m_1(A)m_2(A)m_3(A)}_{m_{1,2,3}^{\text{Conj}}(A)=0.012} \\ &+ \left[ \underbrace{\frac{m_1(A)^2m_2(B)m_3(B)}{m_1(A) + m_2(B) + m_3(B)}}_{x_6(A)\approx 0.009600} \right. \\ &\quad + \underbrace{\frac{m_1(B)m_2(A)^2m_3(B)}{m_1(B) + m_3(B) + m_2(A)}}_{x_5(A)\approx 0.002462} \\ &\quad \left. + \underbrace{\frac{m_1(B)m_2(B)m_3(A)^2}{m_1(B) + m_2(B) + m_3(A)}}_{x_4(A)\approx 0.112696} \right] \\ &+ \left[ \underbrace{\frac{(m_1(A) + m_2(A))m_1(A)m_2(A)m_3(B)}{m_1(A) + m_2(A) + m_3(B)}}_{x_1(A)\approx 0.003429} \right. \\ &\quad + \underbrace{\frac{(m_1(A) + m_3(A))m_1(A)m_2(B)m_3(A)}{m_1(A) + m_3(A) + m_2(B)}}_{x_2(A)\approx 0.050824} \\ &\quad \left. + \underbrace{\frac{(m_2(A) + m_3(A))m_1(B)m_2(A)m_3(A)}{m_1(B) + m_2(A) + m_3(A)}}_{x_3(A)\approx 0.022400} \right] \\ &\approx 0.213411 \end{aligned} \tag{17}$$

Similarly, for the example 2 and using the direct formula

(14) we have for the focal element  $B$

$$\begin{aligned} m_{1,2,3}^{\text{PCR6}}(B) &= \underbrace{m_1(B)m_2(B)m_3(B)}_{m_{1,2,3}^{\text{Conj}}(B)=0.288} \\ &+ \left[ \underbrace{\frac{m_1(B)^2m_2(A)m_3(A)}{m_1(B) + m_2(A) + m_3(A)}}_{x_3(B)\approx 0.025600} \right. \\ &\quad + \underbrace{\frac{m_1(A)m_2(B)^2m_3(A)}{m_1(A) + m_3(A) + m_2(B)}}_{x_2(B)\approx 0.057176} \\ &\quad \left. + \underbrace{\frac{m_1(A)m_2(A)m_3(B)^2}{m_1(A) + m_2(A) + m_3(B)}}_{x_1(B)\approx 0.004571} \right] \\ &+ \left[ \underbrace{\frac{(m_1(B) + m_2(B))m_1(B)m_2(B)m_3(A)}{m_1(B) + m_2(B) + m_3(A)}}_{x_4(B)\approx 0.319304} \right. \\ &\quad + \underbrace{\frac{(m_1(B) + m_3(B))m_1(B)m_2(A)m_3(B)}{m_1(B) + m_3(B) + m_2(A)}}_{x_5(B)\approx 0.029538} \\ &\quad \left. + \underbrace{\frac{(m_2(B) + m_3(B))m_1(A)m_2(B)m_3(B)}{m_1(A) + m_2(B) + m_3(B)}}_{x_6(B)\approx 0.062400} \right] \\ &\approx 0.786589 \end{aligned} \tag{18}$$

It is clear that the results obtained with the direct formula (14) are in agreement with those obtained by the general PCR6 formula (10) when  $S = 3$ .

## VIII. CONCLUSION

In this paper we have developed explicit formulas for the PCR5 and PCR6 fusion of three bayesian BBAs which work with any cardinality of the frame of discernment greater or equal to two. We have verified that our formulas are coherent with general PCR5 and PCR6 formulas. We have also provided the correct PCR5 formula for three BBAs which was erroneous in our 2010 original paper [13]. We hope that these formulas will be helpful for some users of belief functions working only with bayesian belief masses and with only three sources of evidence to combine because these direct formulas are much easier to implement than general PCR5 and PCR6 formulas.

## REFERENCES

- [1] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [2] K. Sentz, S. Ferson, *Combination of Evidence in Dempster-Shafer Theory*, SANDIA Tech. Report, SAND2002-0835, 96 pages, April 2002.
- [3] L.A. Zadeh, *On the validity of Dempster's rule of combination of evidence*, ERL Memo M79/24, Department of EECS, Univ. of California, Berkeley, CA, U.S.A., 1979.
- [4] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, Vol. 7(2), pp. 85–90, 1986.
- [5] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of bayesian fusion rule*, Int. J. of Intelligent Syst., Vol. 29(3), pp. 223–252, 2014.
- [6] F. Smarandache, J. Dezert, *Proportional conflict redistribution rules for information fusion*, Chapter 1 of [10], Vol. 2, 2006.

- [7] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, Chapter 2 of [10], Vol. 2, 2006.
- [8] A. Martin, C. Osswald, J. Dezert, F. Smarandache, *General combination rules for qualitative and quantitative beliefs*, Journal of Advances in Information Fusion (JAIF), Vol. 3(2), pp. 67–89, Dec. 2008.
- [9] T. Dezert, J. Dezert, F. Smarandache, *Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments*, Journal of Advances in Information Fusion (JAIF), Vol. 16, No. 1, June 2021.
- [10] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion (collected works)*, American Research Press, Volumes. 1–4, 2004–2015. <https://www.onera.fr/fr/staff/jean-dezert/references>
- [11] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. Math. Statist., Vol. 38, pp. 325–339, 1967.
- [12] A.P. Dempster, *A generalization of Bayesian inference*, J. of the Royal Statistical Society, Ser. B, Vol. 30, pp. 205–245, 1968.
- [13] F. Smarandache, J. Dezert, *Importance of Sources using the Repeated Fusion Method and the Proportional Conflict Redistribution Rules #5 and #6*, HAL-00471839, 2010, also in [10], Vol. 4. <https://hal.science/hal-00471839>



Part 2:

## **Applications of DSMT**



# Building Change Detection in Satellite Stereo Imagery Based on Belief Functions

Jiaojiao Tian<sup>a</sup>, Peter Reinartz<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Remote Sensing Technology Institute, German Aerospace Center, 82234 Oberpfaffenhofen, Germany.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

Emails: jiaojiao.tian@dlr.de, peter.reinartz@dlr.de, jean.dezert@onera.fr

Originally published as: J. Tian, P. Reinartz, J. Dezert, *Building Change Detection in Satellite Stereo Imagery Based on Belief Functions*, in Proc. of Joint Urban Remote Sensing Event (JURSE 2015), Lausanne, Switzerland, 30 March–1 April 2015, and reprinted with permission.

**Abstract**—3D Building change detection has become a popular research topic along with the improvement of image quality and computer science. When only building changes are of interest, both the multi-temporal images and Digital Surface Models provide valuable but not comprehensive information in the change detection procedure. Therefore, in this paper, belief functions have been adopted for fusing information from these two sources. In the first step, two change indicators are proposed by focusing on building changes. Both indicators have been projected to a sigmoid curve, in which both the concordance and discordance indexes are considered. In order to fuse the concordance and discordance indexes and further fuse the two change indicators, two belief functions are considered. One is the original Dempster-Shafer Theory (DST), and the most recent one is Dezert-Smarandache Theory (DSMT). This paper shows how these belief-based frameworks can help in building change detection problem. Besides using different belief functions in obtaining the global BBAs, four decision-making criteria are tested to extract final building change masks. The results have been validated by compared to the manually extracted change reference mask.

**Keywords:** belief functions, DSMT, satellite imaging, building change detection.

## I. INTRODUCTION

Accurate and efficient detection of changes is of great importance for urban monitoring, which is also an important research field in remote sensing. Change detection methods on large scale land cover monitoring have been intensively studied and reviewed [1], [2]. Along with the ascending of image spectral and temporal resolution, the expectation on automatic change detection has progressively increased, not only on results accuracy, but also on the efficiency and robustness of the methods. Moreover, change detection for a specific target of interest, like buildings is becoming an important research topic. In small scale 2D change detection, which is performed based on only 2D multi-temporal spectral images, problems arise due to misdetections caused by irrelevant changes. The influence of these irrelevant changes is growing as higher resolution images showing more details. Therefore, in this paper, we will further work on satellite multispectral and stereo images, which provides both spectral and height change information.

Adopting satellite stereo imagery for 3D change detection is an exciting and challenging task. Benefiting from improved data quality and advanced computer vision technique, the quality of the generated Digital Surface Models (DSMs) has been largely improved and it is possible to detect changes even for small objects, like single buildings. On the other

side, the DSMs may still exhibit some outliers resulting in occlusions within the stereo/multi views. Several approaches have been proposed for DSM assisted change detection [3], [4], [5], [6]. According to our previous research results, the belief functions introduced in DST allow to work more efficiently and robustly in urban building change detection with very high resolution satellite images [7]. So far, only a basic DS fusion model has been proposed in [6] to define the Basic Belief Assignments (BBAs) thanks to a sigmoid curve considering only the concordance index. Improvement of this DS fusion model for BBAs construction is proposed in this paper to achieve better performance by considering both the concordance and discordance indexes. Since DSMT [8] has been developed in last years as an interesting alternative to DST to circumvent problems of Dempster-Shafer's (DS) rule of combination [9], we also investigate the possibility of using the Proportional Conflict Redistribution Rule #6 (PCR6) of DSMT in our application.

## II. BASICS OF BELIEF FUNCTIONS

Detailed presentations of DST and DSMT can be found in [8], [9] and [10]. Let  $\Theta$  be a frame of discernment of a problem under consideration.  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  consists of a list of  $N$  exhaustive and mutually exclusive elements  $\theta_i$ ,  $i = 1, 2, \dots, N$ . Each  $\theta_i$  represents a possible state related to the problem we want to solve. The assumption of exhaustivity and mutual exclusivity of elements of  $\Theta$  is classically referred as *Shafer's model* of the frame  $\Theta$ . A BBA also called a belief mass function (or just a mass for short), is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  from the power set<sup>1</sup> of  $\Theta$  denoted  $2^\Theta$  to  $[0, 1]$ , that verifies [10]:

$$m(\emptyset) = 0, \quad \text{and} \quad \sum_{X \in 2^\Theta} m(X) = 1; \quad (1)$$

$m(X)$  represents the mass of belief exactly committed to  $X$ . An element  $X \in 2^\Theta$  is called a focal element if and only if  $m(X) > 0$ . In DST, the combination (fusion) of several independent sources of evidences is done with Dempster-Shafer<sup>2</sup> (DS) rule of combination, assuming that the sources are not in total conflict<sup>3</sup>. DS combination of two independent BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$ , denoted symbolically by  $DS(m_1, m_2)$ , is

<sup>1</sup>The power set is the set of all subsets of  $\Theta$ , empty set included.

<sup>2</sup>Although the rule has been proposed originally by Dempster, we call it Dempster-Shafer rule because it has been widely promoted by Shafer in DST.

<sup>3</sup>otherwise DS rule is mathematically not defined because of 0/0 indeterminacy.

defined by  $m^{DS}(\emptyset) = 0$ , and for all  $X \in 2^\Theta \setminus \{\emptyset\}$  by:

$$m^{DS}(X) = \frac{1}{1 - K^{DS}} \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2), \quad (2)$$

where the total degree of conflict  $K^{DS}$  is given by

$$K^{DS} \triangleq \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2). \quad (3)$$

A discussion on the validity of DS rule and its incompatibility with Bayes fusion rule for combining Bayesian BBAs can be found in [9], [11], [12]. To circumvent the problems of DS rule, Smarandache and Dezert (see [8], Vol. 2, Chap. 1), then Martin and Osswald (see [8], Vol. 2, Chap. 2) have developed in DSMT [8] two fusion rules called PCR5 and PCR6 based on the proportional conflict redistribution (PCR) principle which consists

- 1) to apply the conjunctive rule;
- 2) calculate the total or partial conflicting masses;
- 3) then redistribute the (total or partial) conflicting mass proportionally on non-empty sets according to the integrity constraints one has for the frame  $\Theta$ .

This PCR principle transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is not degraded. Because the proportional transfer can be done in two different ways, this has yielded to two different fusion rules. It has been proved in [13] that only PCR6 rule is compatible with frequentist probability estimation, and that is why we recommend its use in the applications. PCR5 and PCR6 rules simplify greatly and coincide for the combination of two sources. In this case, the PCR6 combination is obtained by taking  $m^{PCR6}(\emptyset) = 0$ , and for all  $X \neq \emptyset$  in  $2^\Theta$  by

$$m^{PCR6}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right], \quad (4)$$

where all denominators in Eq. (4) are different from zero. If a denominator is zero, that fraction is discarded.

### III. BUILDING CHANGE DETECTION MODELS

#### A. Choice of the frame of discernment

We now use two sources (indicators) of evidences to solve our problem. As a preparation step, the indicators and focal elements have to be introduced. Two data sources are used for building change detection. One is the satellite images, which contain 2D spectral information. Here we use the Iteratively Reweighted Multivariate Alteration Detection (IRMAD) [14] to highlight changes from the spectral images. The other is the robust height difference which can be calculated from the two Digital Surface Models (DSMs) [6]. Detail of the DSM generation procedure and the characters of the DSMs quality have been described in [5]. As it has been explained in [6], we suppose that new, demolished or changed buildings

exhibit both height changes and spectral changes. The seasonal changes will only influence the spectral images. Therefore, for building change detection, we consider the following three classes (hypotheses) to define our frame of discernment satisfying Shafer's model:  $\Theta = \{\theta_1 \triangleq \text{Pixel} \in \text{BuildingChange}, \theta_2 \triangleq \text{Pixel} \in \text{OtherChange}, \theta_3 \triangleq \text{Pixel} \in \text{NoChange}\}$ .

#### B. Sigmoidal model for BBA construction

BBAs construction is a prerequisite for the combination of sources of evidence. In our previous works [6], the BBAs were built based on sigmoid curves related with the concordance index only. In this paper, we improve our model to construct the BBAs thanks to sigmoidal models for both concordance and discordance indexes following idea proposed in [15]. As explained in [6], the original sigmoid curve is defined as

$$f_{(\tau, T)}(x) = 0.99 / (1 + e^{-\frac{x-T}{\tau}}), \quad (5)$$

where  $x$  is the original value of each indicator. Two parameters  $T$  and  $\tau$  are used to control the symmetry point and the slope of the sigmoid function. The symmetry point indicates a certainty of 50%. The construction of BBAs is explained in [15] and adopted in this paper. In [15] these two parameters  $T$  and  $\tau$  are manually given to sigmoid curve. Here, the multi-level Otsu's thresholding method [16] is used to get symmetry points for both concordance index and discordance index. Otsu's algorithm defines that an image is composed of objects and background. A discriminant analysis is performed by minimizing the intra-class variance. When three classes are of interest, two threshold values are expected. Otsu's method can be extended to

$$\sigma_\omega^2(T_1, T_2) = \omega_1 \sigma_1^2(T_1, T_2) + \omega_2 \sigma_2^2(T_1, T_2) + \omega_3 \sigma_3^2(T_1, T_2). \quad (6)$$

The weights  $\omega_i$  are the probabilities obtained from the image histogram that are separated by the thresholds  $T_1$  and  $T_2$ .  $\sigma_i$  are the variances of the three classes.  $T_1$  and  $T_2$  can be used as the symmetry points of discordance and concordance index respectively. Thus, using height change index as example, the BBAs for discordance and concordance height change index are presented as  $a_{\Delta H}$  and  $b_{\Delta H}$

$$a_{\Delta H} = f_{\tau, T_1}(\Delta H), \quad \text{and} \quad b_{\Delta H} = f_{-\tau, T_2}(\Delta H). \quad (7)$$

The factor  $\tau$  is calculated with a sample value ( $\Delta H = 1$ ,  $a_{\Delta H} = 0.1$ ), which means 1 meter height change indicates 10% probability to be building changes. The BBAs for discordance and concordance image change index are built similarly. Differences appearing in 2D images give a concordance indication for all changes, which include the building changes and other changes ( $\theta_1 \cup \theta_2$ ). In this paper the changes from images are named  $\Delta Img$ .

#### C. BBAs construction using concordance and discordance

The BBAs related with the concordance and discordance indexes are combined to get the global BBA related to each source of evidence. These global BBAs will then be used as input for solving the change detection problem thanks to their combination. In the Tables I and II, we present the two ways of construction of the BBAs of the sources of evidence based either on DS or on PCR6 rules of combination for the height

TABLE I. BBA CONSTRUCTION FOR HEIGHT CHANGE INDICATOR  $\Delta H$ . [ $K_{\Delta H} = a_{\Delta H} b_{\Delta H}$ ]

Focal Elem.	$m_1(\cdot)$	$m'_1(\cdot)$	$m_1^{DS}(\cdot)$	$m_1^{PCR6}(\cdot)$
$\theta_1$	$a_{\Delta H}$	0	$\frac{a_{\Delta H}(1-b_{\Delta H})}{1-K_{\Delta H}}$	$a_{\Delta H}(1-b_{\Delta H}) + \frac{a_{\Delta H}K_{\Delta H}}{a_{\Delta H}+b_{\Delta H}}$
$\theta_2$	0	0	0	0
$\theta_3$	0	0	0	0
$\theta_1 \cup \theta_2$	0	0	0	0
$\theta_2 \cup \theta_3$	0	$b_{\Delta H}$	$\frac{(1-a_{\Delta H})b_{\Delta H}}{1-K_{\Delta H}}$	$(1-a_{\Delta H})b_{\Delta H} + \frac{b_{\Delta H}K_{\Delta H}}{a_{\Delta H}+b_{\Delta H}}$
$\theta_1 \cup \theta_2 \cup \theta_3$	$1-a_{\Delta H}$	$1-b_{\Delta H}$	$\frac{(1-a_{\Delta H})(1-b_{\Delta H})}{1-K_{\Delta H}}$	$(1-a_{\Delta H})(1-b_{\Delta H})$

 TABLE II. BBA CONSTRUCTION FOR IMAGE CHANGE INDICATOR  $\Delta Img$ . [ $K_{\Delta Img} = a_{\Delta Img} b_{\Delta Img}$ ]

Focal Elem.	$m_2(\cdot)$	$m'_2(\cdot)$	$m_2^{DS}(\cdot)$	$m_2^{PCR6}(\cdot)$
$\theta_1$	0	0	0	0
$\theta_2$	0	0	0	0
$\theta_3$	0	$b_{\Delta Img}$	$\frac{(1-a_{\Delta Img})b_{\Delta Img}}{1-K_{\Delta Img}}$	$(1-a_{\Delta Img})b_{\Delta Img} + \frac{b_{\Delta Img}K_{\Delta Img}}{a_{\Delta Img}+b_{\Delta Img}}$
$\theta_1 \cup \theta_2$	$a_{\Delta Img}$	0	$\frac{a_{\Delta Img}(1-b_{\Delta Img})}{1-K_{\Delta Img}}$	$a_{\Delta Img}(1-b_{\Delta Img}) + \frac{a_{\Delta Img}K_{\Delta Img}}{a_{\Delta Img}+b_{\Delta Img}}$
$\theta_2 \cup \theta_3$	0	0	0	0
$\theta_1 \cup \theta_2 \cup \theta_3$	$1-a_{\Delta Img}$	$1-b_{\Delta Img}$	$\frac{(1-a_{\Delta Img})(1-b_{\Delta Img})}{1-K_{\Delta Img}}$	$(1-a_{\Delta Img})(1-b_{\Delta Img})$

change indicator (i.e. the first source of evidence) and the image change indicator (i.e. the second source of evidence). In Table I,  $m_1(\cdot)$  and  $m'_1(\cdot)$  represent the concordance and discordance BBAs from  $\Delta H$ , whereas in Table II  $m_2(\cdot)$  and  $m'_2(\cdot)$  represent the concordance and discordance BBAs from images.

Here for comparison of the two belief functions, these two BBAs are fused with both DS and PCR6 fusion rules. The fusion rules for height change indicator and image change indicator are explained in Table I and Table II. In Table I, the  $m_1$  and  $m'_1$  represent the concordance and discordance BBAs from  $\Delta H$ . In Table II we use  $m_2$  and  $m'_2$  to represent the concordance and discordance BBAs from images.

#### D. BBAs combination for building change detection

From the previous step of BBAs modelings, each pixel will get two sets of BBAs to combine resulting from Table I and II. More precisely, we will have to combine either  $\{m_1^{DS}(\cdot), m'_1^{DS}(\cdot)\}$  if DS rule is preferred for the BBA modeling, or  $\{m_1^{PCR6}(\cdot), m'_1^{PCR6}(\cdot)\}$  if PCR6 rule is adopted. These BBAs have been represented by  $a_1, b_1, c_1$  and  $a_2, b_2, c_2$  in Table III.

TABLE III. FUSION MODELS FOR BUILDING CHANGE DETECTION.

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$	$m_{12}^{DS}(\cdot)$	$m_{12}^{PCR6}(\cdot)$
$\theta_1$	$a_1$	0	$\frac{a_1(b_1+b_3)}{1-a_1b_2}$	$a_1(b_1+b_3) + \frac{a_1a_1b_2}{a_1+b_2}$
$\theta_2$	0	0	$\frac{a_2b_1}{1-a_1b_2}$	$a_2b_1$
$\theta_3$	0	$b_2$	$\frac{(a_2+a_3)b_2}{1-a_1b_2}$	$(a_2+a_3)b_2 + \frac{b_2a_1b_2}{a_1+b_2}$
$\theta_1 \cup \theta_2$	0	$b_1$	$\frac{a_3b_1}{1-a_1b_2}$	$a_3b_1$
$\theta_2 \cup \theta_3$	$a_2$	0	$\frac{a_2b_3}{1-a_1b_2}$	$a_2b_3$
$\Theta$	$a_3$	$b_3$	$\frac{a_3b_3}{1-a_1b_2}$	$a_3b_3$

Based on different BBAs and fusion methods, four sets of global BBAs can be computed from Table III.

$$\begin{aligned}
 G_1 &= DS\{m_1^{DS}(\cdot), m'_1^{DS}(\cdot)\}, \\
 G_2 &= PCR6\{m_1^{DS}(\cdot), m'_1^{DS}(\cdot)\}, \\
 G_3 &= DS\{m_1^{PCR6}(\cdot), m'_1^{PCR6}(\cdot)\}, \\
 G_4 &= PCR6\{m_1^{PCR6}(\cdot), m'_1^{PCR6}(\cdot)\}.
 \end{aligned} \tag{8}$$

After the fusion step, each pixel in the images will get a certain degree of belief for all focal elements. Based on the

these BBAs, a final decision can be made. DST and DSMT have different approaches to get this final decision. In this paper four decision criteria are tested. More precisely, we have evaluated the maximum of global BBAs (Max\_Bel), maximum of plausibility (Max\_Pl), maximum of betting probabilities (Max\_BetP) and the maximum of DSMT (Max\_DSMT), see [8] (Vol. 3, Chap. 3) and [10] for the mathematical definitions of  $Bel(\cdot)$ ,  $Pl(\cdot)$ ,  $BetP(\cdot)$  and  $DSMT(\cdot)$  functions.

#### IV. EXPERIMENTS

The two proposed BBAs modelings and fusion methods (based on DS and PCR6 rules) have been tested on one real dataset. The dataset and the results from each step are detailed in this section.

##### A. Datasets

The experimental dataset for this research work are displayed in Fig.1. It consists of two pairs of IKONOS stereo imagery captured at February 2006 and May 2011 respectively. As a pre-processing step, all data have been correctly radiometrically and geographically co-registered as described in [6]. As shown in Fig. 1, this is a normal building change example. Several buildings have been built on flat surface. The generated DSMs are displayed in Fig. 1c and d.

##### B. Results and evaluation

As the first step, BBAs from image change and height change are extracted and refined based on DS fusion and PCR6 fusion rules. The four sets of global BBAs are prepared corresponding to Eq. (8). Among them the BBA for the focal element  $\theta_1$  (Building change) are shown in Fig. 2. The accuracy of these BBAs have been evaluated by area under Receiver Operating Characteristic curve (AUC). The AUC has been recorded on this figure as the caption of each subfigure. An advantage of PCR6 can be proved here. It has to be noted that the AUCs obtained here are much higher than using only height (AUC = 0.9299) [6] or spectral information (AUC=0.8823), and generally better than the fusion result described in [6] (AUC=0.9621).

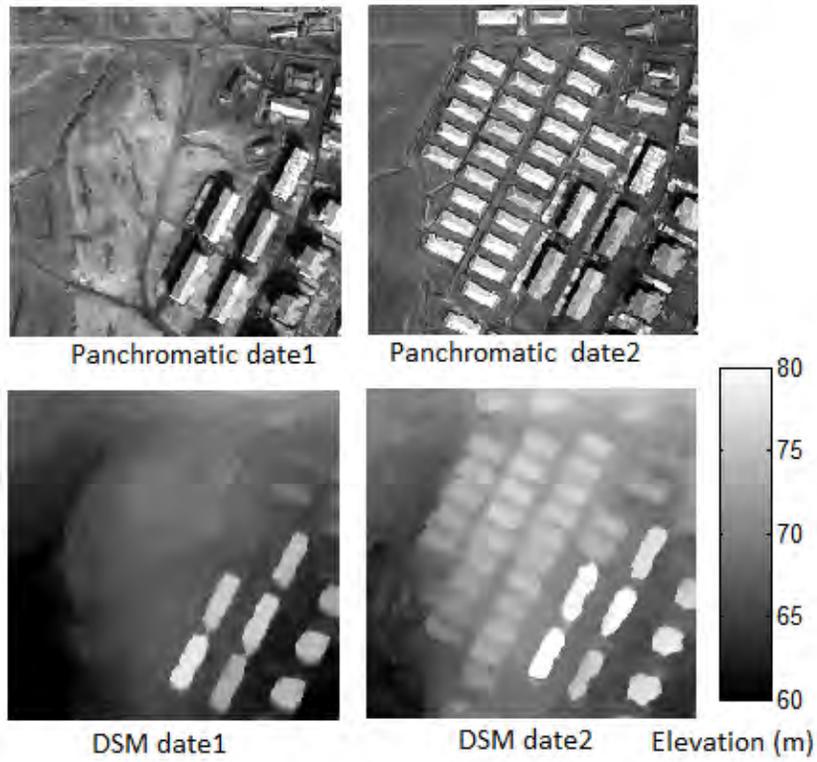


Fig. 1. Experimental dataset: a) panchromatic image from date1; b) panchromatic image from date2; c) DSM from date1; (d) DSM from date2.

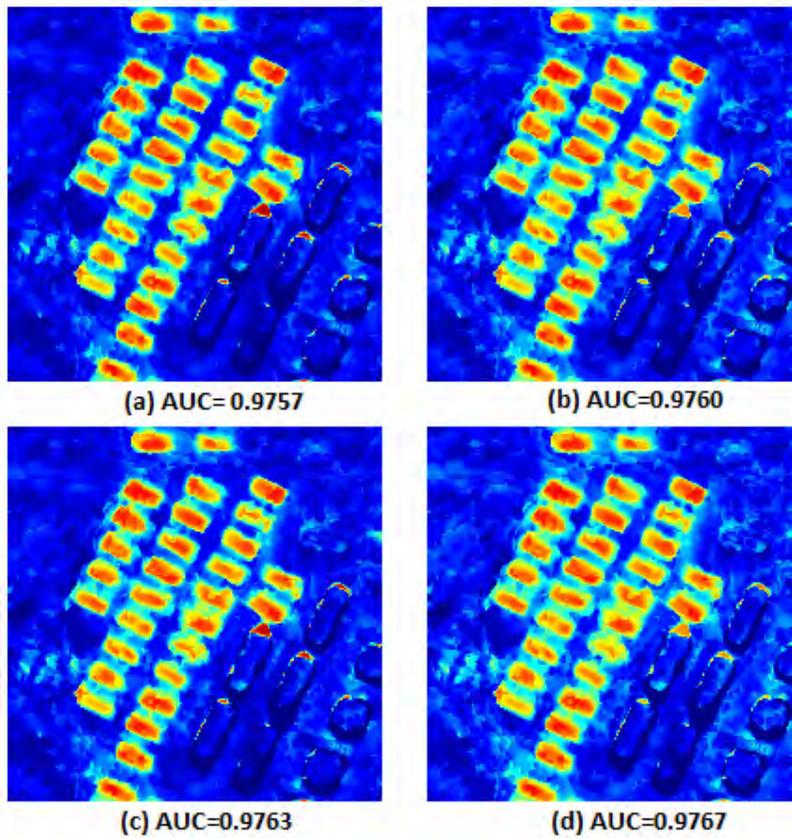


Fig. 2. Four global BBAs sets (a)  $G_1$ ; (b)  $G_2$ ; (c)  $G_3$ ; (d)  $G_4$ .

Besides the AUC comparison, the building change masks extracted from these four global BBAs sets are compared and evaluated. Each global BBA set can generate four building change mask based on these four decision make criteria. These building change masks are evaluated based on Kappa statistic (KA) and true detected rate (TR). In this paper  $TR = \frac{\text{detected positive}}{\text{true positive}} \times 100$  (in %). The comparison results of TR and KA values are shown in Table IV. From Table IV, one sees that,  $G_3$  and  $G_4$  are more advantageous than  $G_1$  and  $G_2$ . However, the highest KA is obtained by  $G_1$  by taking the Max\_Pl. However, in this paper, only the reference data for building changes are available. For better understanding these four global BBAs and decision making criteria, reference data of all three focal elements  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  are required.

TABLE IV. CHANGE MASKS EVALUATION FROM FOUR GLOBAL BBAS.

	$G_1$		$G_2$		$G_3$		$G_4$	
	TR [%]	KA	TR[%]	KA	TR [%]	KA	TR [%]	KA
Max_Bel	93.35	0.7729	93.35	0.7729	93.39	0.7725	93.39	0.7724
Max_Pl	93.23	0.7768	93.23	0.7762	93.23	0.7763	93.25	0.7756
Max_BetP	93.28	0.7747	93.32	0.7762	93.32	0.7745	93.32	0.7741
Max_DSMP	93.30	0.7739	93.30	0.7734	93.30	0.7737	93.34	0.7734

## V. CONCLUSIONS

Belief functions are good choices for DSM assisted change detection. Firstly, once the BBA construction is well done, it can be robustly used for other images in other regions efficiently. Secondly, this fusion approach matches well with the characteristics of our research topic. Since height information is important for separating high/low level objects. Satellite images directly highlight all changes on the land surface. None of these two sources of information can easily and directly lead to a reliable decision on building changes, which matches with the initial idea of belief functions. Generally speaking, both DST and DSMT frameworks offer the possibility to reach a high accuracy result, and PCR6 looks advantageous when a larger conflict exists between the different sources of evidence. More experiments are under progress to provide a finer quantitative comparative analysis in a forthcoming publication.

## REFERENCES

- [1] A. Singh, *Digital change detection techniques using remotely-sensed data*, Int. J. Remote Sens., Vol. 10, no. 6, pp. 989–1003, 1989.
- [2] D. Lu, P. Mausel, E. Brondizio, E. Moran, *Change detection techniques*, Int. J. Remote Sens., Vol. 25, no. 12, pp. 2365–2407, 2004.
- [3] F. Rottensteiner, J. Trinder, S. Clode, K. Kubik, *Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection*, Information Fusion, Vol. 6, no. 4, pp. 283–300, 2005.
- [4] N. Champion, F. Rottensteiner, L. Matikainen, X. Liang, J. Hyyppä, and B. Olsen, *A test of automatic building change detection approaches*, in Proc. of CMRT09, pp. 03–04, 2009.
- [5] J. Tian, P. Reinartz, P. d’Angelo, M. Ehlers, *Region-based automatic building and forest change detection on Cartosat-1 stereo imagery*, ISPRS J. Photogramm. and Remote Sensing, Vol. 79, pp. 226–239, 2013.
- [6] J. Tian, S. Cui, P. Reinartz, *Building change detection based on satellite stereo imagery and digital surface models*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 52, no. 1, pp. 406–417, 2014.
- [7] J. Tian, P. Reinartz, *Comparison of two fusion based building change detection methods using satellite stereo imagery and DSMs*, ISPRS-Int. Archives the Photogramm., Remote Sens. and Spatial Information Sciences, Vol. XL-7, no. W1, pp.103–108, 2013.

- [8] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, Vol. 1–3, American Research Press, Rehoboth, NM, USA, 2004–2009. [Online] Available: [www.onera.fr/fr/staff/jean-dezert/references](http://www.onera.fr/fr/staff/jean-dezert/references)
- [9] J. Dezert, A. Tchamova, *On the validity of Dempster’s fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. Intell. Syst., Vol. 29, no. 3, pp. 223–252, 2014.
- [10] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [11] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Int. Conf. on Information Fusion (Fusion 2012), pp. 655–660, Singapore, July 2012. [Online]. Available: <http://fs.gallup.unm.edu/DSMT.htm>
- [12] A. Tchamova, J. Dezert, *On the behavior of Dempster’s rule of combination and the foundations of Dempster-Shafer theory*, in Proc. of IEEE IS’2012, pp. 108–113, Sofia, Bulgaria, Sept. 6–8, 2012.
- [13] F. Smarandache, J. Dezert, *(On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of Int. Conf. on Information Fusion (Fusion 2013), pp. 1119–1126, Istanbul, Turkey, July 9–12, 2013.
- [14] A.A. Nielsen, *The regularized iteratively reweighted MAD method for change detection in multi-and hyperspectral data*, IEEE Trans. on Image Processing, Vol. 16, no. 2, pp. 463–478, 2007.
- [15] J. Dezert, J.-M. Tacnet, *Sigmoidal model for belief function-based ELECTRE TRI method*, in Belief Functions: Theory and Applications, pp. 401–408, 2012.
- [16] N. Otsu, *A threshold selection method from gray-level histograms*, IEEE Trans. on Syst., Man, Cybern., Vol. 9, no. 1, pp. 62–66, 1975.



# Refined Building Change Detection in Satellite Stereo Imagery Based on Belief Functions and Reliabilities

Jiaojiao Tian<sup>a</sup>, Jean Dezert<sup>b</sup>, Peter Reinartz<sup>a</sup>

<sup>a</sup>Remote Sensing Technology Institute, German Aerospace Center, 82234 Oberpfaffenhofen, Germany.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

Emails: jiaojiao.tian@dlr.de, jean.dezert@onera.fr, peter.reinartz@dlr.de

Originally published as: J. Tian, J. Dezert, P. Reinartz, *Refined Building Change Detection in Satellite Stereo Imagery Based on Belief Functions and Reliabilities*, in Proc. of IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI2015), Extended Studies Center, San Diego State Univ., San Diego, CA, USA, Sept. 14–16, 2015, and reprinted with permission.

**Abstract**—Digital Surface Models (DSMs) generated from satellite stereo imagery provide valuable but not comprehensive information for building change detection. Therefore, belief functions have been introduced to solve this problem by fusing DSM information with changes extracted from images. However, miss-detection can not be avoided if the DSMs are containing large region of wrong height values. A refined workflow is thereby proposed by adopting the initial disparity map to generate a reliability map. This reliability map is then built in the fusion model. The reliability map has been tested in both Dempster-Shafer Theory (DST), and Dezert-Smarandache Theory (DSMT) frameworks. The results have been validated by comparing to the manually extracted change reference mask.

**Keywords:** belief functions, DSMT, satellite imaging, building change detection.

## I. INTRODUCTION

In our previous research [1] [2], belief functions have performed very well for 3D building change detection. As we have mentioned, the accuracy of 2D change detection is limited due to the misdetections caused by irrelevant changes. These irrelevant changes have a larger effect on very high resolution (VHR) images since many details of building changes are expected. The DSMs generated from satellite stereo imagery can largely help to solve this problem. However, the DSMs may still exhibit some outliers resulting in occlusions within the stereo/multi views and due to matching mistakes. In this case, change information from spectral information of the original stereo imagery can and should be used together with height changes to eventually highlight building changes. For this purpose proper fusion theories and approaches are needed.

In paper [2], the belief functions introduced in the Dempster-Shafer Theory (DST) [3] [4], and extended in Dezert-Smarandache Theory (DSMT) [5] are used to deal with the uncertainty information delivered from the DSMs. In [2] the possibility of using Dempster's fusion rule and the Proportional Conflict Redistribution Rule #6 (PCR6) of DSMT in our application have been tested. Though improvements have been proven by comparing to the method stated in [1], false alarms can not be avoided in case of large regions of wrong height change values. Thereupon, in this paper the reliability map is adopted as an additional source of evidence to correct the basic Belief Assignments (BBAs) and thus refine the fusion model.

This paper is organized as follow. Firstly, the belief functions and building change detection fusion models are

briefly reviewed. Then, the reliability discounting techniques are presented and the reliability map is generated. Later, the final four global BBAs are described together with the four decision criteria with which the final change detection mask can be generated. In the end, these refined fusion models are tested on two sets of satellite real data.

## II. BELIEF FUNCTION BASED BUILDING CHANGE DETECTION

### A. Basics of belief functions

Detailed presentations of DST and DSMT can be found in [5], [6] and [3]. Let  $\Theta$  be a frame of discernment of a problem under consideration.  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  consists of a list of  $N$  exhaustive and mutually exclusive elements  $\theta_i$ ,  $i = 1, 2, \dots, N$ . Each  $\theta_i$  represents a possible state related to the problem we want to solve. The assumption of exhaustivity and mutual exclusivity of elements of  $\Theta$  is classically referred as *Shafer's model* of the frame  $\Theta$ . A BBA also called a belief mass function (or just a mass for short), is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  from the power set<sup>1</sup> of  $\Theta$  denoted  $2^\Theta$  to  $[0, 1]$ , that verifies [3]:

$$m(\emptyset) = 0, \quad \text{and} \quad \sum_{X \in 2^\Theta} m(X) = 1. \quad (1)$$

$m(X)$  represents the mass of belief exactly committed to  $X$ . An element  $X \in 2^\Theta$  is called a focal element if and only if  $m(X) > 0$ . In DST, the combination (fusion) of several independent sources of evidences is done with Dempster-Shafer<sup>2</sup> (DS) rule of combination, assuming that the sources are not in total conflict<sup>3</sup>. DS combination of two independent BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$ , denoted symbolically by  $DS(m_1, m_2)$ , is defined by  $m^{DS}(\emptyset) = 0$ , and for all  $X \in 2^\Theta \setminus \{\emptyset\}$  by:

$$m^{DS}(X) = \frac{1}{1 - K^{DS}} \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2), \quad (2)$$

<sup>1</sup>The power set is the set of all subsets of  $\Theta$ , empty set included.

<sup>2</sup>Although the rule has been proposed originally by Dempster, we call it Dempster-Shafer rule because it has been widely promoted by Shafer in DST.

<sup>3</sup>otherwise DS rule is mathematically not defined because of 0/0 indeterminacy.

where the total degree of conflict  $K^{DS}$  is given by

$$K^{DS} \triangleq \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2). \quad (3)$$

A discussion on the validity of DS rule and its incompatibility with Bayes fusion rule for combining Bayesian BBAs can be found in [6], [7], [8]. To circumvent the problems of DS rule, Smarandache and Dezert (see [5], Vol. 2, Chap. 1), then Martin and Osswald (see [5], Vol. 2, Chap. 2) have developed in DSMT [5] two fusion rules called PCR5 and PCR6 based on the proportional conflict redistribution (PCR) principle which consists

- 1) to apply the conjunctive rule;
- 2) calculate the total or partial conflicting masses;
- 3) then redistribute the (total or partial) conflicting mass proportionally on non-empty sets according to the integrity constraints one has for the frame  $\Theta$ .

This PCR principle transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is not degraded. Because the proportional transfer can be done in two different ways, this has yielded to two different fusion rules. It has been proved in [9] that only PCR6 rule is compatible with frequentest probability estimation, and that is why we recommend its use in the applications. PCR5 and PCR6 rules simplify greatly and coincide for the combination of two sources. In this case, the PCR6 combination is obtained by taking  $m^{PCR6}(\emptyset) = 0$ , and for all  $X \neq \emptyset$  in  $2^\Theta$  by

$$m^{PCR6}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right], \quad (4)$$

where all denominators in Eq. (4) are different from zero. If a denominator is zero, that fraction is discarded. If a denominator, e.g.,  $m_1(X) + m_2(Y)$  tends towards 0, then also the conflicting mass  $m_1(X)m_2(Y)$  that is transferable tends to zero because  $m_1(X)$  and  $m_2(Y)$  tend to zero (since they are positive), therefore the redistribution masses also tend to zero. That reflects the continuity of PCR6.

### B. BBAs for Building change detection

1) *Choice of the frame of discernment*: Focusing on building change detection, two change indicators, one from images and one from DSMs are used. Changes from spectral images are highlighted by using the Iteratively Reweighted Multivariate Alteration Detection (IRMAD) [10]. Consequently height changes from DSMs are shown after robust height difference [1]. Three classes are considered to define the frame of discernment satisfying Shafer's model:

$$\Theta = \{\theta_1 \triangleq \text{Pixel} \in \text{BuildingChange}, \\ \theta_2 \triangleq \text{Pixel} \in \text{OtherChange}, \\ \theta_3 \triangleq \text{Pixel} \in \text{NoChange}\}, \quad (5)$$

and

$$\theta_1 \cap \theta_2 \cap \theta_3 = \emptyset. \quad (6)$$

Based on the three classes, the set of focal elements  $FE$  that are of interest in our application is:

$$FE = \{\theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3\}. \quad (7)$$

2) *BBAs construction*: Paper [2] constructed the sigmoidal model for both concordance and discordance indexes. The details and advantages of this approach are described in [11]. The concordance index measures the concordance of change indicator and BBA in the assertion, while the discordance measures the opposition of change indicator to the BBAs in the assertion. The original sigmoid curve is defined as

$$f_{(\tau, T)}(x) = 0.99 / (1 + e^{-\frac{x-T}{\tau}}), \quad (8)$$

where  $x$  is the original value of each indicator. Two parameters  $T$  and  $\tau$  are used to control the symmetry point and the slope of the sigmoid function. The symmetry point indicates a certainty of 50%. In [11] these two parameters  $T$  and  $\tau$  are manually given. Here, the multi-level Otsu's thresholding method [12] is used for automatically getting the symmetry points for both concordance index and discordance index. Otsu's algorithm defines that an image is composed of objects and background. A discriminant analysis is performed by minimizing the intra-class variance. When three classes are of interest, two threshold values are expected. Otsu's method can be extended to

$$\sigma_\omega^2(T_1, T_2) = \omega_1 \sigma_1^2(T_1, T_2) + \omega_2 \sigma_2^2(T_1, T_2) + \omega_3 \sigma_3^2(T_1, T_2). \quad (9)$$

The weights  $\omega_i$  are the probabilities obtained from the image histogram that are separated by the thresholds  $T_1$  and  $T_2$ .  $\sigma_i$  is the standard deviation of the  $i$ -th class, for  $i = 1, 2, 3$ .  $T_1$  and  $T_2$  can be used as the symmetry points of discordance and concordance index respectively. Thus, using height change index as example, the BBAs for discordance and concordance height change index are functions of values  $a_{\Delta H}$  and  $b_{\Delta H}$  defined by

$$a_{\Delta H} = f_{\tau, T_1}(\Delta H), \quad \text{and} \quad b_{\Delta H} = f_{-\tau, T_2}(\Delta H). \quad (10)$$

The factor  $\tau$  is calculated with a sample value ( $\Delta H = 1$ ,  $a_{\Delta H} = 0.1$ ), which means 1 meter height change indicates 10% probability to be building changes. The BBAs for discordance and concordance image change index are built similarly. Differences appearing in 2D images give a concordance indication for all changes, which include the building changes and other changes ( $\theta_1 \cup \theta_2$ ). In this paper the changes from images are named  $\Delta Img$ .

In the Tables I and II, we present the two ways of construction of the BBAs from the sources of evidence based either on DS or on PCR6 rules of combination for the height change indicator (i.e. the first source of evidence) and the image change indicator (i.e. the second source of evidence). In Table I,  $m_1(\cdot)$  and  $m'_1(\cdot)$  represent the concordance and discordance BBAs from  $\Delta H$ , whereas in Table II  $m_2(\cdot)$  and  $m'_2(\cdot)$  represent the concordance and discordance BBAs from images.

TABLE I. BBA CONSTRUCTION FOR HEIGHT CHANGE INDICATOR  $\Delta H$ . [ $K_{\Delta H} = a_{\Delta H} b_{\Delta H}$ ]

Focal Elem.	$m_1(\cdot)$	$m'_1(\cdot)$	$m_1^{DS}(\cdot)$	$m_1^{PCR6}(\cdot)$
$\theta_1$	$a_{\Delta H}$	0	$\frac{a_{\Delta H}(1-b_{\Delta H})}{1-K_{\Delta H}}$	$a_{\Delta H}(1-b_{\Delta H}) + \frac{a_{\Delta H}K_{\Delta H}}{a_{\Delta H}+b_{\Delta H}}$
$\theta_2$	0	0	0	0
$\theta_3$	0	0	0	0
$\theta_1 \cup \theta_2$	0	0	0	0
$\theta_2 \cup \theta_3$	0	$b_{\Delta H}$	$\frac{(1-a_{\Delta H})b_{\Delta H}}{1-K_{\Delta H}}$	$(1-a_{\Delta H})b_{\Delta H} + \frac{b_{\Delta H}K_{\Delta H}}{a_{\Delta H}+b_{\Delta H}}$
$\theta_1 \cup \theta_2 \cup \theta_3$	$1-a_{\Delta H}$	$1-b_{\Delta H}$	$\frac{(1-a_{\Delta H})(1-b_{\Delta H})}{1-K_{\Delta H}}$	$(1-a_{\Delta H})(1-b_{\Delta H})$

 TABLE II. BBA CONSTRUCTION FOR IMAGE CHANGE INDICATOR  $\Delta Img$ . [ $K_{\Delta Img} = a_{\Delta Img} b_{\Delta Img}$ ]

Focal Elem.	$m_2(\cdot)$	$m'_2(\cdot)$	$m_2^{DS}(\cdot)$	$m_2^{PCR6}(\cdot)$
$\theta_1$	0	0	0	0
$\theta_2$	0	0	0	0
$\theta_3$	0	$b_{\Delta Img}$	$\frac{(1-a_{\Delta Img})b_{\Delta Img}}{1-K_{\Delta Img}}$	$(1-a_{\Delta Img})b_{\Delta Img} + \frac{b_{\Delta Img}K_{\Delta Img}}{a_{\Delta Img}+b_{\Delta Img}}$
$\theta_1 \cup \theta_2$	$a_{\Delta Img}$	0	$\frac{a_{\Delta Img}(1-b_{\Delta Img})}{1-K_{\Delta Img}}$	$a_{\Delta Img}(1-b_{\Delta Img}) + \frac{a_{\Delta Img}K_{\Delta Img}}{a_{\Delta Img}+b_{\Delta Img}}$
$\theta_2 \cup \theta_3$	0	0	0	0
$\theta_1 \cup \theta_2 \cup \theta_3$	$1-a_{\Delta Img}$	$1-b_{\Delta Img}$	$\frac{(1-a_{\Delta Img})(1-b_{\Delta Img})}{1-K_{\Delta Img}}$	$(1-a_{\Delta Img})(1-b_{\Delta Img})$

### III. RELIABILITY DISCOUNTING

The reliability discounting has been described and discussed in the references [13] and [14]. Briefly said, if an additional knowledge about the reliability ( $\alpha$ ) of certain indicator ( $X$ ) is available, it can be adopted to refine the initial BBAs.  $\alpha$  would be a value ranging from 0 to 1. And  $\alpha = 1$  means fully reliable, while  $\alpha = 0$  means the indicator is totally unreliable. Based on Shafer's discounting model [3], the reliability discounting factor  $\alpha$  is introduced to discount any BBA  $m(\cdot)$  defined on the power set  $2^\Theta$  as follows:

$$\begin{cases} m_\alpha(X) = \alpha \cdot m(X), \text{ for } X \neq \Theta, \\ m_\alpha(\Theta) = \alpha \cdot m(\Theta) + (1 - \alpha). \end{cases} \quad (11)$$

In the DSM assisted building change detection, false alarms are detected if wrong heights are present in DSM for large regions [1]. And these wrong heights are mostly introduced not in the stereoscope images matching procedure, but in the gaps filling step. In our DSM generation procedure, the height of un-matched pixels are interpolated using the height values of neighborhood pixels. Therefore, a reliable height value can be achieved for small gaps. When large gaps turn up in the disparity map, for example, a whole building roof, the height of that building can not be correctly interpolated. Thus, the percentage of available correctly matched neighborhood pixels inside a predefined region can be used to generate the height reliability. Fig. 1 shows an example of the generated reliability map. Fig. 1a is the gaps mask. The gaps region of the disparity map is represented with black color. Pixels with proper elevation values are displayed with white color. It can be observed, based on our approach that pixels in the center of a gap get lower reliability factor values than pixels next to the gap boundary (see Fig.1b).

In the building change detection procedure, the reliability map of two DSMs ( $\alpha_{DSM1}$  and  $\alpha_{DSM2}$ ) are calculated respectively. They are then fused together to generate a final reliability map  $\alpha_{\Delta H}$  for the height change indicator.

$$\alpha_{\Delta H} = \alpha_{DSM1} \cdot \alpha_{DSM2}. \quad (12)$$

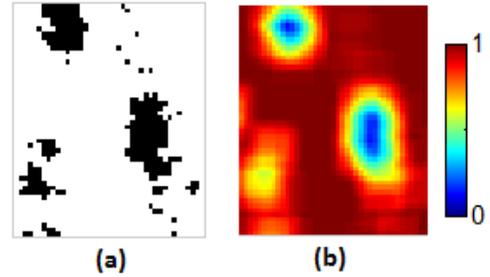


Fig. 1. Reliability map (b) generated from the gaps mask (a).

### IV. GLOBAL BBAs AND CHANGE DETECTION

#### A. Global BBAs generation

The BBAs related with the concordance and discordance indexes are combined to get the global BBA regarding to each source of evidence. These global BBAs will then be used as input for solving the change detection problem thanks to their combination. From the previous step of BBAs modelings, each pixel will get two sets of BBAs to combine results from Table I and II. More precisely, we will have to combine either  $\{m_1^{DS}(\cdot), m_2^{DS}(\cdot)\}$  if DS rule is preferred for the BBA modeling, or  $\{m_1^{PCR6}(\cdot), m_2^{PCR6}(\cdot)\}$  if the PCR6 rule is adopted. These BBAs from Table I and II have been represented by  $a_1, b_1, c_1$  and  $a_2, b_2, c_2$ . In this paper, the mass values  $a_1, b_1$ , and  $c_1$  are further discounted by the generated reliability map  $\alpha_{\Delta H}$  and denoted respectively as  $A_1, B_1$ , and  $C_1$ . More precisely, one computes

$$\begin{cases} A_1 = \alpha_{\Delta H} \cdot a_1, \\ B_1 = \alpha_{\Delta H} \cdot b_1, \\ C_1 = \alpha_{\Delta H} \cdot c_1 + (1 - \alpha_{\Delta H}). \end{cases} \quad (13)$$

In this application, only the reliability map for height change indicators is generated. The reliability map for image change indicators can also be constructed according to the change objects of interested. For instance, vegetation mask can be used to discount the reliability of building changes. However, this paper focuses on the reliability of height information.

When the reliability map of image changes is available, it could be used as the same way as height change reliability map. Table III and Table IV describe the final building change detection models based either on DS or on PCR6 rules. Here, the discounted height change indicators is denoted as  $m_{1\alpha\Delta H}(\cdot)$ .

TABLE III. DS FUSION MODEL FOR BUILDING CHANGE DETECTION.

Focal Elem.	$m_{1\alpha\Delta H}(\cdot)$	$m_2(\cdot)$	$m_{12}^{DS}(\cdot)$
$\theta_1$	$A_1$	0	$\frac{A_1(b_1+b_3)}{1-A_1b_2}$
$\theta_2$	0	0	$\frac{A_2b_1}{1-A_1b_2}$
$\theta_3$	0	$b_2$	$\frac{(A_2+a_3)b_2}{1-A_1b_2}$
$\theta_1 \cup \theta_2$	0	$b_1$	$\frac{A_3b_1}{1-A_1b_2}$
$\theta_2 \cup \theta_3$	$A_2$	0	$\frac{A_2b_3}{1-A_1b_2}$
$\Theta$	$A_3$	$b_3$	$\frac{A_3b_3}{1-A_1b_2}$

TABLE IV. PCR6 FUSION MODEL FOR BUILDING CHANGE DETECTION.

Focal Elem.	$m_{1\alpha\Delta H}(\cdot)$	$m_2(\cdot)$	$m_{12}^{PCR6}(\cdot)$
$\theta_1$	$A_1$	0	$A_1(b_1 + b_3) + \frac{A_1 A_1 b_2}{A_1 + b_2}$
$\theta_2$	0	0	$A_2 b_1$
$\theta_3$	0	$b_2$	$(A_2 + a_3)b_2 + \frac{b_2 A_1 b_2}{A_1 + b_2}$
$\theta_1 \cup \theta_2$	0	$b_1$	$A_3 b_1$
$\theta_2 \cup \theta_3$	$A_2$	0	$A_2 b_3$
$\Theta$	$A_3$	$b_3$	$A_3 b_3$

$m_{1\alpha\Delta H}(\cdot)$  can be obtained from the discounting of the fusion results presented in Table I. Thus they have been denoted respectively as  $m_{1\alpha\Delta H}^{DS}(\cdot)$  and  $m_{1\alpha\Delta H}^{PCR6}(\cdot)$ . This discounted height change indicators are fused in the second step with image change indicator  $m_2(\cdot)$  to generate the final global BBAs. From the tables III and IV, four sets of global BBAs can be computed based on different BBAs and fusion methods as follows:

$$\begin{aligned}
 G_1 &= DS\{m_{1\alpha\Delta H}^{DS}(\cdot), m_2^{DS}(\cdot)\}, \\
 G_2 &= PCR6\{m_{1\alpha\Delta H}^{DS}(\cdot), m_2^{DS}(\cdot)\}, \\
 G_3 &= DS\{m_{1\alpha\Delta H}^{PCR6}(\cdot), m_2^{PCR6}(\cdot)\}, \\
 G_4 &= PCR6\{m_{1\alpha\Delta H}^{PCR6}(\cdot), m_2^{PCR6}(\cdot)\}.
 \end{aligned} \tag{14}$$

For example, if both the BBA modeling procedure and global BBAs are constructed based on DS fusion rule, the generated global BBA is recorded as  $G_1$ .

### B. Change mask generation

After the fusion step, each pixel in the images will get a certain degree of belief for all focal elements. The value of global BBAs in  $\theta_1$  gives a direct building change probability map. A change mask can be generated after giving a threshold value. However, BBAs on the partial ignorance and full ignorance set should also be considered in the decision making procedure. DST and DSMT propose different approaches to take the final decision. In this work, the same decision criteria as used in [2] are tested. They are: 1) maximum of global BBAs (Max\_Bel), 2) maximum of plausibility (Max\_Pl), 3) maximum of betting probabilities (Max\_BetP) and 4) the maximum of DSMP (Max\_DSMP). The reader can refer to [3] and [5] (Vol. 3, Chap. 3) for the mathematical definitions of  $Bel(\cdot)$ ,  $Pl(\cdot)$ ,  $BetP(\cdot)$  and  $DSMP(\cdot)$  functions.

## V. EXPERIMENTS

The improved building change detection fusion models have been tested on satellite images. The datasets and the experiments are described in this section.

### A. Datasets

The experimental datasets consist of two pairs of IKONOS stereo imagery captured in February 2006 and May 2011 respectively shown in Fig. 2 and 3. The first two images in each figure are the panchromatic images of two dates. (c) and (d) are the generated DSMs. They have been generated based on the method explained in [15]. The colors represent the height range in this test region.

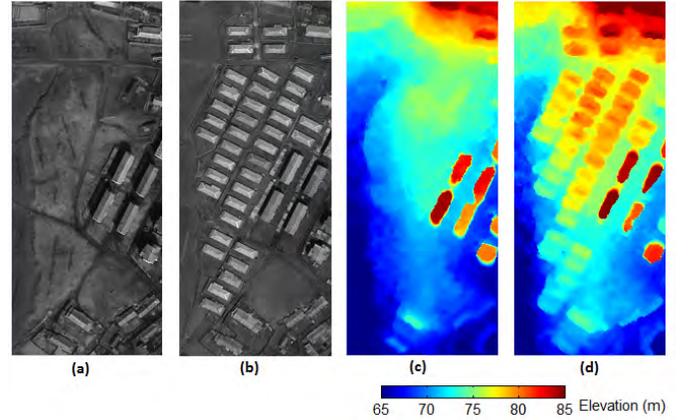


Fig. 2. Experimental dataset: a) panchromatic image from date1; b) panchromatic image from date2; c) DSM from date1; d) DSM from date2.

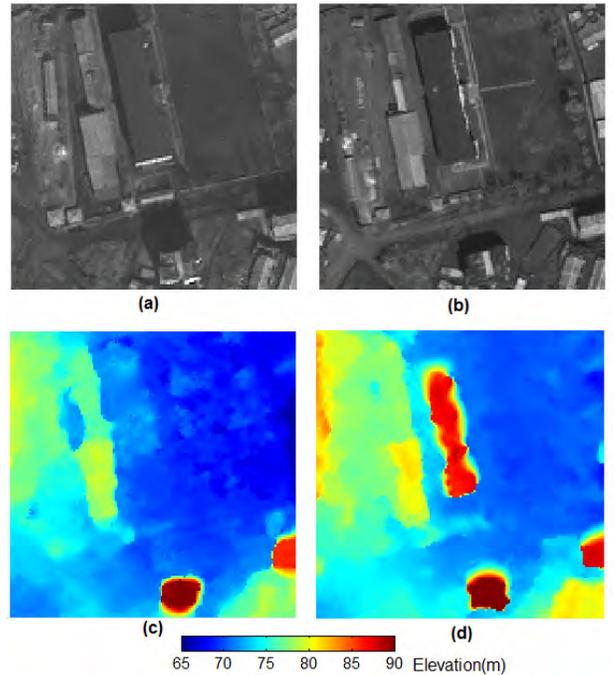


Fig. 3. Datasets of the 2nd test region; a) panchromatic image from date1; b) panchromatic image from date2; c) DSM from date1; d) DSM from date2.

The spatial co-registration is achieved through camera model parameter corrections before the DSM generation procedure [15]. The radiometrical co-registration method has been described in [1]. Fig. 2 shows a normal building change example. Several buildings have been built on flat surface. The generated DSMs are displayed in Fig. 2c and d. In the second example (shown in Fig. 3), a large percentage of pixels on the roof of the large building in the center appear as gaps in the disparity map. In the filling procedure, the large size of the gap in the date1 data lead to the missing of this building in the DSM (Fig. 3c).

### B. Results and evaluation

The refined DS fusion model and PCR6 fusion model have been applied to both datasets respectively. To show the improvement obtained by our method, we have compared its results with the original results we can obtain with the method in [2]. Firstly, the global BBAs of  $\theta_1$  are compared and displayed in Fig. 4 below.

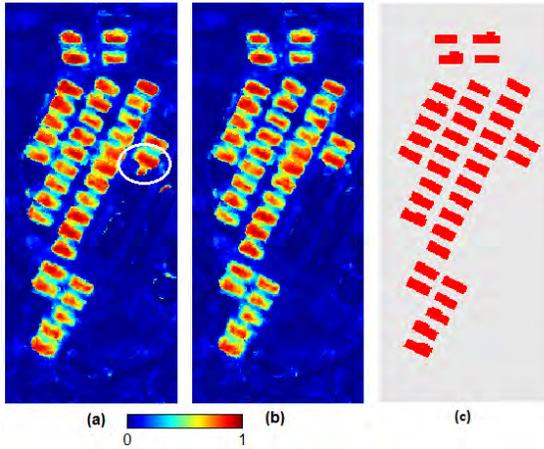


Fig. 4. Global Building change BBAs (a) Initial result; (b) Refined result; (c) Ground truth.

Fig. 4(a) corresponds to the original<sup>4</sup> result, and Fig. 4(b) shows the refined result based on  $G_1(\theta_1)$ . By comparing to the ground truth (Fig. 4(c)), the improvements can be clearly observed in the building boundary regions, especially the building marked with a white circle. In the initial result, the pixels next to this building are falsely detected as BuildingChange.

To evaluate quantitatively the performances of the different fusion approaches, the extracted BBAs from both approaches (original and refined) are compared to the manually extracted change reference masks. The results are analyzed in terms of Receiver Operating Characteristic (ROC) curve [16]. A larger area under the ROC curve (AUC) indicates a better accuracy of the building change map. The numerical evaluation results are described in Table V. The obtained AUC values prove a general improvement after reliability discounting is applied.

In addition to the AUC comparison, the building change masks extracted from these four global BBAs sets are compared and evaluated. Each global BBA set can generate four building change masks based on these four decision criteria.

<sup>4</sup>obtained without reliability discounting, as presented in [2].

TABLE V. QUALITY COMPARISON OF GLOBAL BBA (BUILDING CHANGE).

	Test Region 1		Test Region 2	
	Original	Refined	Original	Refined
$G_1$	0.9811	0.9833	0.9509	0.9950
$G_2$	0.9829	0.9839	0.9485	0.9931
$G_3$	0.9815	0.9837	0.9512	0.9955
$G_4$	0.9835	0.9844	0.9487	0.9939

These building change masks are compared with the masks from paper [2] based on Kappa statistic (KA). The comparison results of Test region 1 are shown in Table VI. Limited by the reference data we can get, only the building change frame is evaluated here. One sees the reliability discounting map helps to improve the result accuracy in all fusion and decision approaches.

In the second test region, there is actually no building changes. The purpose of showing this test region is to further prove the advantage of the extracted reliability map. Fig. 5 shows the extracted reliability discounting map of the height changes. The window size we selected for this test region is  $9 \times 9$ . By using this reliability map, final fusion result of  $G_1(\theta_1)$  is achieved and shown in Fig. 6(a). As a comparison, the  $G_1(\theta_1)$  of the initial fusion model is displayed in Fig. 6(b). This is the same building that we have discussed in paper [1]. It can be noted in Fig. 3, this building exists in both panchromatic images of two dates. However, only the DSM from date1 contains the correct height of this building. In Fig. 3c, this building can not be recognized. Therefore, a very high BBA would be achieved in the height change indicator. A high value in  $m_1(\cdot)$  leads to a high global BBAs in building changes (as shown in Fig. 6(a)). Thus this building would be falsely detected as building changes. However, after discounting this region has much lower global BBAs (see Fig. 6(b)), and can be further correctly detected as NoChange.

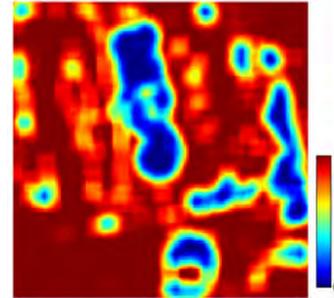


Fig. 5. Generated height change reliability map of the test region 2.

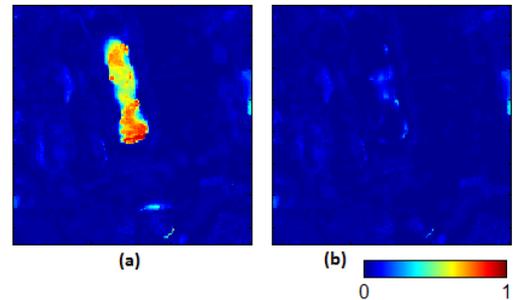


Fig. 6. Global Building change BBAs (a) Initial result; (b) Refined result.

TABLE VI. CHANGE MASKS EVALUATION FROM FOUR GLOBAL BBAS.

	$G_1$		$G_2$		$G_3$		$G_4$	
	Original	Refined	Original	Refined	Original	Refined	Original	Refined
Max_Bel	0.9271	0.9324	0.9271	0.9324	0.9266	0.9322	0.9265	0.9321
Max_Pl	0.9291	0.9342	0.9288	0.9339	0.9287	0.9339	0.9284	0.9336
Max_BetP	0.9283	0.9335	0.9282	0.9334	0.9279	0.9333	0.9278	0.9333
Max_DSMP	0.9281	0.9333	0.9280	0.9331	0.9278	0.9331	0.9276	0.9330

## VI. CONCLUSIONS

Building change detection is a difficult topic, especially when the building changes happen together with other irrelevant changes. Our previous research has evidenced the performance of the belief functions in DSM assisted change detection [2]. In this paper, the change detection accuracy is further improved by adopting an additional reliability map. Height has proved to be an important feature for building change detection. However, the DSMs from satellite images do not always provide reliable height information, due to the occlusion and matching errors. The wrong height information will thus bring false alarms to the change detection procedure. Therefore, the original unfilled disparity maps are adopted to generate an height change reliability map, which is further used in the fusion models.

Our first experimental results have shown that this reliability map can improve the quality of all four global BBAs, and further influences the final change detection results from four decision criteria. However, the two test regions were quite small to draw a definitive conclusion that is why more experiments will be performed on a wider variety of regions with different types of backgrounds. A detailed statistical analysis and comparisons of the results with other techniques is under progress and they will be presented in a forthcoming publication.

Generally speaking, both DST and DSMT frameworks offer the possibility to reach a high accuracy result. The workflow proposed in this paper enables an automatic building change detection procedure. Other reliability maps from images would be further adopted in future work. Furthermore, besides building changes, more change objects will be considered in the fusion model.

## REFERENCES

- [1] J. Tian, S. Cui, and P. Reinartz, "Building change detection based on satellite stereo imagery and digital surface models," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 406–417, 2014.
- [2] J. Tian, P. Reinartz, and J. Dezert, "Building change detection in satellite stereo imagery based on belief functions," in *2015 Joint Urban Remote Sensing Event (JURSE)*. IEEE, 2015, pp. 1–4.
- [3] G. Shafer, *A mathematical theory of evidence*. Princeton university press Princeton, 1976.
- [4] A. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 325–339, 1967.
- [5] F. Smarandache and J. Dezert, *Advances and Applications of DSMT for Information Fusion*. American Research Press, Rehoboth, NM, U.S.A., 2004–2015, vol. 1–4. [Online]. Available: <http://www.onera.fr/staff/jean-dezert?page=2>
- [6] J. Dezert and A. Tchamova, "On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule," *Int. J. Intell. Syst.*, vol. 29, no. 3, pp. 223–252, 2014.
- [7] J. Dezert, P. Wang, and A. Tchamova, "On the validity of Dempster-Shafer theory," in *Proc. of FUSION 2012*, 2012, pp. 655–660. [Online]. Available: <http://fs.gallup.unm.edu/DSMT.htm>
- [8] A. Tchamova and J. Dezert, "On the behavior of Dempster's rule of combination and the foundations of Dempster-Shafer theory," in *Proc. of IS 2012*, 2012, pp. 108–113.
- [9] F. Smarandache and J. Dezert, "On the consistency of PCR6 with the averaging rule and its application to probability estimation," in *Proc. of FUSION 2013*, 2013, pp. 1119–1126.
- [10] A. A. Nielsen, "the regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 463–478, 2007.
- [11] J. Dezert and J.-M. Tacnet, "Sigmoidal model for belief function-based electre tri method," in *Belief Functions: Theory and Applications*, 2012, pp. 401–408.
- [12] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, 1975.
- [13] D. Mercier, B. Quost, and T. Denœux, "Contextual discounting of belief functions," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Springer, 2005, pp. 552–562.
- [14] F. Smarandache, J. Dezert, and J.-M. Tacnet, "Fusion of sources of evidence with different importances and reliabilities," in *Proc. of FUSION2010*. IEEE, 2010, pp. 1–8.
- [15] P. d'Angelo and P. Reinartz, "DSM based orientation of large stereo satellite image blocks," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 39, no. B1, pp. 209–214, 2012.
- [16] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clinical chemistry*, vol. 39, no. 4, pp. 561–577, 1993.

# Generic Object Recognition Based on the Fusion of 2D and 3D SIFT Descriptors

Miaomiao Liu<sup>a</sup>, Xinde Li<sup>a</sup>, Jean Dezert<sup>b</sup>, Chaomin Luo<sup>c</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing, China.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>c</sup>ECE Dept., Univ. of Detroit Mercy, Detroit, MI, USA.

Emails: xindeli@seu.edu.cn, jean.dezert@onera.fr, luoch@udmercy.edu

Originally published as: M. Liu, X. Li, J. Dezert, C. Luo, *Generic Object Recognition Based on the Fusion of 2D and 3D SIFT Descriptors*, in Proc. of Fusion 2015, Washington D.C, USA, July 6-9, 2015, and reprinted with permission.

**Abstract**—This paper proposes a new generic object recognition (GOR) method based on the multiple feature fusion of 2D and 3D SIFT (scale invariant feature transform) descriptors drawn from 2D images and 3D point clouds. We also use trained Support Vector Machine (SVM) classifiers to recognize the objects from the result of the multiple feature fusion. We analyze and evaluate different strategies for making this multiple feature fusion applied to real open-datasets. Our results show that this new GOR method has higher recognition rates than classical methods, even if one has large intra-class variations, or high inter-class similarities of the objects to recognize, which demonstrates the potential interest of this new approach.

**Keywords:** generic object recognition, point cloud, 2D SIFT, 3D SIFT, Feature fusion, BoW, SVM, belief functions, PCR.

## I. INTRODUCTION

Generic object recognition (GOR) in real environment plays a significant role in computer vision and artificial intelligence. It has important applications in intelligent monitoring, robotics, medical image processing, etc [1]–[3]. Contrariwise to specific object recognition<sup>1</sup>, GOR is much more difficult to accomplish. Mainly because the generic features of objects which express the common properties in the same class and help to make the difference between classes need to be found out, instead of defining characteristics of particular category as used in specific object recognition (SOR) methods. The current main techniques for GOR are based on local feature extraction algorithms on 2D images, typically the 2D SIFT (scale invariant feature transform) descriptors [4], [5]. However, 2D images lose the 3D information of the objects, and are susceptible to change due to various external illumination conditions. To solve this drawback, 3D SIFT descriptors based on volumes [3], [6]–[10], and 3D descriptors based on point cloud model [11]–[13] have been proposed recently by several researchers because point cloud model of object is obtained from the depth images which only depends on the geometry of the objects. Such point cloud model has nothing to do with the brightness and reflection features of the objects. That is the main reason why we are also interested by these technique in this paper. 3D SIFT descriptors have been applied successfully in motion

recognition of consecutive video frames by Scovanner et al. [6]. They show good performance in medical image processing [3], [7]–[9] as well. Object recognition has also be done with 3D SIFT in complex Computed Tomography (CT) for airport baggage inspection and security by Flitton et al. [10].

The object recognition algorithms based on single feature only often generate erroneous object recognitions, specially if there are big intra-class variations and some inter-class high similarities, or if there exist important changes in pose and appearance of objects. In these conditions, the use of a single feature is insufficient to make a reliable recognition and classification. To overcome this serious drawback, new recognition algorithms based on multiple features and fusion algorithms have been proposed recently in the literature [14]–[17]. Compared with the recognition algorithm using single feature only, the feature fusion algorithms combine multiple features information which can improve substantially the recognition rate.

In this paper, we propose a new method for GOR based on feature fusion of 2D and 3D SIFT descriptors, which consists of two main phases: 1) a training phase, and 2) a testing phase. In the both phases, we consider two types of inputs:

- 1) The first type of input is a database with 3D point cloud model representation of different objects from different categories (classes). In this work, our database has been just obtained from the web<sup>2</sup>. It is characterized by 3D SIFT descriptors adapted (in this paper) for point cloud – see the next section for details.
- 2) As second input, we use the same database with 2D images including some objects that are characterized by their 2D SIFT descriptors.

From these two inputs, the 2D and 3D SIFT feature descriptors are transformed into the corresponding Bag of Words (BoW) feature vector [18]. In the training phases, these two BoW feature vectors (drawn from the 2D and 3D SIFT) describing the object are used to train Support Vector Machines (SVMs) [19] to get the prediction functions. After this training phase, the system is used to recognize unknown objects in the testing phase. These two BoW feature vectors

<sup>1</sup>such as face recognition [1] (SOR) where only certain objects or certain categories need to be recognized, which can be accomplished by training mass samples.

<sup>2</sup><http://rgbd-dataset.cs.washington.edu/dataset.html>

describing the object are used to make the object recognition in the testing phase. In this paper, we test:

- 1) the feature-level fusion strategy, where we combine (fuse) directly the two BoW-based feature vectors and we feed the trained SVM with the fused vector to get the final recognition result.
- 2) the decision-level fusion strategy, where each of the two BoW-based feature vectors feeds its corresponding trained SVM to get the corresponding recognition result separately. Then we test different fusion rules to combine these two recognition results to get the final recognition result.

The paper is organized as follows. The recognition algorithm is described in details in section II. Section III evaluates the performances of this new method on real datasets. Conclusions with perspectives are given in section IV.

## II. NEW GENERIC OBJECT RECOGNITION METHOD

This new method of object recognition consists in three main steps (features extraction and representation, features fusion, and classifier design) that we present in details in this section. To achieve the good recognition of objects, we propose to combine 2D scale-invariant feature transform (2D SIFT) characterizing the object features, with 3D SIFT (based on point clouds model). We need at first to recall the principle of 2D SIFT [4], [5], and we explain improved 3D SIFT descriptors applied in point cloud.

### Step 1: Features extraction and representation

Feature extraction and representation are necessary for any object recognition algorithm. In many situations the object recognition task is very difficult because it is possible that some (partial) similarities exist in different classes of objects, as well as (partial) dissimilarities in the same class of objects. So the feature extraction process must be done as efficient as possible in order to help the recognition of objects by making the difference between object classes biggest, and by making the difference in the same class smallest. The objects need also to be represented at a certain level of semantic, using limited training objects to represent the class [2].

#### – 2D SIFT descriptor

In 1999, David Lowe [4] did present for the first time a new method to extract keypoints of objects in images, and to describe their local features that allows to make generic object recognition, for example in computer vision applications. His method has then been improved in [5], and extended to 3D by other authors (see next paragraph). The feature description of the object drawn from a training image is then used to identify the presence (if any) of the object in real (usually cluttered) observed scene. To get good object recognition performances, Lowe proposed a (2D) SIFT (scale-invariant feature transform) that warranties that the features extracted (i.e. the key-points) from the training image are detectable under changes in image orientation, scale, noise

and illumination, and even if partial object occlusions occur in the observed scene. Lowe's SIFT feature descriptor is invariant to uniform scaling, orientation, and partially invariant to illumination changes and robust to local geometric (affine) distortion. The stable key-points locations of SIFT are given by the detection of scale-space extrema in the Difference-of-Gaussian (DoG) function  $D(x, y, \sigma)$  convolved with the image  $I(x, y)$ . More precisely, one defines [5]

$$D(x, y, \sigma) \triangleq L(x, y, k\sigma) - L(x, y, \sigma), \quad (1)$$

where  $L(x, y, k\sigma) \triangleq G(x, y, k\sigma) * I(x, y)$  and  $L(x, y, \sigma) \triangleq G(x, y, \sigma) * I(x, y)$  are Gaussian-blurred images at nearby scale-space  $\sigma$  separated by a constant multiplicative factor<sup>3</sup>  $k$ , and where  $*$  is the convolution operator and  $G(x, y, \sigma)$  is the centered Gaussian kernel defined by

$$G(x, y, \sigma) \triangleq \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \quad (2)$$

The local extreme points of  $D(x, y, \sigma)$  functions (DoG images) define the set of keypoint candidates (the SIFT descriptor). To detect the keypoints, each sample point (pixel) is compared to its eight neighbors in the current image and its nine neighbors in the scale below and above. The sample point under test is considered as a keypoint (local extrema) if its value is larger (or smaller) than all of its 26 neighbors. The localization of a candidate keypoint is done by the 2nd-order Taylor expansion of the DoG scale-space function  $D(x, y, \sigma)$  with the candidate keypoint taken as the origin [5]. However in general there are too many candidate keypoints and we need to identify and remove the bad candidates that have too low contrast<sup>4</sup>, or are poorly localized along an edge. For doing this, a contrast thresholding is applied on  $D(x, y, \sigma)$  to eliminate all the candidate keypoints below a chosen<sup>5</sup> threshold value  $\tau$ . To eliminate the candidate keypoints that are poorly localized along an edge, Lowe [5] uses a thresholding method based on the ratio of the eigenvalues of the Hessian matrix  $\mathbf{H}$  of the DoG function, because for poorly defined extrema in the DoG function the principal curvature across the edge would be much larger than the principal curvature along it. More precisely, if the ratio  $Tr(\mathbf{H})^2/Det(\mathbf{H}) > (r_{th}+1)^2/r_{th}$  then the candidate keypoint is rejected. Here,  $r_{th}$  is a chosen threshold value of the ratio between the largest magnitude eigenvalue of  $\mathbf{H}$  and the smaller one<sup>6</sup>.

Once all the keypoints are determined, one must assign a consistent orientation based on local image properties, from which the keypoint descriptor can be represented, hence achieving invariance to image rotation. For this, the scale of the keypoint is used to choose the Gaussian-blurred image  $L$  with the closest scale. The keypoint descriptor is created by computing at first the gradient magnitude  $m(x, y)$  and its

<sup>3</sup>The choice for  $k = 2^{1/s}$  is justified by Lowe in [4], where  $s$  is an integer number of intervals

<sup>4</sup>because they are sensitive to noise.

<sup>5</sup>We have chosen  $\tau = 0.02$  in our simulations.

<sup>6</sup>In [5], Lowe takes  $r_{th} = 10$ .

orientation  $\theta(x, y)$  at each pixel  $(x, y)$  in the region around the keypoint in this Gaussian-blurred image  $L$  as follows [5]

$$\begin{cases} m(x, y) = \sqrt{L_x^2 + L_y^2}, \\ \theta(x, y) = \tan^{-1}\left(\frac{L_y}{L_x}\right), \end{cases} \quad (3)$$

with  $L_x \triangleq L(x+1, y) - L(x-1, y)$  and  $L_y \triangleq L(x, y+1) - L(x, y-1)$ . In [5], a set of orientation histograms is created on  $4 \times 4$  pixel neighborhoods with 8 directions (bins) each. These histograms are computed from magnitude and orientation values of samples in a  $16 \times 16$  region around the keypoint such that each histogram contains samples from a  $4 \times 4$  subregion of the original neighborhood region. The magnitudes are weighted by a Gaussian function with  $\sigma$  equal to one half the width of the descriptor window. The descriptor then becomes a 128-dimensional feature vector because there are  $4 \times 4 = 16$  histograms each with 8 directions. This vector is then normalized to unit length in order to enhance invariance to affine changes in illumination. Also a threshold of 0.2 is applied to reduce the effects of non-linear illumination, and the vector is again normalized. The figure 1 shows an example of  $4 \times 4$  keypoint descriptor, where the space delimited by the purple ellipse is the neighborhood under consideration.

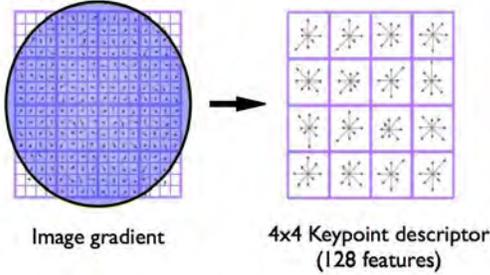


Fig. 1: A  $4 \times 4$  Keypoint descriptor (Credit: J. Hurrelmann).

The simplest method to find the best candidate match for each keypoint would consist in identifying its nearest<sup>7</sup> neighbor in the database of key points from training images. Unfortunately, SIFT-based keypoint matching requires more sophisticated methods because many features from an image will not have any correct match in the training database because of background clutter in observed scene and because of possible missing features in training images, see [5] for details. SIFT method is patented by the University of British Columbia (US Patent 6,711,293 – March 23, 2004) and a demo is available in [20]. Open SIFT codes can be found on the web, for example in [21].

### – 3D SIFT descriptor

The previous 2D SIFT descriptor working with pixels has been extended to 3D using volumes in different manners by different authors [3], [6]–[10]. In this paper, we adapt the 3D SIFT for point cloud inspired by [6], [13]. But all the methods require same functional steps as for 2D SIFT, that

<sup>7</sup>based on Euclidean distance metric.

is 1) Keypoints detection; 2) Key points orientation; and 3) Descriptor representation. We present these steps in detail in the next subsections.

#### 1) Keypoint detection

The scale space of a 3D input point cloud is defined as a 4D function  $L(x, y, z, \sigma) = G(x, y, z, k\sigma) * P(x, y, z)$  obtained by the convolution of a 3D variable-scale centered Gaussian kernel  $G(x, y, z, \sigma)$ , with the input point  $P(x, y, z)$ , where

$$G(x, y, z, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^3} e^{-(x^2+y^2+z^2)/2\sigma^2}, \quad (4)$$

Extending Lowe's approach [5], scale-space  $\sigma$  is separated by a constant multiplicative factor  $k$ , and the candidate keypoints in 4D scale space are taken as the local extrema (maxima or minima) of the multi-scale DoG defined for  $i \in [0, s+2]$  by

$$D(x, y, z, k^i\sigma) = L(x, y, z, k^{i+1}\sigma) - L(x, y, z, k^i\sigma). \quad (5)$$

To find extrema of the multi-scale DoG function, each sample point is compared to its  $27 + 26 + 27 = 80$  neighbors, where 26 neighbors belong to the current scale, and each 27 neighbors in the scale above and below. A keypoint is chosen only if it is larger than all of its neighbors or smaller than all of them. To eliminate the bad candidate keypoints having low contrast, one uses a thresholding method to remove the erroneous points. A contrast threshold is applied on  $D(x, y, z, k^i\sigma)$  to eliminate all the candidate keypoints below a chosen<sup>8</sup> threshold value  $\tau$ .

#### 2) Keypoint orientations

Similarly to 2D SIFT, once all the keypoints are determined in 3D, one must assign a consistent orientation based on local points properties, from which the keypoint descriptor can be represented, hence achieving invariance to object rotation. For this, The two-dimensional histogram is calculated by gathering statistics of the angles between the neighboring points and their center. The keypoint descriptor is created by computing at first the vector magnitude  $m(x, y, z)$  and its orientations  $\theta(x, y, z)$  (azimuth angle) and  $\phi(x, y, z)$  (elevation angle) between each point  $(x, y, z)$  in the region around the keypoint and their center  $(x_c, y_c, z_c)$  as follows<sup>9</sup>

$$\begin{cases} m(x, y, z) &= \sqrt{(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2}, \\ \theta(x, y, z) &= \tan^{-1}((y - y_c)/(x - x_c)), \\ \phi(x, y, z) &= \sin^{-1}((z - z_c)/m(x, y, z)). \end{cases} \quad (6)$$

In 3D point cloud, each point has two values which represent the direction of the region, whereas in 2D case each pixel had only one direction of the gradient.

Extending Lowe's approach in 3D case, in order to find the keypoint orientations we construct a weighted histogram for

<sup>8</sup>We have chosen  $\tau = 0.5$  in our simulations.

<sup>9</sup>In Eq.(6),  $\theta$  and  $\phi$  refer to the original coordinate system. In the paragraph "Descriptor representation" on p. 4, they refer to the rotated coordinate system.  $(x_c, y_c, z_c)$  is not same as  $(x_p, y_p, z_p)$ . The former refers to the center of the keypoint's r-points neighborhood. The latter refers to the keypoint.

the 3D neighborhood around each candidate keypoint. There are different ways for doing this. In this work, a 2D-histogram is produced by grouping the angles in bins which divide  $\theta$  and  $\phi$  into 10 deg angular bins. A regional Gaussian weighting of  $e^{-(2d/R_{\max})^2}$  for the points whose magnitude is  $d$  is applied to the histogram, where  $R_{\max}$  represents the max distance from the center. The sample points at a distance greater than  $R_{\max}$  are ignored. The histogram is smoothed using a Gaussian filter to limit the effect of noise. The dominant azimuth  $\alpha$  and elevation  $\beta$  of the keypoint are determined by the peaks of the 2D-histogram. In order to enhance robustness, peaks in the histogram within 80% of the largest peak are also retained as possible secondary orientations.

### 3) Descriptor representation

Each keypoint  $p$  is described by its location  $\mathbf{p} \triangleq [x_p, y_p, z_p]^t$ , scale  $\sigma_p$ , and orientation angles  $\alpha_p$  and  $\beta_p$ . The descriptor representation associated with a keypoint  $p$  is based on the local spatial characteristics around it to describe its features. To ensure rotation invariance of the descriptor, the  $r$ -points  $p_i$  ( $i = 1, \dots, r$ ) of coordinates  $\mathbf{p}_i \triangleq [x_i, y_i, z_i]^t$  around the keypoint of interest  $p$  are at first transformed (rotated) in the dominant orientation of  $p$  by the following transformation

$$\mathbf{p}'_i = \begin{bmatrix} \cos \alpha_p \cos \beta_p & -\sin \alpha_p & -\cos \alpha_p \sin \beta_p \\ \sin \alpha_p \cos \beta_p & \cos \alpha_p & -\sin \alpha_p \sin \beta_p \\ \sin \beta_p & 0 & \cos \beta_p \end{bmatrix} \cdot \mathbf{p}_i. \quad (7)$$

Then the vector  $\mathbf{n}$  at the key point which is normal to the surface of the  $r$ -points neighborhood is calculated according to the routine available in the open Point Cloud Library (PCL) [22]. For each (rotated) point  $\mathbf{p}'_i$  ( $i = 1, \dots, r$ ) in the  $r$ -points neighborhood of the (rotated) keypoint  $p'$ , we calculate the vector  $\mathbf{p}'\mathbf{p}'_i$  and the magnitude  $m$  and angles  $\theta$  and  $\phi$  according to Eq. (6). The angle  $\delta$  between  $\mathbf{n}$  and  $\mathbf{p}'\mathbf{p}'_i$  is given by

$$\delta = \cos^{-1} \left( \frac{\mathbf{p}'\mathbf{p}'_i \cdot \mathbf{n}}{|\mathbf{p}'\mathbf{p}'_i| \cdot |\mathbf{n}|} \right). \quad (8)$$

Therefore, a keypoint  $p'$  with its neighbor  $p'_i$  is represented by the 4-tuple  $(m, \theta, \phi, \delta)$ . To reduce the computational time, instead of dividing the neighborhood into  $n \times n \times n$  subregions (with  $n = 4$  as in Lowe's 2D SIFT descriptor), we take directly the entire neighborhood, which means that we have  $n = 1$ . The histogram used to generate the 3D descriptor at the keypoint  $p'$  is derived by splitting  $(\theta, \phi, \delta)$  space into 45 deg bins, and adding up the number of points with the Gaussian weighting of  $e^{-(2m/R_{\max})^2}$ . So the dimension of our 3D SIFT descriptor is  $n \times n \times n \times 4 \times 4 \times 8 = 128$  (as for the 2D SIFT descriptor described previously), because  $n = 1$ ; the azimuth angle  $\theta \in [0, 360]$  deg which is split into 8 bins of 45 deg; the elevation angle  $\phi \in [-90, 90]$  deg which is split into 4 bins of 45 deg; and  $\delta \in [0, 180]$  deg which is also split into 4 bins of 45 deg. Each 3D SIFT descriptor is normalized to unity.

The 2D and 3D SIFT descriptors summarize efficiently the useful information contained in 2D and 3D images. Instead of working directly with whole images, it is usually more interesting (in terms of computational burden reduction) to

work directly with 2D and 3D SIFT descriptors, specially if real-time object recognition is necessary. Generally, the objects characterized by 2D and 3D SIFT descriptors have different number of keypoints which makes the feature fusion (FF) problem for object recognition very challenging. For example, for a simple object like an apple, we can get 45 keypoints using 3D SIFT descriptor, and 38 keypoints using 2D SIFT descriptor. To overcome this problem, we adopt the Bag of Words (BoW) model [18] to gather the statistics of the 2D and 3D SIFT descriptors to describe the objects.

### – BoW model for features vector

In the BoW feature model, the feature descriptors of all the interest points are quantized by clustering them into a pre-specified<sup>10</sup> number of clusters. Instead of using  $k$ -means algorithm as in [2], we use the  $k$ -means++ method [23] which selects more effectively the initial cluster centers to complete this step. The resultant cluster centers are now called *visual words*, while the collection of these cluster centers is referred to as the *visual word vocabulary*. Once our vocabulary is computed, the descriptors are matched to each *visual word* based on the Euclidean distance and the frequency of the visual words in image and in point cloud is accumulated into a histogram, which is the BoW feature vector of the image and of the point cloud. So each object in 2D image and in 3D point cloud is described by a  $1 \times 300$  BoW-based feature vector denoted respectively  $\mathbf{BoW}_{2D}$  and  $\mathbf{BoW}_{3D}$ . These two BoW-based feature vectors will be used for feeding the trained SVM classifiers to get the final object recognition.

### Step 2: Classifier design

Once the object description is completed, SVMs are trained to learn objects categories and to perform the object classification. SVM is a supervised and discriminative machine learning method providing usually good performance. Through offline training of pre-limited samples, we seek a compromise between model complexity and learning ability, to get a good discriminant function [19]. Linear SVM classifier is applied for its efficiency and it is a typical classifier for two categories problems. In many real-life applications, we are face to multi-category classification problems and we use trained 1V1 SVMs between classes to set up a multi-category classifier. The training process is done as follows: for training samples belonging to the  $i$ th category, we make a pairwise SVM training with respect to all the other classes. So, we get  $C_n^2 = n(n-1)/2$  1V1 SVM classifiers for training samples of  $n$  categories.

### Step 3: Features fusion strategies

When the two BoW-based features vectors of the object to recognize have been computed from 2D and 3D SIFT descriptors, we have to use them to achieve the object recognition thanks to the trained SVMs from the BoW-based features vectors of known objects of our data base. In this paper, we

<sup>10</sup>In our simulations, we took  $K = 300$ .

present briefly the following different strategies that we have tested:

- 1) *The direct feature-level fusion strategy*: this feature-level fusion is for feeding SVM classifiers in training phase and then making object recognition. With this strategy we combine (fuse) directly the two BoW-based feature vectors  $\mathbf{BoW}_{2D}$  and  $\mathbf{BoW}_{3D}$ , and we feed the trained (global) SVM classifiers with the fused vector to get the final recognition. The principle of our method based on this strategy is summarized in Fig 2.

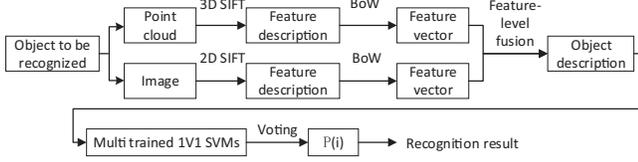


Fig. 2: Direct feature-level fusion strategy.

- 2) *The decision-level fusion strategy*: each BoW-based feature vector  $\mathbf{BoW}_{2D}$  and  $\mathbf{BoW}_{3D}$  feeds a specific trained SVM to get separately the corresponding recognition result. Then we test different fusion rules to combine these two recognition results to get the final fusioned recognition result. In this work we have evaluated the performances of the following rules:

- Average weighted fusion rule,
- PCR6 fusion rule of DSMT [24],
- Murphy's rule of combination [26].

The principle of our method based on this strategy is summarized in Fig 3.

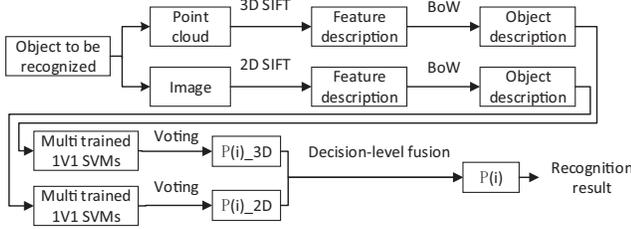


Fig. 3: Decision-level fusion strategy.

### 1) The direct feature-level fusion strategy

This strategy consists of the following steps:

- 1-a) For any object to classify, we extract its 2D and 3D SIFT descriptors associated with each keypoint. So we get  $N_{2D}$  2D SIFT descriptors of size  $1 \times 128$  if one has extracted  $N_{2D}$  keypoints from the 2D image under test, and we get  $N_{3D}$  3D SIFT descriptors of size  $1 \times 128$  if one has extracted  $N_{3D}$  keypoints from the 3D point cloud under test.
- 1-b) From the  $N_{2D}$  2D SIFT descriptors of size  $1 \times 128$ , we compute  $1 \times 300$  BoW feature vectors  $\mathbf{BoW}_{2D}$ , and from the  $N_{3D}$  3D SIFT descriptors of size  $1 \times 128$ , we

compute  $1 \times 300$  BoW feature vectors  $\mathbf{BoW}_{3D}$  thanks to the BoW model representation [18].

- 1-c) The direct feature-level fusion is done by stacking the BoW-based feature vectors  $\mathbf{BoW}_{2D}$  and  $\mathbf{BoW}_{3D}$  to get a  $1 \times 600$  vector  $\mathbf{BoW}_{2D,3D} \triangleq [\mathbf{BoW}_{2D}, \mathbf{BoW}_{3D}]$ .
- 1-d) The feature-level fused vector  $\mathbf{BoW}_{2D,3D}$  is fed in all 1v1 trained SVMs to get the corresponding discriminant results. The probability  $P(i)$  of the object to belong to the category  $c_i$  ( $i = 1, 2, \dots, n$ ) is estimated by voting.
- 1-e) The object is associated to the category (or class) having the largest probability, that is:

$$\text{Class}(\text{Object}) = \arg \max_{1 \leq i \leq n} \{P(i)\}. \quad (9)$$

### 2) The decision-level fusion strategy

As stated before, with this strategy each BoW-based feature vector  $\mathbf{BoW}_{2D}$  and  $\mathbf{BoW}_{3D}$  feeds a specific trained SVM to get separately the corresponding recognition result. Then different fusion rules can be used to combine these two recognition results to get the final fusioned recognition result.

- 2-a) **The average weighted fusion rule**: This very simple rule consists of a voting procedure. The  $\mathbf{BoW}_{2D}$  and  $\mathbf{BoW}_{3D}$  vectors feed separately all corresponding 1v1 trained SVMs to get the discriminant results, and we compute the corresponding number of votes  $vote[i]$  for each class  $c_i$ ,  $i = 1, 2, \dots, n$ . We will denote  $vote_{2D}[i]$  the distribution of votes drawn from 2D SIFT, and  $vote_{3D}[i]$  the distribution of votes drawn from 3D SIFT. The probability  $P_{2D}(i)$  of the object to belong to the class  $c_i$  based on 2D SIFT descriptors is estimated by  $P_{2D}(i) = vote_{2D}[i] / \sum_{i=1}^n vote_{2D}[i]$ , similarly we have  $P_{3D}(i) = vote_{3D}[i] / \sum_{i=1}^n vote_{3D}[i]$ . Then the voting results drawn from SVMs feeded with 2D and 3D SIFT are averaged to obtain the fusion result.
- 2-b) **PCR6 combination rule**: The BBA (Basic Belief Assignment)  $m_1(\cdot)$  and  $m_2(\cdot)$  are built from the empirical probability obtained by voting procedure described in 2-a). The elements of the frame of discernment  $\Theta$  are the  $n$  different classes  $c_1, c_2, \dots, c_n$ . To get the final result, the BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$  are fused using the PCR6 combination rule<sup>11</sup> [24], defined by  $m_{PCR6}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Theta$ ,

$$m_{PCR6}(X) \triangleq \sum_{\substack{x_1, x_2 \in 2^\Theta \\ x_1 \cap x_2 = X}} m_1(x_1)m_2(x_2) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right]. \quad (10)$$

<sup>11</sup>PCR6 formula coincides with the formula of PCR5 fusion rule here because one considers only two BBA's to combine. If more than two BBA's have to be fused altogether, we advise to use PCR6 rather than PCR5 - see [25] for a theoretical justification.

where all denominators in Eq.(10) are different from zero. If a denominator is zero, that fraction is discarded. All propositions/sets are in a canonical form.

- 2-c) **Murphy's rule:** Taking the feature-level fusion of 2D and 3D SIFT as a separate feature, together with the 2D and 3D SIFT, there are three features. Then the BBA  $m_1(\cdot)$ ,  $m_2(\cdot)$  and  $m_3(\cdot)$  are built from the empirical probability obtained by the voting procedure. The vote results of the features are combined based on the Murphy rule<sup>12</sup> [26].

### III. SIMULATION RESULTS

#### A. The experimental setup

We evaluate the recognition algorithm on a large-scale multi-view object dataset collected using an RGB-D camera [27]. This dataset contains color, depth images and point clouds of 300 physically distinct everyday objects taken from different viewpoints. The objects belong to one of 51 categories and contain three viewpoints. To test the recognition ability of our features, we test category recognition on objects that were not present in the training set. At each trial, we randomly choose one test object from each category and train classifiers on the remaining objects. We randomly choose 100 training samples and 60 test samples for each category. The object recognition rate (ORR) is calculated by

$$ORR = n_r / N. \quad (11)$$

where  $n_r$  is the number of objects correctly recognized, and  $N$  is the total number of test samples.

#### B. Experiment results and analysis

##### B.1 Accuracy of our 3D SIFT descriptor

In this simulation, we choose six categories with significant intra-class variations and high inter-class similarities. The objects to recognize are *apple*, *tomato*, *banana*, *pitcher*, *cereal\_box*, and *kleenex*. The Point Feature Histogram (PFH) [11] and PFHRGB methods in open PCL [22] outperform the existed 3D features based on point clouds [28]. In order to verify the advantages of the proposed 3D SIFT for GOR, we compare these tree feature descriptors under the same conditions. Keypoints are detected using SIFTKeypoint module in open PCL [22] for each feature descriptors. Then the vectors of different feature descriptors of the keypoints are calculated. The object recognition rates (ORR) that we get are shown in Table I.

Type of feature descriptor	ORR (in %)
PFH based on [11]	81.39
PFHRGB based on [22]	84.17
3D SIFT based on this paper	91.11

TABLE I: Object recognition rates (ORR) of three descriptors.

The PFHRGB descriptor is an improved PFH feature descriptor enriched with color information which allows to improve object recognition rate. As shown in Table 1, compared

<sup>12</sup>Because results of the fusion with Dempster's rule are very close to results with Murphy's rule in our applications, we do not report them in our analysis.

with PFH and PFHRGB, the object recognition rate we get with our 3D SIFT descriptor adapted for point cloud gains 6.94% w.r.t. PFHRGB and 9.72% w.r.t. PFH.

#### B.2 Performances of feature fusion strategies

Here, we evaluate the performance (i.e. the ORR) of the different features fusion strategies presented in Section II (Step 3). We have chosen 10 categories (*apple*, *tomato*, *banana*, *pitcher*, *cereal\_box*, *kleenex*, *camera*, *coffee\_mug*, *calculator*, *cell\_phone*) having significant intra-class variations and high inter-class similarities. We compare our four fusion approaches: the direct feature-level fusion and the three decision-level fusions (by average weighted fusion, PCR6, and Murphy's rule). The results are shown in Fig. 4.

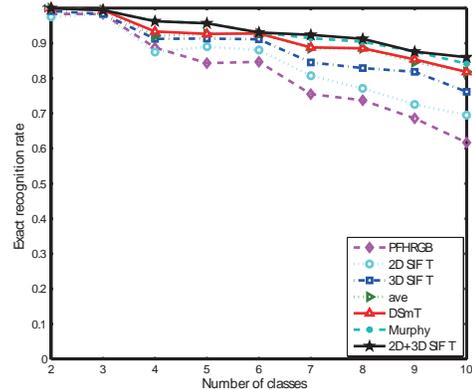


Fig. 4: Performances of the four feature fusion strategies.

where the legend of curves of Fig.4 must be read as follows: DSMT means PCR6 rule in fact, 2D+3D SIFT means the direct feature-level fusion of 2D and 3D SIFT, and *ave* means the average weighted feature fusion rule. The horizontal axis represents the total number of categories that we have tested. Due to the variability of the objects, the information provided by a single feature is too imprecise, uncertain and incomplete for getting good ORR. As shown in Fig.4, ORR obtained with the different feature fusion strategies are better than the ORR obtained with the best single descriptor. The results of average weighted fusion and PCR6 are close, but are lower than the other two fusion methods. Feature-level fusion of 2D and 3D SIFT is taken as the third feature for Murphy's rule. However, compared with the feature-level fusion, the performances of Murphy's rule do not improve. So, the direct feature-level fusion performs best among these fusion strategies, and the following experiments are completed based on the direct feature-level fusion. One clearly sees that 3D SIFT proposed in this work significantly outperforms 2D SIFT and PFHRGB descriptors for GOR. As shown in Fig.4, ORR decreases with the increasing of the number of categories because of the design of the multi-category classifier which consists of many 1V1 SVM classifiers. Each classification error will be accumulated to the final voting results, leading to an increasing of recognition errors.

### B.3 Robustness to intra-class variation and inter-class similarities

In this study, we compare the ORR performances in different classes having high similarity (e.g., apple and tomato), and in the same class but having strong variation (e.g., pitcher object) as in Figs. 5 and 6 below. We evaluate the accuracy



Fig. 5: Apple and Tomato.



Fig. 6: Pitchers.

of PFHRGB, 2D SIFT, 3D SIFT and the feature-level fusion of 2D and 3D SIFT under the same conditions. Training and testing samples are the same as in the first experiment. Our simulation results are shown in Table II.

Feature descriptor	PFHRGB	2D SIFT	3D SIFT	2D+3D SIFT
ORR(apple)	61.67	53.33	71.67	65.00
ORR(tomato)	100	98.33	91.67	100
ORR(banana)	91.67	93.33	93.33	100
ORR(pitcher)	70.00	95.00	96.67	98.33
ORR(cereal_box)	91.67	98.33	95.00	95.00
ORR(kleenex)	90.00	90.00	100	100
Averaged ORR	84.17	88.06	91.11	93.06

TABLE II: ORR (in %) of different classes.

As we see from Table II, using 3D SIFT increases the ORR of 3.05% w.r.t. 2D SIFT. This shows that the introduction of the depth information improve the quality of object recognition. Three different objects of the pitcher class are shown in Figure 6. As we see, there are great differences within such class. 3D SIFT achieves ORR with 96.67% accuracy, much superior to the 70% obtained with PFHRGB. Apple and tomato displayed in Figure 5 look highly similar even if they belong to two distinct classes. 3D SIFT provides much better ORR than the other descriptors. As shown in Table II, our GOR method based on feature-level fusion of 2D and 3D SIFT offer better robustness to intra-class variations and inter-class similarities, and 3D SIFT gives higher accuracy than the other single descriptors.

### B.4 Robustness to changes of the angle of view

In this experiment, we evaluate the performance of our GOR method when applied under different observation conditions, more precisely when the objects are observed under three very distinct angles of view (30 deg, 45 deg and 60 deg). Training samples are the same as the Experiment 1. Randomly select 60 objects from each view to be as the test samples. So for each view, there are 360 test samples from 6 categories. The experimental results are shown in Fig. 7.

From Fig. 7, one sees that ORR with 3D SIFT is relatively accurate and stable compared with PFHRGB descriptor. The direct feature-level fusion strategy (with  $ORR > 90\%$ ) offers much better ORR than using the best single descriptor, which indicates that the combination of 2D and 3D SIFT is effective

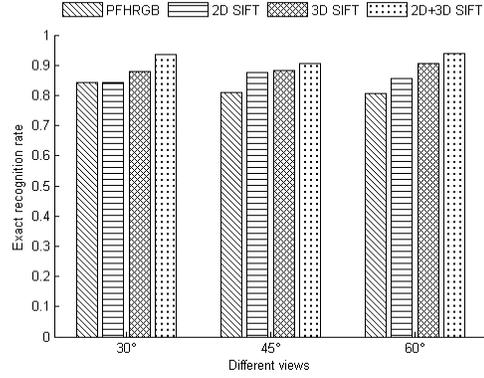


Fig. 7: ORR Performances under 3 angles of view.

and robust for category recognition even under very distinct angles of view.

### B.5 Robustness to size scaling

The training samples are the same as in the first experiment. To evaluate the robustness of our method to size scaling (zooming), the test samples are zoomed out to 1/2, 1/3 and 1/4. As shown in Table III.

Feature descriptor	PFHRGB	2D SIFT	3D SIFT	2D+3D SIFT
ORR (no Zoom)	84.17	88.06	91.11	93.06
ORR (Zoom=1/2)	74.44	77.50	76.67	82.78
ORR (Zoom=1/3)	63.33	64.17	65.28	68.89
ORR (Zoom=1/4)	61.39	46.94	61.67	63.05

TABLE III: Averaged ORR (in %) for different zoomings.

As one sees in Table III, our GOR method with fusion is superior to the algorithm based on single descriptor. However, the ORR of each feature descriptor has decreased. Especially when zoomed to 1/4, the accuracy of ORR with 2D SIFT is only 46.94%. The main reason is that part of the images, such as apple (whose original size is only  $84 \times 82$ ) after scaling, reduces the number of useful keypoints. The feature-level fusion algorithm still provides an averaged ORR of 63.05%.

### B.6 Computational time evaluation

The computational times (CT) of the different feature descriptors have been evaluated with an i7-3770@3.4GHz CPU, under x64 Win7 operating system and are shown in Table IV. The training and test samples are the same as in the first experiment. Because the Point cloud model contains a larger amount of data and richer information than image, therefore CT using point cloud is relatively long, which is normal. The largest proportion of CT in the whole recognition process is the feature extraction and description. 3D SIFT includes keypoints detection and description. If the points' number of the object is  $n$ , the time complexity of keypoints detection is  $O(octaves \cdot scale \cdot k \cdot n)$ . Because the pyramid layers  $octaves$ , scale of each layer  $scale$  and neighborhood of key points  $k$  are constant, the time complexity is  $O(n)$ . For the detected  $m$  keypoints, the time complexity of calculating the descriptors

of the key points is  $O(mn)$ . So the time complexity of 3D SIFT is  $O(mn + n)$ , ignoring lower-order item, the time complexity is  $O(mn)$ . As seen in Table IV, the CT of 3D SIFT has diminished of 34.75% w.r.t. PFHRGB, and the CT performance with fusion of 2D and 3D SIFT turns out to be faster (22.07%) than PFHRGB, and the ORR performance is substantially improved.

Feature descriptors	CT of 360 test samples (in s)	CT of each test sample (in s)
PFHRGB	3404.628	9.4573
3D SIFT	2221.608	6.1711
2D+3D SIFT	2653.272	7.3702

TABLE IV: Computational times for feature descriptors.

#### IV. CONCLUSIONS

Because there are many complex objects in the real scenes we observe in the nature and because of possible large intra-class variations and high inter-class similarities, the generic object recognition (GOR) task is very hard to achieve in general. In this paper we have proposed a new GOR method based on 2D and 3D SIFT descriptors that allows to calculate multiple feature vectors which are combined with different strategies, and feed SVM classifier for making object recognition. The evaluation of the performances based on real open-datasets has shown the superiority of our new 3D SIFT descriptor adapted for point cloud with respect to the existing 3D features such as PFHRGB. Our GOR method based on feature fusion of 2D and 3D SIFT works better than the one using best single feature. For now, if the environment substantially changes, we have to retrain the system. To overcome this problem we will also consider background segmentation within GOR in future works. Also, we would like to reduce the computational time needed for feature extraction and description in maintaining good recognition rate, and we want to explore more feature fusion strategies to improve (if possible) the recognition performances.

#### ACKNOWLEDGMENT

This work was supported by NNSF of China (No. 61175091), Qing Lan Project of Jiangsu Province, Aeronautical Science Foundation of China (20140169002), and Six Major Top-talent Plan of Jiangsu Province.

#### REFERENCES

- [1] Y. Lei, M. Bennamoun, M. Hayat, Y. Guo, *An efficient 3D face recognition approach using local geometrical signatures*, Pattern Recognition, Vol. 47(2), pp. 509–524, 2014.
- [2] X.-D. Li, X. Zhang, B. Zhu, X.-Z. Dai, *A Visual Navigation Method for Robot Based on a GOR and GPU Algorithm*, Robot, Vol. 34(4), pp. 466–475, 2012 (in Chinese).
- [3] S. Allaire, J.J. Kim, S.L. Breen, D.A. Jaffray, V. Pekar, *Full orientation invariance and improved feature selectivity of 3D SIFT with application to medical image analysis*, Proc. IEEE CVPR Workshops, Anchorage, AK, USA, 23–28 June 2008.
- [4] D.G. Lowe, *Object recognition from local scale-invariant features*, Proc. of IEEE CCV Conf., Vol. 2, pp. 1150–1157, Corfu, Greece, Sept. 1999.
- [5] D.G. Lowe, *Distinctive Image Features from Scale-Invariant Key points*, Int. J. of Computer Vision, Vol. 60(2), pp. 91–110, 2004.

- [6] P. Scovanner, S. Ali, M. Shah, *A 3-dimensional SIFT descriptor and its application to action recognition*, Proc. of 15th ACM MM Conf., pp. 357–360, Augsburg, Germany, Sept. 23–29, 2007.
- [7] W. Cheung, G. Hamarneh, *N-SIFT: N-dimensional scale invariant feature transform for matching medical images*, Proc. of 4th IEEE Int. Symp. on Biomedical Imaging, pp. 720–723, Arlington, VA, USA, 2007.
- [8] R.N. Dalvi, I. Hacihaliloglu, R. Abugharbieh, *3D ultrasound volume stitching using phase symmetry and Harris corner detection for orthopaedic applications*, Proc. of SPIE (Medical Imaging 2010), Vol. 7623, San Diego, CA, USA, 2010.
- [9] M. Niemeijer, et al., *Registration of 3D spectral OCT volumes using 3D SIFT feature point matching*, Proc. SPIE Vol. 7259 (Medical Imaging 2009), Lake Buena Vista, FL, USA, 27 March 2009.
- [10] G.T. Flitton, T.P. Breckon, N. Megherbi, *Object Recognition using 3D SIFT in Complex CT Volumes*, Proc. of BMV Conf., pp. 1–12, Aberystwyth, UK, Aug 31–Sept. 3rd, 2010.
- [11] R.B. Rusu, N. Blodow, Z.C. Marton, M. Beetz, *Aligning point cloud views using persistent feature histograms*, Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Syst., pp. 3384–3391, Nice, France, 2008.
- [12] R.B. Rusu, N. Blodow, M. Beetz, *Fast point feature histograms (FPFH) for 3D registration*, Proc. of IEEE Int. Conf. on Robotics and Autom., pp. 3212–3217, Kobe, Japan, 2009.
- [13] S. Lazebnik, C. Schmid, J. Ponce, *A sparse texture representation using local affine regions*, IEEE Trans. on PAMI, Vol. 27(8), pp. 1265–1278, 2005.
- [14] X.-D. Li, J.-D. Pan, J. Dezert, *Automatic Aircraft Recognition using DSMT and HMM*, Proc. of Fusion 2014, Salamanca, Spain, July 2014.
- [15] L. Bo, K. Lai, X. Ren, D. Fox, *Object recognition with hierarchical kernel descriptors*, Proc. of CVPR IEEE Conf., pp. 1729–1736, Colorado Springs, CO, USA, June 2011.
- [16] L. Bo, X. Ren, D. Fox, *Depth kernel descriptors for object recognition*, Proc. of IEEE/RSJ IROS Conf., pp. 821–826, San Francisco, CA, USA, Sept. 2011.
- [17] M. Mirdanias, A.S. Prihatmanto, E. Rijanto, *Object Recognition System in Remote Controlled Weapon Station using SIFT and SURF Methods*, J. of Mechatronics, Elect. Power, and Vehicular Techn., Vol. 4(2), pp. 99–108, 2013.
- [18] J. Sivic, A. Zisserman, *Video google: A text retrieval approach to objects matching in videos*, Proc. of 9th CCV Conf, pp. 1470–1477, 2003.
- [19] B.E. Boser, I.M. Guyon, V.N. Vapnik, *A training algorithm for optimal margin classifiers*, Proc. of the 5th ACM Workshop on Comput. learning theory, pp. 144–152, Pittsburgh, PA, USA, 1992.
- [20] SIFT demo program (Version 4, July 2005). <http://www.cs.ubc.ca/~lowe/keypoints/>
- [21] R. Hess, *An Open Source SIFT Library*, ACM MM, 2010. <http://robwhess.github.io/opensift/>
- [22] R.B. Rusu, S. Cousins, *3D is here: Point cloud library (PCL)*, Proc. of IEEE Int. Conf. on Robotics and Autom., pp. 1–4, Shanghai, China, 2011.
- [23] D. Arthur, S. Vassilvitskii, *K-means++: The advantages of careful seeding*, Proc. of SODA '07, pp. 1027–1035, 2007.
- [24] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, ARP, Rehoboth, NM, U.S.A., Vol. 1–4, 2004–2015. <https://www.onera.fr/fr/staff/jean-dezert/references>
- [25] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, Proc. of Fusion 2013, Istanbul, Turkey, July 2013.
- [26] C.K. Murphy, *Combining Belief Functions when Evidence Conflicts*, Decision Support System, Vol. 29(1), pp. 1–9, 2000.
- [27] K. Lai, L.-F. Bo, X.-F. Ren, D. Fox, *A Large-Scale Hierarchical Multi-View RGB-D Object Dataset*, Proc. of IEEE Int. Conf. on Robotics and Autom., pp. 1817–1824, Shanghai, China, 2011.
- [28] L.A. Alexandre, *3D Descriptors for Object and Category Recognition: a Comparative Evaluation*, Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS 2012, October 7–12, Vilamoura, Portugal, 2012.

# On the Quality Estimation of Optimal Multiple Criteria Data Association Solutions

Jean Dezert<sup>a</sup>, Kaouthar Benameur<sup>a</sup>, Laurent Ratton<sup>b</sup>, Jean-François Grandin<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Thales Systèmes Aéroportés, 2 av. Gay Lussac, 78990 Élancourt, France.

Emails: jean.dezert@onera.fr, kaouthar.benameur@onera.fr,  
laurent.ratton@fr.thalesgroup.com, jean-francois.grandin@fr.thalesgroup.com

Originally published as: J. Dezert, K. Benameur, L. Ratton, J.-F. Grandin, *On the Quality Estimation of Optimal Multiple Criteria Data Association Solutions*, in Proc. of Fusion 2015, Washington D.C, USA, July 6–9, 2015, and reprinted with permission.

**Abstract**—In this paper, we present a method to estimate the quality (trustfulness) of the solutions of the classical optimal data association (DA) problem associated with a given source of information (also called a criterion). We also present a method to solve the multi-criteria DA problem and to estimate the quality of its solution. Our approach is new and mixes classical algorithms (typically Murty’s approach coupled with Auction) for the search of the best and the second best DA solutions, and belief functions (BF) with PCR6 (Proportional Conflict Redistribution rule # 6) combination rule drawn from DSMT (Dezert-Smarandache Theory) to establish the quality matrix of the global optimal DA solution. In order to take into account the importances of criteria in the fusion process, we use weighting factors which can be derived by different manners (ad-hoc choice, quality of each local DA solution, or inspired by Saaty’s Analytic Hierarchy Process (AHP)). A simple complete example is provided to show how our method works and for helping the reader to verify by him or herself the validity of our results.

**Keywords:** Data association, Multi-criteria analysis, belief functions, PCR6, DSMT.

## I. INTRODUCTION

Efficient algorithms for modern multisensor-multitarget tracking (MS-MTT) systems [1], [2] require to estimate and predict the states (position, velocity, etc) of the targets evolving in the surveillance area covered by a set of sensors. These estimation and prediction are based on sensors measurements and dynamical models assumptions. In the monosensor context, MTT requires classically to solve the data association (DA) problem to associate the available measurements at a given time with the predicted states of the targets to update their tracks using filtering techniques (Kalman filter, Particle filter, etc). In the multisensor MTT context, we need to solve more difficult multi-dimensional assignment problems under constraints. Fortunately, efficient algorithms have been developed in the operational research and tracking communities for formalizing and solving these optimal assignments problems (see the related references detailed in the sequel).

Before going further, it is necessary to recall briefly the basis of DA problem and the methods to solve it. This problem can be formulated as follows: We have  $m > 1$  targets  $T_i$

( $i = 1, \dots, m$ ), and  $n > 1$  measurements<sup>1</sup>  $z_j$  ( $j = 1, \dots, n$ ) at a given time  $k$ , and a  $m \times n$  rewards (gain/payoff) matrix  $\Omega = [\omega(i, j)]$  whose elements  $\omega(i, j) \geq 0$  represent the payoff (usually homogeneous to the likelihood) of the association of target  $T_i$  with measurement  $z_j$ , denoted  $(T_i, z_j)$ . The data association problem consists in finding the global optimal assignment of the targets with some measurements by maximizing<sup>2</sup> the overall gain in such a way that no more than one target is assigned to a measurement, and reciprocally.

Without loss of generality, we can assume  $\omega(i, j) \geq 0$  because if some elements  $\omega(i, j)$  of  $\Omega$  were negative, we can always add the constant value<sup>3</sup> to all elements of  $\Omega$  to work with a new payoff matrix  $\Omega' = [\omega'(i, j)]$  having all elements  $\omega'(i, j) \geq 0$ , and we get same optimal assignment solution with  $\Omega$  and with  $\Omega'$ . Moreover, we can also assume without loss of generality  $m \leq n$  because otherwise we can always swap the roles of targets and measurements in the mathematical problem definition by working directly with  $\Omega^t$  instead, where the superscript  $t$  denotes the transposition of the matrix. The optimal assignment problem consists of finding the  $m \times n$  binary association matrix  $\mathbf{A} = [a(i, j)]$  which maximizes the global rewards

$$R(\Omega, \mathbf{A}) \triangleq \sum_{i=1}^m \sum_{j=1}^n \omega(i, j) a(i, j), \quad (1)$$

$$\text{subject to } \begin{cases} \sum_{j=1}^n a(i, j) = 1 & (i = 1, \dots, m), \\ \sum_{i=1}^m a(i, j) \leq 1 & (j = 1, \dots, n), \\ a(i, j) \in \{0, 1\}. \end{cases} \quad (2)$$

The association indicator value  $a(i, j) = 1$  means that the corresponding target  $T_i$  and measurement  $z_j$  are associated, and  $a(i, j) = 0$  means that they are not associated ( $i = 1, \dots, m$  and  $j = 1, \dots, n$ ).

<sup>1</sup>In a multi-sensor context targets can be replaced by tracks provided by a given tracker associated with a type of sensor, and measurements can be replaced by another tracks set. In different contexts, possible equivalents are assigning personnel to jobs or assigning delivery trucks to locations.

<sup>2</sup>In some problems, the matrix  $\Omega = [\omega(i, j)]$  represents a cost matrix whose elements are the negative log-likelihood of association hypotheses. In this case, the data association problems consists in finding the best assignment that minimizes the overall cost.

<sup>3</sup>equals to the absolute value of the minimum of  $\Omega$ .

The solution of the optimal assignment problem stated in (1)–(2) is well reported in the literature and several efficient methods have been developed in the operational research and tracking communities to solve it. The most well-known algorithms are Kuhn-Munkres (or Hungarian) algorithm [3], [4] and its extension to rectangular matrices proposed by Bourgeois and Lassalle in [5], Jonker-Volgenant method [6], and Auction [7]. More sophisticated methods using Murty’s method [8], and some variants [9], [10], [11], [12], [13], [14], [15], are also able to provide not only the best assignment, but also the  $m$ -best assignments. We will not present in details all these classical methods because they have been already well reported in the literature [16], [17].

The purpose of this paper is to propose a solution for two important problems related with the aforementioned Data Association issue:

• **Problem 1 (mono-criterion):** Suppose that the DA reward  $\Omega_1$  has been established based on a unique criterion  $C_1$  then we want to evaluate the quality<sup>4</sup> of each association (pairing) provided in the optimal solution by one of the aforementioned algorithms. The choice of the algorithm does not matter as soon as they are able to provide the optimal DA solution represented by a binary matrix  $\mathbf{A}_1$  (assumed to be unique here for convenience). So based on  $\Omega_1$  and  $\mathbf{A}_1$ , we want to estimate the quality matrix  $\mathbf{Q}_1$  of the optimal pairing solutions given in  $\mathbf{A}_1$ . This quality matrix will be useful to select optimal association pairings that have sufficient quality to be used to update the tracking filters, and not to use the optimal data associations that have a poor quality, which will save computational time and avoid to potentially degrade tracking performances.

• **Problem 2 (multi-criteria):** We assume that we have different Rewards matrices  $\Omega_1, \dots, \Omega_K$  ( $K > 1$ ), established from different criteria from which we can draw optimal DA solutions  $\mathbf{A}_1, \dots, \mathbf{A}_K$  with their corresponding quality matrices  $\mathbf{Q}_1, \dots, \mathbf{Q}_K$  (obtained by the method used for solving Problem 1). We assume that each criterion  $C_k$ ,  $k = 1, \dots, K$  has its own importance with respect to the others which is expressed either by a given relative importance  $K \times K$  matrix  $\mathbf{M}$ , or directly by a weighting  $M \times 1$  vector  $\mathbf{w}$ . The problem 2 consists in finding the optimal (i.e. the one generating the best global quality) DA solution based on all information drawn from the independent multiple criteria we have, that is from  $\mathbf{Q}_1, \dots, \mathbf{Q}_K$  and  $\mathbf{M}$  (or  $\mathbf{w}$ ) in a well-justified and comprehensive manner.

This paper is organized as follows: in section 2 we present a method for solving problem 1 which uses both 1st-best and 2nd-best DA solutions provided by Murty’s algorithm. Our method is based on Belief Functions (BF), the Proportional Conflict Redistribution fusion rule #6 (PCR6) developed in Dezert-Smarandache Theory (DSMT) framework [19], and the pignistic probability transform. Section 3, proposes a solution

<sup>4</sup>In this paper, the quality of a pairing of the optimal DA solution refers to a confidence score which corresponds to a degree of trustfulness one grants to this pairing for taking the decision to use it, or not.

for Problem 2 exploiting Saaty’s AHP method, BF and also Murty’s algorithm. Section 4 presents a full simple detailed example to show how the method works for readers who want to check by themselves our results. Section 5 will conclude this paper with perspectives.

## II. SOLUTION OF PROBLEM 1 (MONO-CRITERION)

This solution has already been addressed in details in [21] and we will just briefly present here the main ideas for making this paper self containing. In problem 1, we want to establish a confidence level (i.e. a quality indicator) of the pairings of the optimal data association solution. More precisely, we are searching for an answer to the question: how to measure the quality of the pairings  $a(i, j) = 1$  provided in the optimal assignment solution  $\mathbf{A}$ ? The necessity to establish a quality indicator is motivated by the following three main practical reasons:

- 1) In some practical tracking environment with the presence of clutter, some association decisions ( $a(i, j) = 1$ ) are doubtful. For these unreliable associations, it is better to wait for new information (measurements) instead of applying the hard data association decision, and making potentially serious association mistakes.
- 2) In some multisensor systems, it can be also important to save energy consumption for preserving a high autonomy of the system. For this goal, only the most trustful specific associations provided in the optimal assignment have to be used instead of all of them.
- 3) The best optimal assignment solution is not necessarily unique. In such situation, the establishment of quality indicators may help in selecting one particular optimal assignment solution among multiple possible choices.

It is worth noting that the 1st-best, as well as the 2nd-best, optimal assignment solutions are unfortunately not necessarily unique. Therefore, we need to take into account the possible multiplicity of assignments in the analysis of the problem. The multiplicity index of the best optimal assignment solution is denoted  $\beta_1 \geq 1$ , and the multiplicity index of the 2nd-best optimal assignment solution is denoted  $\beta_2 \geq 1$ , and we will denote the sets of corresponding assignment matrices by  $\mathcal{A}_1 = \{\mathbf{A}_1^{(k_1)}, k_1 = 1 \dots, \beta_1\}$  and by  $\mathcal{A}_2 = \{\mathbf{A}_2^{(k_2)}, k_2 = 1 \dots, \beta_2\}$ . Here are three simple examples with different multiplicities in solutions:

**Example 1:** If we take  $\Omega = \begin{bmatrix} 8 & 1 & 2 \\ 5 & 3 & 3 \end{bmatrix}$ , then  $\beta_1 = 2$  and  $\beta_2 = 1$  because the 1st best and 2nd best DA solutions are

$$\mathbf{A}_1^{k_1=1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{A}_1^{k_1=2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

**Example 2:** If we take  $\Omega = \begin{bmatrix} 6 & 3 & 9 \\ 1 & 4 & 1 \end{bmatrix}$ , then  $\beta_1 = 1$  and  $\beta_2 = 2$  because the 1st best and 2nd best DA solutions are

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{A}_2^{k_2=1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{A}_2^{k_2=2} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

**Example 3:** If we take  $\Omega = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ , then  $\beta_1 = 2$  and  $\beta_2 = 2$  because the 1st best and 2nd best DA solutions are

$$\mathbf{A}_1^{k_1=1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_1^{k_1=2} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \\ \mathbf{A}_2^{k_2=1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2^{k_2=2} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

To establish the quality of the specific associations (pairings)  $(i, j)$  satisfying  $a_1(i, j) = 1$  belonging to the optimal assignment matrix  $\mathbf{A}_1$ , we propose to use both  $\mathbf{A}_1$  and 2nd-best assignment solution  $\mathbf{A}_2$ . The basic idea is to use the values  $a_1(i, j) = 1$  in the best, and  $a_2(i, j)$  in the 2nd-best assignments to identify the change (if any) of the optimal pairing  $(i, j)$ . In fact, we assume<sup>5</sup> that higher quality of an entry in a quality matrix suggests that its association in an optimal solution is more stable across those good solutions. The connection between the stability of an association across the good solutions and the stability over an error in measurement is done through the components of the reward matrices (the inputs of our method) which must take into account the measurement uncertainties. Based on this assumption, our quality indicator will be defined using both the stability of the pairing and its relative impact in the global reward. This proposed method works also when the 2nd-best assignment solution  $\mathbf{A}_2$  is not unique (as shown in examples 2 and 3). Our method helps to select the best (most trustful) optimal assignment in case of multiplicity of  $\mathbf{A}_1$  matrices. We do not claim that the definition of the quality matrix proposed in this work is the best proposal. However, we propose a new comprehensive way of solving this problem from a practical standpoint.

To take into account efficiently the reward values of each specific association given in the best assignment  $\mathbf{A}_1$  and in the 2nd-best assignment  $\mathbf{A}_2^{k_2}$  for estimating the quality of DA solutions, we propose to use the following construction of quality indicators depending on the type of matching:

- When  $a_1(i, j) = a_2^{k_2}(i, j) = 0$ , one has full agreement on “non-association”  $(T_i, z_j)$  in  $\mathbf{A}_1$  and in  $\mathbf{A}_2^{k_2}$  and this non-association  $(T_i, z_j)$  has no impact on the global rewards  $R_1(\Omega, \mathbf{A}_1)$  and  $R_2(\Omega, \mathbf{A}_2^{k_2})$ , and it will be useless. Therefore, we can set its quality arbitrarily to any arbitrary value, typically we take  $q^{k_2}(i, j) = 0$  because these values are not useful at all for the application (i.e. tracking) standpoint.
- When  $a_1(i, j) = a_2^{k_2}(i, j) = 1$ , one has a full agreement on the association  $(T_i, z_j)$  in  $\mathbf{A}_1$  and in  $\mathbf{A}_2^{k_2}$ . his association  $(T_i, z_j)$  has however different impacts in the global rewards values  $R_1(\Omega, \mathbf{A}_1)$  and  $R_2(\Omega, \mathbf{A}_2^{k_2})$ . To qualify the quality

<sup>5</sup>This assumption has however not been proven formally yet and its validity is a challenging open-question left for future research works.

of this association  $(T_i, z_j)$ , we define the two basic belief assignments (BBA's) on  $X \triangleq (T_i, z_j)$  and  $X \cup \neg X$  (the ignorance), for  $s = 1, 2$  as follows:

$$\begin{cases} m_s(X) = a_s(i, j) \cdot \omega(i, j) / R_s(\Omega, \mathbf{A}_s), \\ m_s(X \cup \neg X) = 1 - m_s(X). \end{cases} \quad (3)$$

Applying the conjunctive fusion rule (here one has no conflicting mass), we get

$$\begin{cases} m(X) = m_1(X)m_2(X) + m_1(X)m_2(X \cup \neg X) \\ \quad + m_1(X \cup \neg X)m_2(X), \\ m(X \cup \neg X) = m_1(X \cup \neg X)m_2(X \cup \neg X). \end{cases} \quad (4)$$

Applying the pignistic transformation<sup>6</sup> [20], we get finally  $BetP(X) = m(X) + \frac{1}{2} \cdot m(X \cup \neg X)$  and  $BetP(\neg X) = \frac{1}{2} \cdot m(X \cup \neg X)$ . Therefore, we choose as quality indicator for the association  $(T_i, z_j)$  the value  $q^{k_2}(i, j) \triangleq BetP(X) = m(X) + \frac{1}{2} \cdot m(X \cup \neg X)$ .

- When  $a_1(i, j) = 1$  and  $a_2^{k_2}(i, j) = 0$ , one has a disagreement (conflict) on the association  $(T_i, z_j)$  in  $\mathbf{A}_1$  and in  $(T_i, z_{j_2})$  in  $\mathbf{A}_2^{k_2}$ , where  $j_2$  is the measurement index such that  $a_2(i, j_2) = 1$ . To qualify the quality of this non-matching association  $(T_i, z_j)$ , we define the two following basic belief assignments (BBA's) of the propositions  $X \triangleq (T_i, z_j)$  and  $Y \triangleq (T_i, z_{j_2})$

$$\begin{cases} m_1(X) = a_1(i, j) \cdot \frac{\omega(i, j)}{R_1(\Omega, \mathbf{A}_1)}, \\ m_1(X \cup Y) = 1 - m_1(X), \end{cases} \quad (5)$$

and

$$\begin{cases} m_2(Y) = a_2(i, j_2) \cdot \frac{\omega(i, j_2)}{R_2(\Omega, \mathbf{A}_2^{k_2})}, \\ m_2(X \cup Y) = 1 - m_2(Y). \end{cases} \quad (6)$$

Applying the conjunctive fusion rule, we get  $m(X \cap Y = \emptyset) = m_1(X)m_2(Y)$  and

$$\begin{cases} m(X) = m_1(X)m_2(X \cup Y), \\ m(Y) = m_1(X \cup Y)m_2(Y), \\ m(X \cup Y) = m_1(X \cup Y)m_2(X \cup Y). \end{cases} \quad (7)$$

Because we need to work with a normalized combined BBA, we can choose different rules of combination (say either Dempster-Shafer's rule, Dubois-Prade's rule, Yager's rule [19], etc). In this work, we propose to use the Proportional Conflict Redistribution rule no. 6 (PCR6) proposed originally in DSmt framework [19] because it has been proved very efficient in practice [28], [29]. Hence with PCR6, we get:

$$\begin{cases} m(X) = m_1(X)m_2(X \cup Y) + m_1(X) \cdot \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)}, \\ m(Y) = m_1(X \cup Y)m_2(Y) + m_2(X) \cdot \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)}, \\ m(X \cup Y) = m_1(X \cup Y)m_2(X \cup Y). \end{cases} \quad (8)$$

Applying the pignistic probability transformation, we get finally  $BetP(X) = m(X) + \frac{1}{2} \cdot m(X \cup Y)$  and

<sup>6</sup>We have chosen here BetP for its simplicity and because it is widely known, but DSmtP could be used instead for expecting better performances [19].

$BetP(Y) = m(Y) + \frac{1}{2} \cdot m(X \cup Y)$ . Therefore, we choose the quality indicators as follows:  $q^{k_2}(i, j) = BetP(X)$ , and  $q^{k_2}(i, j_2) = BetP(Y)$ .

The absolute quality factor  $Q_{abs}(\mathbf{A}_1)$  of the optimal assignment given in  $\mathbf{A}_1$  conditioned by  $\mathbf{A}_2^{k_2}$ , for any  $k_2 \in \{1, 2, \dots, \beta_2\}$  is defined as

$$Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2}) \triangleq \sum_{i=1}^m \sum_{j=1}^n a_1(i, j) q^{k_2}(i, j). \quad (9)$$

The absolute average quality factor  $Q_{aver}(\mathbf{A}_1)$  per association of the optimal assignment given in  $\mathbf{A}_1$  conditioned by  $\mathbf{A}_2^{k_2}$ , for any  $k_2 \in \{1, 2, \dots, \beta_2\}$  is defined by

$$Q_{aver}(\mathbf{A}_1, \mathbf{A}_2^{k_2}) = \frac{1}{m} \cdot Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2}). \quad (10)$$

where  $m$  is the number of "1" in the optimal DA matrix  $\mathbf{A}_1$  (i.e. the number of targets).

To take into account the eventual multiplicities (when  $\beta_2 > 1$ ) of the 2nd-best assignment solutions  $\mathbf{A}_2^{k_2}$ ,  $k_2 = 1, 2, \dots, \beta_2$ , we need to combine the  $Q_I(\mathbf{A}_1, \mathbf{A}_2^{k_2})$  values. Several methods can be used for this, in particular we can use either:

– **A weighted averaging approach:** The quality indicator components  $q(i, j)$  of the quality matrix  $\mathbf{Q}$  are then obtained by averaging the qualities obtained from each comparison of  $\mathbf{A}_1$  with  $\mathbf{A}_2^{k_2}$ . More precisely, one will take

$$q(i, j) \triangleq \sum_{k_2=1}^{\beta_2} w(\mathbf{A}_2^{k_2}) q^{k_2}(i, j), \quad (11)$$

where  $w(\mathbf{A}_2^{k_2})$  is a weighting factor in  $[0, 1]$ , such that  $\sum_{k_2=1}^{\beta_2} w(\mathbf{A}_2^{k_2}) = 1$ . Since all assignments  $\mathbf{A}_2^{k_2}$  have the same global reward value  $R_2$ , then we suggest to take  $w(\mathbf{A}_2^{k_2}) = 1/\beta_2$ . A more elaborate method would consist of using the quality indicator of  $\mathbf{A}_2^{k_2}$  based on the 3rd-best solution, which can be itself computed from the quality of the 3rd assignment solution based on the 4th-best solution, and so on by a similar mechanism.

– **A belief-based approach:** (see [18] for basics on belief functions): A second method would express the quality by a belief interval  $[q^{\min}(i, j), q^{\max}(i, j)]$  in  $[0, 1]$  instead of single real number  $q(i, j)$  in  $[0, 1]$ . More precisely, one can compute the belief and plausibility bounds of the quality by taking  $q^{\min}(i, j) \equiv Bel(a_1(i, j)) = \min_{k_2} q^{k_2}(i, j)$  and  $q^{\max}(i, j) \equiv Pl(a_1(i, j)) = \max_{k_2} q^{k_2}(i, j)$ . Hence for each possible pair  $(i, j)$ , one can define a basic belief assignment (BBA)  $m_{ij}(\cdot)$  on the frame of discernment  $\Theta \triangleq \{T = \text{trustful}, -T = \text{not trustful}\}$ , which characterizes the quality of the pairing  $(i, j)$  in the optimal assignment solution  $\mathbf{A}_1$ , as follows

$$\begin{cases} m_{ij}(T) = q^{\min}(i, j), \\ m_{ij}(-T) = 1 - q^{\max}(i, j), \\ m_{ij}(T \cup -T) = q^{\max}(i, j) - q^{\min}(i, j). \end{cases} \quad (12)$$

Because only the optimal associations<sup>7</sup>  $(i, j)$  such that  $a_1(i, j) = 1$  are useful in tracking algorithms to update the tracks, we do not need to pay attention (compute and store) the qualities of components  $(i, j)$  such that  $a_1(i, j) = 0$ . In fact all components  $(i, j)$  such that  $a_1(i, j) = 0$  should be set to zero by default in  $\mathbf{Q}$  matrix.

**Example 4:** Let's consider the rewards matrix

$$\mathbf{\Omega} = \begin{bmatrix} 1 & 11 & 45 & 30 \\ 17 & 8 & 38 & 27 \\ 10 & 14 & 35 & 20 \end{bmatrix}.$$

We get one 1st best ( $\beta_1 = 1$ ) and four 2nd best ( $\beta_2 = 4$ ) DA solutions with their respective qualities as follows:

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \Rightarrow R_1(\mathbf{\Omega}, \mathbf{A}_1) = 86,$$

$$\mathbf{A}_2^{k_2=1} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \Rightarrow R_2(\mathbf{\Omega}, \mathbf{A}_2^{k_2=1}) = 82,$$

$$\mathbf{Q}(\mathbf{A}_1, \mathbf{A}_2^{k_2=1}) \approx \begin{bmatrix} 0 & 0 & 0.59 & 0 \\ 0 & 0 & 0 & 0.41 \\ 0 & 0.65 & 0 & 0 \end{bmatrix},$$

$$\mathbf{A}_2^{k_2=2} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow R_2(\mathbf{\Omega}, \mathbf{A}_2^{k_2=2}) = 82,$$

$$\mathbf{Q}(\mathbf{A}_1, \mathbf{A}_2^{k_2=2}) \approx \begin{bmatrix} 0 & 0 & 0.89 & 0 \\ 0 & 0 & 0 & 0.56 \\ 0 & 0.45 & 0 & 0 \end{bmatrix},$$

$$\mathbf{A}_2^{k_2=3} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \Rightarrow R_2(\mathbf{\Omega}, \mathbf{A}_2^{k_2=3}) = 82,$$

$$\mathbf{Q}(\mathbf{A}_1, \mathbf{A}_2^{k_2=3}) \approx \begin{bmatrix} 0 & 0 & 0.89 & 0 \\ 0 & 0 & 0 & 0.76 \\ 0 & 0.52 & 0 & 0 \end{bmatrix},$$

$$\mathbf{A}_2^{k_2=4} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \Rightarrow R_2(\mathbf{\Omega}, \mathbf{A}_2^{k_2=4}) = 82,$$

$$\mathbf{Q}(\mathbf{A}_1, \mathbf{A}_2^{k_2=4}) \approx \begin{bmatrix} 0 & 0 & 0.59 & 0 \\ 0 & 0 & 0 & 0.56 \\ 0 & 0.35 & 0 & 0 \end{bmatrix},$$

Note that the absolute quality factors are :

$$\begin{aligned} Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=1}) &\approx 1.66, & Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=2}) &\approx 1.91, \\ Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=3}) &\approx 2.19, & Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=4}) &\approx 1.51. \end{aligned}$$

Therefore, we can see that

$$\begin{aligned} Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=3}) &> Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=2}) \\ &> Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=1}) > Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2=4}), \end{aligned}$$

<sup>7</sup>found using Murty's algorithm.

which makes perfectly sense because  $\mathbf{A}_1$  has more matching pairings with  $\mathbf{A}_2^{k_2=3}$  than with others 2nd-best assignments  $\mathbf{A}_2^{k_2}$  ( $k_2 \neq 3$ ). These pairings have also the strongest impact in the global reward value. Therefore, the quality matrix  $\mathbf{Q}$  differentiates the quality of each pairing in the optimal assignment  $\mathbf{A}_1$  as expected. This method provides an effective and comprehensive solution to estimate the quality of each specific association provided in the optimal assignment solution  $\mathbf{A}_1$ . The averaged qualities per association are:

$$\begin{aligned} Q_{aver}(\mathbf{A}_1, \mathbf{A}_2^{k_2=1}) &\approx 0.55, & Q_{aver}(\mathbf{A}_1, \mathbf{A}_2^{k_2=2}) &\approx 0.63, \\ Q_{aver}(\mathbf{A}_1, \mathbf{A}_2^{k_2=3}) &\approx 0.73, & Q_{aver}(\mathbf{A}_1, \mathbf{A}_2^{k_2=4}) &\approx 0.50. \end{aligned}$$

The global quality matrix is then given by (using the averaging approach)

$$\begin{aligned} \mathbf{Q}(\mathbf{A}_1, \mathcal{A}_2) &= \frac{1}{\beta_2} \sum_{k_2=1}^{\beta_2} \mathbf{Q}(\mathbf{A}_1, \mathbf{A}_2^{k_2}) \\ &\approx \begin{bmatrix} 0 & 0 & 0.74 & 0 \\ 0 & 0 & 0 & 0.57 \\ 0 & 0.49 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The global quality indexes  $Q_{abs}(\mathbf{A}_1, \mathcal{A}_2)$  and  $Q_{aver}(\mathbf{A}_1, \mathcal{A}_2)$  are then approximately equal to 1.8 and 0.6 respectively.

One can also improve the estimation of the quality matrix by using the absolute quality factor of each solution  $\mathbf{Q}(\mathbf{A}_1, \mathbf{A}_2^{k_2})$ , for  $k_2 = 1, \dots, \beta_2$  to define the normalized weighting factors as follows:

$$\mathbf{w} = [w_{k_2}, k_2 = 1, \dots, \beta_2]',$$

with  $w_{k_2} \triangleq \frac{Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2})}{K}$ , and where the normalization factor  $K$  is given by  $K = \sum_{k_2=1}^{\beta_2} Q_{abs}(\mathbf{A}_1, \mathbf{A}_2^{k_2})$ . In this example, we get the weights

$$\begin{aligned} \mathbf{w} &= [w_1 \ w_2 \ w_3 \ w_4]' \approx \left[ \frac{1.66}{7.27} \ \frac{1.91}{7.27} \ \frac{2.19}{7.27} \ \frac{1.51}{7.27} \right]' \\ &= [0.2283 \ 0.2627 \ 0.3012 \ 0.2077]'. \end{aligned}$$

The global quality matrix is then given by (using the averaging approach)

$$\begin{aligned} \mathbf{Q}(\mathbf{A}_1, \mathcal{A}_2) &= \sum_{k_2=1}^{\beta_2} w_{k_2} \mathbf{Q}(\mathbf{A}_1, \mathbf{A}_2^{k_2}) \\ &\approx \begin{bmatrix} 0 & 0 & 0.76 & 0 \\ 0 & 0 & 0 & 0.58 \\ 0 & 0.49 & 0 & 0 \end{bmatrix}. \end{aligned}$$

If we prefer to use the Belief Interval Measure (BIM) instead of the previous averaging approach, we will get in this example the following imprecise qualities values:

Optimal. assignments	BIM
(1, 3)	$\approx [0.59, 0.89]$
(2, 4)	$\approx [0.41, 0.76]$
(3, 2)	$\approx [0.35, 0.65]$

Based on the comparisons of (pessimistic) lower bounds, or (optimistic) upper bounds of BIM, we observe that we get

a consistent ordering of the qualities of the optimal solutions (same ordering as with the averaging method).

### III. SOLUTION OF THE 2ND PROBLEM (MULTI-CRITERIA)

In this section, we evaluate the global DA association solution, with estimation of its quality, based on the knowledge of the qualities of multiple optimal DA solutions established separately based on distinct association criteria  $C_k$ ,  $k = 1, \dots, K$ . More precisely, given the set of quality matrices  $\mathbf{Q}^k$  ( $k = 1, \dots, K$ ) defined by the components  $q^k(i, j)$  according to Eq.(11), how to establish the global optimal DA solution with its overall quality matrix  $\mathbf{Q}$ ? Moreover, we want to take into account the importance of each criteria (when defined) in the establishment of the solution.

In fact this 2nd problem is linked to the previous one and the method developed for solving our first problem will also help to solve this second problem as it will be shown in the following. Our solution is based on four distinct steps:

• **Step 1:** Estimation of the normalized weighting vector  $\mathbf{w}$  of the criteria: Two simple approaches are proposed to establish the normalized criteria ranking (weighting) vector.

- 1) **Direct method:** The weightings factors can be directly established either by an external factors or by the system designer. If these weightings factors are not available, we propose to compute them from the qualities indicators derived by the method used to solve the 1st problem (see the previous section). For example, if we consider  $K$  criteria providing quality factors  $Q_{abs}^k(\mathbf{A}_1(C_k), \mathcal{A}_2(C_k))$ ,  $k = 1, 2, \dots, K$ , then we compute the normalized  $K \times 1$  weighting vector  $\mathbf{w} = [w_1 w_2 \dots w_K]'$  with the  $k$ -th component given by

$$w_k \triangleq \frac{Q_{abs}^k(\mathbf{A}_1(C_k), \mathcal{A}_2(C_k))}{\sum_{j=1}^K Q_{abs}^j(\mathbf{A}_1(C_j), \mathcal{A}_2(C_j))}, \quad (13)$$

where  $Q_{abs}^k(\mathbf{A}_1(C_k), \mathcal{A}_2(C_k))$  is the absolute quality factor obtained from the quality matrix  $\mathbf{Q}^k(\mathbf{A}_1, \mathcal{A}_2)$  of the optimal DA for the criteria  $C_k$ .

- 2) **Saaty's method:** This method is part of Saaty's AHP method widely used for multi-criteria decision analysis in operational research [22], [23], [24], and it has been connected with information fusion and belief functions in [25], [26], [27]. The relative importance of one criterion over another must be expressed by the system designer using a pairwise  $K \times K$  comparison matrix (also called knowledge matrix)  $\mathbf{M} = [m_{pq}]$  where the element  $m_{pq}$  of the matrix defines the importance of criteria  $C_p$  with respect to the criteria  $C_q$ , with  $p, q \in \{1, 2, \dots, K\}$ . For example, see [25] for details, let's consider only  $K = 3$  criteria, if the comparison matrix is given by

$$\mathbf{M} = \begin{bmatrix} (1/1) & (1/3) & (4/1) \\ (3/1) & (1/1) & (5/1) \\ (1/4) & (1/5) & (1/1) \end{bmatrix},$$

it means that the element  $m_{13} = 4/1$  indicates that the criteria  $C_1$  is four times as important as the criteria  $C_3$  for the system designer (or decision-maker), etc. From this pairwise matrix, Saaty demonstrated that the ranking of the priorities of the criteria can be obtained from the normalized eigenvector, denoted  $\mathbf{w}$ , associated with the principal/max eigenvalue of the matrix  $\mathbf{M}$ , denoted  $\lambda$ . In our example, one gets  $\lambda = 3.0857$  and  $\mathbf{w} = [0.2797 \ 0.6267 \ 0.0936]^T$  which shows that  $C_2$  criterion is the most important criterion with the weight 0.6267, then the criterion  $C_1$  is the second most important criterion with weight 0.2797, and finally  $C_3$  criterion is the least important criterion with weight 0.0936.

- **Step 2:** Combined estimation of the qualities of each target association

Once the normalized weighting vector  $\mathbf{w}$  of the criteria has been obtained, we need at first to compute the combined/weighted estimation of the qualities of each target association with the  $n$  available measurements. This is done by building the following  $n \times K$  matrix

$$\mathbf{Q}_i \triangleq [\mathbf{q}_i(C_1) \dots \mathbf{q}_i(C_K)], \quad (14)$$

where each column  $\mathbf{q}_i(C_k)$  of the matrix  $\mathbf{Q}_i$  corresponds to the transpose of the  $i$ -th row of the quality matrix  $\mathbf{Q}^k(\mathbf{A}_1, \mathbf{A}_2)$ .

Then following AHP approach, we multiply this  $n \times K$  matrix  $\mathbf{Q}_i$  by the normalized criteria ranking  $K \times 1$  vector  $\mathbf{w}$  (obtained either from the direct method of Saaty's one) to get the combined estimation of the qualities of each target association. More precisely, for the  $i$ -th target, we obtain the following  $n \times 1$  vector

$$\mathbf{q}_i \triangleq \mathbf{Q}_i \mathbf{w}. \quad (15)$$

- **Step 3:** Search for the optimal global assignment based on combined qualities derived from the criteria.

From the set of  $m$  vectors  $\mathbf{q}_i$  ( $i=1,2,\dots,m$ ) we need to solve now a new optimal DA association problem with the (global)  $m \times n$  rewards matrix defined by

$$\mathbf{\Omega}_G \triangleq [\mathbf{q}_1 \ \mathbf{q}_2 \ \dots \ \mathbf{q}_m]^T. \quad (16)$$

Murty's algorithm is then used again here to get the optimal DA solution(s) providing the best global reward, and to generate also all the 2nd-best solutions that are necessary to estimate its quality in Step 4.

- **Step 4:** Estimation of the quality of the optimal DA solution.

We use the method described in Section 2 for solving the problem 1 to estimate the quality of the optimal DA solution. If several 1st-best DA solutions occur, we choose the solution generating the highest  $Q_{abs}$  quality index.

#### IV. A SIMPLE ILLUSTRATIVE EXAMPLE

For the sake of simplicity, let's consider the following example with  $m = 3$  targets,  $n = 5$  measurements, and 3 criteria  $C_1$ ,  $C_2$  and  $C_3$  associated with the (randomly chosen) rewards matrices:

$$\mathbf{\Omega}(C_1) = \begin{bmatrix} 100 & 20 & 33 & 5 & 27 \\ 11 & 80 & 25 & 37 & 62 \\ 38 & 2 & 24 & 78 & 46 \end{bmatrix},$$

$$\mathbf{\Omega}(C_2) = \begin{bmatrix} 87 & 35 & 43 & 20 & 95 \\ 28 & 83 & 25 & 10 & 29 \\ 10 & 7 & 72 & 41 & 29 \end{bmatrix},$$

$$\mathbf{\Omega}(C_3) = \begin{bmatrix} 25 & 78 & 49 & 60 & 9 \\ 30 & 26 & 79 & 20 & 49 \\ 20 & 20 & 3 & 47 & 81 \end{bmatrix}.$$

##### A. Qualities of optimal data associations

Applying the method described in section 1, we easily obtain the following quality matrices of optimal DA solutions:

- For criterion  $C_1$ , one gets  $\beta_1 = 1$  and  $\beta_2 = 1$ , and the following 1st best and 2nd best DA solutions

$$\mathbf{\Omega}(C_1) \Rightarrow \begin{cases} \mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \\ \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \end{cases}$$

providing the 1st and 2nd best global rewards  $R(\mathbf{\Omega}(C_1), \mathbf{A}_1) = 258$  and  $R(\mathbf{\Omega}(C_1), \mathbf{A}_2) = 240$ . Applying the method described in Section 2, we obtain the following quality matrix related with the optimal DA based on criterion  $C_1$ :

$$\mathbf{Q}^1 \approx \begin{bmatrix} 0.82 & 0 & 0 & 0 & 0 \\ 0 & 0.52 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.76 & 0 \end{bmatrix}.$$

- For criterion  $C_2$ , one gets  $\beta_1 = 1$  and  $\beta_2 = 1$ , and the following 1st best and 2nd best DA solutions

$$\mathbf{\Omega}(C_2) \Rightarrow \begin{cases} \mathbf{A}_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \\ \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \end{cases}$$

providing the 1st and 2nd best global rewards  $R(\mathbf{\Omega}(C_2), \mathbf{A}_1) = 250$  and  $R(\mathbf{\Omega}(C_2), \mathbf{A}_2) = 242$ . Applying the method described in Section 2, we obtain the following quality matrix related with the optimal DA based on criterion  $C_2$ :

$$\mathbf{Q}^2 \approx \begin{bmatrix} 0 & 0 & 0 & 0 & 0.51 \\ 0 & 0.78 & 0 & 0 & 0 \\ 0 & 0 & 0.74 & 0 & 0 \end{bmatrix}.$$

- For criterion  $C_3$ , one gets  $\beta_1 = 1$  and  $\beta_2 = 1$ , and the following 1st best and 2nd best DA solutions

$$\Omega(C_3) \Rightarrow \begin{cases} \mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{A}_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \end{cases}$$

providing the 1st and 2nd best global rewards  $R(\Omega(C_3), \mathbf{A}_1) = 238$  and  $R(\Omega(C_3), \mathbf{A}_2) = 220$ . Applying the method described in Section 2, we obtain the following quality matrix related with the optimal DA based on criterion  $C_3$ :

$$\mathbf{Q}^3 \approx \begin{bmatrix} 0 & 0.53 & 0 & 0 & 0 \\ 0 & 0 & 0.78 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.79 \end{bmatrix}.$$

### B. Multicriteria-based DA solution with its quality

- Case 1: If we assume that all criteria have the same weights in the search of optimal DA solution, then we take the normalized weighting vector as  $\mathbf{w} = [1/3 \ 1/3 \ 1/3]'$ . Therefore, the weighted average  $\Omega_G = \sum_{k=1}^{K=3} w_k \mathbf{Q}^k$  of the quality matrices  $\mathbf{Q}^1$ ,  $\mathbf{Q}^2$  and  $\mathbf{Q}^3$  gives us the following rewards matrix

$$\Omega_G \approx \begin{bmatrix} 0.27 & 0.17 & 0 & 0 & 0.17 \\ 0 & 0.43 & 0.26 & 0 & 0 \\ 0 & 0 & 0.25 & 0.25 & 0.26 \end{bmatrix}.$$

Now we solve the DA association problem to maximize the global quality reward using Murty's algorithm and we get the following 1st best and 2nd best DA solutions:

$$\Omega_G \Rightarrow \begin{cases} \mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \end{cases}$$

with the 1st and 2nd best global rewards  $R(\Omega_G, \mathbf{A}_1) \approx 0.97$  and  $R(\Omega_G, \mathbf{A}_2) \approx 0.96$ . Applying the method described in Section II to estimate the quality of this optimal DA solution, we obtain the following quality matrix:

$$\mathbf{Q} \approx \begin{bmatrix} 0.74 & 0 & 0 & 0 & 0 \\ 0 & 0.84 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.50 \end{bmatrix}.$$

- Case 2: If we use the prior information given by absolute quality indicators to build the normalized weighting vector, we get

$$Q_{abs}^1 = \sum_{i=1}^m \sum_{j=1}^n \mathbf{Q}^1(i, j) \approx 2.11,$$

$$Q_{abs}^2 = \sum_{i=1}^m \sum_{j=1}^n \mathbf{Q}^2(i, j) \approx 2.04,$$

$$Q_{abs}^3 = \sum_{i=1}^m \sum_{j=1}^n \mathbf{Q}^3(i, j) \approx 2.10,$$

and we have  $Q_{abs}^1 + Q_{abs}^2 + Q_{abs}^3 = 6.2672$ . So that, the normalized weights are given by

$$\mathbf{w} = [w_1 \ w_2 \ w_3]' = \left[ \frac{2.1154}{6.2672} \ \frac{2.0426}{6.2672} \ \frac{2.1091}{6.2672} \right]' \approx [0.3375 \ 0.3260 \ 0.3365]'$$

The weighted average  $\Omega_G = \sum_{k=1}^{K=3} w_k \mathbf{Q}^k$  of the quality matrices  $\mathbf{Q}^1$ ,  $\mathbf{Q}^2$  and  $\mathbf{Q}^3$  give us now the following rewards matrix

$$\Omega_G \approx \begin{bmatrix} 0.27 & 0.17 & 0 & 0 & 0.16 \\ 0 & 0.43 & 0.26 & 0 & 0 \\ 0 & 0 & 0.24 & 0.25 & 0.26 \end{bmatrix}.$$

Now we solve the DA association problem to maximize the global quality reward and we get the following 1st best and 2nd best DA solutions:

$$\Omega_G \Rightarrow \begin{cases} \mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\ \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \end{cases}$$

with the 1st and 2nd best global rewards  $R(\Omega_G, \mathbf{A}_1) \approx 0.97$  and  $R(\Omega_G, \mathbf{A}_2) \approx 0.96$ . Applying the method described in Section 2 to estimate the quality of this optimal DA solution, we obtain the following quality matrix:

$$\mathbf{Q} \approx \begin{bmatrix} 0.74 & 0 & 0 & 0 & 0 \\ 0 & 0.84 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.50 \end{bmatrix}.$$

Because the normalized weights based on the absolute quality indicators, in this example, are all close to 1/3, we obtain the result of the multicriteria-based optimal DA and its quality close to what we get when assuming equi-importance of the criteria in the fusion process, which is normal.

To qualify qualitatively the quality of the pairings in the optimal DA solution, we split the quality range [0;1] into three subintervals as follows<sup>8</sup>

$$\begin{aligned} \text{Low quality :} & \quad \text{if } q(i, j) \in [0; 1/3], \\ \text{Medium quality :} & \quad \text{if } q(i, j) \in [1/3; 2/3], \\ \text{High quality :} & \quad \text{if } q(i, j) \in [2/3; 1]. \end{aligned}$$

<sup>8</sup>Of course, other repartitions could be used instead depending on the what would prefer the system designer.

Based on this qualitative scale, we finally get for our example the final multicriteria-based DA solution

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

with the qualitative quality matrix

$$\mathbf{Q}^{qualitative} = \begin{bmatrix} High & - & - & - & - \\ - & High & - & - & - \\ - & - & - & - & Medium \end{bmatrix},$$

where the notation “-” means “that the quality evaluation does not apply”, or is interpreted (by default) as “the worst quality”.

**Remark:** It is worth to note that this approach provides in general not the same results as if one would combine (and weight) directly the original reward matrices of each criterion. In this example, the weighted global reward matrix  $\mathbf{\Omega}_{direct} = \sum_{k=1}^K w_k \mathbf{\Omega}(C_k)$  would be equal to

$$\mathbf{\Omega}_{direct} \approx \begin{bmatrix} 69.07 & 45.39 & 41.75 & 29.31 & 40.85 \\ 22.93 & 61.43 & 44.46 & 22.79 & 47.44 \\ 23.13 & 9.98 & 30.78 & 55.75 & 53.53 \end{bmatrix},$$

corresponding to the quality matrix of optimal DA solution

$$\mathbf{Q}_{direct} \approx \begin{bmatrix} 0.73 & 0 & 0 & 0 & 0 \\ 0 & 0.84 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.47 & 0 \end{bmatrix}.$$

One sees that these high quality solutions are fully consistent with the high quality solutions of our method. However, the medium quality solution (we get (3,4) pairing from the direct optimal assignment versus (3,5) assignment obtained by our method) mismatch. This reflects an ambiguity in the choice of the assignment of target  $T_3$ . Therefore, such assignment is unreliable because of its low quality, and should not be used to update the track of this target.

## V. CONCLUSION

In this paper, we have proposed two methods based on belief functions for establishing: 1) the quality of pairings given by optimal data association (or assignment) solution using a chosen algorithm (typically Murty’s algorithm coupled with Auction algorithm) with respect to a given criterion, and 2) the quality of the multicriteria-based optimal data association solution. Our methods are independent of the choice of the algorithm used in finding the optimal assignment solution, and, in case of multiple optimal solutions, they provide also a way to select the best optimal assignment solution (the one having the highest absolute quality factor). The methods developed in this paper are general in the sense that they can be applied to different types of association problems corresponding to different sets of constraints. This method can be extended to SD-assignment problems as well. As perspectives, we would like to extend our approach to the n-D assignment context, and then evaluate its performances in a realistic multi-target tracking scenario.

## REFERENCES

- [1] Y. Bar-Shalom, P. Willett, X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, Storrs, CT, USA, 2011.
- [2] D.L. Hall, C.Y. Chong, J. Llinas, M. Liggins II, *Distributed Data Fusion for Network-Centric Operations*, CRC Press, 2013.
- [3] H.W. Kuhn, *The hungarian method for the assignment problem*, Naval Research Logistic Quarterly, Vol. 2, pp. 83–97, 1955.
- [4] J. Munkres, *Algorithms for the assignment and transportation problems*, SIAM J., Vol.5 (1), pp. 32–38, March 1957.
- [5] F. Bourgeois, J.C. Lassalle, *An extension of the Munkres algorithm for the assignment problem to rectangular matrices*, Comm. of the ACM, Vol. 14(12), pp. 802–804, Dec. 1971.
- [6] R. Jonker, A. Volgenant, *A shortest augmenting path algorithm for dense and sparse linear assignment problems*, Journal of Computing, Vol. 38(4), pp. 325–340, Dec. 1987.
- [7] D.P. Bertsekas, *The auction algorithm: A distributed relaxation method for the assignment problem*, Ann. of Oper. Res., Vol. 14(1), pp. 105–123, Dec. 1988.
- [8] K.G. Murty, *An algorithm for ranking all the assignments in order of increasing cost*, Operations Research, Vol. 16(3), pp. 682–687, 1968.
- [9] C.R. Chegireddy, H.W. Hamacher, *Algorithms for finding K-best perfect matching*, Discrete Applied Mathematics, Vol. 18, pp. 155–165, 1987.
- [10] R. Danchick, G.E. Newnam, *A fast method for finding the exact N-best hypotheses for multitarget tracking*, IEEE Trans. on AES, Vol. 29(2), pp. 555–560, April 1993.
- [11] M.L. Miller, H.S. Stone, I.J. Cox, *Optimizing Murty’s ranked assignment method*, IEEE Trans. on AES, Vol. 33(3), pp. 851–862, July 1997.
- [12] M. Pascoalo, M.E. Captivo, J. Climaco, *A note on a new variant of Murty’s ranking assignments algorithm*, 4OR Journal, Vol. 1, pp. 243–255, 2003.
- [13] Z.J. Ding, D. Vandervies, *A modified Murty’s algorithm for multiple hypothesis tracking*, Proc. of SPIE Signal and Data Proc. of Small Targets, Vol. 6236, May 2006.
- [14] E. Fortunato, et al., *Generalized Murty’s algorithm with application to multiple hypothesis tracking*, Proc. of Fusion 2007, pp. 1–8, Québec, July 9–12 2007.
- [15] X. He, R. Tharmarasa, M. Pelletier, T. Kirubarajan, *Accurate Murty’s Algorithm for Multitarget Top Hypothesis*, Proc. of Fusion 2011, Chicago, IL, USA, July 5–8 2011.
- [16] H.W. Hamacher, M. Queyranne, *K-best solutions to combinatorial optimization problems*, Ann. of Oper. Res., No. 4, pp. 123–143, 1985/6.
- [17] M. Dell’Amico, P. Toth, *Algorithms and codes for dense assignment problems: the state of the art*, Discrete Appl. Math., Vol. 100, pp. 17–48, 2000.
- [18] G. Shafer, *A mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, NJ, USA, 1976.
- [19] F. Smarandache, J. Dezert, *Advances and applications of DSMT for information fusion*, Volumes 1–4, American Research Press, 2004–2015. <http://www.gallup.unm.edu/~smarandache/DSMT.htm>
- [20] P. Smets, R. Kennes, *The transferable belief model*, Artificial Intelligence, Vol. 66(2), pp. 191–234, 1994.
- [21] J. Dezert, K. Benameur, *On the Quality of Optimal Assignment for Data Association*, Proc. of Belief 2014, Oxford, UK, Sept. 26–28, 2014.
- [22] T.L. Saaty, *A scaling method for priorities in hierarchical structures*, J. of Math. Psych., Vol. 15, pp. 59–62, 1977.
- [23] T.L. Saaty, *The Analytical Hierarchy Process*, McGraw Hill, 1980.
- [24] E.H. Forman, S.I. Gass, *The analytical hierarchy process: an exposition*, Oper. Res., Vol. 49(4), pp. 469–487, 2001.
- [25] J. Dezert, J.-M. Tacnet, M. Batton-Hubert, F. Smarandache, *Multicriteria decision making based on DSMT/AHP*, Proc. of Int. Workshop on Belief Functions, Brest, France, April 2–4, 2010.
- [26] J. Dezert, J.-M. Tacnet, *Evidential Reasoning for Multi-Criteria Analysis based on DSMT-AHP*, Proc. of ISAHP 2011, Italy, June 2011.
- [27] J.-M. Tacnet, J. Dezert, M. Batton-Hubert, *AHP and Uncertainty Theories for Decision Making using the ER-MCDA Methodology*, Proc. of ISAHP 2011, Italy, June 2011.
- [28] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, Proc. of Fusion 2013 Conf., Istanbul, Turkey, July 9–12, 2013.
- [29] J. Dezert, A. Tchamova, *On the validity of Dempster’s fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intelligent Syst., Special Issue: Advances in Intelligent Systems, Vol. 29, No. 3, pp.223–252, March 2014.

# Environment Perception Using Grid Occupancy Estimation with Belief Functions

Jean Dezert, Julien Moras, Benjamin Pannetier

The French Aerospace Lab

ONERA/DTIM/EVF

F-91761 Palaiseau, France.

Emails: jean.dezert@onera.fr, julien.moras@onera.fr, benjamin.pannetier@onera.fr

Originally published as: J. Dezert, J. Moras, B. Pannetier, *Environment Perception Using Grid Occupancy Estimation with Belief Functions*, in Proc. of Fusion 2015, Washington D.C, USA, July 6–9, 2015, and reprinted with permission.

**Abstract**—Grid map offers a useful representation of the perceived world for mobile robotics navigation. It will play a major role for the security (obstacle avoidance) of next generations of terrestrial vehicles, as well as for future autonomous navigation systems. In a grid map, the occupancy of each cell representing a small piece of the surrounding area of the vehicle must be estimated at first from sensors measurements, and then it must also be classified into different classes in order to get a complete and precise perception of the dynamic environment where the vehicle moves. So far, the estimation and the grid map updating have been done using fusion techniques based on the probabilistic framework, or on the classical belief function framework thanks to an inverse model of the sensors and Dempster-Shafer rule of combination. Recently we have shown that PCR6 rule (Proportional Conflict Redistribution rule #6) proposed in DSMT (Dezert-Smarandache) Theory did improve substantially the quality of grid map with respect to other techniques, specially when the quality of available information is low, and when the sources of information appear as conflicting. In this paper, we go further and we analyze the performance of the improved version of PCR6 with Zhang’s degree of intersection. We will show through different realistic scenarios (based on a 4-layers LIDAR sensor) the benefit of using this new rule of combination in a practical application.

**Keywords:** Information fusion, grid map, cell occupancy, perception, belief functions, DSMT, PCR6, ZPCR6.

## I. INTRODUCTION

Occupancy Grids (OG) are often used for intelligent vehicle environment perception and navigation, which requires techniques for data fusion, localization and obstacle avoidance. As OGs manage a representation of the environment that does not make any assumption on the geometrical shape of the detected elements, they provide a general framework to deal with complex perception conditions. In our previous works, we did focus on the use of a multi-echo and multi-layer LIDAR system in order to characterize the dynamic surrounding environment of a vehicle driving in common traffic conditions. The perception strategy involved map estimation and scan grids [1], [2] based either on the classical bayesian framework, or on classical evidential framework based on Dempster-Shafer theory (DST) [3] of belief functions. The map grid acts as a filter that accumulate information and allows to detect moving objects. A comparative analysis of performances of these approaches has already been published recently in [4].

In dynamic environments, it is crucial to have a good modeling of the information flow in the data fusion process

in order to avoid adding wrong implicit prior knowledge that will need time to be forgotten. In this context, evidential OG are particularly interesting to make a good management of the information since it is possible to explicitly make the distinction between non explored and moving cells.

The idea of using the probabilistic framework to estimate the grid occupancy has been popularized by Elfes in his pioneered works in 1990’s [8]. Later, the idea has been extended with the fuzzy logic theory framework by Oriolo et al. [10], and in parallel with the belief function (evidential) framework as well [11]–[15]. Most of the aforementioned research works dealt only with acoustic sensors (i.e SONAR). Recently, DSMT has also been applied for the perception of the environment with acoustic sensors as reported in [16]–[18].

The aim of this paper is to analyze the performance of the improved version of PCR6 taking into account Zhang’s degree of intersection of focal elements (called ZPCR6 rule) which has been presented in details in the companion paper [7] in a realistic perception problem using a 4-layers LIDAR sensor. We show how the environment perception with non acoustic sensors can be done, and compare the performances of different fusion rules (Bayesian, Dempster-Shafer, PCR6 and ZPCR6) in terms of accuracy of grid map estimation.

This paper is organized as follows. After a short presentation of the basics of belief functions and rules of their combination based on DST and DSMT in the next section, we will present the inverse sensor models in section III with the construction of the basic belief assignments (BBA). In section IV, we present an illustrating scenario for environment perception including a mobile object with a platform equipped with a LIDAR, and we compare our new realistic simulation results with those obtained by the probabilistic and the classical belief-based approaches. We will show how static and mobile objects are extracted from the occupancy grid map using digital image processing. Finally, conclusion and outline perspectives are given in section V.

## II. BASICS OF BELIEF FUNCTIONS AND THEIR FUSION

Dempster-Shafer’s theory (DST) of evidence has been developed by Shafer in 1976 from Dempster’s works [3]. DST is known also as the theory of belief functions and it is mainly characterized by a frame of discernment (FoD), sources of evidence represented by basic belief assignment (BBA), belief

(Bel) and plausibility (Pl) functions, and Dempster's rule and denoted DS rule in the sequel<sup>1</sup> of combination. DST has been modified and extended into Dezert-Smarandache theory [6] (DSmT) to work with quantitative or qualitative BBA and to combine the sources of evidence in a more efficient way thanks to new proportional conflict redistribution (PCR) fusion rules – see [19]–[22] for discussion and examples. We briefly recall in the next subsections the basics of the theory of belief functions

### A. Belief functions

Let consider a finite discrete FoD  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ , with  $n > 1$ , of the fusion problem under consideration and its fusion space  $G^\Omega$  which can be chosen either as the power-set  $2^\Omega$ , the hyper-power set<sup>2</sup>  $D^\Omega$ , or the super-power set  $S^\Omega$  depending on the model that fits with the problem [6]. A BBA associated with a given source of evidence is defined as the mapping  $m(\cdot) : G^\Omega \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in G^\Omega} m(A) = 1$ . The quantity  $m(A)$  is called mass of belief of  $A$  committed by the source of evidence. Belief and plausibility functions are defined by

$$\text{Bel}(A) = \sum_{\substack{B \subseteq A \\ B \in G^\Omega}} m(B) \quad \text{and} \quad \text{Pl}(A) = \sum_{\substack{B \cap A \neq \emptyset \\ B \in G^\Omega}} m(B) \quad (1)$$

The degree of belief  $\text{Bel}(A)$  given to a subset  $A$  quantifies the amount of justified specific support to be given to  $A$ , and the degree of plausibility  $\text{Pl}(A)$  quantifies the maximum amount of potential specific support that could be given to  $A$ . If for some  $A \in G^\Omega$ ,  $m(A) > 0$  then  $A$  is called a focal element of the BBA  $m(\cdot)$ . When all focal elements are singletons and  $G^\Omega = 2^\Omega$  then the BBA  $m(\cdot)$  is called a Bayesian BBA [3] and its corresponding belief function  $\text{Bel}(\cdot)$  is homogeneous to a (possibly subjective) probability measure, and one has  $\text{Bel}(A) = P(A) = \text{Pl}(A)$ , otherwise in general one has  $\text{Bel}(A) \leq P(A) \leq \text{Pl}(A)$ ,  $\forall A \in G^\Omega$ . The vacuous BBA representing a totally ignorant source is defined as  $m_v(\Omega) = 1$ .

### B. Fusion rules

Many rules have been proposed in the literature in the past decades (see [6], Vol. 2 for a detailed list of fusion rules) to combine efficiently several distinct sources of evidence represented by the BBA's  $m_1(\cdot)$ ,  $m_2(\cdot)$ ,  $\dots$ ,  $m_s(\cdot)$  ( $s \geq 2$ ) defined on same fusion space  $G^\Omega$ . In this paper, we focus only on DS rule because it has been historically proposed in DST and it is still widely used in applications, and on the PCR rule no. 6 (i.e. PCR6) proposed in DSmT because it provides a very interesting alternative of DS rule, even if PCR6 is more complex to implement in general than DS rule.

In DST framework, the fusion space  $G^\Omega$  equals the power-set  $2^\Omega$  because Shafer's model of the frame  $\Omega$  is assumed, which means that all elements of the FoD are exhaustive and

exclusive. The combination of the BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$ , is done by :  $m_{1,2}^{DS}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Omega$

$$m_{1,2}^{DS}(X) \triangleq \frac{1}{1 - m_{1,2}(\emptyset)} \sum_{\substack{X_1, X_2 \in 2^\Omega \\ X_1 \cap X_2 = X}} \prod_{i=1}^2 m_i(X_i) \quad (2)$$

where the numerator of (2) is the mass of belief on the conjunctive consensus on  $X$ . The denominator  $1 - m_{1,2}(\emptyset)$  is a normalization constant. The total degree of conflict between the two sources of evidences is classically defined by

$$m_{1,2}(\emptyset) \triangleq \sum_{\substack{X_1, X_2 \in 2^\Omega \\ X_1 \cap X_2 = \emptyset}} \prod_{i=1}^2 m_i(X_i) \quad (3)$$

According to Shafer [3], the two sources are said in total conflict if  $m_{1,2}(\emptyset) = 1$ . In this case the combination of the sources by DS rule cannot be done because of the mathematical 0/0 indeterminacy. The vacuous BBA  $m_v(\Omega) = 1$  is a neutral element for DS rule. This rule is commutative and associative, and the formula (2) can be easily generalized for the combination of  $s > 2$  sources of evidences. DS rule remains the milestone fusion rule of DST.

The doubts of the validity of DS rule has been discussed by Zadeh in 1979 [28]–[30] based on a very simple example with two highly conflicting sources of evidences. Since 1980's, many criticisms have been done about the behavior and the justification of such DS rule. More recently, Dezert et al. in [19], [20] have put in light other problematic behaviors of DS rule even in low conflicting cases and showed serious flaws in logical foundations of DST [21]. To overcome the limitations and problems of DS rule of combination, a new family of PCR rules have been developed in DSmT framework. We present the most elaborate one, i.e. the PCR6 fusion rule, which has been used in our perception application for grid occupancy estimation.

In PCR rules, instead of following the DS normalization (the division by  $1 - m_{1,2}(\emptyset)$ ), we transfer the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is entirely preserved. The general principle of PCR consists: 1) to apply the conjunctive rule, 2) to calculate the total or partial conflicting masses; 3) then redistribute the (total or partial) conflicting mass proportionally on non-empty sets according to the integrity constraints one has for the frame  $\Omega$ . Because the proportional transfer can be done in different ways, there exist several versions of PCR rules of combination. PCR6 fusion rule has been proposed by Martin and Osswald in [6] Vol. 2, Chap. 2, as a serious alternative to PCR5 fusion rule proposed originally by Smarandache and Dezert in [6] Vol. 2, Chap. 1. Martin and Osswald had proposed PCR6 based on intuitive considerations and they had shown through different simulations that PCR6 was more stable than PCR5 in term of decision for combining  $s > 2$  sources of evidence. When only two sources are combined, PCR6 and PCR5 fusion rules coincide, but they differ as soon as more than two sources

<sup>1</sup>DS acronym standing for *Dempster-Shafer* since Dempster's rule has been widely promoted by Shafer in the development of his mathematical theory of evidence [3].

<sup>2</sup>which corresponds to a Dedekind's lattice, see [6] Vol. 1.

have to be combined altogether. Recently, it has been proved in [22] that only PCR6 rule is consistent with the averaging fusion rule which allows to estimate the empirical (frequentist) probabilities involved in a discrete random experiment. For Shafer's model of FoD<sup>3</sup>, PCR6 fusion of two BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$  is defined by  $m_{1,2}^{PCR6}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Omega$

$$m_{1,2}^{PCR6}(X) = \sum_{\substack{X_1, X_2 \in 2^\Omega \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2) + \sum_{\substack{Y \in 2^\Omega \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right] \quad (4)$$

where all denominators in (4) are different from zero. If a denominator is zero, that fraction is discarded. All propositions/sets are in a canonical form [6]. Very basic Matlab codes of PCR rules can be found in [6], [23] and from the toolboxes repository on the web [27]. Like the averaging fusion rule, the PCR6 fusion rule is commutative but not associative. The vacuous belief assignment is a neutral element for this rule.

The PCR6 rule of combination (as well as DS rule) use only part of the whole information available (i.e. the values of the masses of belief only), and they don't exploit the cardinalities of focal elements entering in the fusion process. Because the cardinalities of focal elements are fully taken into account in the computation of the measure of degree of intersection between sets, we have recently proposed to improve PCR6 rules using this measure in the companion paper [7]. The basic idea is to replace any conjunctive product by its discounted version thanks to the measure of degree of intersection  $D$  when the intersection of focal elements is not empty. The product of partial (or total) conflicting masses are not discounted by the measure of degree of intersection because the degree of intersection between two (or more) conflicting focal elements always equals zero, that is if  $X \cap Y = \emptyset$ , then  $D(X, Y) = 0$ . In [7], we have shown in different examples why Zhang's *degree of intersection* [31], denoted  $D^Z(X_1, \dots, X_s)$ , is more interesting than classical Jaccard's degree.  $D^Z(X_1, \dots, X_s)$  is mathematically defined by

$$D^Z(X_1, \dots, X_s) \triangleq \frac{|X_1 \cap X_2 \cap \dots \cap X_s|}{|X_1| \cdot |X_2| \cdot \dots \cdot |X_s|} \quad (5)$$

where  $|X_1 \cap X_2 \cap \dots \cap X_s|$  is the cardinality of the intersection of the focal elements  $X_1, X_2, \dots, X_s$ , and  $|X_1|, |X_2|, \dots, |X_s|$  their cardinalities. The improved version of PCR6 with Zhang's degree of intersection (called ZPCR6 rule) is easy to get and it corresponds to the following formula<sup>4</sup>

<sup>3</sup>that is when  $G^\Omega = 2^\Omega$ , and assuming all elements exhaustive and exclusive.

<sup>4</sup>The general ZPCR6 formula for  $s > 2$  sources is detailed in [7].

$$m_{1,2}^{ZPCR6}(X) = \frac{1}{K_{1,2}^{ZPCR6}} \left[ \sum_{\substack{X_1, X_2 \in 2^\Omega \\ X_1 \cap X_2 = X}} D^Z(X_1, X_2) m_1(X_1) m_2(X_2) + \sum_{\substack{Y \in 2^\Omega \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right] \right] \quad (6)$$

where  $K_{1,2}^{ZPCR6}$  is a normalization constant such that  $\sum_{X \in 2^\Omega} m_{1,2}^{ZPCR6}(X) = 1$ . As for PCR6, one has  $m_{1,2}^{ZPCR6}(\emptyset) = 0$  and ZPCR6 is commutative but not associative. The advantage of ZPCR6 over PCR6 and DS rules is its ability to respond to the inputs in a more effective way has clearly shown in very interesting examples detailed in [7]. Due to space limitation, these examples will not be presented and discussed here again.

### C. Discounting

A discounting effect can be applied on a mass function  $m(\cdot)$  if a piece of information has its reliability lowered. In this case, a new mass function  $m_\alpha(\cdot)$  (with  $\alpha \in [0, 1]$ ) is computed from  $m(\cdot)$  and a part of the mass of each element of the FoD is transferred to the whole FoD  $\Omega$  which represents the total ignorance.

$$m_\alpha(A) = \begin{cases} (1 - \alpha) \cdot m(A) & \text{if } A \neq \Omega \\ (1 - \alpha) \cdot m(A) + \alpha & \text{if } A = \Omega \end{cases} \quad (7)$$

### III. EVIDENTIAL OCCUPANCY GRID

The basic idea of an Occupancy Grid (OG) is to divide the surrounding environment (the ground plane of 2D world) into a set of cells (denoted  $C^i$ ,  $i \in [0, n]$ ) in order to estimate their occupancy state. In a probabilistic framework, the aim is to estimate the probabilities  $P(O^i | z_{1:t})$  and  $P(F^i | z_{1:t})$  given a set of measures  $z_{1:t}$  from the beginning up to the current time  $t$ .  $O^i$  (resp.  $F^i$ ) denotes the occupied (resp. free) state of the cell  $C^i$ . Finally, a decision rule is applied in order to select the most likely state for each cell.

For Evidential approach, occupancy grid represents the information using a mass function over the frame of discernment (FoD)  $\Omega = \{F, O\}$ . So the mass functions used in grid have the structure

$$m_t = [ m_t(\emptyset) \quad m_t(F) \quad m_t(O) \quad m_t(\Omega) ] \quad (8)$$

The occupancy mass function can be used during the fusion process, then the decision can be taken using pignistic transform [26] to get a probability measure and use the same decision rule. An interesting part of evidential occupancy grid is that the FoD can be more complex, and as the fusion is done cell by cell the fusion scheme will be still valid.

Occupancy grids can be classified into two categories depending on the use of a forward or inverse sensor model. The forward model relies on Bayes inference. Since this approach takes into account the conditional dependency of the cells of

the map, it is well adapted to a sensor that observes a large domain of cells with only one reading measurement (e.g. an ultrasonic sonar). However, it requires heavy processing that can be handled by optimized approximation.

The inverse model approach is well adapted to narrow fields of measure by sensors (e.g. LIDAR). It is composed of two separate steps. First, a snapshot map of the sensor reading is built using an inverse sensor model  $P(O^i|z_t)$ . This model can take into account the conditional dependency between the sensor reading and the occupancy of the seen cells. Then, a fusion process (denoted  $\odot$ ) is done with the previous map  $P(O^i|z_{1:t-1})$  as an independent opinion poll fusion:

$$P(O^i|z_{1:t}) = P(O^i|z_t) \odot P(O^i|z_{1:t-1}) \quad (9)$$

In the probabilistic framework, the usual fusion operation between states  $A$  and  $B$  coming from independent measurement, use independent opinion poll [34] :

$$P(A) \odot P(B) = \frac{P(A)P(B)}{P(A)P(B) + (1 - P(A))(1 - P(B))} \quad (10)$$

Inverse approaches have very efficient implementations (e.g. log-odd) that make them popular in mobile robotics [8], [9], [25]. Maps built using inverse models are usually less accurate, since they just take into account the dependency of the cells observed in one reading, but it is a good approximation with accurate and high resolution sensors observing a limited number of cells at a time. Moreover, when the sensor is multi-echo and multi-layer, the conditional dependency of the seen cells can be modeled in an efficient way.

#### A. Fusion strategy with the inverse model

When dealing with the inverse model approach, an estimate of the pose of the robot has to be available, and a map grid  $G^M$  has to be handled. This grid is defined in a world-referenced frame (so it does not move with the robot) and it is updated when a new sensor reading is available. Because of the likely evolution of the world in a dynamic environment, the OG update has to be completed by a remanence strategy. The fusion architecture is based on a prediction-correction paradigm to fuse one or several sensors observations.

*a) Prediction step:* The prediction step computes the predicted map grid at time  $t$  from the map grid estimated at time  $t - 1$ . Depending on the available information, this step can be very refined as done in [24]. Because we don't have specific information on the velocity of the objects (or cells), the prediction step is done by the classical discounting technique. The confidence in past data is controlled by a remanence factor  $\alpha \in [0; 1]$ . The prediction stage is therefore governed by

$$G_t^M = \text{discount}(\hat{G}_{t-1}^M, \alpha) \quad (11)$$

*b) Correction step:* The correction step consists in the combination of the previously estimated map grid with the grid built from the current measures thanks to the inverse model sensor (see more details in [1], [2]). This one is called ScanGrid  $G_t^S$ . As this information is referenced in the sensor frame, a 2D warping is applied to reshape this grid into the

fusion frame. To perform this operation, the current pose  $q_t$  is estimated using a GPS sensor and the rigid homogeneous transformation matrix  $H_t$  is computed. When GPS becomes unavailable, the CAN (Controller Area Network) bus is used to get the robot odometric data. The motion matrix  $H_t$  and the extrinsic calibration matrix  $C$  are used to compute a remapping function  $f(x, y)$  according to Eq. (12) below

$$f(x, y) = C \cdot H_t \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (12)$$

Finally, the ScanGrid is remapped with  $f$  and fused with the previous map grid according to the general formula

$$G_t^M(i, j) = \hat{G}_t^M(i, j) \odot G_t^S(f(i, j)) \quad (13)$$

where the grid  $G_t^S$  represents the BBA produced by the sensor model. This BBA is created in respect to sensor data (e.g. LIDAR point here) and a sensor model to infer an instant occupancy grid. With the probabilistic approach, it refers to the occupancy probability  $P_t^S(O)$ . With the evidential approach it refers to the occupancy mass function  $m_t^S = [m_t^S(\emptyset) \ m_t^S(F) \ m_t^S(O) \ m_t^S(\Omega)]$ . The grid  $\hat{G}_t^M$  refers to the previous MapGrid  $G_{t-1}^M$  predicted at current time using Eq. (11). In the next parts, for each approach considered, the fusion rule  $\odot$  used in Eq. (13) is different. Bayesian approach uses Eq. (10), DS approach uses Eq. 2, PCR6 approach uses Eq. (4), and ZPCR6 approach Eq. (6).

#### B. Discounting in Occupancy Grids

The main advantage of using discounting is to provide a simple way to model the presence of dynamic object in the scene. This model allows to make a prediction without information on the dynamic at the cell level (or at the object level) which is generally not directly available from sensors, and merely difficult to estimate without greedy time-computing algorithms [24] (especially when the evidential framework is adopted). The main issue with the discounting effect is that it makes impossible to build persistent static map. Indeed, cells not viewed by the sensor will quickly converge to the ignorance state, so this strategy cannot be used to build the map of a building for instance. If we are interested to build static map in presence of moving objects, the discounting function is then not recommended. We will see why in the next part of the paper where in this case Bayesian and DS fusion rules will not be very efficient. To handle this case, it is recommended to use either PCR6 or ZPCR6 rules.

## IV. SIMULATION RESULTS

In this section, we present simulation results of grid occupancy estimation in a realistic scenario based on different rules of combination (Bayesian fusion, Dempster-Shafer rule, PCR6 and ZPCR6 fusion rules).

A. Basic simulation

**Setup:** In order to present the basic behavior of the different combination rules studied, we have realized at first some simple 1D-simulations, where we consider a grid cell crossed by a moving object. In this case, the state of the cell changes from free-state to occupied-state at time  $t_1$  and from occupied-state to free-state at time  $t_2$ . The figures 1–4 show the results of these simulations under different conditions.

On each subfigure, we show on the top raw the real state of the cell (i.e. the ground-truth). The second raw shows the sensor data simulated that correspond to the BBA of the state of the cell. This mass function is built according to the state of the cell, the level of confidence of the sensor and can be

eventually perturbed with additional noises. FA indicates the rate of False Alarms and ND the rate of Non Detections. We will consider different level of confidence for  $m_{SG}(O)$  when the cell is occupied and  $m_{SG}(F)$  when the cell is free. The subfigures at the bottom represent the level of belief of the cell state obtained with Bayesian fusion, Dempster-Shafer (DS) fusion, PCR6 and ZPCR6 fusion rules.

**Effect of discounting:** Fig.1 shows the results of the classical chain using a discounting factor  $\alpha = 0.05$  while Fig. 2 is the same case without discounting ( $\alpha = 0$ ). With discounting, all the fusion rules behave similarly. Without discounting, a lag appears with Bayesian and DS fusion rules. The lag is seriously reduced with PCR6 and ZPCR6.

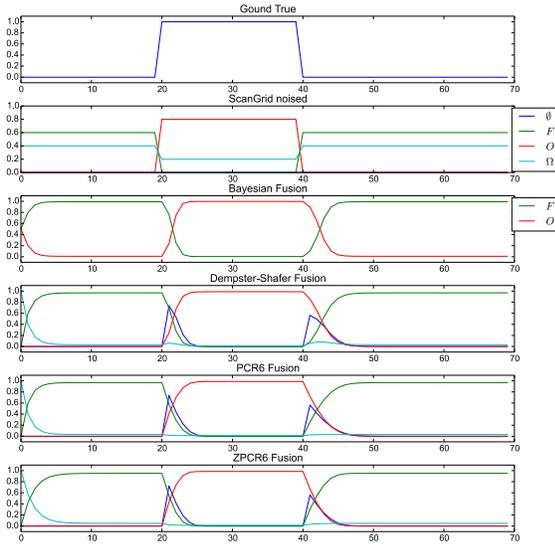


Fig. 1. Case with discounting ( $\alpha = 0.05$ ).

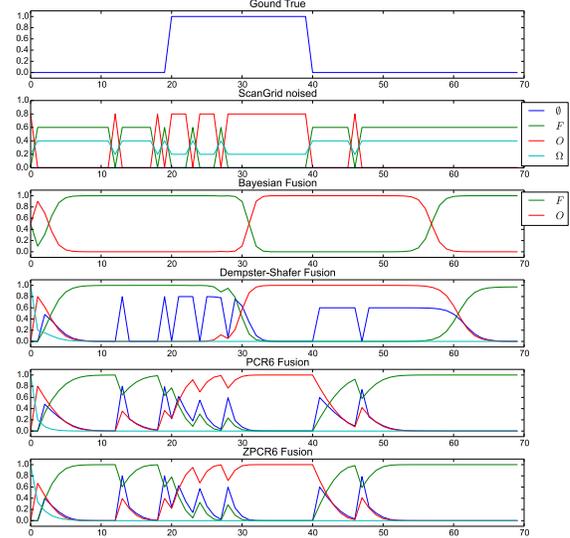


Fig. 3. Case with noise (FA=10%, ND=10%) and  $\alpha = 0$ .

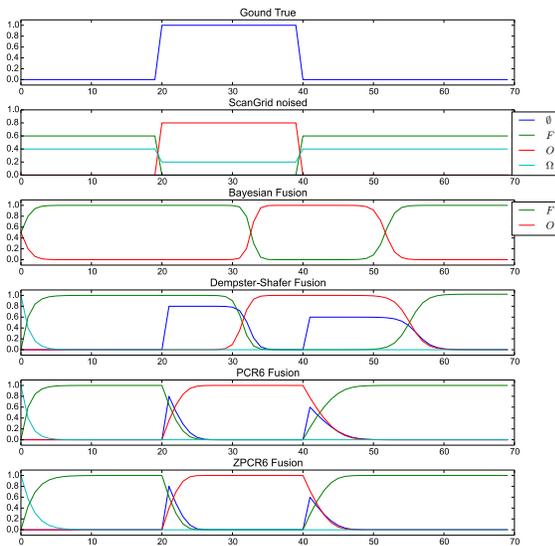


Fig. 2. Case without discounting ( $\alpha = 0$ ).

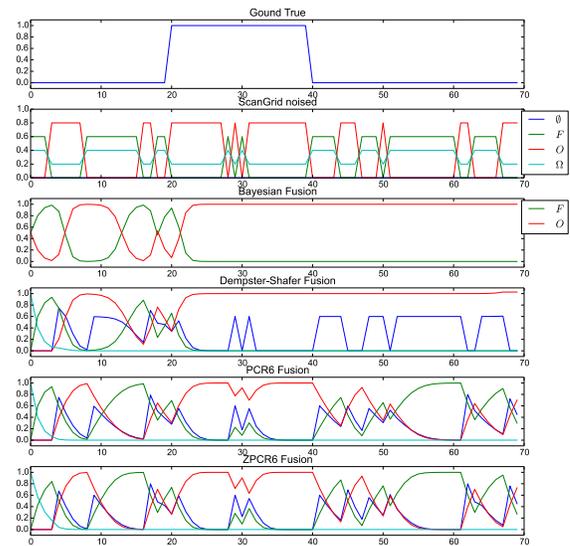


Fig. 4. Case with noise (FA=30%, ND=15%) and  $\alpha = 0$ .

**Performances analyses:** To evaluate the performance of our method, we did perform 10000 Monte Carlo runs for each simulation in order to estimate the false alarm and non detection rates. In order to make the decision, the pignistic probability has been computed and a MAP estimator has been used. Each simulation (from No 0 to No 8) correspond to different conditions (the discounting level, the rate of noise impacting sensor observations) reported in Table I.

Simu N°	discounting $\alpha$	sensor noise $ND/FA$	sensor belief $m_{SG}(O)/m_{SG}(F)$
0	0.05	0	0.8/0.6
1	0	0	0.8/0.6
2	0.05	10/10	0.8/0.6
3	0	10/10	0.8/0.6
4	0.05	15/30	0.8/0.69
5	0	15/30	0.8/0.68
6	0	15/30	0.6/0.4
7	0	25/50	0.6/0.4
8	0	25/50	0.4/0.2

TABLE I  
PARAMETERS OF SIMULATIONS.

For each simulation presented in the Table I, we obtain the performances (rates of FA and ND in %) shown in Table II.

Simu #	Bayesian		DS		PCR6		ZPCR6	
	ND	FA	ND	FA	ND	FA	ND	FA
0	10.0	6.0	10.0	6.0	10.0	6.0	10.0	4.0
1	65.0	24.0	60.0	32.0	10.0	6.0	10.0	4.0
2	11.2	9.2	10.5	10.0	10.0	9.6	11.1	7.8
3	77.7	15.2	73.5	18.9	11.5	6.7	11.3	7.5
4	9.2	28.0	8.2	31.5	8.4	28.9	9.8	25.8
5	33.0	62.7	26.9	65.8	8.4	28.8	10.1	24.7
6	31.3	63.9	26.0	67.3	9.3	38.7	8.8	32.3
7	15.1	76.9	11.5	79.4	5.7	64.0	7.5	55.2
8	7.1	83.9	5.1	85.1	1.9	87.3	3.0	76.2

TABLE II  
RATES OF FALSE ALARM AND NON DETECTION (IN %).

Simulations 0 and 1 illustrated by Figures 1 and 2, correspond to the noise-free situation. By removing the discounting operator, Bayesian and DS approach have a lag in the detection of the change of state that impacts clearly their performances. The PCR6 and ZPCR6 approach are not concerned by this effect because of PCR of conflict. Simulations 2 and 3 (see Fig. 3) include 10% of wrong measurement caused by noises. The fusion rules behave similarly as for simulations 0 and 1, but the performances are a bit lower which shows the effect of noisy measurements in the estimation process. For simulations 4, 5 and 6, the noise reaches 15% for ND and 30% for FA which is important. As we see in Fig. 4, the Bayesian and DS fusion rules are not able to detect the second state change, during the simulation time. This induces the bad false alarm rates. In the last simulations 7 and 8 the noise is very important (about 25% of ND and 50% of FA). In these conditions, all the methods have poor false alarm rates but the PCR6 and ZPCR6 keep good non detection rates. Globally, we see an improvement of the performances when using ZPCR6, specially for the reduction of the FA rates.

## B. LIDAR simulation

In this simulation, the DS and PCR6 fusion rules are compared on a 2D occupancy grid problem close to real application for robot perception. The simulation was realized using the Robot Operating System (ROS) [32] environment and the Gazebo [33] simulator is used here to simulate a Hokuyo LIDAR and a moving object as shown on Figure 5. The simulated sensor has a FoV (Field of View) about  $270^\circ$  and a max range about 10m. The rate of the scan is 20Hz and the ranges of the LIDAR point is corrupted with a Gaussian noise  $\mathcal{N}(0, 0.1)$ .

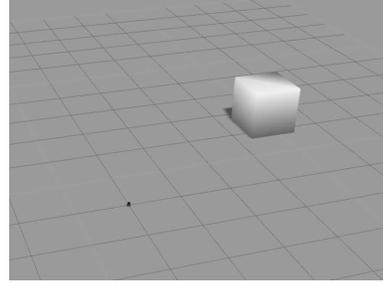


Fig. 5. Gazebo simulation: the box turns around the LIDAR sensor.

Figure 6 shows a simulated scan. The beams that do not hit obstacle within the range are considered as max range (as done in the real Hokuyo sensor). The moving object is a box which has a circular trajectory and moves at 6 rpm around the LIDAR. A *ground true* grid is computed according the real position of the box and its geometry at each scan time. The grid used is a square of 10 m by 10 m with a resolution of 0.1 m and the ScanGrid BBA are set to  $m_{SG}(O) = 0.8$ ,  $m_{SG}(\Omega) = 0.2$  for occupied cells and  $m_{SG}(F) = 0.6$ ,  $m_{SG}(\Omega) = 0.4$  for free cells.

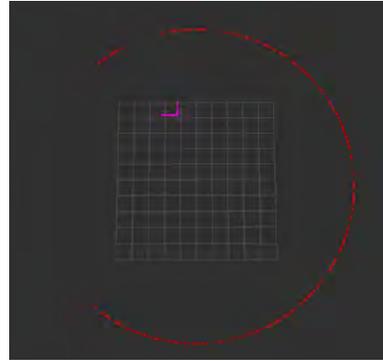


Fig. 6. Bird view of one LIDAR scan.

In order to quantify the results, we compute some metrics. However, because of occlusion, only the cells located on the edges of the box can be considered, that is why we don't consider global metrics. We consider here the two following metrics: 1) the number of correct occupied cell (proportional to recall in our case), and 2) the number of conflicting cells close to the box. The first describes the ability of the method to add

objects into the map, and also by analogy to remove object from the map. The second describes the ability of the method to detect moving objects by generating conflict, this ability is important and is one of the improvement of the evidential grid with respect to the classical Bayesian grid.

Figures 7 shows the result over one turn. The number of cells detected for both metrics depends a lot on the position around the sensor. This can be explained because, in some place, the LIDAR sensor is able to see two edges of the box, in other situations the LIDAR sensor detects just one edge, and when the box is behind (on the back of) the sensor it is out of the field of view of the LIDAR. On figure 7, we can see that the number of occupied cell with the ZPCR6 is more than those with PCR6 and DS fusion rules. Contrarily to the ZPCR6 and PCR6, the DS fusion without discounting can not handle well the quick change of states in the map. The x-axis of Figures 7 is the time stamp of the LIDAR scans, and the y-axis is the number of occupied cells.

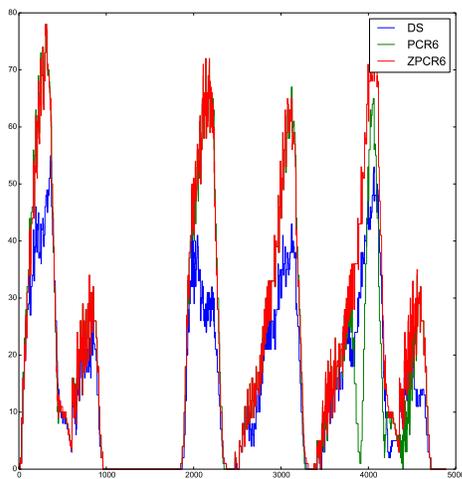


Fig. 7. 2D LIDAR Simulation: Number of correct occupied cells (blue=DS rule, green=PCR6 rule, red=ZPCR6 rule).

### C. Real data processing

A real experimentation was realized using an Hokuyo UTM-30LX sensor. This experimentation takes place in an office in which a person was walking into. The evidential occupancy grid fusion node was implemented within the ROS environment. The grid has the same size and resolution as in the previous example. The BBA used in the sensor model has been set to  $m_{SG}(O) = 0.8$ ,  $m_{SG}(\Omega) = 0.2$  for occupied cells and  $m_{SG}(F) = 0.8$ ,  $m_{SG}(\Omega) = 0.2$  for free cells. No discounting was applied.

Figures 8–10 present the occupancy grid estimation using DS, PCR6 and ZPCR6 rules for a typical snapshot of the sequence. The color of cells denotes the state having the highest mass value: green for  $F$  (free state), red for  $O$  (occupied state), and black for  $\Omega$  (full uncertainty). For convenience, we have also displayed in blue all the cells that carry a conflicting mass  $m(\emptyset) > 0$  before applying the normalization step of DS rule,

or before applying the proportional conflict redistribution with PCR6, or both with ZPCR6.

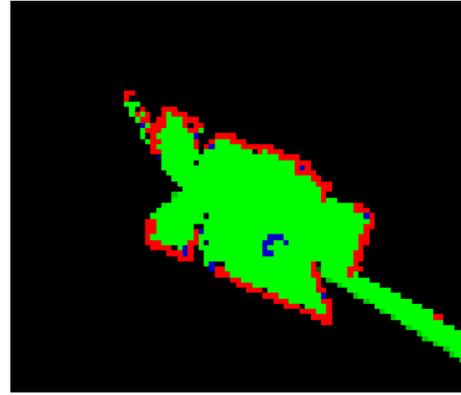


Fig. 8. Snapshot 1 - with DS fusion.

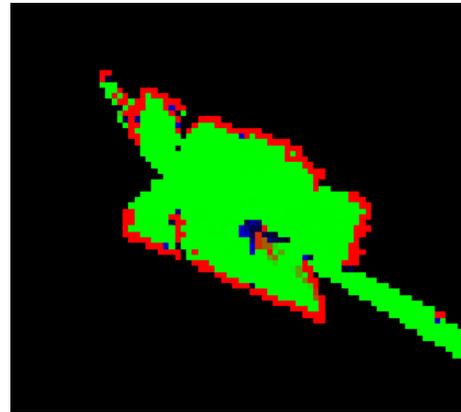


Fig. 9. Snapshot 1 - with PCR6 fusion.

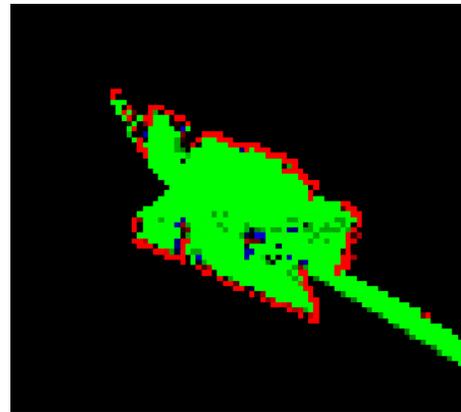


Fig. 10. Snapshot 1 - with ZPCR6 fusion.

Figure 8 shows the result using DS rule. The room scanned by the sensor is correctly mapped and its bounds (mainly walls and doors) are clearly identified by the red pixels. The free space (green pixels) is correctly detected in the room except near the people that is labeled as free (with conflicting cell

shown in blue for convenience). The people moving around the desk in the office room is only detected from conflicting cells when he stops to walk several times. Figure 9 shows the PCR6 result at the same time stamps. In this case, the people is rightly detected as shown by the red pixels (occupied cells) inside the green area (the office room). A conflict cell is created when he starts walking in the room. The static part of the room is also detected (as with DS fusion rule). Figure 10 shows the ZPCR6 result at the same time stamps. The results of this rule are close to the PCR6 rule but the level of ignorance (the mass on  $\Omega$ ) is higher on the cells behind the person. This can be understood because the mass  $m(\Omega)$  is weighed by a factor 0.5 during the transition.

## V. CONCLUSIONS AND PERSPECTIVES

In this work we have presented a novel application of the belief functions which significantly improves the map building process for intelligent vehicles environment perception and grid map estimation. This work shows the importance of defining an accurate sensor model. We have considered the uncertainties of the LIDAR measurements and used the ZPCR6 rule of DSMT to model and combine sensor information. Our new method differs of Bayesian approach by allowing support for more than one proposition at a time, rather than a single hypothesis. It is an interval-based approach, as defined by the lower and upper probability bounds  $[Bel, Pl]$  allowing the lack of measurement to be modeled adequately. This new method based on ZPCR6 rule differs from the classical evidential approach based on DS rule and improves in theory the results based PCR6, and more substantially the results of DS rule. Our experimental results with the LIDAR confirm the improvement of the accuracy of this new grid estimation method w.r.t previous methods, but the improvement obtained with ZPCR6 over PCR6 is not so important because of the too simplistic structure of the chosen frame of discernment. As research perspectives, we will try to implement these fusion rules in 3D occupancy grid (Octomap based) and use a stereo camera with dense disparity map computation as sensor source. Also we would like to deal with refined frames of discernment to ameliorate the precision of the perception and to emphasize the advantages of ZPCR6 rule.

## REFERENCES

- [1] J. Moras, V. Cherfaoui, P. Bonnifait, Credibilist occupancy grids for vehicle perception in dynamic environments, 2011 IEEE Int. Conf on Robotics and Automation, Shanghai, China, pp. 84–89, May 2011.
- [2] J. Moras, V. Cherfaoui, P. Bonnifait, Moving Objects Detection by Conflict Analysis in Evidential Grids, 2011 IEEE Intelligent Vehicles Symposium IV, pp. 1122–1127, June 2011.
- [3] G. Shafer, A Mathematical Theory of Evidence. Princeton: Princeton University Press, 1976.
- [4] J. Moras, V. Cherfaoui, P. Bonnifait, Evidential Grids Information Management in Dynamic Environments, Proc. of Fusion 2014 Int. Conf, Salamanca, Spain, July 2014.
- [5] J. Moras, J. Dezert, B. Pannetier, Grid occupancy estimation for environment perception based on belief functions and PCR6, Proc. of SPIE (Signal Processing, Sensor/Information Fusion, and Target Recognition XXIV), Baltimore, MD, USA, 20–22 April 2015.
- [6] F. Smarandache, J. Dezert, Advances and applications of DSMT for information fusion (Collected works), American Research Press, USA Vol.1–4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [7] F. Smarandache, J. Dezert, Modified PCR Rules of Combination with Degrees of Intersections, (Submitted to Fusion 2015), Washington D.C., USA, July 2015.
- [8] A. Elfes, Using occupancy grids for mobile robot perception and navigation, Computer, Vol. 22, no. 6, pp. 46–57, 1989.
- [9] H.P. Moravec, Sensor fusion in certainty grids for mobile robots, AI Magazine, Summer: 116–121, 1988.
- [10] G. Oriolo, G. Ulivi, M. Vendittelli, On-line map building and navigation for autonomous mobile robots, Proc. 1995 IEEE Int. Conf. on Robotics and Automation, pp. 2900–2906, Nagoya, Japan, 1995.
- [11] K. Hughes, R. Murphy, Ultrasonic robot localization using Dempster-Shafer theory, in Proc. of SPIE on Neural and Stochastic Methods in Image and Signal Processing, 1992, pp. 2–11, 1992.
- [12] P. Tirumalai, B.G. Schunk, R.C. Jain, Evidential reasoning for building environment maps, IEEE Trans. on SMC, Vol. 25(1), pp. 10–20, 1995.
- [13] F. Gambino, G. Oriolo, G. Ulivi, A comparison of three uncertainty calculus techniques for ultrasonic map building, in: Proc. 1996 SPIE Int. Symp. on Aerospace/Defense Sensing and Control, pp. 249–260, Orlando, FL, USA, 1996.
- [14] D. Pagac, E. Nebot, H. Durrant-Whyte, An evidential approach to probabilistic map-building, Proc. of IEEE Int. Conf. on Robotics and Automation, Vol. 1, pp. 745–750, Minneapolis, MN, USA, April 1996.
- [15] T. Reineking, J. Clemens, Evidential FastSLAM for grid mapping, Proc. of Fusion 2013 Conf., pp. 789–796, Istanbul, Turkey, July 2013.
- [16] X. Li, X. Huang, M. Wang, J. Xu, H. Zhang, DSMT Coupling with PCR5 for Mobile Robots Map Reconstruction, Proc. of 2006 Int. Conf. on mechatronics and Automation, Luoyang, China, June 26–28, 2006.
- [17] P. Li, X. Huang, S. Yang, J. Dezert, SLAM and path planning of mobile robot using DSMT, Journal of Software Engineering, Vol. 7, No. 2, pp. 46–67, 2013.
- [18] J. Zhou, J. Duan, G. Yang, Occupancy Grid Mapping Based on DSMT for Dynamic Environment Perception, International Journal of Robotics and Automation (IJRA), Vol. 2, No. 4, pp. 129–139, December 2013.
- [19] J. Dezert, P. Wang, A. Tchamova, On the validity of Dempster-Shafer theory, Proc. of Fusion 2012 Int. Conf., Singapore, July 9–12, 2012.
- [20] A. Tchamova, J. Dezert, On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory, 6th IEEE Int. Conf. on Int. Syst. (IS '12), Sofia, Bulgaria, Sept. 6–8, 2012.
- [21] J. Dezert, A. Tchamova, On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule, International Journal of Intelligent Systems, Vol. 29(3), pp. 223–252, March 2014.
- [22] F. Smarandache, J. Dezert, On the consistency of PCR6 with the averaging rule and its application to probability estimation, Proc. of Fusion 2013, Istanbul, Turkey, July 2013.
- [23] F. Smarandache, J. Dezert, J.-M. Tacnet, Fusion of sources of evidence with different importances and reliabilities, Proc. of Fusion 2010 Int. Conf., Edinburgh, UK, July 26–29, 2010.
- [24] M.K. Tay, et al., The Bayesian occupation filter, in Probabilistic Reasoning and Decision Making in Sensory-Motor Systems, pp. 77–98, Springer, 2008.
- [25] T. Weiss, B. Schiele, K. Dietmayer, Robust Driving Path Detection in Urban and Highway Scenarios Using a Laser Scanner and Online Occupancy Grids, 2007 IEEE Intelligent Vehicles Symposium, pp. 184–189, June 2007.
- [26] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, IJAR, Vol. 38, pp. 133–147, Feb. 2005.
- [27] <http://bfasp.iutlan.univ-rennes1.fr/wiki/index.php/Toolboxes>
- [28] L.A. Zadeh, On the validity of Dempster's rule of combination, Memo M79/24, Univ. of California, Berkeley, CA, U.S.A., 1979.
- [29] L.A. Zadeh, Book review: A mathematical theory of evidence, The AI Magazine, Vol. 5, No. 3, pp. 81–83, 1984.
- [30] L.A. Zadeh, A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination, The AI Magazine, Vol. 7 (2), pp. 85–90, 1986.
- [31] L. Zhang, Representation, independence and combination of evidence in Dempster-Shafer theory, in Advances in the Dempster-Shafer theory of evidence, John Wiley & Sons, New York, pp. 51–95, 1994.
- [32] M. Quigley et al., ROS: an open-source Robot Operating System, in 2009 ICRA Workshop on Open Source Software, 2009.
- [33] N. Koenig, A. Howard, Design and Use Paradigms for Gazebo, An Open-Source Multi-Robot Simulator, in 2004 IEEE/RSJ Int. Conf. on Intell. Robots and Syst., pp. 2149–2154, Sept. 2004.
- [34] R.A. Jacobs, Methods for combining expert's probability assessments, Neural Computation, Vol. 7 (5), pp. 867–888, 1995.

# Grid Occupancy Estimation for Environment Perception Based on Belief Functions and PCR6

Julien Moras, Jean Dezert, Benjamin Pannetier

The French Aerospace Lab – ONERA/DTIM/EVF, F-91761 Palaiseau, France.

julien.moras@onera.fr, jean.dezert@onera.fr, benjamin.pannetier@onera.fr

Originally published as: J. Moras, J. Dezert, B. Pannetier, *Grid Occupancy Estimation for Autonomous Vehicle Perception*, SPIE defense & Security Conference, Baltimore, MD, USA, April 20–24, 2015, and reprinted with permission.

**Abstract**—In this contribution, we propose to improve the grid map occupancy estimation method developed so far based on belief function modeling and the classical Dempster’s rule of combination. Grid map offers a useful representation of the perceived world for mobile robotics navigation. It will play a major role for the security (obstacle avoidance) of next generations of terrestrial vehicles, as well as for future autonomous navigation systems. In a grid map, the occupancy of each cell representing a small piece of the surrounding area of the robot must be estimated at first from sensors measurements (typically LIDAR, or camera), and then it must also be classified into different classes in order to get a complete and precise perception of the dynamic environment where the robot moves. So far, the estimation and the grid map updating have been done using fusion techniques based on the probabilistic framework, or on the classical belief function framework thanks to an inverse model of the sensors. Mainly because the latter offers an interesting management of uncertainties when the quality of available information is low, and when the sources of information appear as conflicting. To improve the performances of the grid map estimation, we propose in this paper to replace Dempster’s rule of combination by the PCR6 rule (Proportional Conflict Redistribution rule #6) proposed in DSMT (Dezert-Smarandache) Theory. As an illustrating scenario, we consider a platform moving in dynamic area and we compare our new realistic simulation results (based on a LIDAR sensor) with those obtained by the probabilistic and the classical belief-based approaches.

**Keywords:** Grid map, cell occupancy, perception, belief functions, DSMT, PCR6, robotics.

## I. INTRODUCTION

Occupancy Grids (OG) are often used for robot environment perception and navigation, which requires techniques for data fusion [1], localization [2] and obstacle avoidance [3]. As OGs manage a representation of the environment that does not make any assumption on the geometrical shape of the detected elements, they provide a general framework to deal with complex perception conditions. In our previous works, we did focus on the use of a multi-echo and multi-layer LIDAR system in order to characterize the dynamic surrounding environment of a robot navigating in an unrestricted area. The perception strategy involved map estimation and scan grids [4], [5] based either on the classical Bayesian framework, or on classical evidential framework based on Dempster-Shafer theory (DST) [6] of belief functions. The map grid acts as a filter that accumulate information and allows to detect moving objects.

A comparative analysis of performances of these approaches has already been published recently in [7].

In dynamic environments, it is crucial to have a good modeling of the information flow in the data fusion process in order to avoid adding wrong implicit prior knowledge that will need time to be forgotten. In this context, evidential OG are particularly interesting to make a good management of the information since it is possible to explicitly make the distinction between non explored and moving cells. In this paper, we explore the use of Dezert-Smarandache Theory [8] (DSMT) as an alternative approach of the classical DST to provide better accurate estimation of the grid map occupancy for robot perception.

The idea of using the probabilistic framework to estimate the grid occupancy has been popularized by Elfes in his pioneered works in 1990’s [9]–[13]. Later, the idea has been extended with the fuzzy logic theory framework by Oriolo et al. [15]–[21], and in parallel with the belief function (evidential) framework as well [22]–[29]. Most of the aforementioned research works dealt only with acoustic sensors only (i.e SONAR). Recently, DSMT has also been applied for the perception of the environment with acoustic sensors as reported in [30]–[32], [34]–[36].

Our main contribution is to propose a new method to make the perception with non acoustic sensors, and to compare the performances of the Proportional Conflict Redistribution rule no. 6 (PCR6) of DSMT with respect to Dempster-Shafer’s (DS) rule of combination in terms of accuracy of grid map estimation. In our application, we work with a LIDAR sensor on board on a robot moving in dynamic environment.

This paper is organized as follows. After a short presentation of the basics of belief functions and rules of their combination based on DST and DSMT in the next section, we will present the inverse sensor models in section III with the construction of the basic belief assignments (BBA). In section IV, we present an illustrating scenario for environment perception including a mobile object with a platform equipped with a LIDAR, and we compare our new realistic simulation results with those obtained by the probabilistic and the classical belief-based approaches. We will show how static and mobile objects are extracted from the occupancy grid map using digital image processing. Finally, conclusion and outline perspectives are given in section V.

## II. EVIDENTIAL FRAMEWORK

Dempster-Shafer's theory (DST) of evidence has been developed by Shafer in 1976 from Dempster's works [6]. DST is known also as the theory of belief functions and it is mainly characterized by a frame of discernment (FoD), sources of evidence represented by basic belief assignment (BBA), belief (Bel) and plausibility (Pl) functions, and the Dempster's rule, denoted as DS rule of combination in the sequel<sup>1</sup>. DST has been modified and extended into Dezert-Smarandache theory [8] (DSMT) to work with quantitative or qualitative BBA and to combine the sources of evidence in a more efficient way thanks to new proportional conflict redistribution (PCR) fusion rules – see [37]–[40] for discussion and examples. We briefly recall in the next subsections the basics of the theory of belief functions.

### A. Belief functions

Let's consider a finite discrete FoD  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ , with  $n > 1$ , of the fusion problem under consideration and its fusion space  $G^\Omega$  which can be chosen either as the power-set  $2^\Omega$ , the hyper-power set<sup>2</sup>  $D^\Omega$ , or the super-power set  $S^\Omega$  depending on the model that fits with the problem [8]. A BBA associated with a given source of evidence is defined as the mapping  $m(\cdot) : G^\Omega \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in G^\Omega} m(A) = 1$ . The quantity  $m(A)$  is called mass of belief of  $A$  committed by the source of evidence. Belief and plausibility functions are defined by

$$\text{Bel}(A) = \sum_{\substack{B \subseteq A \\ B \in G^\Omega}} m(B) \quad \text{and} \quad \text{Pl}(A) = \sum_{\substack{B \cap A \neq \emptyset \\ B \in G^\Omega}} m(B) \quad (1)$$

The degree of belief  $\text{Bel}(A)$  given to a subset  $A$  quantifies the amount of justified specific support to be given to  $A$ , and the degree of plausibility  $\text{Pl}(A)$  quantifies the maximum amount of potential specific support that could be given to  $A$ . If for some  $A \in G^\Omega$ ,  $m(A) > 0$  then  $A$  is called a focal element of the BBA  $m(\cdot)$ . When all focal elements are singletons and  $G^\Omega = 2^\Omega$  then the BBA  $m(\cdot)$  is called a Bayesian BBA [6] and its corresponding belief function  $\text{Bel}(\cdot)$  is homogeneous to a (possibly subjective) probability measure, and one has  $\text{Bel}(A) = P(A) = \text{Pl}(A)$ , otherwise in general one has  $\text{Bel}(A) \leq P(A) \leq \text{Pl}(A)$ ,  $\forall A \in G^\Omega$ . The vacuous BBA representing a totally ignorant source is defined as  $m_v(\Omega) = 1$ .

### B. Fusion rules

Many mathematical rules have been proposed in the literature over the decades (see [8], Vol. 2 for a detailed list of fusion rules) to combine efficiently several distinct sources of evidence represented by the BBA's  $m_1(\cdot)$ ,  $m_2(\cdot)$ ,  $\dots$ ,  $m_s(\cdot)$  ( $s \geq 2$ ) defined on same fusion space  $G^\Omega$ . In this paper, we focus only on DS rule because it has been historically proposed

in DST and it is still widely used in applications, and on the PCR rule no. 6 (i.e. PCR6) proposed in DSMT because it provides a very interesting alternative of DS rule, even if PCR6 is more complex to implement in general than DS rule.

In DST framework, the fusion space  $G^\Omega$  equals the power-set  $2^\Omega$  because Shafer's model of the frame  $\Omega$  is assumed, which means that all elements of the FoD are exhaustive and exclusive. The combination of the BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$ , is done by :  $m_{1,2}^{DS}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Omega$

$$m_{1,2}^{DS}(X) \triangleq \frac{1}{1 - m_{1,2}(\emptyset)} \sum_{\substack{X_1, X_2 \in 2^\Omega \\ X_1 \cap X_2 = X}} \prod_{i=1}^2 m_i(X_i) \quad (2)$$

where the numerator of (2) is the mass of belief on the conjunctive consensus on  $X$ . The denominator  $1 - m_{1,2}(\emptyset)$  is a normalization constant, where the total degree of conflict denoted  $m_{1,2}(\emptyset)$  between the two sources of evidences is defined by

$$m_{1,2}(\emptyset) \triangleq \sum_{\substack{X_1, X_2 \in 2^\Omega \\ X_1 \cap X_2 = \emptyset}} \prod_{i=1}^2 m_i(X_i) \quad (3)$$

According to Shafer [6], the two sources are said in total conflict if  $m_{1,2}(\emptyset) = 1$ . In this case the combination of the sources by DS rule cannot be done because of the mathematical 0/0 indeterminacy. The vacuous BBA  $m_v(\Omega) = 1$  is a neutral element for DS rule. This rule is commutative and associative, and the formula (2) can be easily generalized for the combination of  $s > 2$  sources of evidences. DS rule remains the milestone fusion rule of DST.

The doubts of the validity of DS rule has been discussed by Zadeh in 1979 [47]–[49] based on a very simple example with two highly conflicting sources of evidences. Since 1980's, many criticisms have been done about the behavior and the justification of such DS rule. More recently, Dezert et al. in [37], [38] have put in light other counter-intuitive behaviors of DS rule even in low conflicting cases and showed serious flaws in logical foundations of DST [39]. To overcome the limitations and problems of DS rule of combination, a new family of PCR rules have been developed in DSMT framework. We present the most elaborate one, i.e. the PCR6 fusion rule, which has been used in our perception application for grid occupancy estimation.

In PCR rules, instead of following the DS normalization (the division by  $1 - m_{1,2}(\emptyset)$ ), we transfer the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is entirely preserved. The general principle of PCR consists: 1) to apply the conjunctive rule, 2) to calculate the total or partial conflicting masses; 3) then redistribute the (total or partial) conflicting mass proportionally on non-empty sets according to the integrity constraints one has for the frame  $\Omega$ . Because the proportional transfer can be done in different ways, there exist several versions of PCR rules of combination. PCR6 fusion rule has been proposed by Martin and Osswald in

<sup>1</sup>DS acronym standing for *Dempster-Shafer* since Dempster's rule has been widely promoted by Shafer in the development of his mathematical theory of evidence [6].

<sup>2</sup>which corresponds to a Dedekind's lattice, see [8] Vol. 1.

[8] Vol. 2, Chap. 2, as a serious alternative to PCR5 fusion rule proposed originally by Smarandache and Dezert in [8] Vol. 2, Chap. 1. Martin and Osswald had proposed PCR6 based on intuitive considerations and they had shown through different simulations that PCR6 was more stable than PCR5 in term of decision for combining  $s > 2$  sources of evidence. When only two sources are combined, PCR6 and PCR5 fusion rules coincide, but they differ as soon as more than two sources have to be combined altogether. Recently, it has been proved in [40] that only PCR6 rule is consistent with the averaging fusion rule which allows to estimate the empirical (frequentist) probabilities involved in a discrete random experiment.

For Shafer's model of FoD<sup>3</sup>, the PCR6 combination of two BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$  is defined by  $m_{1,2}^{PCR6}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Omega$

$$m_{1,2}^{PCR6}(X) = \sum_{\substack{X_1, X_2 \in 2^\Omega \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2) + \sum_{\substack{Y \in 2^\Omega \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right] \quad (4)$$

where all denominators in (4) are different from zero. If a denominator is zero, that fraction is discarded. All propositions/sets are in a canonical form [8]. Very basic Matlab codes of PCR rules can be found in [8], [41] and from the toolboxes repository on the web [46]. Like the averaging fusion rule, the PCR6 fusion rule is commutative but not associative. The vacuous belief assignment is a neutral element for this rule.

### C. Discounting

A discounting effect can be applied on a mass function  $m(\cdot)$  if a piece of information has its reliability lowered. In this case, a new mass function  $m_\alpha(\cdot)$ , (with  $\alpha \in [0, 1]$ ) is computed from  $m(\cdot)$  and a part of the mass of each element of the FoD is transferred to the whole FoD  $\Omega$  which represents the total ignorance.

$$m_\alpha(A) = \begin{cases} (1 - \alpha) \cdot m(A) & \text{if } A \neq \Omega \\ (1 - \alpha) \cdot m(A) + \alpha & \text{if } A = \Omega \end{cases} \quad (5)$$

### D. Pignistic transformation

Finally, the pignistic transformation *BetP* [45] allows to compute a probability measure from a mass function by distributing proportionally the mass of the subsets on their focal elements:

$$\forall A \in \Omega, \text{BetP}(A) \triangleq \sum_{B \in 2^\Omega} \frac{|A \cap B|}{|B|} \cdot m(B) \quad (6)$$

where  $|A \cap B|$  is the cardinal of the subset  $A \cap B$ , and  $|B|$  is the cardinal of subset the  $B$ .

However, this transformation is not bijective (a part of the information is lost). So, one can find an infinity of mass functions with the same pignistic probability. This issue is inherent in the nature of probabilities which are not able to distinguish randomness from (epistemic) uncertainty.

<sup>3</sup>i.e. when  $G^\Omega = 2^\Omega$ , and assuming all elements exhaustive and exclusive.

## III. EVIDENTIAL OCCUPANCY GRID

The basic idea of an Occupancy Grid (OG) is to divide the surrounding environment (the ground plane of 2D world) into a set of cells (denoted  $C^i$ ,  $i \in [0, n]$ ) in order to estimate their occupancy state. In a probabilistic framework, the aim is to estimate the probabilities  $P(O^i|z_{1:t})$  and  $P(F^i|z_{1:t})$  given a set of measures  $z_{1:t}$  from the beginning up to the current time  $t$ .  $O^i$  (resp.  $F^i$ ) denotes the occupied (resp. free) state of the cell  $C^i$ . Finally, a decision rule is applied in order to select the most likely state for each cell.

For Evidential approach, occupancy grid represents the information using a mass function over the frame of discernment (FoD)  $\Omega = \{F, O\}$ . So the mass functions used in grid have the structure

$$m_t = [ m_t(\emptyset) \quad m_t(F) \quad m_t(O) \quad m_t(\Omega) ] \quad (7)$$

The occupancy mass function can be used during the fusion process, then the decision can be taken using pignistic transform to get a probability measure and use the same decision rule. An interesting part of evidential occupancy grid is that the FoD can be more complex, and as the fusion is done cell by cell the fusion scheme will be still valid .

Occupancy grids can be classified into two categories depending on the use of a forward, or inverse, sensor model. The forward model relies on Bayes inference. Since this approach takes into account the conditional dependency of the cells of the map, it is well adapted to a sensor that observes a large domain of cells with only one reading measurement (e.g. a ultrasonic SONAR). However, it requires heavy processing that can be handled by optimized approximation [42] or GPU computing [43].

The inverse model approach is well adapted to narrow fields of measures sensors (e.g. LIDAR). It is composed of two separate steps. First, a snapshot map of the sensor reading is built using an inverse sensor model  $P(O^i|z_t)$ . This model can take into account the conditional dependency between the sensor reading and the occupancy of the seen cells. Then, a fusion process (denoted  $\odot$ ) is done with the previous map  $P(O^i|z_{1:t-1})$  as an independent opinion poll fusion:

$$P(O^i|z_{1:t}) = P(O^i|z_t) \odot P(O^i|z_{1:t-1}) \quad (8)$$

In the probabilistic framework, the usual fusion operation between states  $A$  and  $B$  coming from independent measurement, use independent opinion poll [52] :

$$P(A) \odot P(B) = \frac{P(A) \cdot P(B)}{P(A) \cdot P(B) + (1 - P(A)) \cdot (1 - P(B))} \quad (9)$$

Inverse approaches have very efficient implementations (e.g. log-odd) that make them popular in mobile robotics [10], [14], [44]. Maps built using inverse models are usually less accurate, since they just take into account the dependency of the cells observed in one reading, but it is a good approximation with accurate and high resolution sensors observing a limited number of cells at a time. Moreover, when the sensor is multi-echo and multi-layer, the conditional dependency of the seen cells can be modeled in an efficient way.

### A. Fusion strategy with the inverse model

When dealing with the inverse model approach, an estimate of the pose of the robot has to be available and map grid  $G^M$  has to be handled. This grid is defined in a world-referenced frame (so it does not move with the robot) and is updated when a new sensor reading is available. Because of the likely evolution of the world in a dynamic environment, the OG update has to be completed by a remanence strategy. The fusion architecture follows then a prediction-correction paradigm and can be used to fuse one or several sensors observations.

*a) Prediction step:* The prediction step computes the predicted map grid at time  $t$  from the map grid estimated at time  $t - 1$ . Depending on the available information, this step can be very refined like done in [43]. As we consider here that no specific information on the velocity of the objects (or cells) is available, the prediction is done by discounting. The confidence in aged data is controlled by a remanence factor  $\alpha \in [0; 1]$ . The prediction stage is therefore governed by

$$G_t^M = \text{discount} \left( \hat{G}_{t-1}^M, \alpha \right) \quad (10)$$

*b) Correction step:* The correction step consists in the combination of the previously estimated map grid with the grid built from the current measures thanks to the inverse model sensor (see more details in [4], [5]). This one is called ScanGrid  $G_t^S$ . As this information is referenced in the sensor frame, a 2D warping is applied to reshape this grid into the fusion frame. To perform this operation, the current pose  $q_t$  is estimated using a GPS sensor and the rigid homogeneous transformation matrix  $H_t$  is computed. When GPS becomes unavailable, the CAN (Controller Area Network) bus is used to get the robot odometric data. The motion matrix  $H_t$  and the extrinsic calibration matrix  $C$  are used to compute a remapping function  $f(x,y)$  according to Eq.(11) below

$$f(x, y) = C \cdot H_t \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (11)$$

Finally, the ScanGrid is remapped with  $f$  and fused with the previous map grid.

$$G_t^M(i, j) = \hat{G}_t^M(i, j) \odot G_t^S(f(i, j)) \quad (12)$$

The grid  $G_t^S$  represents the BBA produced by the sensor model. This BBA is created in respect to sensor data (e.g. LIDAR point here) and a sensor model to infer an instant occupancy grid. For probabilistic approach, it refers to the occupancy probability  $P_t^S(O)$ , for evidential approach it refers for a occupancy mass function  $m_t^S = [m_t^S(\emptyset) \ m_t^S(F) \ m_t^S(O) \ m_t^S(\Omega)]$ . The grid  $\hat{G}_t^M$  refers to the previous MapGrid  $G_{t-1}^M$  predicted at current time using Eq.(10). In the following parts, for each approach considered, the rule  $\odot$  used in Eq.(12) is different. Bayesian approach uses Eq.(9), DS approach uses Eq.(2) and PCR6 approach uses Eq.(4).

### B. Discounting in Occupancy Grids

The main advantage of using discounting is to provide a simple way to model the presence of dynamic object in the scene. This model allows to make a prediction without information on the dynamic at the cell level (or at the object level) which is generally not directly available from sensors and merely difficult to estimate without greedy time-computing algorithms [43] (especially when the evidential framework is adopted). The main issue with the discounting effect is that it makes impossible to build persistent static map. Indeed, cells not viewed by the sensor will quickly converge to the ignorance state. Therefore, this strategy cannot be used to build the map of a building for instance. If we are interested to build static map in presence of moving objects, the discounting function is then not recommended. We will see why in the next part of the paper where in this case Bayesian and DS fusion rules will not be very efficient. To handle this case, we will show why it is recommended to use the PCR6 rule.

## IV. RESULTS

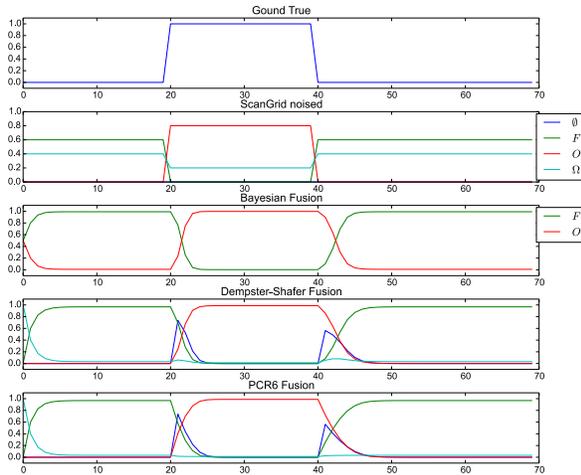
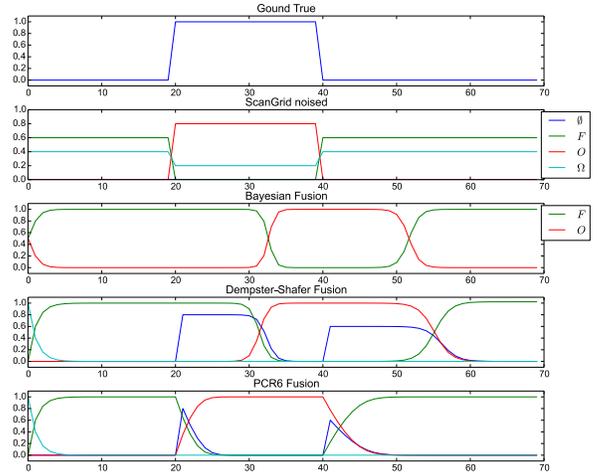
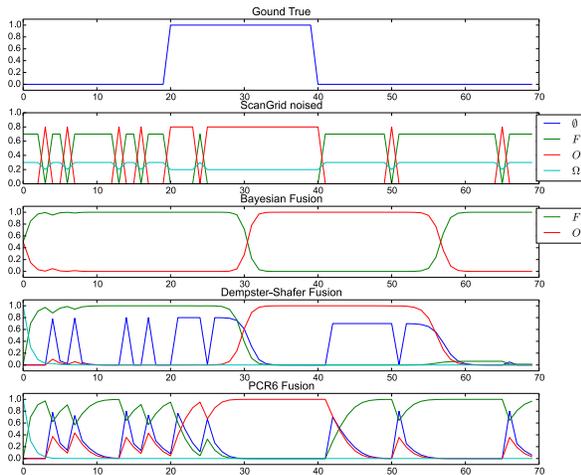
In this section, we present simulation results of grid occupancy estimation in a realistic scenario based on different rules of combination (Bayesian fusion, Dempster-Shafer rule, and PCR6 fusion rule).

### A. Basic simulation

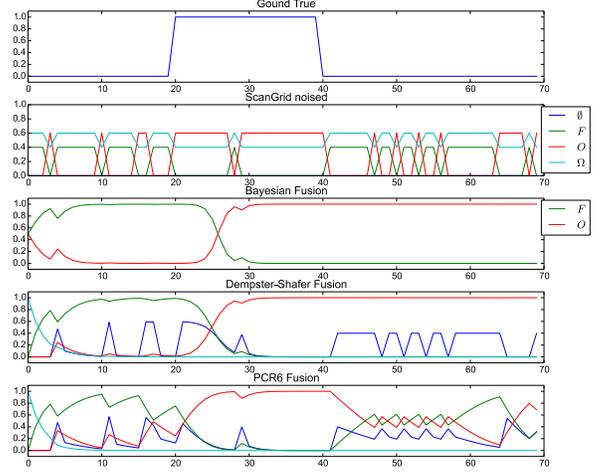
*a) Setup:* In order to present the basic behavior of the different combination rules studied, we have realized at first some simple 1D-simulations, where we consider a grid cell crossed by a moving object. In this case, the state of the cell changes from free-state to occupied-state at time  $t_1$  and from occupied-state to free-state at time  $t_2$ . Figure 1 shows the results of these simulations under different conditions. On each subfigure, we show on the top plot the real state of the cell (i.e. the groundtruth). The second row of each subplot shows the sensor data simulated that corresponds to the BBA of the state of the cell. This mass function is built according to the state of the cell, the level of confidence of the sensor and can be eventually perturbed with additional noises. FA indicates the rate of False Alarms, and ND the rate of Non Detections. We will consider different levels of confidence for  $m_{SG}(O)$  when the cell is occupied, and  $m_{SG}(F)$  when the cell is free. The bottom plot of each subplot represents the level of belief of the cell state obtained with Bayesian fusion, Dempster-Shafer (DS) fusion and the PCR6 fusion rules respectively.

*Effect of discounting:* Figure 1a presents the results of the classical chain using a discounting factor  $\alpha = 0.05$  while figure 1b presents the same case without discounting ( $\alpha = 0$ ). If the discounting is applied, all the fusion rules behave similarly, but if the discounting is not used, a lag appears with Bayesian and DS fusion rules. The lag effect is seriously reduced with PCR6 rule.

*Performances analyses:* The performance of our method is summarized in Table I. For each simulation, 10000 Monte Carlo runs have been performed, in order to estimate the false alarm and non detection rates. In order to make the


 (a) Case with discounting ( $\alpha = 0.05$ ).

 (b) Case without discounting ( $\alpha = 0$ ).


(c) Case with noise (FA=10%, ND=10%) and no discounting.



(d) Case with noise (FA=30%, ND=15%) and no discounting.

Figure 1: Evolution of the belief in a cell crossed by an obstacle observed by a sensor.

decision, the pignistic probability has been computed and a MAP estimator has been used. For each simulation presented in the Table I, we also mention the discounting level, the rate of noise impacting sensor observations.

Simulations 0 and 1, illustrated by the figure 1a and the figure 1b, correspond to the noise-free situation. By removing the discounting operator, Bayesian and DS approach have a lag in the detection of the change of state that impacts clearly their performances. The PCR6 approach is much less concerned by this effect because of the proportional conflict redistribution process. Simulations 2 and 3 (see figure 1c) include 10% of wrong measurement caused by noises. The fusion rules behave similarly as for simulation 0 and 1, but the performances are

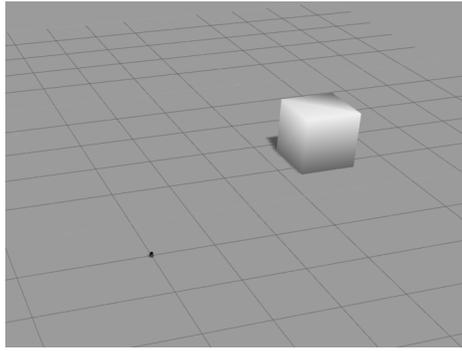
a bit lower which reflects the effect of noisy measurements in the grid estimation process. For simulations 4, 5 and 6, the noise reaches 15% for ND and 30% for FA which is quite strong. As we see in figure 1d, the Bayesian and DS fusion rules are not able to detect the second state change, during the simulation time. This induces bad False alarm rates. In the last simulations 7 and 8, the noise is about 25% of ND and 50% of FA. In these conditions, all the methods have poor false alarm rates but the PCR6 keeps good (low) non detection rates.

### B. LIDAR simulation

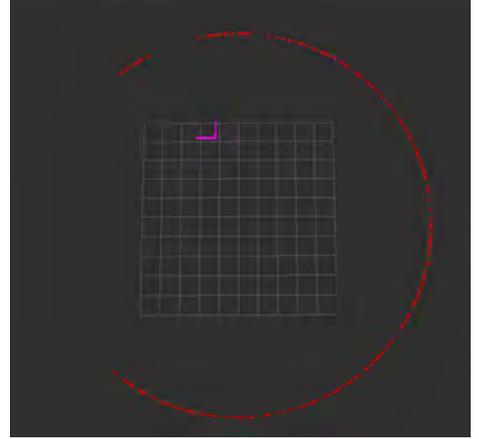
In this simulation, the DS and PCR6 fusion rules are compared on a 2D occupancy grid problem close to real

N°	discounting	key time	sensor noise	sensor belief	Bayesian		DS		PCR6	
	$\alpha$	$t_1/t_2$	$ND/FA$	$m_{SG}(O)/m_{SG}(F)$	ND	FA	ND	FA	ND	FA
0	0.05	20/40	0	0.8/0.6	10.0	6.0	10.0	6.0	10.0	6.0
1	0	20/40	0	0.8/0.6	65.0	24.0	60.0	32.0	10.0	0.6
2	0.05	20/40	10/10	0.8/0.6	11.2	9.2	10.5	10.0	10.0	9.6
3	0	20/40	10/10	0.8/0.6	77.7	15.2	73.5	18.9	11.5	6.7
4	0.05	20/40	15/30	0.8/0.6	9.2	28.0	8.2	31.5	8.4	28.9
5	0	20/40	15/30	0.8/0.6	33.0	62.7	26.9	65.8	8.4	28.8
6	0	20/40	15/30	0.6/0.4	31.3	63.9	26.0	67.3	9.3	38.7
7	0	20/40	25/50	0.6/0.4	15.1	76.9	11.5	79.4	5.7	64.0
8	0	20/40	25/50	0.4/0.2	7.1	83.9	5.1	85.1	1.9	87.3

Table I: Comparison of false alarm and non detection rates (%).



(a) View of the Gazebo simulation: the box turns around the LIDAR sensor.



(b) Bird view of one LIDAR scan.

Figure 2: Simulation setup.

application for robot perception. The simulation was realized using the Robot Operating System (ROS) [50] environment and the Gazebo [51] simulator is used here to simulate a Hokuyo LIDAR and a moving object as shown on Figure 2a. The simulated sensor has a FoV (Field of View) about  $270^\circ$  and a max range about 10m. The rate of the scan is 20Hz and the ranges of the LIDAR point are corrupted with a Gaussian noise  $\mathcal{N}(0, 0.1)$ .

Figure 2b shows a simulated LIDAR scan. The beams that do not hit obstacle within the range are considered as max range (as done in the real Hokuyo sensor). The moving object is a box which has a circular trajectory and moves at 6 rpm around the LIDAR. A *ground true* grid is computed according the real position of the box and its geometry at each scan time. The grid used is a square of 10 m by 10 m with a resolution of 0.1 m, and the ScanGrid BBA are set to  $m_{SG}(O) = 0.8$ ,  $m_{SG}(\Omega) = 0.2$  for occupied cells and  $m_{SG}(F) = 0.6$ ,  $m_{SG}(\Omega) = 0.4$  for free cells.

In order to quantify the results, we compute some metrics. However, because of occlusion, only the cells located on the edges of the box can be considered, that is why we don't consider global metrics. We consider here the two following metrics: 1) the number of correct occupied cell (proportional to recall in our case), and 2) the number of conflicting cells

close to the box. The first describes the ability of the method to add objects into the map and also by analogy to remove object from the map. The second describes the ability of the method to detect moving objects by generating conflict. This ability is important and is one of the improvement of evidential grid against classical Bayesian grid estimation.

Figures 3 & 4 show the result over one turn. The number of cells detected for both metrics depends a lot on the position around the sensor. This can be explained because, in some place, the LIDAR sensor is able to see two edges of the box. In other situations, the LIDAR sensor detects just one edge, and when the box is behind (on the back of) the sensor it is out of the field of view of the LIDAR. On Fig. 3, we can see that the number of occupied cells with the PCR6 fusion is greater than with the DS fusion. Contrarily to the PCR6, the DS fusion without discounting cannot estimate well the quick changes of states in the map. From the motion standpoint, Fig. 4 shows that the PCR6 approach keeps the same level to generate conflict in presence of moving object (similar to DS fusion). The x-axis of Figures 3 & 4 is the time stamp of the LIDAR scans, and the y-axis is the number of cells in different states (occupied for Fig. 3, or with conflict for Fig. 4).

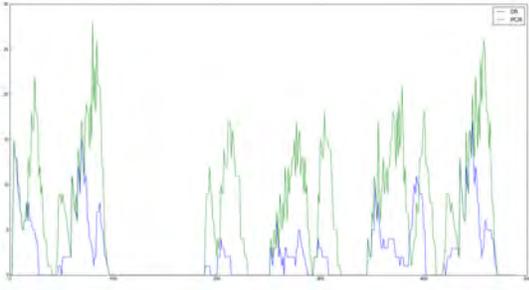


Figure 3: 2D LIDAR simulation: Number of correct occupied cells (green=PCR6,blue=DS).

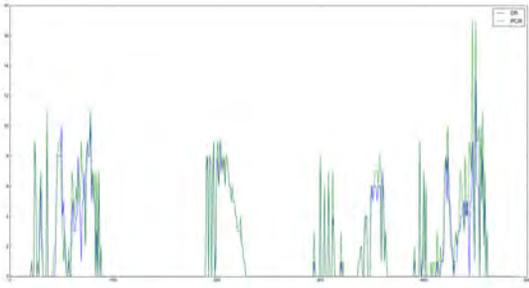


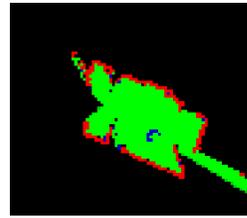
Figure 4: 2D LIDAR simulation: Number of conflicted cells into the box shape (green=PCR6,blue=DS).

### C. Real data processing

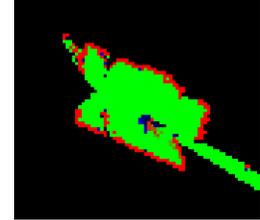
A real experimentation was realized using an Hokuyo UTM-30LX sensor. This experimentation takes place in an office in which a person was walking into. The evidential occupancy grid fusion node was implemented within the ROS environment. The grid has the same size and resolution as in the previous example. The BBA used in the sensor model has been set to  $m_{SG}(O) = 0.8$ ,  $m_{SG}(\Omega) = 0.2$  for occupied cells and  $m_{SG}(F) = 0.86$ ,  $m_{SG}(\Omega) = 0.2$  for free cells. No discounting was applied.

Figure 5 presents the occupancy grid estimation using DS and PCR6 rules of combination and for two typical snapshots of the sequence. The color of cells denotes the state having the highest mass value: green for  $F$  (free state), red for  $O$  (occupied state), and black for  $\Omega$  (full uncertainty). For convenience, we have also displayed in blue all the cells that carry a conflicting mass  $m(\emptyset) > 0.1$  before applying the normalization step of DS rule, or before applying the proportional conflict redistribution with PCR6. Figure 5a and 5c show the result using DS rule. The room scanned by the sensor is correctly mapped and its bounds (mainly walls and doors) are clearly identified by the red pixels. The free space (green pixels) is correctly detected in the room except near the people that is labeled as free (with conflicting cell shown in blue for convenience). The people moving around the desk in the office room is only detected from conflicting cells when

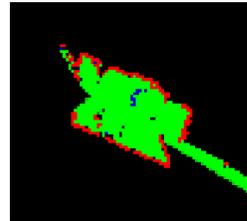
he stops to walk several times. Figure 5b and 5d show the PCR6 result at the same time stamps. In this case, the people is correctly detected as shown by the red pixels (occupied cells) inside the green area (the office room). A conflict cell is created when he starts walking in the room. The static part of the room is also detected (as with DS fusion rule).



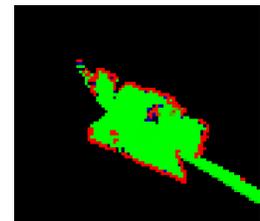
(a) Snapshot 1 - DS fusion.



(b) Snapshot 1 - PCR6 fusion.



(c) Snapshot 2 - DS fusion.



(d) Snapshot 2 - PCR6 fusion.

Figure 5: Result of evidential occupancy grid in real experimentation.

## V. CONCLUSIONS AND PERSPECTIVES

In this work we have presented a novel application of the belief functions which significantly improves the map building process for robots environment perception and grid map estimation. This work shows the importance of defining an accurate sensor model. We have considered the uncertainties of the LIDAR measurements and used the PCR6 rule of DSmT to model and combine sensor information. Our new method differs of Bayesian approach by allowing support for more than one proposition at a time, rather than a single hypothesis. It is an interval-based approach, as defined by the lower and upper probability bounds  $[Bel, Pl]$  allowing the lack of measurement to be modeled adequately. This new method differs also from the classical evidential approach because the PCR6 rule is used instead of DS rule. Experimental results with the LIDAR confirm the improvements of the accuracy of this new grid estimation method with respect to previous methods. As perspectives, we will try to implement this fusion rule in 3D occupancy grid (Octomap based) and use a stereo camera with dense disparity map computation as sensor source. In future works, we will consider in this perception context more classes into the frame of discernment and we will also test the improved PCR6 rule of combination including Zhang's degree of intersection.

## REFERENCES

- [1] P.S.P. Stepan, M.K.M. Kulich, L.P.L. Preucil, *Robust data fusion with occupancy grid*, IEEE Trans. on SMC, Part C, Vol. 35, No. 1, 2005.
- [2] H. Zhao, et al., *SLAM in a dynamic large outdoor environment using a laser scanner*, in 2008 IEEE Int. Conf. on Robotics and Automation, pp. 1455–1462, May 2008.
- [3] A. Cherubini, F. Chaumette, *Visual navigation with obstacle avoidance*, in 2011 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 1593–1598, Sept. 2011.
- [4] J. Moras, V. Cherfaoui, P. Bonnifait, *Credibilist occupancy grids for vehicle perception in dynamic environments*, 2011 IEEE Int. Conf on Robotics and Automation, Shanghai, China, pp. 84–89, May 2011.
- [5] J. Moras, V. Cherfaoui, P. Bonnifait, *Moving Objects Detection by Conflict Analysis in Evidential Grids*, 2011 IEEE Intelligent Vehicles Symposium IV, pp. 1122–1127, June 2011.
- [6] G. Shafer, *A Mathematical Theory of Evidence*, Princeton: Princeton University Press, 1976.
- [7] J. Moras, V. Cherfaoui, P. Bonnifait, *Evidential Grids Information Management in Dynamic Environments*, Proc. of Fusion 2014 Int. Conf, Salamanca, Spain, July 2014.
- [8] F. Smarandache, J. Dezert, *Advances and applications of DSMT for information fusion (Collected works)*, American Research Press, USA Vol.1, 2, 3 & 4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [9] A. Elfes, *Occupancy Grid: A Probabilistic Framework for Robot Perception and Navigation*, Ph.D. Thesis, Carnegie Mellon Univ., PA, USA, 1989.
- [10] A. Elfes, *Using occupancy grids for mobile robot perception and navigation*, Computer, Vol. 22, no. 6, pp. 46–57, 1989.
- [11] A. Elfes, *Occupancy grids: A Stochastic spatial representation for active robot perception*, in Proc. of the Conference on Uncertainty in AI, pp. 60–70, 1990.
- [12] A. Elfes, *Occupancy grids: A stochastic spatial representation for active robot perception*, in Autonomous Mobile Robots: Perception, Mapping, and Navigation (S.S. Iyengar and A. Elfes, eds.), IEEE Computer Society Press, pp. 60–71, 1991.
- [13] A. Elfes, *Multi-source spatial data fusion using Bayesian reasoning*, in: M.A. Abidi, R.C. Gonzales (Eds.), *Data Fusion in Robotics and Machine Intelligence* (Chapter 3), Academic Press, New York, 1992.
- [14] H.P. Moravec, *Sensor fusion in certainty grids for mobile robots*, AI Magazine, Summer: 116–121, 1988.
- [15] G. Oriolo, G. Ulivi, M. Vendittelli, *On-line map building and navigation for autonomous mobile robots*, Proc. 1995 IEEE Int. Conf. on Robotics and Automation, pp. 2900–2906, Nagoya, Japan, 1995.
- [16] G. Oriolo, G. Ulivi, M. Vendittelli, *Fuzzy maps: A new tool for mobile robot perception and planning*, Journal of Robotic Systems, Vol. 14, No. 3, pp. 179–197, 1997.
- [17] K. Konolige, *Improved occupancy grids for map building*, Autonomous Robots, Vol. 4, No. 4, pp. 351–367, 1997.
- [18] G. Oriolo, G. Ulivi, *Real time map building and navigation for autonomous robots in unknown environments*, IEEE Trans. on SMC, Vol. 28, No. 3, pp. 316–332, 1998.
- [19] G. Oriolo, G. Ulivi, M. Vendittelli, *Real-time map building and navigation for autonomous robots in unknown environments*, IEEE Transactions on SMC, Vol. 5, 1999.
- [20] S. Noykov, C. Roumenin, *Occupancy grids building by SONAR and mobile robot*, Robotics and Auton. Syst., Vol. 55, pp. 162–175, 2007.
- [21] I.-H. Li, C.-C. Hsu, S.-S. Lin, *Map Building of Unknown Environment Based on Fuzzy Sensor Fusion of Ultrasonic Ranging Data*, International Journal of Fuzzy Systems, Vol. 16, No. 3, pp. 368–377, September 2014.
- [22] K. Hughes, R. Murphy, *Ultrasonic robot localization using Dempster-Shafer theory*, in Proc. of SPIE on Neural and Stochastic Methods in Image and Signal Processing, pp. 2–11, 1992.
- [23] P. Tirumalai, B.G. Schunk, R.C. Jain, *Evidential reasoning for building environment maps*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 25, No. 1, pp. 10–20, 1995.
- [24] F. Gambino, G. Oriolo, G. Ulivi, *A comparison of three uncertainty calculus techniques for ultrasonic map building*, in: Proc. 1996 SPIE Int. Symp. on Aerospace/Defense Sensing and Control, pp. 249–260, Orlando, FL, USA, 1996.
- [25] D. Pagac, E. Nebot, H. Durrant-Whyte, *An evidential approach to probabilistic map-building*, Proc. of IEEE Int. Conf. on Robotics and Automation, Vol. 1, pp. 745–750, Minneapolis, MN, USA, April 1996.
- [26] D. Pagac, E. Nebot, H. Durrant-Whyte, *An evidential approach to map-building for autonomous vehicles*, IEEE Trans. on Robotics and Automation, Vol. 14, No. 4, pp. 623–629, 1998.
- [27] M. Ribo, A. Pinz, *A comparison of three uncertainty calculi for building SONAR-based occupancy grids*, Robotics and Autonomous Systems, Vol. 31, pp. 201–209, 2001.
- [28] T. Reineking, J. Clemens, *Evidential FastSLAM for grid mapping*, Proc. of Fusion 2013 Conf., pp. 789–796, Istanbul, Turkey, July 2013.
- [29] J. Clemens, T. Reineking, *Multi-Sensor Fusion Using Evidential SLAM for Navigating a Probe through Deep Ice*, Proc. Of Belief 2014, Oxford, UK, Sept. 2014.
- [30] X. Li, X. Huang, M. Wang, J. Xu, H. Zhang, *DSMT Coupling with PCR5 for Mobile Robots Map Reconstruction*, Proc. of 2006 Int. Conf. on mechatronics and Automation, Luoyang, China, June 26–28, 2006.
- [31] X. Li, X. Huang, M. Wang, *Robot Map Building from SONAR Sensors and DSMT*, Information & Security : An Int. J., Bulg. Acad. of Sci., Vol. 20, pp. 104–121, Sofia, March 2006.
- [32] X. Huang, X. Li, J. Dezert, M. Wang, *A Fusion Machine Based on DSMT and PCR5 for Robot's Map Reconstruction*, Int. Journal of information Acquisition, Vol. 3, No.3, pp. 201–211, Sept. 2006.
- [33] X. Li, X. Huang, M. Wang, J. Dezert, *A fusion Machine based on DSMT and PCR5 for robot's map reconstruction*, International Journal of Information Acquisition, Vol. 3, No. 3, pp. 1–11, 2006.
- [34] X. Li, X. Huang, J. Dezert, L. Duan, M. Wang, *A successful application of DSMT in SONAR grid map building and comparison with DST-based approach*, Int. Journal of Innovative Computing, Information and Control, Vol. 3, No. 3, June 2007.
- [35] P. Li, X. Huang, S. - Yang, J. Dezert, *SLAM and path planning of mobile robot using DSMT*, J. of Software Eng., Vol. 7, No. 2, pp. 46–67, 2013.
- [36] J. Zhou, J. Duan, G. Yang, *Occupancy Grid Mapping Based on DSMT for Dynamic Environment Perception*, International Journal of Robotics and Automation (IJRA), Vol. 2, No. 4, pp. 129–139, December 2013.
- [37] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Proc. of Fusion 2012 Int. Conf., Singapore, July 9–12, 2012.
- [38] A. Tchamova, J. Dezert, *On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory*, 6th IEEE Int. Conf. on Int. Syst. (IS'12), Sofia, Bulgaria, Sept. 6–8, 2012.
- [39] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intelligent Systems, Special Issue: Advances in Intelligent Systems, Vol. 29(3), pp. 223–252, March 2014.
- [40] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, Proc. of Fusion 2013, Istanbul, Turkey, July 2013.
- [41] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, Proc. of Fusion 2010 Int. Conf., Edinburgh, UK, July 26–29, 2010.
- [42] S. Thrun, *Learning occupancy grid maps with forward sensor models*, Autonomous robots, Vol. 15, No. 2, pp. 111–127, 2003.
- [43] M.K. Tay, K. Mekhnacha, M. Yguel et al., *The Bayesian occupation filter*, in Probabilistic Reasoning and Decision Making in Sensory-Motor Systems (P. Bessière, C. Laugier, and R. Siegwart, eds.), pp. 77–98, Springer Berlin Heidelberg, Springer Ed., 2008.
- [44] T. Weiss, B. Schiele, K. Dietmayer, *Robust Driving Path Detection in Urban and Highway Scenarios Using a Laser Scanner and Online Occupancy Grids*, IEEE Intell. Vehicles Symp., pp. 184–189, June 2007.
- [45] P. Smets, *Decision making in the TBM: the necessity of the pignistic transformation*, IJAR, Vol. 38, pp. 133–147, Feb. 2005.
- [46] <http://bfas.iutlan.univ-rennes1.fr/wiki/index.php/Toolbox>
- [47] L.A. Zadeh, *On the validity of Dempster's rule of combination*, Memo M79/24, Univ. of California, Berkeley, CA, U.S.A., 1979.
- [48] L.A. Zadeh, *Book review: A mathematical theory of evidence*, The AI Magazine, Vol. 5, No. 3, pp. 81–83, 1984.
- [49] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, Vol. 7 (2), pp. 85–90, 1986.
- [50] M. Quigley, K. Conley, B.P. Gerkey et al., *ROS: an open-source Robot Operating System*, in 2009 ICRA Workshop on Open Source Software, 2009.
- [51] N. Koenig, A. Howard, *Design and Use Paradigms for Gazebo - An Open-Source Multi-Robot Simulator*, in 2004 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 2149–2154, Sept. 2004.
- [52] R.A. Jacobs, *Methods for combining expert's probability assessments*, Neural Computation, Vol. 7 (5), pp. 867–888, 1995

# Belief Function Theory Based Decision Support Methods: Application to Torrent Protection Work Effectiveness and Reliability Assessment

Simon Carladous<sup>a,b,c,e</sup>, Jean-Marc Tacnet<sup>a,b</sup>, Jean Dezert<sup>d</sup>, Mireille Batton-Hubert<sup>e</sup>

<sup>a</sup>IRSTEA, Snow Avalanche Engineering and Torrent Control Research Unit,  
Saint Martin d'Hères, France

<sup>b</sup>Grenoble Alpes University, France.

<sup>c</sup>AgroParisTech, Paris, France.

<sup>d</sup>ONERA, The French Aerospace Lab, Palaiseau, France.

<sup>e</sup>ENSMSE, Saint-Étienne, France.

Emails: simon.carladous@irstea.fr, jean-marc.tacnet@irstea.fr, jean.dezert@onera.fr, batton@emse.fr

Originally published as: S. Carladous, J.-M. Tacnet, J. Dezert, M. Batton-Hubert, *Belief Function Theory Based Decision Support Methods: Application to Torrent Protection Work Effectiveness and Reliability Assessment*, in Proc. of European Safety and Reliability Conf. (ESREL 2015), Zurich, Switzerland, Sept. 7–10, 2015, and reprinted with permission.

**Abstract**—Civil engineering protection works mitigate natural risks in mountains, such as torrents. Analysing their effectiveness at several scales is an essential issue in the risk management. Based on expert knowledge, used methods have been developed under risky environment. However, decision is made under uncertainty because of 1) the lack of information and knowledge on natural phenomena and 2) the heterogeneity of available information and 3) the reliability of sources. In this paper, we propose to help decision-makers with advanced multicriteria decision making methods (MCDMs). Combining classical MCDM approaches, belief function, fuzzy sets and possibility theories, they make it possible decisions based on heterogeneous, imprecise and uncertain evaluation of criteria provided by more or less reliable sources in an uncertain context. COWA-ER (Cautious Ordered Weighted Averaging with Evidential Reasoning), Fuzzy-Cautious OWA or ER-MCDA (Evidential Reasoning for Multi Criteria Decision Analysis) are thus applied to several scales of effectiveness assessment.

**Keywords:** Torrent protection, belief functions, MCDM.

## I. INTRODUCTION

Mountain natural phenomena such as torrential floods put people and buildings at risk. Protection works influence both causes and effects of phenomena to limit induced risks. For instance, check-dams control material volume and flow of torrential floods. Their design allow them to reduce sediment production (Figure 1). Defining the strategy for investment and maintenance is an essential issue in the risk management process. It is based on their effectiveness assessment. Decision support tools help assessing their economic efficiency depending on their structural state and functional effects on phenomena (stopping, braking, guiding, etc.) [1].

Cost Benefit Analysis (CBA) is the most used decision-aid method in the natural hazard context. It helps assessing efficiency of potential actions comparing, for several scenarii, investment and maintenance costs with direct and indirect losses [2]. Actually, natural risk analysis is limited to a set of scenarii which can be discussed [3]. However, probability

knowledge (distribution or scenarii) is affected by the lack of information on phenomena, but also by heterogeneity and reliability of available sources (Tacnet 2009). [4].

Concepts of failure mode and effects analysis (FMEA), already used for hydraulic dams [5], are extended to assess the effectiveness of check-dams [6]. Those methodologies elicit the expert reasoning process and consider structural, functional and economic features [1]: indicators formalise information processing to make it repeatable and reproducible [7]. Nevertheless, assessment is based on heterogeneous and imprecise information provided by more or less reliable sources [4].

Methods to represent information imperfection are needed to aid decisions including check-dam effectiveness assessment. Advanced MCDMs combining classical MCDM approaches [8], [9], belief function [10], [11], fuzzy sets [12] and possibility theories [13] have been developed to help decisions under risk or uncertainty such as COWA [14], Fuzzy-Cautious OWA [15] and ER-MCDA [16].

This paper first recalls the context of information imperfection related to check-dams. We secondly introduce the principles of new belief function theory based evolutions of MCDMs. We then apply them to cases related to effectiveness of check-dams. We finally discuss remaining issues for new decision-making methods in risky and uncertain contexts.

## II. EFFECTIVENESS OF PROTECTION WORKS IN AN UNCERTAIN ENVIRONMENT

Assessing effectiveness of existing check-dams is based on their structural state, functional capacity and relative risk reduction. We describe below the decision context and information imperfection all over the decision process.

### A. Formalization of decision context

1) *Several system scales as alternatives:* Protecting exposed elements with check-dams is based on interdependent systems. A check-dam  $E^l$  belongs to a device  $D^o$ . Several

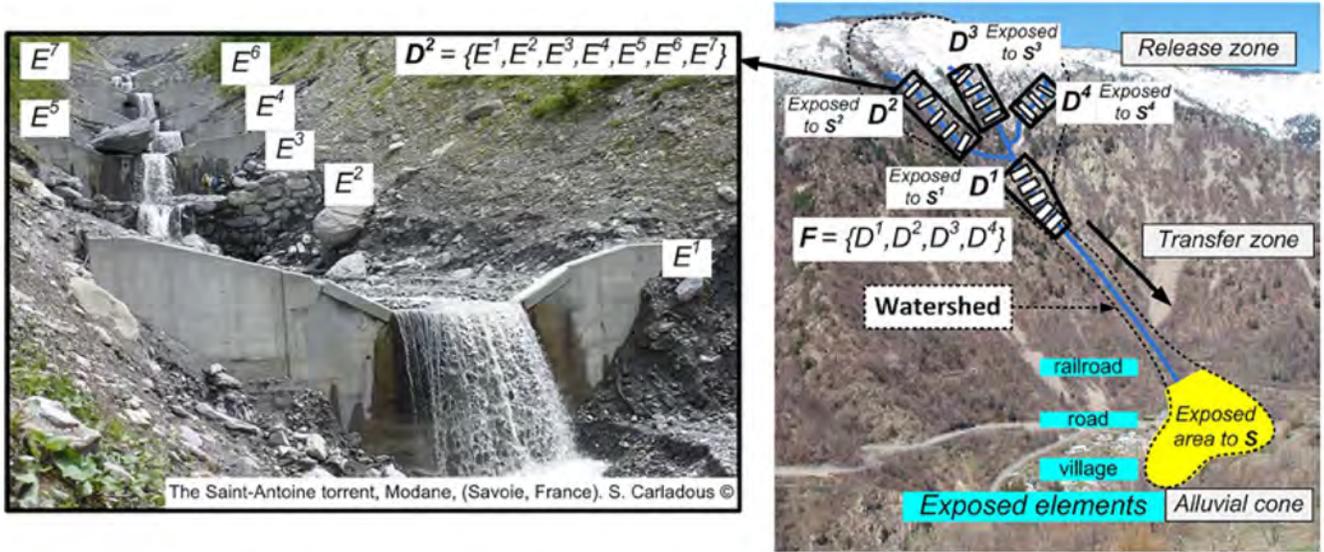


Figure 1. Multi-system formalization for check-dams in a torrential watershed.

devices protect exposed elements at the watershed scale ( $F$ ). Each  $E^l$  is considered as an alternative belonging to a set of  $m$  check dams  $D^o = \{E^1, \dots, E^l, \dots, E^m\}$ .  $D^o$  is a device alternative in the set  $F = \{D^1, \dots, D^o, \dots, D^t\}$ .  $F$  represents all  $t$  devices which protect exposed elements in the watershed (Figure 1).

2) *Possible actions on systems as alternatives*: For each system scale  $E^l$ ,  $D^o$  and  $F$ , several actions  $a_i$  can be proposed: for example, no action ( $a_1$ ), maintenance of check-dams ( $a_2$ ) or building new works ( $a_3$ ).  $E_i^l$ ,  $D_i^o$  and  $F_i$  represent all possible actions on each system scale (resp. a single check-dam, a set of check-dams, all sets in the watershed).

3) *Decision objects and linked problems*: A decision-making problem consists in choosing, ranking or sorting alternatives on the basis of quantitative or qualitative criteria  $g_j$  [9]. Effectiveness is the level of objective achievement [17]. Sorting alternatives  $E^l$ ,  $D^o$  and  $F$  in effectiveness classes (e.g., optimal, correct, partial, deficient) is a recurrent issue. Choosing between several alternatives  $a_i$ , or ranking them, are other practical issues.

#### B. Various information is needed but is imperfect

1) *The states of the nature  $S$  or  $S^o$* : Debris flows and torrential floods with bed-load transport are the two main torrential processes [18]. Choosing a specific criterion of interest for each process is needed (e.g., flow volume or deposit depth).

The states of the nature analysis depends on its location in the watershed. They can be represented by a finite or a continuous set according to available information. For torrential floods, field experts define a finite set  $S = \{S_1, \dots, S_k, \dots, S_n\}$  for  $F$  scale and another set  $S^o = \{S_1^o, \dots, S_k^o, \dots, S_n^o\}$  for  $D^o$  and  $E^l$  scales [1].

2) *The decision-maker (DM) preferences on  $g_j$* : Assessing each alternative in a MCDM context requires three elements from the DM about  $g_j$ : 1) the list of  $g_j$ , 2) weights  $w_j$ : preferences between  $g_j$ , 3)  $g_j$  assessment scale: preferences between alternative evaluations through a total or a partial pre-order [9], [19].

3) *Decision-making and imperfect information*: To compare several alternatives, decision support tools are based on several  $g_j$  evaluations of their consequences (payoffs/gains) under  $S$  (or  $S^o$ ). For example, each  $F_i$  is evaluated given the knowledge on  $S$  and the payoff matrix defined by  $C = [C_{ik}]$  where  $i = 1, \dots, q$  and  $k = 1, \dots, n$  (Eq. (1)). The decision problem consists in choosing the alternative  $F_{i^*} \in F_i$  which maximizes the payoff to the DM. We assume that  $C_{ik}$  assessment can be based on several  $g_j$ .

$$\begin{matrix}
 & S_1 & \dots & S_k & \dots & S_n \\
 F_1 & \left( \begin{array}{cccc} C_{11} & \dots & C_{1k} & \dots & C_{1n} \\ \vdots & & \vdots & & \vdots \\ F_i & C_{i1} & \dots & C_{ik} & \dots & C_{in} \\ \vdots & & \vdots & & \vdots \\ F_q & C_{q1} & \dots & C_{qk} & \dots & C_{qn} \end{array} \right) & = & C, & (1)
 \end{matrix}$$

Whatever the decision context, all decisions relate to imperfection of used information to assess  $S$ ,  $C_{ik}$  and  $g_j$  [4]: **inconsistency** (conflict between sources); **imprecision** (e.g., interval of numerical values); **incompleteness** (lack of information while data exist); **aleatory uncertainty** (aleatory events); **epistemic uncertainty** (lack of knowledge).

Depending on his knowledge about  $S$ , the DM is face on different decision-making problems [14]: **under certainty** (only one  $S_k$  is known); **under risk** (the true  $S$  is unknown but one knows all the probabilities  $p_k = P(S_k)$ ); **under**

**ignorance** (one assumes no knowledge about the true state but that it belongs to  $S$ ); **under uncertainty** (the knowledge on  $S$  is characterized by a belief structure).

### III. NEW BELIEF FUNCTION THEORY BASED EVOLUTION OF MCDMS

Comparing alternatives requires assessment of 1) DM preferences on  $w_j$  and  $g_j$ , 2) information imperfection to evaluate  $S$  and  $g_j$ , 3) MCDM choice to aggregate several  $g_j$  to define  $C$ . In this part, we introduce the principles of methods based on new MCDMs evolutions based on belief function theory.

#### A. Basics of belief functions

Shafer [10] originally proposed the basics of belief functions. One starts with a finite set  $\Theta$  (called the frame of discernment of the decision problem). Each element of  $\Theta$  is a potential answer of the decision problem and they are assumed exhaustive and exclusive. The powerset of  $\Theta$  denoted  $2^\Theta$  is the set of all subsets of  $\Theta$ , empty set included. A body of evidence is a source of information that will help the DM to identify the best element of  $\Theta$ . The interest of belief functions is their ability to model epistemic uncertainties. Each body of evidence is characterized by basic belief assignment (bba), or a mass of belief, which is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  that satisfies  $m(\emptyset) = 0$ , and for all  $A \neq \emptyset \in 2^\Theta$  the condition  $\sum_{A \subseteq \Theta} m(A) = 1$ . The Belief function  $\text{Bel}(\cdot)$  and the plausibility function  $\text{Pl}(\cdot)$  are defined from  $m(\cdot)$  by :

$$\text{Bel}(A) = \sum_{B \subseteq A | B \in 2^\Theta} m(B), \quad (2)$$

$$\text{Pl}(A) = \sum_{B \cap A \neq \emptyset | B \in 2^\Theta} m(B). \quad (3)$$

$\text{Bel}(A)$  and the plausibility function  $\text{Pl}(A)$  are often interpreted as lower and upper bounds of the unknown probability of  $A$ . The vacuous bba defined as  $m_v(\Theta) = 1$  models the full ignorant source of evidence. Shafer [10] proposed Dempster's rule to combine distinct sources of evidence which has been subject to strong debates in fusion community starting from Zadeh's first criticism in 1979. Since the 90's many alternatives have been proposed to combine more or less efficiently belief functions, as well as an extension of belief function in the framework of Dezert-Smarandache Theory (DSMT) as shown and discussed in [11].

According to the DM attitude, credibilities, plausibilities, Smets' Pignistic probability  $BetP$  [20] or Dezert-Smarandache probability  $DSmP_{\epsilon=0}$  [11] (Vol. 3) can be computed to compare alternatives.

#### B. ER-MCDA

Tacnet [4] proposed the ER-MCDA methodology. Its originality consists in the association of different theories. It dissociates imperfect evaluations from their combination in the fusion process considering both evaluation imperfection and heterogeneity, reliability of sources. It uses developments for MCDM based on the combination of Analytic Hierarchic Process (AHP) approach developed by Saaty [8] and DSMT

[11]. AHP allows to build bbas from DM preferences on solutions which are established with respect to several  $g_j$ . DSMT allows to aggregate efficiently the (possibly highly conflicting) bbas based on each criterion. DSMT-AHP method also allows to take into account the different importances of  $g_j$  and/or of the different members of the DM group.

ER-MCDA exploits the following general principles into independent steps:

- The AHP methodology helps to analyze the decision problem through a hierarchical structure and to define the evaluation classes for decision through a common frame of discernment  $\Theta$ .
- The imprecise evaluation and mapping of  $g_j$ : qualitative or quantitative criteria are evaluated through possibility distributions representing both imprecision and uncertainty [13]. Possibility distribution can be derived into bbas [21]. We use a mapping process that projects the bbas expressed on fuzzy sets expressed on  $\Theta$  [12].
- The fusion of mapped evaluations and  $g_j$ : a first fusion process is done for all evaluations of the different sources for a same  $g_j$ . Bbas can be discounted according to the reliability level of each source. We finally get bbas for each  $g_j$  whose weights  $\omega_j$  have been defined according to the classical AHP method. Those  $\omega_j$  are derived into importance discounting factors. Bbas corresponding to each  $g_j$  are then fused a second time to get the final result which is called a decision profile. This profile shows not only the decision to take but provides also an evaluation of the distribution of knowledge on the other levels and uncertainty. It is possible to check if all sources agree about the decision and also to have an idea about the uncertainty of their evaluation. The quality of information leading to decision is linked to the decision itself. The results can be bbas or belief, plausibility values that correspond to pessimistic or optimistic choice of a decision level. With ER-MCDA, one uses PCR6 (Proportional Conflict Redistribution Rule no 6) developed in DSMT [11] (Vol. 3) to palliate disadvantages of the classical Dempster fusion rule discussed in [22]. The importance of criteria is a different concept than the classical reliability concept developed and used in the belief theory context. In order to make a difference between importance of criteria, uncertainty related to the evaluations of criteria and reliability of the different sources, specific methods such as DSMT-AHP [23], [24] have extended Saaty's AHP method.

#### C. COWA-ER and Fuzzy Cautious OWA

Tacnet and Dezert [14] proposed the COWA-ER method for decision-making under uncertainty taking into account imperfect evaluations and unknown beliefs about groups of the possible states of the world. COWA-ER mixes cautiously the principle of Ordered Weighted Averaging (OWA) approach [25] with the fusion of belief functions proposed in DSMT [11]. Fuzzy Cautious OWA [15] is an improvement of COWA-ER using fuzzy sets.

1) *The OWA approach*: To recall it, we take into account the **decision-making** problems introduced in II-B3 and Eq. (1).

**1 – under certainty**: one chooses  $F_{i^*}$  with  $i^* \triangleq \arg \max_i \{C_{ik}\}$ .

**2 – under risk**: as for the CBA (cf Sec. I), for each  $F_i$ , we compute expected payoff  $E[C_i] = \sum_k p_k \cdot C_{ik}$ , then we choose  $F_{i^*}$  with  $i^* \triangleq \arg \max_i \{E[C_i]\}$ .

**3 – under ignorance**: Yager [25] uses the OWA operator as a weighted average of ordered values of a variable. For each  $F_i$  and a given criterion of interest  $g_j$ , one chooses a weighting vector  $W_i = [w_{i1}, w_{i2}, \dots, w_{in}]$  and computes its OWA value  $V_i \triangleq \text{OWA}(C_{i1}, C_{i2}, \dots, C_{in}) = \sum_k w_{ik} \cdot b_{ik}$  where  $b_{ik}$  is the  $k$ th largest element in the collection of payoffs  $C_{i1}, C_{i2}, \dots, C_{in}$ . Then one chooses  $F_{i^*}$  with  $i^* \triangleq \arg \max_i \{V_i\}$ .  $W_i$  depends on the decision attitude of the DM (pessimistic, optimistic, normative/neutral, etc.).

**4 – under uncertainty**: one assumes that a priori knowledge on the frame  $S$  is given by a bba  $m(\cdot) : 2^S \rightarrow [0, 1]$ . This case includes all previous cases depending on the choice of  $m(\cdot)$ . Yager's OWA under uncertainty is based on the derivation of a *generalized expected value*  $C_i$  of payoff for each  $F_i$  as follows:

$$C_i = \sum_{l=1}^r m(X_l) V_{il}, \quad (4)$$

where  $r$  is the number of focal elements of the belief structure  $(S, m(\cdot))$ .  $m(X_l)$  is the mass of belief of  $X_l \in 2^S$ , and  $V_{il}$  is the payoff we get when we select  $F_i$  and the state of the nature lies in  $X_l$ .

For  $F_i$  and a focal element  $X_l$ , instead of using all payoffs  $C_{ik}$ , we consider only the payoffs in the set  $M_{il} = \{C_{ik} | S_k \in X_l\}$  and  $V_l = \text{OWA}(M_{il})$  for some decision-making attitude chosen a priori. Once generalized expected values  $C_i$ ,  $i = 1, 2, \dots, q$  are computed, we compare alternatives through these results.

The principle of this method is simple, but its implementation can be quite greedy in computational resources specially if one wants to adopt a particular attitude for a given level of optimism, specially if the dimension of the frame  $S$  is large.

2) *The COWA-ER approach*: Yager's OWA approach is based on the choice of a given attitude measured by an optimistic index in  $[0, 1]$  to get the weighting vector  $W_i$ . What should be done in practice if we don't know which attitude to adopt? An answer to this question has been proposed in Cautious OWA with Evidential Reasoning (COWA-ER) which exploits the results of the two extreme attitudes jointly (pessimistic and optimistic ones) to take a decision under uncertainty based on the imprecise valuation of alternatives. In COWA-ER, the pessimistic and optimistic OWA are used respectively to construct the intervals of expected payoffs for different alternatives. For example, for  $q$  alternatives, the expected payoffs are:

$$E[C] = \begin{bmatrix} E[C_1] \\ E[C_2] \\ \vdots \\ E[C_q] \end{bmatrix} \subset \begin{bmatrix} [C_1^{\min}, C_1^{\max}] \\ [C_2^{\min}, C_2^{\max}] \\ \vdots \\ [C_q^{\min}, C_q^{\max}] \end{bmatrix}.$$

Therefore, one has  $q$  sources of information before using the belief functions framework. Basically, the COWA-ER methodology requires four steps:

- Step 1: normalization of imprecise values in  $[0, 1]$ ;
- Step 2: conversion of each normalized imprecise value into elementary bba  $m^o(\cdot)$ ;
- Step 3: fusion of bba  $m^i(\cdot)$  with some combination rule (typically the PCR6 rule);
- Step 4: choice of the final decision based on the resulting combined bba.

With COWA-ER, we consider as  $\Theta$ , the finite set of alternatives  $\Theta = \{Z_1, Z_2, \dots, Z_q\}$  and the sources of belief associated with them obtained from the normalized imprecise expected payoff vector  $E^{Imp}[C_i]$ . The modeling for computing a bba associated to hypothesis  $F_i$  from any imprecise value  $[a; b] \subseteq [0; 1]$  is done by:

$$\begin{cases} m_i(F_i) = a, \\ m_i(\bar{F}_i) = 1 - b, \\ m_i(F_i \cup \bar{F}_i) = m_i(\Theta) = b - a, \end{cases} \quad (5)$$

where  $\bar{F}_i$  is the  $F_i$ 's complement in  $\Theta$ .

COWA-ER can help to take a decision if one wants on a group/subset of alternatives satisfying a min of credibility (or plausibility level) selected by the DM. It can also be extended directly for the fusion of several sources of informations when each source can provide a payoffs matrix. We can also discount each source easily if needed.

3) *The Fuzzy-COWA-ER approach*: Unfortunately, COWA-ER has a serious limitation because the computational time depends on the number of alternatives. In COWA-ER, each expected interval is used as an information source, however, these expected intervals are jointly obtained and thus these information sources are relatively correlated. For these reasons, a modified version of COWA-ER, called Fuzzy-COWA-ER (or FCOWA-ER for short) has been developed in [15]. With FCOWA-ER, we consider the 2 columns of the expected payoff  $E[C_i]$  as two information sources, representing pessimistic and optimistic attitudes. The column-wise normalized expected payoff is:

$$E^{Fuzzy}[C] = \begin{bmatrix} N_1^{\min}, N_1^{\max} \\ N_2^{\min}, N_2^{\max} \\ \vdots \\ N_q^{\min}, N_q^{\max} \end{bmatrix},$$

where  $N_i^{\min} \in [0, 1]$  ( $i = 1, \dots, q$ ) represents the normalized value in the column of pessimistic attitude, and  $N_i^{\max} \in [0, 1]$  represents the normalized value in the column of optimistic attitude. The vectors  $[N_1^{\min}, \dots, N_q^{\min}]$  and  $[N_1^{\max}, \dots, N_q^{\max}]$

can be seen as two fuzzy membership functions (FMFs) representing the possibilities of all the alternatives  $F_1, \dots, F_q$ . The FCOWA-ER method requires also four steps:

- Step 1: normalize each column in  $E[C]$ , respectively, to obtain  $E^{Fuzzy}[C]$ ;
- Step 2: conversion of two normalized columns, i.e., two FMFs (Fuzzy Membership Functions) into two bbas  $m_{Pess}(\cdot)$  and  $m_{Opti}(\cdot)$  using the  $\alpha$ -cut approach introduced in [26];
- Step 3: fusion of bbas  $m_{Pess}(\cdot)$  and  $m_{Opti}(\cdot)$  with some combination rule (typically the PCR6 rule);
- Step 4: choice of the final decision based on the resulting combined bba.

In FCOWA-ER, only one combination step is needed. Furthermore, the bba's obtained by using  $\alpha$ -cuts are consonant support (nested in order).

#### IV. APPLICATION TO PROBLEMS OF PROTECTION WORKS EFFECTIVENESS

##### A. Assessment of structural effectiveness of a single check-dam through ER-MCDA

1) *AHP methodology*: The problem consists in choosing the observed structural effectiveness level of a given  $E^l$ . It is assessed through 6 criteria  $g_j$  [27] (Figure 2).  $\omega_j$  in Table I are defined by experts.

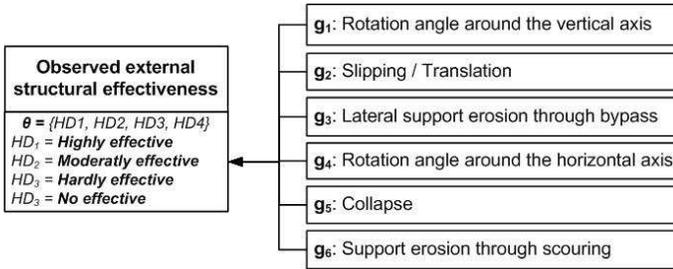


Figure 2. Hierarchical structure to assess observed structural effectiveness of  $E^l$ .

According to *DST* (Dempster-Shafer Theory) framework [10],  $\Theta$  is composed of 4 exclusive elements of effectiveness levels:  $HD_1 = \text{'High'}$ ,  $HD_2 = \text{'Medium'}$ ,  $HD_3 = \text{'Low'}$  and  $HD_4 = \text{'None'}$ .

2) *Imprecise evaluation and mapping of  $g_j$* : For  $E^l$ , we assume evaluations of  $g_j$  by two experts (sources)  $s_1$  and  $s_2$  through possibility distributions (Table I). Through expert elicitation, a set of fuzzy intervals  $L-R$  links each  $g_j$  evaluation scale and  $\Theta$ . Quotations used are extracted from [28] (Table II). Using this mapping process, bbas are established in Table III.

To take into account the reliability of each source, we discount the input masses of Table III by applying the classical Shafer's discounting method [10]. We use here discounting factors  $\alpha_{s1} = 0.7$  and  $\alpha_{s2} = 0.5$ . We obtain 12 discounted bba's (noted  $m'_1$  and  $m'_2$ ) in Table IV.

3) *Two steps of fusion*: The step 1 consists in combining the bbas  $m'_1(\cdot)$  and  $m'_2(\cdot)$  for each  $g_j$  with PCR6 fusion rule (Table V).

In step 2, we apply to each bba of Table V the importance discounting method presented in [29]. We use  $\omega_j$  (Table I) to get the Table VI. After combining its 6 bba with a variant of PCR6 to take into account positive masses on  $\emptyset$ , noted  $PCR6_\emptyset$  and a normalization procedure [29], we finally get the Table VII.

According to it,  $E^l$  is mainly medium effective because the highest belief mass is  $m(HD2)$ . Because  $m(HD3)$ , we can say that  $E^l$  effectiveness is more between low and medium, but not high, nor none.

##### B. Comparing actions on $A_i$ using (F)COWA-ER

1) *Decision problem elicitation*:  $g_j$  is the  $D^o$  effectiveness level. In a *DST* framework, one assumes 7 scenarii such as flood with bedload transport ( $S_3^o$ ) or debris flow ( $S_6^o$ ). One considers 5 possible actions such as repair of all the degraded check-dams ( $a_3$ ) or renewal of all check-dams ( $a_5$ ).

2)  *$C_{ik}$  and  $S^o$  evaluations*: One rates  $C_{ik}$  with an integer between 0 (no effective) and 10 (very high effective) [7]. As Eq. (1), one assumes  $C$  where  $q = 5$  and  $n = 7$  (Eq. (6)).

$$C = \begin{bmatrix} 5 & 3 & 4 & 2 & 3 & 1 & 1 \\ 7 & 4 & 6 & 3 & 4 & 2 & 1 \\ 8 & 5 & 7 & 4 & 5 & 3 & 1 \\ 10 & 7 & 10 & 6 & 7 & 5 & 2 \\ 10 & 9 & 10 & 9 & 10 & 10 & 4 \end{bmatrix}. \quad (6)$$

One considers 4  $X_l$ :  $X_1 = S_1^o \cup S_3^o \cup S_5^o$ ,  $X_2 = S_2^o \cup S_4^o \cup S_5^o \cup S_6^o$ ,  $X_3 = S_7^o$ ,  $X_4 = \Theta$ . One gives  $m(X_1) = 0.6$ ,  $m(X_2) = 0.2$ ,  $m(X_3) = 0.01$  and  $m(X_4) = 0.19$ . Applying the OWA pessimistic and optimistic operators, one can assess the bounds of expected effectiveness levels for each actions given by Eq. (7).

$$E[C] \subset \begin{bmatrix} [2.20; 4.56] \\ [3.00; 6.34] \\ [3.80; 7.33] \\ [5.60; 9.32] \\ [8.60; 9.94] \end{bmatrix}, \quad (7)$$

$$E^{Imp}[C] \approx \begin{bmatrix} [0.22; 0.46] \\ [0.30; 0.64] \\ [0.38; 0.74] \\ [0.56; 0.94] \\ [0.87; 1.00] \end{bmatrix}. \quad (8)$$

3) *Results through COWA-ER*: Steps 1 and 2 make it possible to assess bbas of the 5 actions in the Table VIII passing by the normalized imprecise matrix  $E^{Imp}[C]$  given in Eq. (8).

Step 3 combines the 5 bba's altogether with choice of the PCR6 fusion rule (Table IX).

Choosing the decision-making rule is needed to implement the step 4. Results are compared in the Table X. One sees that based on max of Bel, of BetP, of DSMP or of PI, the best action is always  $D_5$ .

Table I  
CRITERIA EVALUATIONS OF  $E^l$ .

Criterion $g_j$	$\omega_j$	Unity	Expert 1 (s1)			Expert 2 (s2)		
			E	$\Pi(E)$	N(E)	E	$\Pi(E)$	N(E)
$g_1$	0.1	Degree (d)	d=0	1	1	d=0	1	1
$g_2$	0.1	Meter (m)	m=0	1	1	m=0	1	1
$g_3$	0.3	Meter (m)	$2 \leq m \leq 4$	1	1	$1 \leq m \leq 6$	1	1
			$2.5 \leq m \leq 3.5$	1	0.3	$2 \leq m \leq 4$	1	0.7
$g_4$	0.1	Degree (d)	d=0	1	1	$15 \leq m \leq 20$	1	1
						$10 \leq m \leq 20$	1	0.55
$g_5$	0.1	Meter (m)	$0.1 \leq m \leq 0.5$	1	1	m=0	1	1
			$0.2 \leq m \leq 0.4$	1	0.5			
$g_6$	0.3	Meter (m)	$0.2 \leq m \leq 1.2$	1	1	$0.2 \leq m \leq 0.8$	1	1
			$0.4 \leq m \leq 0.8$	1	0.7			
			$0.5 \leq m \leq 0.7$	1	0.3			

Table II  
MAPPING MODELS FOR EACH CRITERION.

Criterion $g_j$	HD1		HD2		HD3		HD4	
	Supp	Noy	Supp	Noy	Supp	Noy	Supp	Noy
$g_1$	$0 \leq d \leq 5$	d=0	$0 \leq d \leq 15$	$5 \leq d \leq 10$	$10 \leq d \leq 30$	$15 \leq d \leq 25$	$25 \leq d$	$30 \leq d$
$g_2$	$0 \leq m \leq 0.3$	m=0	$0 \leq m \leq 1$	$0.3 \leq m \leq 0.7$	$0.7 \leq m \leq 2$	$1 \leq m \leq 1.7$	$1.7 \leq m$	$2 \leq m$
$g_3$	$0 \leq m \leq 0.5$	m=0	$0 \leq m \leq 2.5$	$0.5 \leq m \leq 2$	$2 \leq m \leq 4$	$2.5 \leq m \leq 3.5$	$3.5 \leq m$	$4 \leq m$
$g_4$	$0 \leq d \leq 5$	d=0	$0 \leq d \leq 15$	$5 \leq d \leq 10$	$10 \leq d \leq 30$	$15 \leq d \leq 25$	$25 \leq d$	$30 \leq d$
$g_5$	$0 \leq m \leq 0.2$	m=0	$0 \leq m \leq 0.5$	$0.2 \leq m \leq 0.3$	$0.3 \leq m \leq 1$	$0.5 \leq m \leq 0.8$	$0.8 \leq m$	$1 \leq m$
$g_6$	$0 \leq m \leq 0.3$	m=0	$0 \leq m \leq 1.5$	$0.3 \leq m \leq 1.1$	$1.1 \leq m \leq 2.5$	$1.5 \leq m \leq 2.2$	$2.2 \leq m$	$2.5 \leq m$

Table III  
 $E^l$  BBA'S AFTER MAPPING PROCESS IN A DST FRAMEWORK.

Criterion $g_j$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_6$
	$m_1(\cdot)$					
HD1	1	1	0	0.03125	0.005	
HD2	0	0	0.0875	0	0.78125	0.9912
HD3	0	0	0.825	0	0.1875	0.00037
HD4	0	0	0.0875	0	0	0
	$m_2(\cdot)$					
HD1	1	1	0	0	1	0.1875
HD2	0	0	0.1375	0.1125	0	0.8125
HD3	0	0	0.665	0.8875	0	0
HD4	0	0	0.1975	0	0	0

Table V  
 $E^l$  BBAS AFTER THE STEP 1 OF PCR6-MCDA.

Criterion $g_j$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_6$
	$m_{Step1}(\cdot)$					
HD1	0.85	0.85	0	0.57656	0.35445	0.03820
HD2	0	0	0.0674	0.0198	0.41628	0.81046
HD3	0	0	0.69909	0.25364	0.07927	0.00131
HD4	0	0	0.08351	0	0	0
$\Theta$	0.15	0.15	0.15	0.15	0.15	0.15003

Table VI  
 $E^l$  BBAS AFTER IMPORTANCE DISCOUNTING.

Criterion $g_j$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_6$
	$m_{Step1}(\cdot)$ after importance discounting					
$\emptyset$	0.9	0.9	0.7	0.9	0.9	0.7
HD1	0.085	0.085	0	0.05765	0.03544	0.01146
HD2	0	0	0.02022	0.00198	0.04163	0.24314
HD3	0	0	0.20973	0.02537	0.00793	0.00039
HD4	0	0	0.02505	0	0	0
$\Theta$	0.015	0.015	0.045	0.015	0.015	0.04501

Table IV  
SHAHER'S DISCOUNTING OF INPUT MASSES WITH RELIABILITY FACTORS  
 $\alpha_{s1} = 0.7$  AND  $\alpha_{s2} = 0.5$ .

Criterion $g_j$	$g_1$	$g_2$	$g_3$	$g_4$	$g_5$	$g_6$
	$m'_1(\cdot)$					
HD1	0.7	0.7	0	0.7	0.021875	0.0035
HD2	0	0	0.06125	0	0.546875	0.69384
HD3	0	0	0.57750	0	0.13125	0.00259
HD4	0	0	0.06125	0	0	0
$\Theta$	0.3	0.3	0.3	0.3	0.3	0.3
	$m'_2(\cdot)$					
HD1	0.5	0.5	0	0	0.5	0.09375
HD2	0	0	0.06875	0.05625	0	0.40625
HD3	0	0	0.33250	0.44375	0	0
HD4	0	0	0.09875	0	0	0
$\Theta$	0.5	0.5	0.5	0.5	0.5	0.5

Table VII  
 $E^l$  BBAS AFTER THE STEP 2 OF PCR6-MCDA.

Criterion $g_j$	$m_{PCR6\emptyset}(\cdot)$	$m_{PCR6\emptyset}^{normalized}(\cdot)$
$\emptyset$	0.96601	0
HD1	0.00547	0.16094
HD2	0.01560	0.45901
HD3	0.01141	0.33559
HD4	0.00017	0.00494
$\Theta$	0.00134	0.03951

4) Results through Fuzzy COWA-ER: Step 1 makes it possible to get from Eq. (7) a normalized imprecise matrix  $E^{Fuzzy}[C]$  in Eq. (9).

$$E^{Fuzzy}[C] \approx \begin{bmatrix} [0.26; 0.46] \\ [0.35; 0.64] \\ [0.44; 0.74] \\ [0.65; 0.94] \\ [1.00; 1.00] \end{bmatrix}. \quad (9)$$

Table VIII  
BASIC BELIEF ASSIGNMENTS OF THE 5 ACTIONS.

Alternatives $D_i$	$m_i(D_i)$	$m_i(\bar{D}_i)$	$m_i(D_i \cup \bar{D}_i)$
$D_1$	0.22	0.54	0.24
$D_2$	0.30	0.36	0.34
$D_3$	0.38	0.26	0.36
$D_4$	0.56	0.06	0.38
$D_5$	0.86	0	0.14

Table IX  
FUSION OF THE 5 ELEMENTARY BBAS WITH PCR6.

Focal element	$m_{PCR6}(\cdot)$
$D_1$	0.02835
$D_2$	0.04805
$D_3$	0.07318
$D_4$	0.15185
$D_5$	0.39179
$D_1 \cup D_5$	0.00019
$D_2 \cup D_5$	0.0004
$D_3 \cup D_5$	0.00059
$D_4 \cup D_5$	0.00269
$D_1 \cup D_4 \cup D_5$	0.0012
$D_2 \cup D_3 \cup D_5$	0.00056
$D_2 \cup D_4 \cup D_5$	0.00254
$D_3 \cup D_4 \cup D_5$	0.00372
$D_1 \cup D_2 \cup D_5$	0.00018
$D_1 \cup D_3 \cup D_5$	0.00026
$D_1 \cup D_2 \cup D_3 \cup D_5$	0.00138
$D_1 \cup D_2 \cup D_4 \cup D_5$	0.02194
$D_1 \cup D_3 \cup D_4 \cup D_5$	0.04123
$D_2 \cup D_3 \cup D_4 \cup D_5$	0.09063
$D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5$	0.13927

Table X  
BEL, BETP, DSMP AND PL OF EFFECTIVENESS LEVELS OF ACTIONS ON  $D_i$  BASED ON COWA-ER.

$D_i$	$Bel(D_i)$	$BetP(D_i)$	$DSmP_{\epsilon=0}(D_i)$	$Pl(D_i)$
$D_1$	0.028	0.073	0.037	0.234
$D_2$	0.048	0.106	0.066	0.305
$D_3$	0.073	0.136	0.103	0.351
$D_4$	0.152	0.222	0.221	0.455
$D_5$	0.392	0.463	0.572	0.699

For the step 2, by using a 5  $\alpha$ -cut approach, we convert  $E^{Fuzzy}[C]$  into 2 bbas  $m_{P_{ess}}(\cdot)$  and  $m_{Opti}(\cdot)$ . Step 3 combines them with choice of the PCR6 fusion rule. Results are given in the Table XI.

The Table XII shows the (approximate) values of  $Bel(\cdot)$ ,  $BetP(\cdot)$ ,  $DSmP_{\epsilon=10^{-6}}(\cdot)$  and  $Pl(\cdot)$  based on  $m_{PCR6}(\cdot)$  values of Table XI. One sees that based on max of Bel, of BetP, of DSMP or of Pl, the best action is always  $D_5$  (similar decision as with COWA-ER).

Table XI  
THE 2 BBAS TO COMBINE AND THE RESULT OF PCR6 FUSION

Focal Element	$m_{P_{ess}}(\cdot)$	$m_{Opti}(\cdot)$	$m_{PCR6}(\cdot)$
$D_5$	0.35	0.06	0.3895
$D_4 \cup D_5$	0.21	0.20	0.2847
$D_3 \cup D_4 \cup D_5$	0.09	0.10	0.1033
$D_2 \cup D_3 \cup D_4 \cup D_5$	0.09	0.18	0.1051
$D_1 \cup D_2 \cup D_3 \cup D_4 \cup D_5$	0.26	0.46	0.1174

Table XII  
CREDIBILITY, BETP, DSMP AND PLAUSIBILITY OF EFFECTIVENESS LEVELS OF  $D_i$  BASED ON FCOWA-ER.

$D_i$	$Bel(D_i)$	$BetP(D_i)$	$DSmP(D_i)$	$Pl(D_i)$
$D_1$	0	0.023	0	0.117
$D_2$	0	0.050	0	0.222
$D_3$	0	0.084	0	0.326
$D_4$	0	0.227	0	0.611
$D_5$	0.389	0.616	1	1

### V. CONCLUSIONS AND PERSPECTIVES

In this paper, we have both formalized the decision problem and applied recent advanced MCDMs (ER-MCDA, COWA-ER, FCOWA-ER) to assess effectiveness of torrent protective check-dams in a context of imperfect information and more or less reliable sources. This application, based on expert knowledge, provides a class evaluation related to available knowledge. Others outranking methods such as the Soft-Electre Tri (SET) methodology [30] can also be applied to sort protection systems in predefined effectiveness classes. Defining uncertain states of nature and corresponding belief mass  $m(\cdot)$  remains challenging. Comparing belief functions theory with Bayesian probabilities or Choquet capacities in this actual context is a next step [31]. Effect on results of the fusion rules and order of combinations have also to be compared.

From an operational point of view, next steps will consist in DM and decision problem complete elicitation, criteria, importance, preferences on evaluation scale assessments. Afterwards, these methods will be combined in a global process taking into account all the system scales related to protection system devices.

### VI. ACKNOWLEDGEMENTS

This research has been partially funded by the French Ministry of Agriculture and Forestry (MAAF) and the French Ministry of Environment (MEDDE).

### REFERENCES

[1] S. Carladou, J.-M. Tacnet, N. Eckert, C. Curt, M. Batton-Hubert, *Vers une analyse intégrée de l'efficacité des ouvrages de protection contre les risques naturels en montagne: évaluation économique en complément des volets structurels et fonctionnels*, in C. Curt, L. Peyras, J. Baroth, and A. Chateaneuf (Eds.), 8me Journées Fiabilité, Matériaux et Structures - analyse de risques et fiabilité des systèmes dans leur environnement, Aix-en-Provence, France, pp. 61–74, April 2014.

- [2] M. Bründl, H. Romang, N. Bischof, C. Rheinberger, *The risk concept and its application in natural hazard risk management in Switzerland*, Natural Hazards and Earth System Science, Vol. 9–3, pp. 801–813, 2009
- [3] N. Eckert, C. Keylock, D. Bertrand, E. Parent, T. Faug, F. Favier, M. Naaim, *Quantitative risk and optimal design approaches in the snow avalanche field: Review and extensions*, Cold Regions Science and Technology, Vol. 79–80, pp. 1–19, 2012.
- [4] J.-M. Tacnet, *Prise en compte de l'incertitude dans l'expertise des risques naturels en montagne par analyse multicritères et fusion d'information*, Ph.D. Thesis, Ecole Nationale supérieure des Mines de Saint-Etienne, France, 2009.
- [5] L. Peyras, P. Royet, A. Salmi, M. Salembier, D. Boissier, *Etude de la sûreté de fonctionnement d'un aménagement hydraulique de génie civil*, Revue Européenne de Génie Civil, Vol. 10(5), pp. 615–631, 2006.
- [6] N. Ghariani, C. Curt, J.-M. Tacnet, *Analyse des dysfonctionnements des ouvrages de protection contre les crues torrentielles*, in C. Curt, L. Peyras, J. Baroth, and A. Chateaufort (Eds.), 8me Journées Fiabilité, Matériaux et Structures - analyse de risques et fiabilité des systèmes dans leur environnement, Aix-en-Provence, France, pp. 47–59.
- [7] C. Curt, L. Peyras, D. Boissier, *A Knowledge Formalization and Aggregation-Based Method for the Assessment of Dam Performance*, Computer-Aided Civil and Infrastructure Engineering, Vol. 25, pp. 171–184, 2010.
- [8] T. Saaty, *The analytic hierarchy process*, McGraw Hill, 1980.
- [9] B. Roy, *Méthodologie Multicritère d'Aide à la Décision*, Paris, France: Economica Collection Gestion – Série Production et techniques quantitatives appliquées à la gestion, 1985.
- [10] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [11] F. Smarandache, J. Dezert. *Advances and Applications of DSMT for Information Fusion – Collected Works*, Vol. 1–4. American Research Press, 2004, 2006, 2009, & 2015.
- [12] L.A. Zadeh, *Fuzzy sets*, Information and control, Vol. 8(3), pp. 338–353, 1965.
- [13] L.A. Zadeh, *Fuzzy sets as a basis for a theory of possibility*, Fuzzy sets and systems, Vol. 1, pp. 3–28, 1978.
- [14] J.-M. Tacnet, J. Dezert, *Cautious OWA and evidential reasoning for decision making under uncertainty*, in Proc. of 14th Int. Conf. on Information Fusion (Fusion 2011), pp. 2074–2081, Chicago, IL, USA, 5-8 July 2011.
- [15] D. Han, J. Dezert, J.-M. Tacnet, C. Han, *A fuzzy-cautious OWA approach with evidential reasoning*, in Proc. of 15th Int. Conf. on Information Fusion (Fusion 2012), pp. 278–285, Singapore, July 2012.
- [16] J.-M. Tacnet, M. Batton-Hubert, J. Dezert, *A two-step fusion process for multi-criteria decision applied to natural hazards in mountains*, in Proc. of Belief 2010, International Workshop on the Theory of Belief Functions, 1–2 April 2010.
- [17] AFNOR, *Norme française NF EN ISO 9000*, Management de la qualité, 2005.
- [18] M. Meunier, *Éléments d'hydraulique torrentielle*, No 1 Etude Montagne, CEMAGREF, Grenoble, France, 1991.
- [19] J. Von Neumann, O. Morgenstern, *Theory of Games and Economic Behaviour*, Princeton University Press, New Jersey, USA, 1953
- [20] P. Smets, *Decision making in the TBM: The necessity of the pignistic transformation*, Int. Journal of Approximate Reasoning, Vol. 38, 2005.
- [21] C. Baudrit, D. Guyonnet, D. Dubois, *Postprocessing the hybrid method for addressing uncertainty in risk assessments*, Environmental Engineering, Vol. 131(2), pp. 1750–1754, 2005.
- [22] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. Journal of Intelligent Systems, Vol. 29(3), pp. 223–252, 2014.
- [23] J. Dezert, J.-M. Tacnet, M. Batton-Hubert, F. Smarandache, *Multi-criteria decision making based on DSMT-AHP*, in Proc. of BELIEF 2010 Workshop on the Theory of Belief Functions, 2010.
- [24] J. Dezert, J.-M. Tacnet, *Evidential Reasoning for Multi-Criteria Analysis based on DSMT-AHP*, in Proc. of International Symposium on AHP (ISAHP 2011), Italy, June 2011.
- [25] R. Yager, *Decision making under Dempster-Shafer uncertainties*, Studies in Fuzziness and Soft Computing, Vol. 219, pp. 619–632, 2008.
- [26] M. Florea, A.-L. Jousselme, D. Grenier, E. Bossé, *Approximation techniques for the transformation of fuzzy sets into random sets*, Fuzzy Sets and Systems, Vol. 159, pp. 270–288, 2008.
- [27] J.-M. Tacnet, C. Curt, *Dispositifs de protection contre les risques torrentiels. Indicateurs d'analyse de l'efficacité des ouvrages (Rapport Final – Version Provisoire V 0.3)*, Technical Report, 62 pages, IRSTEA, Grenoble, France, March 2013.
- [28] B. Bouchon-Meunier, *La logique floue et ses applications*, Addison-Wesley, 257 pages, Oct. 1995.
- [29] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of Int. Conf. on Information Fusion (Fusion 2010), Seattle, USA, 2010.
- [30] J. Dezert, J.-M. Tacnet, *Soft ELECTRE TRI outranking method based on belief functions*, in Proc. of 15th Int. Conf. on Information Fusion (Fusion 2012), pp. 607–614, Singapore, 9–12 July 2012.
- [31] M. Cohen, J.-M. Tallon, *Décision dans le risque et l'incertain: l'apport des modèles non-additifs*, Revue d'économie politique, Vol. 110(5), pp. 631–681, 2000.

# The Impact of the Quality Assessment of Optimal Assignment for Data Association in a Multitarget Tracking Context

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Pavlina Konstantinova<sup>c</sup>

<sup>a</sup> The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>c</sup>European Polytechnical University, Pernik, Bulgaria.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, pavlina.konstantinova@gmail.com

Originally published as: J. Dezert, A. Tchamova, P. Konstantinova, *The Impact of the Quality Assessment of Optimal Assignment for Data Association in a Multitarget Tracking Context*, Cybernetics and Information Technologies, Vol. 15, No. 7, pp. 88–98, Dec. 2015, and reprinted with permission.

**Abstract**—The main purpose of this paper is to apply, and to test the performance of the new method, based on belief functions, proposed by Dezert & all to evaluate the quality of individual association pairings provided in the optimal data association solution for improving the performances of multisensor-multitarget tracking systems. The advantages of its implementation in an illustrative realistic surveillance context when some of association decisions are unreliable and doubtful and lead to potentially critical mistakes are discussed. A comparison with the results obtained on the base of Generalized Data Association is made.

**Keywords:** data association, belief functions, PCR6 fusion rule, multitarget tracking.

## I. INTRODUCTION

The problem of Data Association (DA) is a central in the modern multi-target tracking (MTT) systems' design [1,2]. It relates to the process of associating uncertain measurements (observations) to known tracks, and it is conditioned and motivated by the most important function of each surveillance system – to keep and to improve target tracks maintenance performance. In the monosensor context it corresponds to proper sensor observations partitioning (at a given scan) to the predicted states of the targets in such a way that their tracks' updates to be as precise, correct, and reasonable, as possible.

There are several approaches developed to resolve correlation ambiguities and to select the best observation-track pairings, based on different models. Some of them establish reward matrix based on Kinematic only Data Association (KDA) and on a probabilistic framework [3,4]. Some of them rely on Belief Functions (BF) [5-9] and motivate the incorporation of the advanced concepts for Generalized Data Association (GDA) [6-8], allowing the introduction of target attribute (target type, radar cross section, etc.) into the association logic, in order to improve track maintenance performance in complicated situations (closely spaced/crossing targets), when kinematics data are insufficient for coherent decision making. The main peculiarity consists in applying Dezert-Smarandache theory (DSMT) of plausible and paradoxical reasoning [8] to model and to process the utilized attribute data. In most

common case, when surveillance system provides kinematic only information (such as range, azimuth, elevation), obtained during a given scan, the most common way of dealing is to solve the optimal DA solution and to use all solutions (pairings) to update tracks, even if some of the pairings have poor quality. It could yields, in fact, to a bad/wrong track updating, and, as a result, the overall tracking performance could be degraded substantially.

The most recent method proposed by Dezert & Benameur [10] to evaluate the Quality Assessment of Data Association (QADA) encountered in multiple target tracking applications in a mono-criterion context, and recently extended in [11] for the multi-criteria context deal just with the case above. It assumes that the rewards matrix is known and has been obtained by a method chosen by the user. It is based on belief functions for establishing the quality of pairings (interpreted as a confidence score) belonging to the optimal data assignment solution based on its consistency (stability) with respect to all the second best solutions, provided by a chosen algorithm.

The main purpose of our paper is to serve as a preliminary study of MTT performance evaluation based on QADA-KDA approach, and to discuss its advantages in an illustrative multi-target tracking scenario. We will make also comparison between its performance and the results obtained on the base of GDA. The paper is organized as follows. Section II described the problem of DA in the multitarget tracking context. Section III provides the details about the new method [10] for quality assessment of optimal DA solution. In Section IV the simulation scenario and results are presented and discussed. The conclusion is given in Section V.

## II. DATA ASSOCIATION PROBLEM IN MULTITARGET TRACKING CONTEXT

Data Association is very important, and the most decisive step in the multitarget tracking surveillance process. The DA problem consists in finding the global optimal assignment of the targets  $T_i$  ( $i = 1, \dots, m$ ) to some measurements  $z_j$  ( $j = 1, \dots, n$ ) at a given time  $k$  by maximizing the overall gain



ignorance is modelled by the proposition  $X \cup Y$ . Then one obtains:

$$\begin{cases} m_1(X) = a_1(i, j)\omega(i, j)/R_1(\Omega, A_1), \\ m_1(X \cup Y) = 1 - m_1(X). \end{cases} \quad (5)$$

$$\begin{cases} m_2(Y) = a_2(i, j_2)\omega(i, j_2)/R_2(\Omega, A_2), \\ m_2(X \cup Y) = 1 - m_2(Y). \end{cases} \quad (6)$$

Different rules of combination (Dempster-Shafer's, Dubois-Prade's, Yager's [16]) could be chosen to work with a normalized combined BBA. The method [10] recommends to use the Proportional Conflict Redistribution rule no. 6 (PCR6), proposed originally in DSMT framework [8], because it has been proved very efficient in practice. With PCR6, the following fusion result  $m_{PCR6}(\cdot) = m_1(\cdot) \oplus m_2(\cdot)$  is obtained:

$$\begin{cases} m_{PCR6}(X) = m_1(X)m_2(X \cup Y) + m_1(X) \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)}, \\ m_{PCR6}(Y) = m_1(X \cup Y)m_2(Y) + m_2(Y) \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)}, \\ m_{PCR6}(X \cup Y) = m_1(X \cup Y)m_2(X \cup Y). \end{cases} \quad (7)$$

The decision is taken on the base of the pignistic transformation:

$$BetP(X) = m_{PCR6}(X) + \frac{1}{2}m_{PCR6}(X \cup Y), \quad (8)$$

$$BetP(Y) = m_{PCR6}(Y) + \frac{1}{2}m_{PCR6}(X \cup Y). \quad (9)$$

The quality indicators are chosen as:  $q(i, j) = BetP(X)$  and  $q(i, j_2) = BetP(Y)$ . The absolute quality factor becomes:

$$Q_{abs}(A_1, A_2) = \sum_{i=1}^m \sum_{j=1}^n a_1(i, j)q(i, j). \quad (10)$$

#### IV. SIMULATION SCENARIO AND RESULTS

The noise-free multitarget tracking simulation scenario (Fig.1) consists of three air targets moving in parallel from West to East with constant velocity of 100 m/sec and a distance between them 150 m. The stationary sensor is located at the origin. The sampling period is  $T_{scan} = 5$  sec, and the measurement standard deviations are 0.5 deg and 65m for azimuth and range respectively. The surveillance of moving targets is performed during 15 scans. Figure 2 shows the respective noised scenario.

The classical target tracking algorithm was run, consisting in two basic steps: (i) data association to associate proper measurements (distance, angle) with correct targets and (ii) track filtering to update the targets state vectors, once the optimal assignment was found. In our simulation the Converted Measurement Kalman Filter [1] is used.

In this work we will focus our attention on DA step, which is very important, and the most decisive one in the multitarget tracking. The Global Nearest Neighbour (GNN) [1] approach is used in order to make a decision for data association. One obtains the assignment matrix  $AMat(i, j)$ , ( $i = 1, \dots, m; j = 1, \dots, n$ ) based on normalized distances between measurement  $j$  and target  $i$ . In order to eliminate unlikely (kinematics-based) observation-to-track pairings, the classical validation

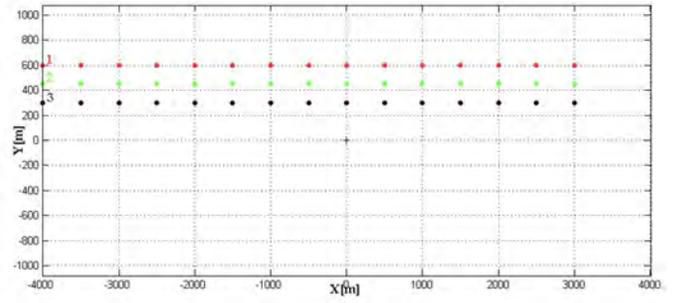


Fig. 1. Noise-free MTT scenario.

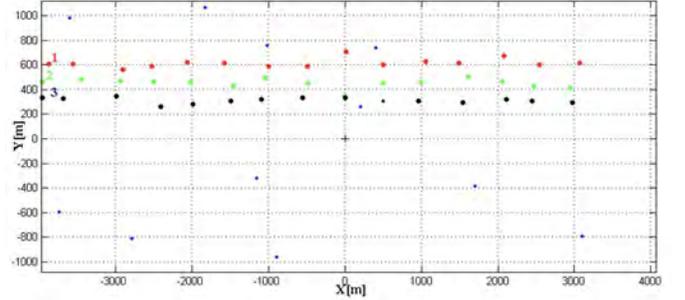


Fig. 2. Noised MTT scenario.

test  $d^2(i, j) \leq \gamma$  is carried on the Mahalanobis distance [1,2]  $d^2(i, j)$  computed from the measurement  $z_j(k)$  at a given time moment  $k$ , and its prediction  $\hat{z}_i(k|k-1)$  by

$$d^2(i, j) \triangleq (z_j(k) - \hat{z}_i(k|k-1))' S^{-1}(k) (z_j(k) - \hat{z}_i(k|k-1)). \quad (11)$$

Assuming given measurement vector's size  $M$ , the quantity  $d^2(i, j)$  could be interpreted as a sum of the squares of  $M$  independent Gaussian random variables with zero means and unit standard deviations. For that reason  $d^2(i, j)$  have  $\chi_M^2$  distribution with  $M$  degrees of freedom and allowable probability of a valid observation falling outside the gate. In our case a probability of 1% is approved, then from the table of the chi-square distribution [2] one obtains the threshold  $\gamma = 9.21$ . In fact, this value represents the biggest possible distance's value associated with observation-to-track pairings. Based on this, one assumes that if  $j$ -th measurement does not fall in the gate of target  $i$ , then the value, associated with this pairing ( $i, j$ ) in the assignment matrix could be set to be enough big (in our case equals to 100), in order to prepare the assignment matrix for the next step. The classical Munkres and Katta-Murty methods [15] are used in order to obtain the first and second best assignment solutions for measurement-to-track associations. By minimizing the sum of the chosen pairings' distances, a binary association matrix  $A = [a(i, j)]$  is obtained. Figure 3 shows the typical MTT performance, based on classical GNN approach with Kinematic only DA (KDA), when one does not utilize additional procedures to improve the quality of DA.

In case of noised measurements, it is evident that at scan number 9 tracks 2 and 3 change their directions, becoming

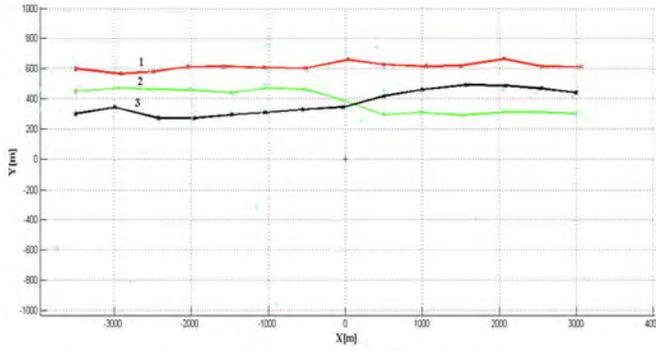


Fig. 3. Typical MTT performance with KDA.

crossing, instead of following their parallel moving behavior. It is because of incorrect determination of the incoming measurements in such a way, that they are too close and the solution of the assignment problem, which is the core of GNN, is impossible to be categorical. The problem consists also in the proximity of the targets (inter-distance of 150 m), and in the bad sensor distance resolution of  $\sigma_D = 65$  m. It leads to wrong GNN association decision.

A criterion for a minimal admissible measurements' distances is chosen here as  $d_{\min} < \sigma_D/2$ . During scan no. 9 one has:  $d_{23} < d_{\min}$ .

In such a critical case, one needs to utilize some additional information in order to avoid associations' miscorrelations. Here the method in [10] is applied. The goal is to estimate the quality of questionable pairings  $(T_2, z_2)$  and  $(T_3, z_3)$ . One obtains the corresponding reward matrix  $\Omega = [\omega(i, j)]$  with elements  $\omega(i, j)$  representing the gain of the associations of target  $T_i$  ( $i = 1, \dots, m$ ) with measurement  $z_j$  ( $j = 1, \dots, n$ ). It is achieved as:  $\omega(i, j) = 10 - AMat(i, j)$ . The reason for this expression relates to the already determined maximal normalized distance  $\gamma = 9.21$ , according to the table of chi-square distribution. The data association deals with finding the global optimal assignment of the targets to some measurements by maximizing the overall gain in such a way that no more than one target is assigned to a measurement, and reciprocally. This is an equivalent measure for optimality, as is the global minimum of the distances.

The algorithm, based on [10] was automatically applied during the scan no. 9, because the minimum distance between observations no. 2 and no. 3 is under the accepted limits  $d(2, 3) = 15.83\text{m} < \sigma_D/2$ . The quality matrix at scan no. 9, containing the quality levels associated with the chosen pairings in the first best solution

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

characterizing the set  $\{(T_1, z_1), (T_2, z_2), (T_3, z_3)\}$  of associa-

tions, with respect to the second best solution

$$A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

characterizing the set  $\{(T_1, z_1), (T_2, z_3), (T_3, z_2)\}$  of associations, is given in Table I below.

 TABLE I  
 QUALITY MATRIX AT SCAN 9.

Obs/Track	1	2	3
1	0.773	0.000	0.000
2	0.000	0.504	0.000
3	0.000	0.000	0.498

It is obvious that according to the first best assignment solution, one has:  $q(T_1, z_1) = 0.773$ ,  $q(T_2, z_2) = 0.504$ , and  $q(T_3, z_3) = 0.498$ . We accept the admissible for a correct association quality threshold to be set to  $q_T = 0.7$ .

Based on the associations quality assessment (Table I), and the accepted quality threshold  $q_T = 0.7$ , one could make the following decision: The only pairing, among those, chosen by Munkres algorithm in the first best assignment is  $(T_1, z_1)$ , because its quality level exceeds the accepted reasonable for correct association quality threshold  $q(T_1, z_1) = 0.773 > q_T$ . Following the decision logic in [10], only  $(T_1, z_1)$  pairing will be used in the updating process, while the second and third tracks will keep going under prediction mode while the next measurements will be available, because  $q(T_2, z_2) = 0.504 < q_T$ , and  $q(T_3, z_3) = 0.498 < q_T$ . The performance of the MTT algorithm, based on the QADA-KDA is shown on Fig.4. It is obvious that the reasoned/informed decision taken at scan no.9, based on QADA-KDA method leads to miscorrelation conflict resolution.

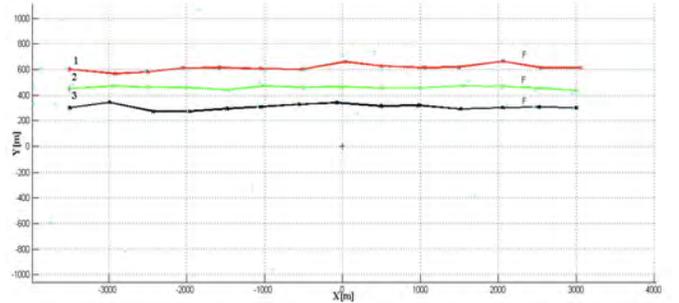


Fig. 4. MTT performance with QADA-KDA.

In order to compare the obtained by QADA result, the simulation was made by applying GDA (Fig.5), when the target attribute (target type) is introduced into the association logic, in order to improve track maintenance performance in the same MTT scenario, with an additional assumption, that targets go from West to East in a group with the following type order {Fighter, Cargo, Fighter}.

GDA-MTT improves the process of DA by utilizing target's type decision (based on confusion matrix  $C = [c_{ij}]$ ) coupled

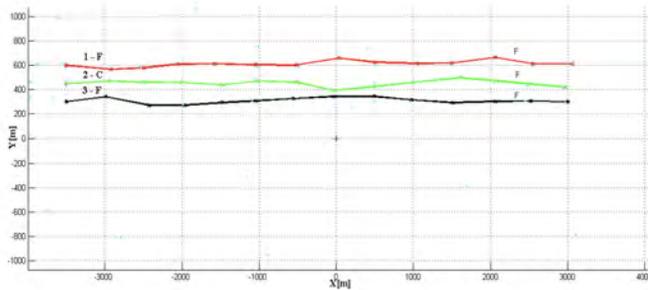


Fig. 5. MTT performance with GDA.

with the classical kinematic measurements. The way of constructing the confusion matrix is based on some underlying decision-making process based on specific attribute features measurements. Its elements represent the probability of decisions  $T_d$  ( $T_1 = \text{Fighter}$ ,  $T_2 = \text{Cargo}$ ) that the target type is  $j$  when its real type is  $i$ , more precisely

$$c_{ij} = P(T_d = T_j | \text{True Target Type} = i).$$

In our simulation we have chosen the following confusion matrix

$$C = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}.$$

GDA is applied at each scan during the whole surveillance process, in order to prevent observation-to-track miscorrelations.

## V. CONCLUSION

This work is a preliminary study of MTT performance evaluation based on the new Quality Assessment of optimal DA method, proposed by Dezert & al. It assures the stability of MTT performance and could be applied in all cases relating to the impossibility of DA to produce an association decision with a high quality. It might concern cases, when only kinematics measurements are available, as well the cases when attribute and kinematic data are both available, because QADA is totally independent of the applied logic to obtain the best DA solution. The work's perspective concerns Monte Carlo based evaluation of different, more critical MTT scenarios in a multi-sensor context.

## REFERENCES

- [1] Y. Bar-Shalom (Ed.), *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, USA, 1990.
- [2] Y. Bar-Shalom, T.E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.
- [3] Y. Bar Shalom, P.K. Willet, X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, Storrs, CT, USA, 2011.
- [4] X. He, R. Tharmarasa, M. Pelletier, T. Kirubarajan, *Accurate Murty's Algorithm for Multitarget Top Hypothesis*, in Proc. of Int. Conf. of Information Fusion (Fusion 2011, Chicago, USA, 2011).
- [5] J. Dezert, F. Smarandache, A. Tchamova, *On the Blackman's Association Problem*, in Proc. of Int. Conf. of Information Fusion (Fusion 2003), Cairn, Australia, 2003.
- [6] A. Tchamova, J. Dezert, T. Semerdjiev, P. Konstantinova, *Target tracking with Generalized Data Association based on the General DSMT Rule of Combination*, in Proc. of Int. Conf. of Information Fusion (Fusion 2004), Stockholm, Sweden, 2004.

- [7] J. Dezert, A. Tchamova, T. Semerdjiev, P. Konstantinova, *Performance Evaluation of Fusion rules for Multitarget Tracking in Clutter based on Generalized Data Association*, in Proc. of Int. Conf. of Information Fusion (Fusion 2005), Philadelphia, USA, 2005.
- [8] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, American Research Press, Vol. 1, 2, 3, 2004, 2006, 2009. <https://www.onera.fr/fr/staff/jean-dezert/references>
- [9] T. Dencoux, N. El Zoghby, V. Cherfaoui, A. Jouglet, *Optimal Object Association in the Dempster-Shafer Framework*, IEEE Trans. on Cybernetics, Vol. 44(12), pp. 2521–2531, Dec. 2014.
- [10] J. Dezert, K. Benameur, *On the Quality of Optimal Assignment for Data Association*, Springer, Lecture Notes in Computer Science, Vol.8764, pp. 374–382, 2014.
- [11] J. Dezert, K. Benameur, L. Ratton, J.-F. Grandin, *On the Quality Estimation of Optimal Multiple Criteria Data Association Solutions*, in Proc. of Int. Conf. of Information Fusion (Fusion 2015), Washington D.C, USA, July 6–9, 2015.
- [12] H.W. Kuhn, *The Hungarian Method for the Assignment Problem*, Naval Research Logistic Quarterly, Vol. 2, pp. 83–97, 1955.
- [13] J. Munkres, *Algorithms for the Assignment and Transportation Problems*, Journal of the Society of Industrial and Applied Mathematics, Vol.5 (1), pp. 32–38, 1957.
- [14] F. Bourgeois, J.C. Lassalle, *An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices*, Comm. of the ACM, Vol.14 (12), pp.802–804, 1971.
- [15] K.G. Murty, *An Algorithm for Ranking all the Assignments in Order of Increasing Cost*, Operations Research Journal, Vol. 16 (3), pp. 682–687, 1968.
- [16] P. Smets, R. Kennes, *The Transferable Belief Model*, Artificial Intelligence, Vol. 66(2), pp. 191–234, 1994.



# Multitarget Tracking Performance based on the Quality Assessment of Data Association

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Pavlina Konstantinova<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>c</sup>European Polytechnical University, Pernik, Bulgaria.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, pavlina.konstantinova@gmail.com

Originally published as: J. Dezert, A. Tchamova, P. Konstantinova, *Multitarget Tracking Performance based on the Quality Assessment of Data Association*, in Proc. of 19th Int. Conf. on Information Fusion, Heidelberg (Fusion2019), Germany, July 5–8, 2016, and reprinted with permission.

**Abstract**—The main objective of this paper is to present, to apply, and to test the effectiveness of the new method, based on belief functions, proposed by Dezert et al. in order to evaluate the quality of the individual association pairings provided in the classical optimal data association solution for improving the performances of multitarget tracking systems in clutter, when some of the association decisions given in the optimal assignment solution are unreliable and doubtful and lead to potentially critical mistake. This evaluation is based on a Monte Carlo simulation for particular difficult maneuvering and non-maneuvering MTT problems in clutter. A comparison with the results obtained on the base of Kinematic only Data Association and Generalized Data Association is made.

**Keywords:** Data association, Belief Functions, PCR6 fusion rule, multitarget tracking.

## I. INTRODUCTION

Data association (DA) is a fundamental and central problem in up-to-date multitarget tracking (MTT) systems [1]-[2]. It entails selecting the most trustable associations between uncertain sensor's measurements and existing targets at a given time. In the presence of dense MTT environment, with false alarms and sensors detection probability less than unity, the problem of DA becomes more complex, because it should contend with many possibilities of pairings, some of which are in practice very doubtful, unreliable, and could lead to critical association mistakes in overall tracking process. To avoid such cases, sometimes it is better to wait for a new measurements during the next scan, instead of taking a hard DA decision, which actually is not always unique.

Several methods have been devised over the years, in order to resolve properly DA problem. They are originating from different models. Some rely on the established reward matrix based on Kinematic only Data Association (KDA) and on a probabilistic framework [3]-[4]. Some other studies are based on Belief Functions (BF) [5]-[9], motivating the incorporation of the advanced concepts for Generalized Data Association (GDA) [6]-[8], where a particular target's attribute is introduced into the association logic in order to compensate the complicated cluttered cases, when kinematics data are insufficient for adequate decision making. Dezert-Smarandache Theory (DSMT) of plausible and paradoxical reasoning [8] is used to model and to process the utilized

attribute data. Although interesting and approved, all these methods currently developed are limited to the following aspect - all of them solve the optimal DA problem and use all optimal observations-to-tracks pairings, selected in the first best DA solution to update tracks, even if some of them have poor quality. In consequence the overall tracking performance could be degraded substantially. In order to deal with this case the most recent method to evaluate the Quality Assessment of Data Association (QADA) encountered in multiple target tracking applications in a mono-criterion context is proposed by Dezert and Benameur [10]. It is extended in [11] for the multi-criteria context. This novel method assumes the reward matrix is known, regardless of the manner in which it is obtained by the user. It is based on BF for achieving the quality of pairings (interpreted as a confidence score) belonging to the optimal data assignment solution based on its consistency (stability) with respect to all the second best solutions, provided by a chosen algorithm.

This paper is an extension of our preliminary study on the effect of applying QADA method in MTT presented in [17]. The main purpose of our paper is to assess the efficiency of QADA method in a critical, conflicting MTT situation. The evaluation is based on a Monte Carlo simulation for particular difficult maneuvering and non-maneuvering MTT problems in clutter. The QADA based MTT performance is compared with the results, obtained for KDA and GDA based MTT, concerning the same scenarios. The paper is organised as follows. In order to achieve a good readability of the paper, we recall in section II the data association problem within the MTT context, and in a section III the details of the new method, proposed by Dezert et al. [10] for quality assessment of pairings, chosen in the optimal DA solution. In section IV we discuss and propose the way in which Kalman filtering could be affected in order to reflect the knowledge we have obtained on the base of QADA method. Two simulation MTT scenarios (with non-maneuvering and maneuvering targets) are presented and the results, obtained on the base of QADA-, KDA-, and GDA based MTT are discussed. Conclusions are made in Section VI.

## II. DATA ASSOCIATION PROBLEM IN MTT CONTEXT

The DA problem consists in finding the global optimal assignments of targets  $T_i, i = 1, \dots, m$  to some measurements  $\mathbf{z}_j, j = 1, \dots, n$  at a given time  $k$  by maximizing the overall gain in such a way, that no more than one target is assigned to a measurement, and reciprocally.

The  $m \times n$  reward (gain/painoff) matrix  $\Omega = [\omega(i, j)]$  is defined by its elements  $\omega(i, j) > 0$ , representing the gain of the association of target  $T_i$  with the measurement  $\mathbf{z}_j$ . These values are usually homogeneous to the likelihood ratios. In our case  $\omega(i, j)$  represents the normalized distances between the measurement  $Z_j$  and target  $T_i$ :  $d^2(i, j) \triangleq (\mathbf{z}_j(k) - \hat{\mathbf{z}}_i(k|k-1))' \mathbf{S}^{-1}(k) (\mathbf{z}_j(k) - \hat{\mathbf{z}}_i(k|k-1)) \leq \gamma$  computed from the measurement  $\mathbf{z}_j(k)$  and its prediction  $\hat{\mathbf{z}}_i(k|k-1)$  computed by the tracker of target  $i$  (see [2] for details), and the inverse of the covariance matrix  $\mathbf{S}(k)$  of the innovation computed by the tracking filter. In this case the DA problem consists in finding the best assignment, minimizing the overall cost.

The optimal DA problem consists in finding the  $m \times n$  binary association matrix  $\mathbf{A} = [a(i, j)]$  with  $a(i, j) \in \{0, 1\}$ , maximizing the global reward  $R(\Omega, \mathbf{A})$ , given by:

$$R(\Omega, \mathbf{A}) \triangleq \sum_{i=1}^m \sum_{j=1}^n \omega(i, j) a(i, j). \quad (1)$$

If  $a(i, j) = 1$ , it means that one has an association between target  $T_i$  and measurement  $\mathbf{z}_j$ . The association indicator value  $a(i, j) = 0$  means that they are not associated.

$$a(i, j) = \begin{cases} 1, & \text{if } \mathbf{z}_j \text{ is associated to track } T_i, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The importance of the assignment problem is quite clear and various successful solutions to its solving already exist. Among the well known are Kuhn-Munkres algorithm (known as Hungarian) [12]-[13], and its extension proposed by Bourgeois and Lassalle in [14] to rectangular matrices. More sophisticated Murty's method [15] provides not only the first best assignment, but also the  $m$ -best assignments in order of increasing cost, as it was shown in the examples in [10]-[11]. The best optimal assignment solution is not necessarily unique, as well as the second best one. Usually in MTT algorithms the first best assignment solution is taken as a hard decision for association. But in some real practical cases of dense multi-target and cluttered environment, DA problem is difficult to solve, because some of the associations decisions  $a(i, j)$  are unreliable, so they could lead to potential mistakes.

For example, in case of incorrect determination of the incoming measurements for two tracks in such a way, that they are too close, the solution of the assignment problem, that is the core of the Global Nearest Neighbour (GNN) approach, is impossible to be sufficiently explicit. In such a case, it will be more cautious not to rely on all the pairings confirmed in the first best solution, no matter than only some of them are trustable enough. Utilizing the already obtained and available  $m$ -best assignments solutions, Dezert et al. [10], [11] provided an appealing method for taking into account this knowledge.

## III. QUALITY ASSESSMENT OF PAIRINGS IN DA

In order to establish the quality of particular associations, associated with the optimal assignment matrix  $\mathbf{A}_1$ , and satisfying the condition  $a_1(i, j) = 1$ , QADA method proposes to utilize both, first and second assignment solutions  $\mathbf{A}_1$  and  $\mathbf{A}_2$ . For a self-containing purpose, this section recalls briefly the principle of QADA that has been already detailed in [10], [11] with a tracking application in [17].

The main idea behind it is to compare the values  $a_1(i, j)$  in  $\mathbf{A}_1$  with the corresponding values  $a_2(i, j)$  in  $\mathbf{A}_2$ , and to identify if there is a change of the optimal pairing  $(i, j)$ . In our MTT context  $(i, j)$  means an association between measurement  $\mathbf{z}_j$  and target  $T_i$ . One establishes a quality indicator associated with this pairing, depending on the stability of the pairing and also, on its relative impact in the global reward. The proposed method works also when the 1<sup>st</sup> and 2<sup>nd</sup> optimal assignments  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are not unique, i.e., there are multiplicities available. The construction of the quality indicator is based on BF theory and Proportional Conflict Redistribution Rule no.6 (PCR6), defined within DSMT [8]. It depends on the type of the pairing matching, as it is described below:

- If  $a_1(i, j) = a_2(i, j) = 0$ , one has a full agreement on the hypothesis 'non-association' of the given pairing  $(T_i, \mathbf{z}_j)$  in  $\mathbf{A}_1$  and  $\mathbf{A}_2$ . This 'non-association' has no impact on the global reward values  $R_1(\Omega, \mathbf{A}_1)$  and  $R_2(\Omega, \mathbf{A}_2)$ , therefore it will be useless to utilize it in DA. Hence, in this case, the quality indicator will be set to zero,  $q(i, j) = 0$ .
- If  $a_1(i, j) = a_2(i, j) = 1$ , one has a full agreement on the hypothesis 'association' of the pairing  $(T_i, \mathbf{z}_j)$  in  $\mathbf{A}_1$  and  $\mathbf{A}_2$ . This 'association'  $(T_i, \mathbf{z}_j)$  has different impacts on the global reward values  $R_1(\Omega, \mathbf{A}_1)$  and  $R_2(\Omega, \mathbf{A}_2)$ . In order to estimate the quality of this matching pairing, one establishes two Basic Belief Assignments (BBAs),  $m_s(\cdot), s = 1, 2$ , according to both sources of information (1<sup>st</sup> and 2<sup>nd</sup> optimal assignments matrices  $\mathbf{A}_1$  and  $\mathbf{A}_2$ ). The frame of discernment consists of a single hypothesis  $X = (T_i, \mathbf{z}_j)$ : measurement  $\mathbf{z}_j$  belongs to the track  $T_i$ . The ignorance is modelled by the proposition  $X \cup \bar{X}$ , where  $\bar{X}$  is the negation of hypothesis  $X$ :

$$\begin{cases} m_s(X) = a_1(i, j) \cdot \omega(i, j) / R_1(\Omega, \mathbf{A}_1), \\ m_s(X \cup \bar{X}) = 1 - m_s(X). \end{cases} \quad (3)$$

Applying the conjunctive rule of combination [8] (Vol. 1), one gets:

$$\begin{cases} m_{12}(X) = m_1(X)m_2(X) + m_1(X)m_2(X \cup \bar{X}) \\ \quad + m_1(X \cup \bar{X})m_2(X), \\ m_{12}(X \cup \bar{X}) = m_1(X \cup \bar{X})m_2(X \cup \bar{X}). \end{cases} \quad (4)$$

The pignistic transformation [16] is applied in order to obtain pignistic probabilities, built on the base of combined belief assignments, such as:  $BetP(X) = m_{12}(X) + \frac{1}{2} \cdot m_{12}(X \cup \bar{X})$  and  $BetP(\bar{X}) = \frac{1}{2} \cdot m_{12}(X \cup \bar{X})$

$\bar{X}$ ). Then one chooses the quality indicator, associated with the pairing  $(i, j)$ , as  $q(i, j) = \text{BetP}(X)$ .

- If  $a_1(i, j) = 1$  and  $a_2(i, j) = 0$ , then a conflict is encountered on the association  $(T_i, \mathbf{z}_j)$  in  $\mathbf{A}_1$  and  $\mathbf{A}_2$ . Then one could find the association  $(T_i, \mathbf{z}_{j_2})$  in  $\mathbf{A}_2$ , where  $j_2$  is the index, such that  $a_2(i, j_2) = 1$ . In order to define the quality of such conflicting association, one establishes two BBAs,  $m_s(\cdot)$ ,  $s = 1, 2$  according to both sources of information ( $\mathbf{A}_1$  and  $\mathbf{A}_2$ ). The frame of discernment consists of two propositions:  $\Theta = \{X = (T_i, \mathbf{z}_j), Y = (T_i, \mathbf{z}_{j_2})\}$ , and the BBAs are defined by [10].

$$\begin{cases} m_1(X) = a_1(i, j) \cdot \frac{\omega(i, j)}{R_1(\Omega, \mathbf{A}_1)}, \\ m_1(X \cup Y) = 1 - m_1(X), \end{cases} \quad (5)$$

$$\begin{cases} m_2(Y) = a_2(i, j_2) \cdot \frac{\omega(i, j_2)}{R_2(\Omega, \mathbf{A}_2)}, \\ m_2(X \cup Y) = 1 - m_2(Y). \end{cases} \quad (6)$$

Applying PCR6 fusion rule [8] (Vol. 3), one gets:

$$\begin{cases} m(X) = m_1(X) \cdot m_2(X \cup Y) + m_1(X) \cdot \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)}, \\ m(Y) = m_1(X \cup Y) \cdot m_2(Y) + m_2(Y) \cdot \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)}, \\ m(X \cup Y) = m_1(X \cup Y)m_2(X \cup Y). \end{cases} \quad (7)$$

Applying again the pignistic transformation, one gets  $\text{BetP}(X) = m(X) + \frac{1}{2} \cdot m(X \cup Y)$  and  $\text{BetP}(Y) = m(Y) + \frac{1}{2} \cdot m(X \cup Y)$ . Hence, the quality indicators here are chosen as:  $q(i, j) = \text{BetP}(X)$  and  $q(i, j_2) = \text{BetP}(Y)$ . The absolute quality factor becomes:  $Q_{abs}(\mathbf{A}, \mathbf{A}_2) = \sum_{i=1}^m \sum_{j=1}^n .a(i, j) \cdot q(i, j)$ .

Once obtained, this quality matrix  $\mathbf{Q} = [q(i, j)]$ ,  $i = 1, \dots, m$ ;  $j = 1, \dots, n$ , where the elements  $q(i, j) \in [0, 1]$  define the quality of particular associations, chosen in the optimal assignment matrix  $\mathbf{A}_1$ . It will be utilized in the next step of the classical MTT algorithm - Kalman filtering (KF).

#### IV. KALMAN FILTERING INFLUENCED BY QADA METHOD

The classical target tracking algorithm was run, consisting of two basic steps: (i) data association to associate the proper measurements (distance, angle) with correct targets and (ii) track filtering to update the targets state vectors, once the optimal assignment is found. In our simulation the Global Nearest Neighbour (GNN) [1] approach is applied in order to make a decision for data associations. GNN approach is a DA method that provides an assignment matrix for quality assessment of data association.

The Converted Measurement Kalman Filter (CMKF) is used for track filtering. We will not recall it in details, which can be found in many standard textbooks [1]-[2], but will make an impact on the manner, in which the obtained quality assessment of pairings in the optimal assignment solution influences the target's state updating.

In order to derive KF equations, the goal is to find an equation computing an a posteriori state estimate  $\hat{\mathbf{x}}(k+1|k+1)$  at time  $(k+1)$  as a linear combination of an a priori estimate

$\hat{\mathbf{x}}(k+1|k)$ , and a weighted difference between the true measurement  $\mathbf{z}(k+1)$  and a measurement prediction:

$$\hat{\mathbf{x}}(k+1|k+1) = \hat{\mathbf{x}}(k+1|k) + \mathbf{W}(k+1)\tilde{\mathbf{z}}(k+1) \quad (8)$$

The difference  $\tilde{\mathbf{z}}(k+1) \triangleq \mathbf{z}(k+1) - \mathbf{H}\hat{\mathbf{x}}(k+1|k)$ , called a measurement innovation (or residual), reflects the discrepancy between the predicted measurement  $\hat{\mathbf{z}}(k+1|k) = \mathbf{H}(k+1)\hat{\mathbf{x}}(k+1|k)$  and the true one  $\mathbf{z}(k+1)$ , where  $\mathbf{H}(k+1)$  is the so-called observation matrix. If  $\tilde{\mathbf{z}}(k+1)$  is equal to zero, it means, that both, the true measurement and predicted one are in full agreement, which is the perfect case. The matrix  $\mathbf{W}(k+1)$  is the filter's gain matrix obtained by minimizing the a posteriori estimate error covariance. It is given by the following formulae, where  $\mathbf{R}$  is the measurement error covariance, and  $\mathbf{P}(k+1|k)$  is the predicted covariance matrix of the state estimate error:

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{P}(k+1|k)\mathbf{H}^T(k+1)\mathbf{S}^{-1}(k+1) \\ &= \mathbf{P}(k+1|k)\mathbf{H}^T(k+1) \end{aligned} \quad (9)$$

$$\cdot [\mathbf{H}(k+1)\mathbf{P}(k+1|k)\mathbf{H}^T(k+1) + \mathbf{R}]^{-1}. \quad (10)$$

From Eqs. (8) and (10) one could conclude, that the value of measurement error covariance  $\mathbf{R}$  influences the gain's value  $\mathbf{W}(k+1)$ , and respectively the state estimate in the way below:

- If the measurement error covariance  $\mathbf{R} \rightarrow \mathbf{0}$ , the true measurement  $\mathbf{z}(k+1)$  is trusted more, and in the same time predicted measurement  $\mathbf{H}\hat{\mathbf{x}}(k+1|k)$  is trusted less.
- If the measurement error covariance  $\mathbf{R}$  increases, the true measurement  $\mathbf{z}(k+1)$  is trusted less, and in the same time predicted measurement  $\mathbf{H}\hat{\mathbf{x}}(k+1|k)$  is trusted more.

Let's now recall again what kind of information one obtains, having in hand the quality matrix, derived by QADA method [10]. It gives us a knowledge about the confidence  $q(i, j)$  in all pairings  $(T_i, \mathbf{z}_j)$ ,  $i = 1, \dots, m$ ;  $j = 1, \dots, n$ , chosen in the first best assignment solution. The smaller quality (confidence) of hypothesis " $\mathbf{z}_j$  belongs to  $T_i$ " means, that the particular measurement error covariance  $\mathbf{R}$  was increased and one should not trust fully in the actual (true) measurement  $\mathbf{z}(k+1)$ .

Having this conclusion in mind, in this work we propose, such a behaviour of the measurement error covariance to be modelled by  $\mathbf{R} = \frac{\mathbf{R}}{q(T_i, \mathbf{z}_j)}$ , for every pairing, chosen in the first best assignment and on the base of corresponding quality value obtained. Then, Kalman filter gain decreases, and as a result, the true measurement  $\mathbf{z}_j(k+1)$  is trusted less in the updated state estimate  $\hat{\mathbf{x}}(k+1|k+1)$ .

The MTT algorithm tested in this paper is based on the classical one (using Kalman Filters based on kinematics measurements) because we are only concerned with impact QADA on the performances of such type of tracking filters for now. Our aim is not to compare this QADA-MTT to other more sophisticate MTT algorithms<sup>1</sup>, but we believe that QADA approach could also be useful for improving performances of more sophisticate MTT algorithms as well. This is left for future research works.

<sup>1</sup>In fact, we will just compare QADA-MTT to KDA-MTT and GDA-MTT based on CMKF in Section V.

## V. SIMULATION SCENARIOS AND RESULTS

Two simulation MTT scenarios - non-maneuvering and maneuvering are presented and the results, obtained on the base of QADA-, KDA-, and GDA based MTT are discussed.

## A. Maneuvering targets simulation scenario

The simulation scenario (Fig. 1) consists of three air targets with two classes. The stationary sensor is located at the origin. The sampling period is  $T_{scan} = 5sec$  and the measurement standard deviations are 0.4 deg and 25m for azimuth and range respectively. The targets go from West to East with the following type order CFC (C=Cargo, F=Fighter) with constant velocity  $100m/sec$ . At the beginning the targets move from different directions. The first target moves from North-West with heading 120 degrees from North. At  $scan\ no. = 8$  the target performs a maneuver until  $scan\ no. = 15$  with transversal acceleration  $+1.495m/s^2$  and settles towards East, moving in parallel according to X axis. The second target moves during the whole scenario in parallel according to X from West to East without maneuvering. The third target at the beginning moves from South-West with heading 60 degrees from North. At  $scan\ no. = 8$  the target performs a maneuver until  $scan\ no. = 15$  with transversal acceleration  $-1.495m/s^2$  and settles towards East, moving in parallel according to X axis. The inter-distance between the targets during scans 15th - 18th (the parallel segment) is approximately 150m. At  $scan\ no. = 18$  to  $scan\ no. = 25$  the first and the third targets make new maneuvers. The first one is directed to North-East and the second - to South-East. The process noise standard deviations for the two nested models for constant velocity IMM (Interacting Multiple Models) filter [1], [3] are  $0.1m/s^2$  and  $7m/s^2$  respectively. The number of false alarms (FA) follows a Poisson distribution and FA are uniformly distributed in the surveillance region.

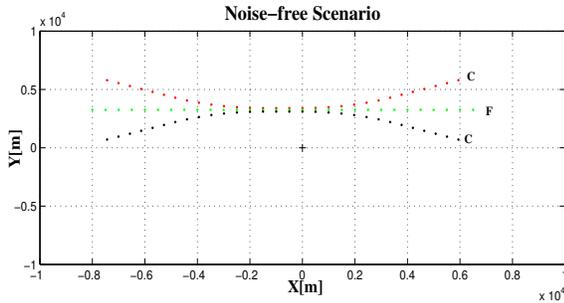


Figure 1. Noise-free maneuvering MTT Scenario.

Fig. 2 shows the respective noised scenario.

GDA-MTT [6], [7] improves DA process by utilizing target's type decision based on the confusion matrix  $\mathbf{C} = [C_{ij}]$  coupled with the classical kinematic measurements, where  $C_{ij} = P(T_d = T_j / TrueTargetType = T_i)$  represents the probability of decisions  $T_d = (T_1 \triangleq Fighter, T_2 \triangleq Cargo)$ , that the target type is  $j$  when its real type is  $i$ . In our simulation  $\mathbf{C} = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$ .

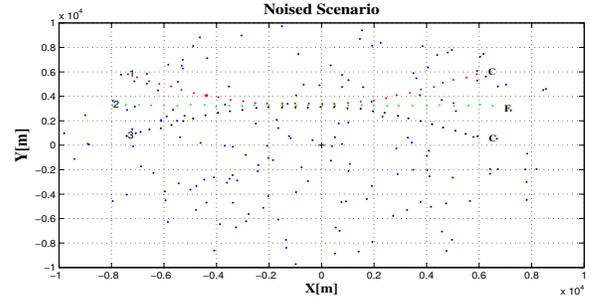


Figure 2. Noised maneuvering MTT Scenario.

Monte Carlo (MC) simulations for the considered MTT scenario are made for 200 MC runs, applying KDA, QADA, and GDA. Our goal is to evaluate, show, and to discuss the effect of Quality Assessment of Optimal Assignment for Data Association on the overall target tracking performance in comparison to results, obtained for the same scenario, by Kinematic only Data Association, and Generalized Data Association based MTT. We use an idealized track initiation in order to prevent uncontrolled impact of this stage on the statistical parameters of the tracking process during Monte Carlo tests of the new developed algorithm. The true targets positions (known in our simulations) for the first two scans are used for tracks initiation.

The evaluation of MTT performance is based on the criteria of tracks' purity, tracks' life, and percentage of miscorrelation. Track's purity criteria examines the ratio between the number of particular performed ( $j$ th observation -  $i$ th track) associations (in case of detected target) over the total number of all possible associations during the tracking scenario. Track's life is evaluated as an average number of scans before track's deletion. In our simulations, a track is cancelled and deleted from the list of tracked tracks, when during 3 consecutive scans it cannot be updated with some measurement because there is no validated measurement in the validation gate. We call this, the "cancelling/deletion condition". The status of the tracked tracks is denoted "alive".

The percentage of miscorrelation examines the relative number of incorrect (observation-to-track) associations during the scans.

The results for less noised case (with 0.2 FA in average in the filter validation gate) are given in Table 1.

Table I  
MANEUVERING SCENARIO: COMPARISON BETWEEN KDA, QADA, GDA BASED MTT PERFORMANCES FOR  $FA = 0.2$ .

	KDA-MTT	QADA-MTT	GDA-MTT
Average Track Life [%]	86.65	<b>92.82</b>	91.06
Average Miscorrelation [%]	7.27	<b>3.69</b>	3.06
Track Purity [%]	77.44	<b>88.20</b>	85.74

QADA-MTT exceeds KDA-MTT according to average track life and track purity, and shows better performance concerning the encountered average track life in comparison to GDA-MTT. Figure 3 shows the most informative knowledge - a

percentage of miscorrelations, encountered during the consecutive scans. One could see, that QADA-MTT shows almost two times better performance in comparison to KDA-MTT, and is close to GDA-MTT performance.

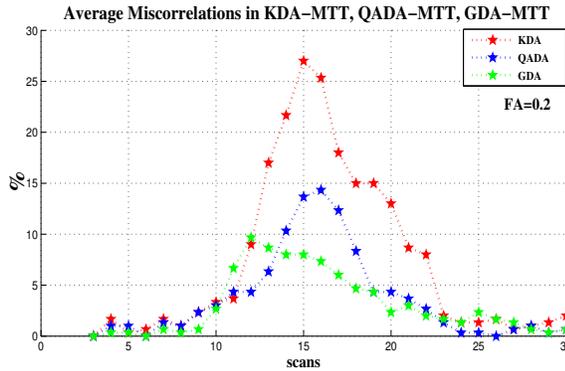


Figure 3. Maneuvering scenario: Average miscorrelations in KDA-MTT, QADA-MTT, GDA-MTT for noised case  $FA = 0.2$

The respective results for the most noised case (with 0.4 FA in average in the filter validation gate) are given in Table 2 below.

Table II  
MANEUVERING SCENARIO: COMPARISON BETWEEN KDA, QADA, GDA BASED MTT PERFORMANCES FOR  $FA = 0.4$

	KDA-MTT	QADA-MTT	GDA-MTT
Average Track Life [%]	74.27	<b>86.61</b>	86.52
Average Miscorrelation [%]	10.58	<b>7.05</b>	4.68
Track Purity [%]	60.42	<b>77.96</b>	79.35

As a whole, the results for  $FA = 0.4$  are deteriorated in comparison to the less noised case, but still QADA-MTT shows stably better performance with respect to KDA-MTT performance. The average track life keeps a little bit higher than in GDA-MTT case.

The Fig.4, showing the percentage of miscorrelations in more difficult noised case, confirms that QADA-MTT overcomes KDA-MTT performance.

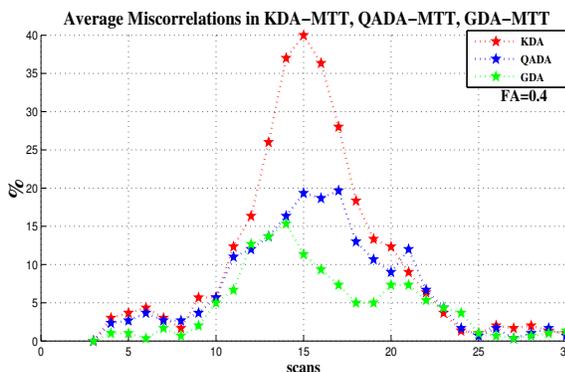


Figure 4. Maneuvering scenario: Average miscorrelations in KDA-MTT, QADA-MTT, GDA-MTT for noised case  $FA = 0.4$

The figures 5 and 6 show typical performances of QADA-MTT and KDA-MTT systems.

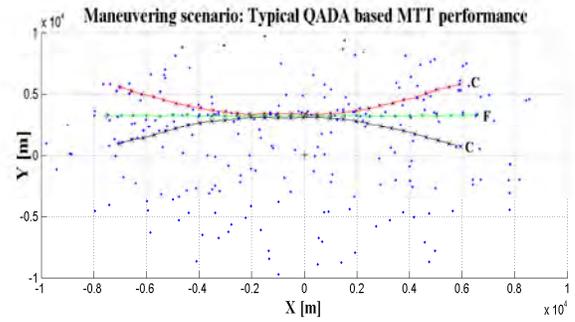


Figure 5. Maneuvering scenario: Typical performance of QADA based MTT.

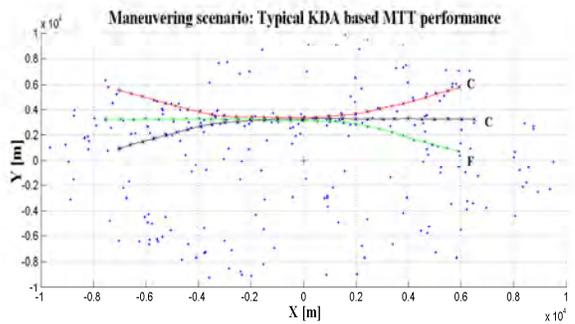


Figure 6. Maneuvering scenario: Typical performance of KDA based MTT.

The figures 7 and 8 show the averaged filtered errors along X (designated by asterisk) and Y (designated by circles) axes, and the distance error associated with the maneuvering track 1 in the considered scenario.

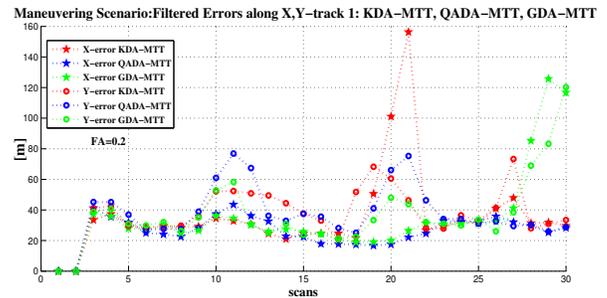


Figure 7. Filtered errors along X,Y for maneuvering track 1 - KDA-MTT, QADA-MTT, GDA-MTT.

For the maneuvering target 1, the errors, along X axis, obtained by using QADA-MTT, are definitely smaller than those, encountered with KDA-MTT. The errors along Y are a little bit bigger than respective errors along X, but as a whole the distance error, encountered by using QADA-MTT are smaller than in KDA-MTT. MC errors are evaluated on the base of the averaged errors associated with all “alive” tracks. Some of the errors occurred (for example in Fig.7 and Fig.8)

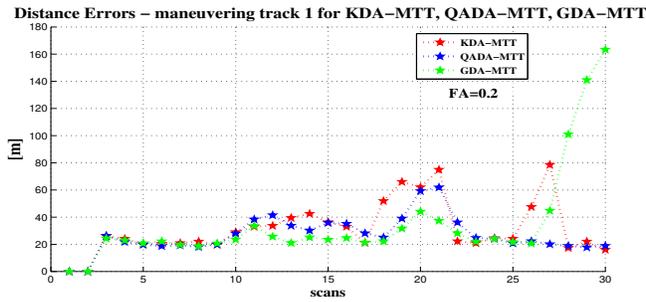


Figure 8. Maneuvering scenario: Distance errors for maneuvering track 1 - KDA-MTT, QADA-MTT, GDA-MTT.

could be explained by the unrealized canceling of tracks at the end of the scenario, when some tracks go toward canceling, but cannot satisfy the canceling condition because of lack of time. As a result they are not cancelled (and not deleted) leading that way to the increasing error.

Figures 9 and 10 show the behaviour of the same errors, but now associated with the near-by non-maneuvering target 2.

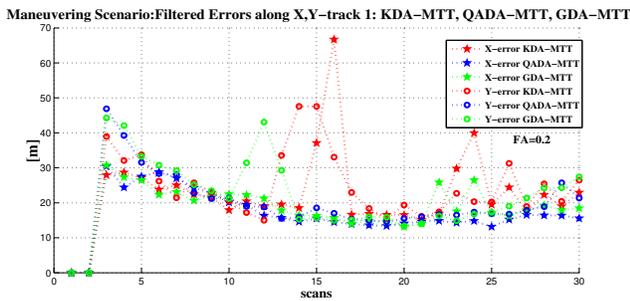


Figure 9. Maneuvering scenario: Filtered errors along X,Y for non-maneuvering track 2 - KDA-MTT, QADA-MTT, GDA-MTT.

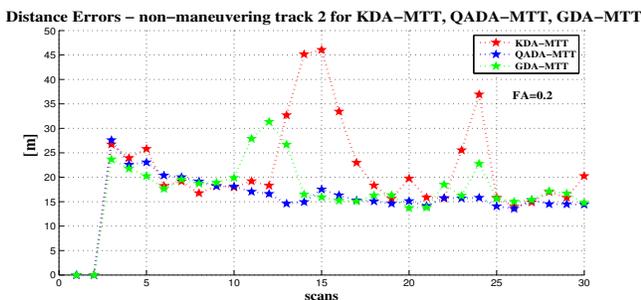


Figure 10. Maneuvering scenario: Distance errors for non-maneuvering track 2 - KDA-MTT, QADA-MTT, GDA-MTT.

For the non maneuvering target 2, the filtered errors along X and Y axes, obtained by using QADA-MTT, are smooth and definitely smaller than those, encountered with KDA-MTT. As a consequence, the associated with QADA-MTT distance error is smaller than in KDA- and GDA-MTT. The errors are calculated on the base only of the “alive” tracks.

### B. Non-maneuvering targets simulation scenario

The noise-free non-maneuvering targets simulation scenario (see Fig.11) consists of three air targets moving in parallel from West to East with the type order CFC (C=Cargo, F=Fighter) with constant velocity of 100m/sec and a distance between them 150m. The stationary sensor is located at the origin. The sampling period is  $T_{scan} = 5sec$ , and the measurement standard deviations are 0.5 deg and 65m for azimuth and range respectively. The surveillance of moving targets is performed during 15 scans. The confusion matrix, utilized by GDA is  $C = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$ . Fig. 12 shows the respective noised scenario.

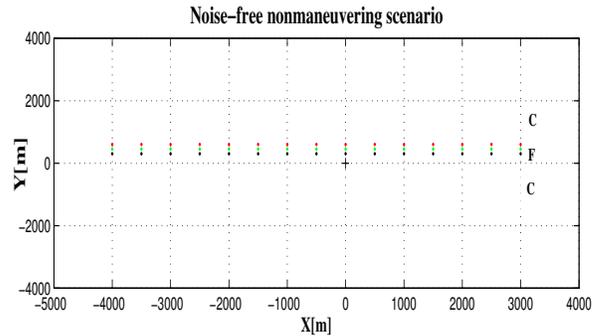


Figure 11. Noise-free non-maneuvering MTT Scenario.

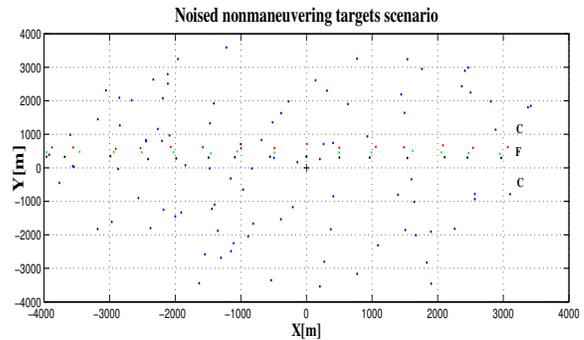


Figure 12. Noised non-maneuvering MTT Scenario.

As reported in Table 3, QADA-MTT shows again almost 2 times better performance, in comparison to KDA-MTT, according to the average miscorrelations, and also better performance regarding the average track life and track purity.

Table III  
NON-MANEUVERING SCENARIO: COMPARISON BETWEEN KDA, QADA, GDA BASED MTT PERFORMANCES FOR  $FA = 0.2$ .

	KDA-MTT	QADA-MTT	GDA-MTT
Average Track Life [%]	89.79	<b>94.21</b>	97.59
Average Miscorrelation [%]	21.36	<b>10.77</b>	5.82
Track Purity [%]	64.46	<b>81.72</b>	90.15

Fig.13 shows the percentage of miscorrelations in less noised case (with 0.2 FA in average per gate).

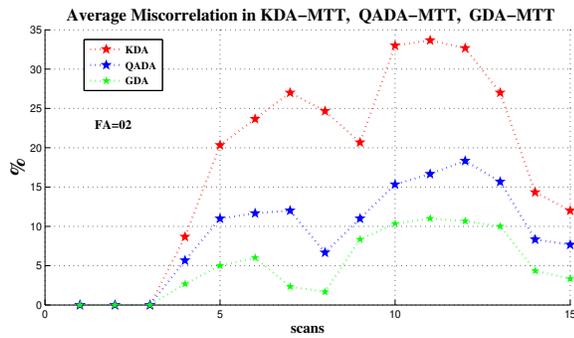


Figure 13. Non-maneuvering scenario: Average miscorrelations in KDA-MTT, QADA-MTT, GDA-MTT.

The same QADA-MTT behaviour is valid in the more dense cluttered environment with 0.4 FA in average per gate (see table 4 and fig. 14).

Table IV

NON-MANEUVERING SCENARIO: COMPARISON BETWEEN KDA, QADA, GDA BASED MTT PERFORMANCES FOR  $FA = 0.4$ .

	KDA-MTT	QADA-MTT	GDA-MTT
Average Track Life [%]	90.72	<b>92.18</b>	96.77
Average Miscorrelation [%]	20.69	<b>12.15</b>	6.26
Track Purity [%]	65.46	<b>77.38</b>	88.82

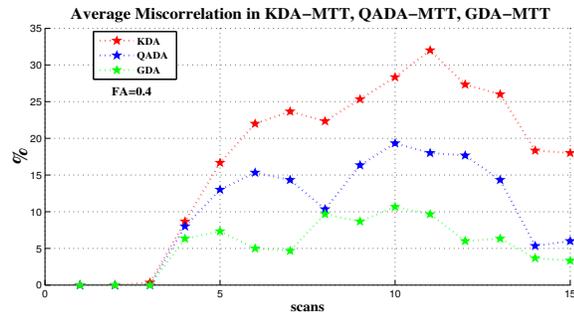


Figure 14. Non-maneuvering scenario: Average miscorrelations in KDA-MTT, QADA-MTT, GDA-MTT.

The figures 15 and 16 show typical performances of QADA-MTT and KDA-MTT systems.

The figures 17–20 show the encountered filtered errors along X and Y axes and the distance errors, associated with the intermediate track 2 for both noised cases (when the number of FA per gate is 0.2 and 0.4).

One observes (for example in Fig.9 and Fig.17) that errors associated with this simpler (non-maneuvering) scenario sometimes appear to be greater than in the previous more complicated (maneuvering) one. It is because the sensor's errors are defined deliberately greater in the non-maneuvering scenario. It provokes a complex situations, where the impact of QADA method is better demonstrated.

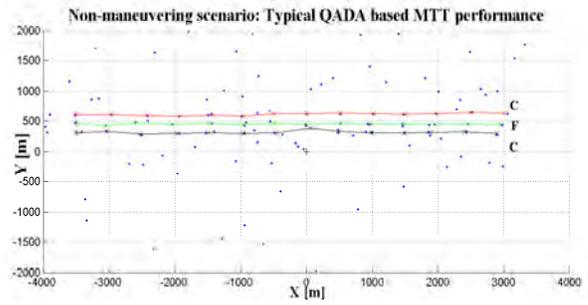


Figure 15. Non-maneuvering scenario: Typical performance of QADA based MTT.

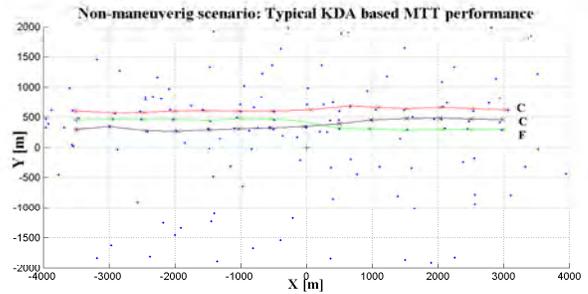


Figure 16. Non-maneuvering scenario: Typical performance of KDA based MTT.

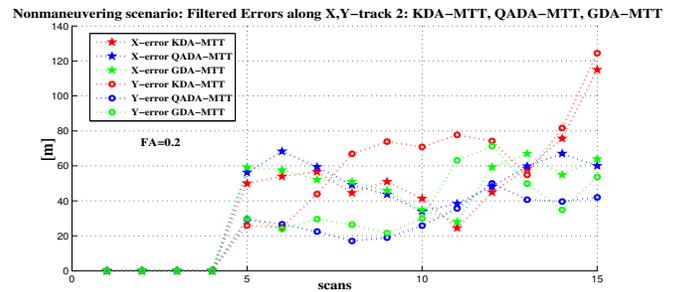


Figure 17. Non-maneuvering scenario: Filtered errors along X,Y for track 2 - KDA-MTT, QADA-MTT, GDA-MTT.

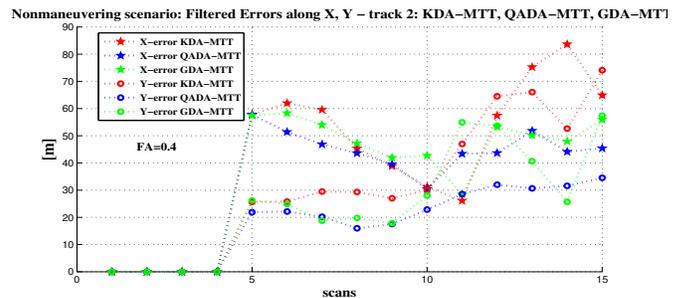


Figure 18. Non-maneuvering scenario: Filtered errors along X,Y for track 2 - KDA-MTT, QADA-MTT, GDA-MTT.

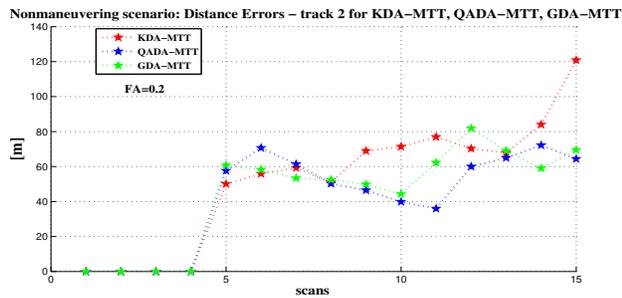


Figure 19. Non-manuevering scenario: Distance errors for non-manuevering track 2 - KDA-MTT, QADA-MTT, GDA-MTT.

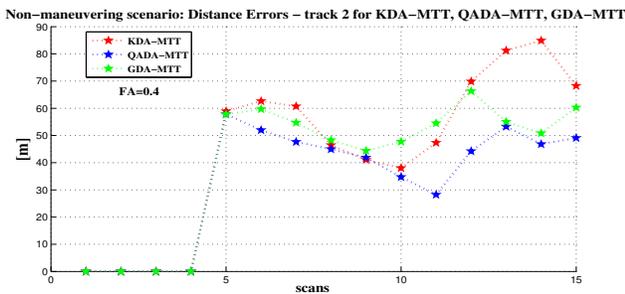


Figure 20. Non-manuevering scenario: Distance errors for non-manuevering track 2 - KDA-MTT, QADA-MTT, GDA-MTT.

## VI. CONCLUSIONS

This work assesses the efficiency of MTT performance in cluttered conflicting situations, based on the recent QADA method. The QADA based MTT performance is compared with the results, obtained for KDA and GDA based MTT, concerning two (maneuvering and non-maneuvering targets) scenarios. Our Monte Carlo simulation results show that QADA-MTT performs better than KDA-MTT for all measures of performances in all scenarios under low or heavy clutter conditions with target detection probabilities less than one, which is the main result of this paper.

Concerning the comparison of performances of QADA-MTT (using kinematics measurements only) with respect to GDA-MTT, we observe that the performances of GDA-MTT are slightly better than those of QADA-MTT. This conclusion is not very surprising because GDA-MTT uses more information (kinematics and attributes) than KDA-MTT or QADA-MTT (which are based on kinematics measurements only). Therefore, the ability of GDA-MTT to provide better tracking performances is what we naturally expect. However, we must emphasize that QADA method could also be used to improve GDA-MTT as well in a similar manner as it has been used to improve the performances of KDA-MTT. This possible improvement of GDA-MTT with QADA is under investigation and will be reported in a forthcoming publication.

Taking in mind, that MTT problems as a general do not able to utilize additional target attribute information, (i.e. when only kinematic measurements are available), applying QADA

instead of KDA leads to better MTT performance, because of its ability to estimate the quality of the individual pairings given in the optimal assignment solution. QADA is totally independent of the applied logic to obtain the best DA solution. Hence, it could be applied successfully in all cases when attribute or/and kinematic data are available.

## REFERENCES

- [1] Y. Bar Shalom (Ed.), *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, 1990.
- [2] Y. Bar Shalom, T. Fortmann, *Tracking and Data Association*, Academic Press, 1988.
- [3] Y. Bar Shalom, P.K. Willett, X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, Storrs, CT, USA, 2011.
- [4] X. He, R. Tharmarasa, M. Pelletier, T. Kirubarajan, *Accurate Murty's Algorithm for Multitarget Top Hypothesis*, In: Proceedings of Int. Conference of Information Fusion, Chicago, USA, 2011.
- [5] J. Dezert, F. Smarandache, A. Tchamova, *On the Blackman's Association Problem*, in Proc. of Int. Conference on Information Fusion (Fusion 2003), Cairn, Australia, 2003.
- [6] A. Tchamova, J. Dezert, T. Semerdjiev, P. Konstantinova, *Target tracking with Generalized Data Association based on the General DSMT Rule of Combination*, in Proc. on Int. Conference of Information Fusion (Fusion 2004), Stockholm, Sweden, 2004.
- [7] J. Dezert, A. Tchamova, T. Semerdjiev, P. Konstantinova, *Performance Evaluation of Fusion rules for Multitarget Tracking in Clutter based on Generalized Data Association*, in Proc. of Int. Conference on Information Fusion (Fusion 2005), Philadelphia, USA, 2005.
- [8] F. Smarandache and J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, Volumes 1, 2, 3 & 4, ARP, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [9] T. Denœux, N. El Zoghby, V. Cherfaoui, A. Jouglet, *Optimal Object Association in the Dempster-Shafer Framework*, IEEE Trans. on Cybernetics, Vol.44, No.12, pp. 2521–2531, Dec. 2014.
- [10] J. Dezert, K. Benameur, *On the Quality of Optimal Assignment for Data Association*, Springer, Lecture Notes in Computer Science, Vol.8764, pp. 374–382, 2014.
- [11] J. Dezert, K. Benameur, L. Ratton, J.F. Grandin, *On the Quality Estimation of Optimal Multiple Criteria Data Association Solutions*, in Proc. of Int. Conference on Information Fusion (Fusion 2015), Washington D.C, USA, July 6-9, 2015.
- [12] H.W. Kuhn, *The Hungarian Method for the Assignment Problem*, Naval Research Logistic Quarterly, Vol. 2, pp. 83–97, 1955.
- [13] J. Munkres, *Algorithms for the Assignment and Transportation Problems*, Journal of the Society of Industrial and Applied Mathematics, Vol. 5 1), pp. 32–38, 1957.
- [14] F. Bourgeois, J.C. Lassalle, *An Extension of the Munkres Algorithm for the Assignment Problem to Rectangular Matrices*, Comm. of the ACM, Vol. 14(12), pp. 802–804, 1971.
- [15] K.G. Murty, *An Algorithm for Ranking all the Assignments in Order of Increasing Cost*, Operations Research Journal, Vol. 16(3), pp. 682–687, 1968.
- [16] P. Smets, R. Kennes, *The transferable Belief Model*, Artificial Intelligence, Vol. 66(2), pp. 191–234, 1994.
- [17] J. Dezert, A. Tchamova, P. Konstantinova, *The Impact of the Quality Assessment of Optimal Assignment for Data Association in Multitarget Tracking Context*, Cybernetics and Information Technologies Journal, Vol. 15, No. 7, pp.88–98, 2015.

# Performance Evaluation of improved QADA-KF and JPDAF for Multitarget Tracking in Clutter

Jean Dezert<sup>a</sup>, Pavlina Konstantinova<sup>b</sup>, Albena Tchamova<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>European Polytechnical University, Pernik, Bulgaria.

<sup>c</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

Emails: jean.dezert@onera.fr, pavlina.konstantinova@gmail.com, tchamova@bas.bg

**Abstract**—This paper presents a performance evaluation of two types of multi-target tracking algorithms: 1) classical Kalman Filter based algorithms for multi-target tracking improved with Quality Assessment of Data Association (QADA) method using optimal data association, and 2) the Joint Probabilistic Data Association Filter (JPDAF). QADA technique is improved by using new basic belief assignment (bba) modelling, and also modified by means of the new Belief interval distance applied for computing the quality indicator associated with the pairings in the optimal data association solution. The evaluation is based on Monte Carlo simulations for maneuvering multiple-target tracking (MTT) problem in clutter.

**Keywords:** Data association, JPDAF, Belief Functions, PCR6 fusion rule, QADA, Multitarget Tracking.

## I. INTRODUCTION

The main function of each radar surveillance system is to keep targets tracks maintenance. It becomes a crucial and challenging problem especially in complicated situations of closely spaced or crossing targets. The main objective of multiple-target tracking (MTT) is to estimate jointly, at each observation time moment, the number of targets continuously moving in a given region and their trajectories from the noisy sensor data.

Data Association (DA) is a central problem in MTT systems' design [1], [2]. It relates to the process of associating uncertain measurements (observations) to the tracked tracks. In the presence of a dense MTT environment, with false alarms and sensor detection probabilities less than unity, the problem of DA becomes more complex, because it should contend with many possibilities of pairings, some of which are in practice very imprecise, unreliable, and could lead to critical association mistakes in the overall tracking process.

In order to deal with these complex associations the most recent method to evaluate the Quality Assessment of Data Association (QADA) encountered in multiple target tracking applications in a mono-criterion context was proposed by Dezert and Benameur [4], and extended in [5] for the multicriteria context. It is based on belief functions (BF) for achieving the quality of pairings belonging to the optimal data assignment solution based on its consistency with respect to all the second best solutions, provided by a chosen algorithm. Recently, in

[6], [19] the authors discussed and proposed the way in which Kalman filter (KF) could be enhanced in order to reflect the knowledge obtained based on the QADA method, called QADA-KF method. QADA assumes that the reward matrix is known, regardless of the manner in which it is obtained by the user. In this paper QADA method is improved by using new BBA modelling, and also modified by means of the new Belief Interval distance (BID) [18] applied for computing the quality indicator associated with the pairing in the optimal DA solution. The results are compared with those obtained by using Pignistic Probabilities [16]. We propose and test the performance of two versions of QADA-KF. The first one utilizes the assignment matrix, provided by the Global Nearest Neighbor (GNN) method, called QADA-GNN KF approach. The second one utilizes the assignment matrix, provided by the Probabilistic Data Association (PDA) method, called QADA-PDA KF method. These two QADA-KF methods are compared with the well-known Joint Probabilistic Data Association Filter (JPDAF) [7], [8], [9] which is an extension of the Probabilistic Data Association Filter (PDAF) [1] to a fixed and known number of targets. JPDAF uses joint association events and joint association probabilities in order to avoid conflicting measurement-to-track assignments by making a soft (probabilistic) assignment of all validated measurements to multiple targets.

The main objective of this paper is to: (1) improve QADA method by using new bba modelling; (2) modify the improved QADA method by means of the new Belief interval distance for computing the quality indicator; (3) compare the performances of: (a) classical MTT algorithms based on the GNN approach for data association, utilizing Kinematic only Data (KDA) based MTT; (b) QADA-GNN KF based MTT; (c) QADA-PDA KF based MTT; (d) JPDAF based MTT.

The evaluation is based on a Monte Carlo simulation for particular difficult maneuvering MTT problem in clutter. This paper is organized as follows. Section II is devoted to the improved QADA method. Section III discusses the Kalman Filter improved by QADA. The two variants of the assignment matrix, utilized by QADA are discussed in Section IV. In Section V the JPDAF is described and discussed. A particular simulation MTT scenario and results are presented for the

KDA, QADA-GNN KF, QADA-PDA KF, and JPDAF in Section VI. Conclusions are made in Section VII.

## II. THE IMPROVED QUALITY ASSESSMENT OF OPTIMAL DATA ASSOCIATION

### A. Improvement of QADA bba modelling

DA is a decisive step in MTT systems [1], [2]. It consists in finding the global optimal assignments of targets  $T_i$ ,  $i = 1, \dots, m$  to some measurements  $z_j$ ,  $j = 1, \dots, n$  at a given time  $k$  by maximizing the overall gain in such a way, that no more than one target is assigned to a measurement, and reciprocally. The  $m \times n$  reward matrix  $\Omega = [\omega(i, j)]$  is defined by its elements  $\omega(i, j) > 0$ , representing the gain of the association of target  $T_i$  with the measurement  $z_j$ .

The first and the second best assignments matrices  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are used [4], in order to establish the quality of the specific associations (pairings) satisfying the condition  $a_1(i, j) = 1$ . The main idea behind QADA method is to compare the values  $a_1(i, j)$  in  $\mathbf{A}_1$  with the corresponding ones  $a_2(i, j)$  in  $\mathbf{A}_2$ , and to identify the change (if any) of the optimal pairing  $(i, j)$ . In our MTT context,  $(i, j)$  means that measurement  $z_j$  is associated with target  $T_i$ . A quality indicator is established, depending on both the stability of the pairing and its relative impact on the global reward. The proposed method works also when the 1-st and 2-nd best optimal assignment  $\mathbf{A}_1$  and  $\mathbf{A}_2$  and are not unique, i.e. there are multiplicities available. The construction of the quality indicators is based on Belief Functions (BF) theory and Proportional Conflict Redistribution fusion rule no. 6 (PCR6), defined within DSMT theory [16]. It depends on the type of pairing matching in the way, described below:

- In case, when  $a_1(i, j) = a_2(i, j) = 0$ , one has a full agreement on “non-association” of the given pairing  $(i, j)$  in  $\mathbf{A}_1$  and  $\mathbf{A}_2$ . This “non-association” has no impact on the global reward values  $R_1(\Omega, \mathbf{A}_1)$  and  $R_2(\Omega, \mathbf{A}_2)$ , so it will be useless to utilize it in DA. Hence, the quality indicator value is set to  $q(i, j) = 1$ .
- In case, when  $a_1(i, j) = a_2(i, j) = 1$ , one has a full agreement on “association” of the given pairing  $(i, j)$  in  $\mathbf{A}_1$  and  $\mathbf{A}_2$ . This “association” has different impacts on the global reward values  $R_1(\Omega, \mathbf{A}_1)$  and  $R_2(\Omega, \mathbf{A}_2)$ . In order to estimate the quality of this matching association, one establish two basic belief assignments (BBA),  $m_s(\cdot)$  ( $s = 1, 2$ ) according to the both sources of information ( $\mathbf{A}_1$  and  $\mathbf{A}_2$ ). The Frame of Discernment (FoD), one reasons on, consists of a single hypothesis  $X = (T_i, z_j)$ : measurement  $z_j$  belongs to track  $T_i$ . The ignorance is modeled by the proposition  $X \cup \bar{X}$ , where  $\bar{X}$  is a negation of hypothesis  $X$ . The BBA  $m_s(\cdot)$  is defined by

$$\begin{cases} m_s(X) = a_s(i, j) \cdot \omega(i, j) / R_s(\Omega, \mathbf{A}_s), \\ m_s(X \cup \bar{X}) = 1 - m_s(X). \end{cases}$$

Applying the conjunctive rule of combination of  $m_1(X)$  with  $m_2(X)$  we get

$$\begin{cases} m_{12}(X) = m_1(X)m_2(X) + m_1(X)m_2(X \cup \bar{X}) \\ \quad + m_1(X \cup \bar{X})m_2(X), \\ m_{12}(X \cup \bar{X}) = m_1(X \cup \bar{X})m_2(X \cup \bar{X}). \end{cases} \quad (1)$$

In our previous works [6], [17], [19], we did propose to use the pignistic transform BetP to establish the quality indicator.

- In case, when  $a_1(i, j) = 1$  and  $a_2(i, j) = 0$ , then a disagreement (conflict) on the association  $(T_i, z_j)$  in  $\mathbf{A}_1$  and  $\mathbf{A}_2$  is detected. One could find the association  $(T_i, z_{j_2})$  in  $\mathbf{A}_2$ , where  $j_2$  is the measurement index, such that  $a_2(i, j_2) = 1$ . In order to define the quality of such conflicting association  $(T_i, z_j)$ , one establishes two basic belief assignments (BBA),  $m_s(\cdot)$  ( $s = 1, 2$ ) according to the both sources of information ( $\mathbf{A}_1$  and  $\mathbf{A}_2$ ). The FoD, one reasons on, consists of the following two propositions:  $X = (T_i, z_j)$ , and  $Y = (T_i, z_{j_2})$ . The ignorance is modeled by the proposition  $X \cup Y$ . In our previous works [4], we did define the BBAs by:

$$\begin{cases} m_1(X) = a_1(i, j) \cdot \omega(i, j) / R_1(\Omega, \mathbf{A}_1), \\ m_1(X \cup Y) = 1 - m_1(X), \end{cases} \quad (2)$$

$$\begin{cases} m_2(Y) = a_2(i, j_2) \cdot \omega(i, j_2) / R_2(\Omega, \mathbf{A}_2), \\ m_2(X \cup Y) = 1 - m_2(Y). \end{cases} \quad (3)$$

This modeling in fact does not work efficiently in some cases and that is why we need to revise it to make the QADA approach working more efficiently. For example, let's consider only one target  $T$  and two validated measurements  $z_1$  and  $z_2$  with the following payoff matrix  $\Omega = [100 \ 1]$ . The two possible associations are represented by  $\mathbf{A}_1 = [1 \ 0]$  providing a reward  $R_1(\Omega, \mathbf{A}_1) = 100$ , and  $\mathbf{A}_2 = [0 \ 1]$  providing a reward  $R_2(\Omega, \mathbf{A}_2) = 1$ . In this simple case, one has  $\Theta = \{X = (T, z_1), Y = (T, z_2)\}$ . In applying formulas (2)–(4), one gets

$$\begin{cases} m_1(X) = a_1(i, j) \cdot \omega(i, j) / R_1(\Omega, \mathbf{A}_1) = 1 \times \frac{100}{100} = 1 \\ m_1(X \cup Y) = 1 - m_1(X) = 0 \end{cases} \quad (4)$$

$$\begin{cases} m_2(Y) = a_2(i, j_2) \cdot \omega(i, j_2) / R_2(\Omega, \mathbf{A}_2) = 1 \times \frac{1}{1} = 1, \\ m_2(X \cup Y) = 1 - m_2(Y). \end{cases} \quad (5)$$

The conjunctive combination rule gives:

$$\begin{aligned} m_{12}(X) &= m_1(X)m_2(X \cup Y) = 1 \times 0 = 0, \\ m_{12}(Y) &= m_1(X \cup Y)m_2(Y) = 0 \times 1 = 0, \\ m_{12}(X \cup Y) &= m_1(X \cup Y)m_2(X \cup Y) = 0 \times 0 = 0, \\ m_{12}(\emptyset) &= m_1(X)m_2(Y) = 1 \times 1 = 1. \end{aligned}$$

Applying PCR6 fusion rule [16] (Vol.3):

$$\begin{cases} m_{12}^{PCR6}(\emptyset) = 0, \\ m_{12}^{PCR6}(X) = m_{12}(X) + m_1(X) \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)} = \frac{1}{2}, \\ m_{12}^{PCR6}(Y) = m_{12}(Y) + m_2(Y) \frac{m_1(X)m_2(Y)}{m_1(X)+m_2(Y)} = \frac{1}{2}, \\ m_{12}^{PCR6}(X \cup Y) = m_{12}(X \cup Y) = 0, \end{cases} \quad (6)$$

which yields (using the Pignistic transformation)  $BetP(X) = 0.5$  and  $BetP(Y) = 0.5$ . This result is counter-intuitive (not realistic) because in this very simple case one knows that  $(T, z_1)$  is obviously the best data association solution. To circumvent this serious problem, we propose to modify the bba modeling by taking a new model of bba construction as follows:

$$\begin{cases} m_1(X) = a_1(i, j) \cdot \frac{\omega(i, j)}{R_1(\Omega, A_1) + R_2(\Omega, A_2)}, \\ m_1(X \cup Y) = 1 - m_1(X), \end{cases} \quad (7)$$

$$\begin{cases} m_2(Y) = a_2(i, j_2) \cdot \frac{\omega(i, j_2)}{R_1(\Omega, A_1) + R_2(\Omega, A_2)}, \\ m_2(X \cup Y) = 1 - m_2(Y). \end{cases} \quad (8)$$

If we apply this modeling on the previous example, we obtain

$$\begin{cases} m_1(X) = a_1(i, j) \cdot \frac{\omega(i, j)}{R_1(\Omega, A_1) + R_2(\Omega, A_2)} = 1 \times \frac{100}{101} \approx 0.99, \\ m_1(X \cup Y) = 1 - m_1(X) = 0.01, \end{cases}$$

$$\begin{cases} m_2(Y) = a_2(i, j_2) \cdot \frac{\omega(i, j_2)}{R_1(\Omega, A_1) + R_2(\Omega, A_2)} = 1 \times \frac{1}{101} \approx 0.01, \\ m_2(X \cup Y) = 1 - m_2(Y) = 0.99. \end{cases}$$

Hence, one gets now

$$\begin{aligned} m_{12}(X) &= m_1(X)m_2(X \cup Y) = 0.9801, \\ m_{12}(Y) &= m_1(X \cup Y)m_2(Y) = 0.0001, \\ m_{12}(X \cup Y) &= m_1(X \cup Y)m_2(X \cup Y) = 0.0099, \\ m_{12}(\emptyset) &= m_1(X)m_2(Y) = 0.0099. \end{aligned}$$

Applying PCR6 redistribution principle, one gets finally

$$\begin{aligned} m_{12}^{PCR6}(X) &= 0.9801 + (0.99 \times 0.0099)/1 = 0.989901, \\ m_{12}^{PCR6}(Y) &= 0.0001 + (0.01 \times 0.0099)/1 = 0.010099, \end{aligned}$$

which yields  $BetP(X) = 0.989901$  and  $BetP(Y) = 0.010099$ . This result fits now perfectly with what we expect, that is  $X = (T, z_1)$  is obviously the best data association solution.

### B. Improvement of quality indicator calculating by using Belief Interval (BI) distance

In [11], [20] the Euclidean belief interval distance between two bbas  $m_1(\cdot)$  and  $m_2(\cdot)$  is defined on the powerset of a given  $\Theta = \{\theta_1, \dots, \theta_n\}$  as follows

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))}, \quad (9)$$

where  $N_c = 1/2^{|\Theta|-1}$  is a normalization factor to have  $d_{BI}(m_1, m_2) \in [0, 1]$ , and  $d_W(BI_1(X), BI_2(X))$  is the Wasserstein's distance [22] between belief

intervals  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)] = [a_1, b_1]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)] = [a_2, b_2]$ . More specifically,

$$d_W([a_1, b_1], [a_2, b_2]) = \sqrt{\left[\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2}\right]^2 + \frac{1}{3} \left[\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2}\right]^2}. \quad (10)$$

In [20], we have proved that  $d_{BI}(x, y)$  is a true distance metric because it satisfies the properties of non-negativity ( $d_{BI}(x, y) \geq 0$ ), non-degeneracy ( $d_{BI}(x, y) = 0 \Leftrightarrow x = y$ ), symmetry ( $d_{BI}(x, y) = d_{BI}(y, x)$ ), and the triangle inequality ( $d_{BI}(x, y) + d_{BI}(y, z) \geq d_{BI}(x, z)$ ), for any bba  $x, y$  and  $z$  defined on  $2^\Theta$ . The choice of Wasserstein's distance in  $d_{BI}$  definition is justified by the fact that Wasserstein's distance is a true distance metric and it fits well with our needs because we have to compute a distance between  $[Bel_1(X), Pl_1(X)]$  and  $[Bel_2(X), Pl_2(X)]$ .

For notation convenience, we denote  $m_X$  the categorical bba having only  $X$  as focal element, where  $X \neq \emptyset$  is an element of the powerset of  $\Theta$ . More precisely,  $m_X$  is the particular (categorical) bba defined by  $m_X(X) = 1$  and  $m_X(Y) = 0$  for any  $Y \neq X$ . Such basic bba plays an important role in our new decision scheme because its corresponding belief interval reduces to the degenerate interval  $[1, 1]$  which represents the certainty on  $X$ . The basic principle of the new decision scheme we propose is very simple and intuitively makes sense. It consists in selecting as the final decision (denoted by  $\hat{X}$ ) the element of the powerset for which the belief interval distance between the bba  $m(\cdot)$  and  $m_X$ ,  $X \in 2^\Theta \setminus \{\emptyset\}$  is the smallest one. Therefore, take as the final decision  $\hat{X}$  given by

$$\hat{X} = \arg \min_{X \in 2^\Theta \setminus \{\emptyset\}} d_{BI}(m, m_X). \quad (11)$$

where  $d_{BI}(m, m_X)$  is computed according to (9).  $m(\cdot)$  is the bba under test, and  $m_X$  the categorical bba focused on  $X$  defined above.

This decision scheme is very general in the sense that the decision making can be done on any type of element of the power-set  $2^\Theta$ , and not necessarily only on the elements (singletons) of the FoD. This method not only provides the final decision  $\hat{X}$  to make, but also it evaluates how good this decision is with respect to its alternatives if we define the quality indicator  $q(\hat{X})$  as follows

$$q(\hat{X}) \triangleq 1 - \frac{d_{BI}(m, m_{\hat{X}})}{\sum_{X \in 2^\Theta \setminus \{\emptyset\}} d_{BI}(m, m_X)}. \quad (12)$$

One sees that the quality indicator  $q(\hat{X})$  of the decision  $\hat{X}$  will become maximum (equal to one) when the distance between the bba  $m(\cdot)$  and  $m_{\hat{X}}$  is zero, which means that the bba  $m(\cdot)$  is focused in fact only on the element  $X$ . The higher  $q(\hat{X})$  is, the more confident in the decision  $\hat{X}$  we should be.

Of course, if a decision must be made with some extra constraint defined by a (or several) condition(s), denoted  $c(X)$ , then we must take into account  $c(X)$  in (11), that is

$$\hat{X} = \arg \min_{X \in 2^\Theta \setminus \{\emptyset\} | c(X) \text{ is true}} d_{BI}(m, m_X).$$

and also in the derivation of quality indicator by taking  $\sum_{X \in 2^{\Theta} \setminus \{\emptyset\}} c(X) \text{ is true } d_{BI}(m, m_X)$  as denominator in (12). Theoretically any other strict distance metric, for instance Jousselme's distance [23], [24], could be used instead of  $d_{BI}(\cdot, \cdot)$ . We have chosen  $d_{BI}$  distance because of its ability to provide good and reasonable behavior [20] as will be shown. When there exists a tie between multiple decisions  $\{\hat{X}_j, j > 1\}$ , then the prudent decision corresponding to their disjunction  $\hat{X} = \cup_j \hat{X}_j$  should be preferred (if allowed), otherwise the final decision  $\hat{X}$  is made by a random selection among the elements of  $\{\hat{X}_j, j > 1\}$ .

### III. QADA BASED KALMAN FILTER

The aim of this paper is to compare the performance of the JPDAF based MTT algorithm with the classical MTT algorithm, using the CMKF based on kinematics measurements, but improved by the QADA method.

In [6], the authors discuss and propose the way in which Kalman filter (KF) could be improved in order to reflect the knowledge obtained based on the QADA method.

Let's briefly recall what kind of information is obtained, having in hand the quality matrix, derived by QADA, in the MTT context. It gives knowledge about the confidence  $q(i, j)$  in all pairings  $(T_i, z_j)$ ,  $i = 1, \dots, m$ ;  $j = 1, \dots, n$ , chosen in the first best assignment solution. The smaller quality (confidence) of hypothesis " $z_j$  belongs to  $T_i$ " means, that the particular measurement error covariance  $R$  was increased and the filter should not trust fully in the actual (true) measurement  $z(k+1)$ .

Having this conclusion in mind, the authors propose, such a behavior of the measurement error covariance to be modeled by  $R = R/q(T_i, z_j)$ , for every pairing, chosen in the first best assignment and based on the corresponding quality value obtained. Then, when the Kalman filter gain decreases the true measurement  $z_j(k+1)$  is trusted less in the updated state estimate  $\hat{x}(k+1|k+1)$ .

### IV. BUILDING ASSIGNMENT MATRIX FOR QADA

QADA assumes the reward matrix is known, regardless of the manner in which it is obtained by the user. In this paper we propose two versions of QADA-KF. The first one utilizes the assignment matrix built from the single normalized distances, provided by the Global Nearest Neighbor method, called QADA-GNN KF method. The second one utilizes the assignment matrix, built from the posterior association probabilities, provided by the Probabilistic Data Association (PDA) method, called QADA-PDA KF method.

#### A. Assignment matrix based on GNN method

The GNN method finds and propagates the single most likely hypothesis during each scan to update KF. It is a hard (i.e., binary) decision approach, as compared to the JPDAF which is a soft (i.e., probabilistic) decision approach using all validated measurements with their probabilities of association. GNN method was applied in [6] and [17] to obtain the assignment matrix, utilized in QADA. In this case

the elements of assignment matrix  $\omega(i, j)$  ( $i = 1, \dots, m$ ,  $j = 1, \dots, n$ ) represent the normalized distances  $d(i, j) = [(z_j(k) - \hat{z}_i(k|k-1))' S^{-1}(k)(z_j(k) - \hat{z}_i(k|k-1))]^{1/2}$  between the validated measurement  $z_j(k)$  and target  $T_i$  satisfying the condition  $d^2(i, j) < \gamma$ . The distance  $d(i, j)$  is computed from the measurement  $z_j(k)$  and its prediction  $\hat{z}_i(k|k-1)$  (see [1] for details), and the inverse of the covariance matrix  $S(k)$  of the innovation, computed by the tracking filter. The threshold  $\gamma$ , for which the probability of given observation to fall in the gate is 0.99, could be defined from the table of the Chi-square distribution with  $M$  degrees of freedom and allowable probability of a valid observation falling outside the gate. In this case the DA problem consists in finding the best assignment that minimizes the overall cost.

#### B. Assignment matrix based on PDA method

The Probabilistic Data Association (PDA) method [1] calculates the association probabilities for validated measurements at a current time moment to the target of interest. PDA assumes the following hypotheses according to each validated measurement:

- $H_i(k)$ :  $z_j$  is a measurement, originated from the target  $T_i$  of interest,  $i = 1, \dots, m$ ;
- $H_0(k)$ : no one of the validated measurement originated from the target of interest.

If  $N$  observations fall within the gate of track  $i$ ,  $N+1$  hypotheses will be formed. The probability of  $H_0$  is proportional to  $p_{i0} = \lambda_{FA}^N (1 - P_d)$ , and the probability of  $H_j$  ( $j = 1, \dots, N$ ) is proportional to

$$p_{ij} = \frac{\lambda_{FA}^{N-1} P_g P_d \cdot e^{-\frac{d_{ij}^2}{2}}}{(2\pi)^{M/2} \cdot \sqrt{|S_{ij}|}}. \quad (13)$$

where  $P_g$  is the a priori probability that the correct measurement is in the validation gate [1];  $P_d$  is the target detection probability;  $\lambda_{FA}$  is the spatial density of false alarms (FA). The probabilities  $p_{ij}$  can be rewritten as [1]

$$p_{ij} = \begin{cases} \frac{b}{b + \sum_{i=1}^N \alpha_{ii}} & \text{for } j = 0 \text{ (no valid obs.)}, \\ \frac{\alpha_{ij}}{b + \sum_{i=1}^N \alpha_{ii}} & \text{for } 1 \leq j \leq N, \end{cases} \quad (14)$$

where

$$b \triangleq (1 - P_g P_d) \lambda_{FA} (2\pi)^{M/2} \cdot \sqrt{|S_{ij}|}, \quad (15)$$

and

$$\alpha_{ij} \triangleq P_d \cdot e^{-\frac{d_{ij}^2}{2}}. \quad (16)$$

The assignment matrix used in QADA method is established from all  $p_{ij}$  given by (14) related with all association hypotheses. This matrix will have  $m$  rows (where  $m$  is the number of all targets of interest), and  $N+1$  columns for the hypotheses generated. The  $(N+1)$ -th column will include the values  $p_{i0}$  associated with  $H_0(k)$ .

## V. JOINT PROBABILISTIC DATA ASSOCIATION FILTER

The Joint Probabilistic Data Association Filter (JPDAF) is an extension of the Probabilistic Data Association Filter (PDAF) for tracking multiple targets in clutter [1], [2], [10], [11], [12]. This Bayesian tracking filter uses the probabilistic assignment of all validated measurements belonging to the target gate to update its estimate. The preliminary version of JPDAF was proposed by Bar-Shalom in 1974 [13], then updated and finalized in [7], [8], [9]. The assumptions of JPDAF are the following:

- the number  $N_T$  of established targets in clutter is known;
- all the information available from the measurements  $Z^k$  up to time  $k$  is summarized by the sufficient statistic  $\hat{x}^t(k)$  (the approximate conditional mean), and covariance  $P^t(k|k)$  for each target  $t$ ,  $t = 1, \dots, N_T$ ;
- the real state  $x^t(k)$  of a target  $t$  at time  $k$  is modeled by a Gaussian pdf  $\mathcal{N}(x^t(k); \hat{x}^t(k), P^t(k|k))$ ;
- each target  $t$  follows its own dynamic model;
- each target generates at most one measurement at each observation time and there are no merged measurements;
- each target is detected with some known detection probability  $P_d^t$ ;
- the false alarms (FA) are uniformly distributed in surveillance area and their number follows a Poisson pmf with FA density  $\lambda_{FA}$ .

In JPDAF, the measurement to target association probabilities are computed jointly across the targets and only for the latest set of measurements. This appealing theoretical approach however can give rise to very high combinatorics complexity if there are several persistent interferences, typically when several targets are crossing or if they move closely during several consecutive scans. Moreover, some track coalescence effects may also appear which degrades substantially the JPDAF performances as it will shown in section VI. These limitations of JPDAF have already been reported in [14]. Let's consider a cluster (a cluster is a group of targets which have some measurements in common in their validation gates, i.e. non-empty intersections) of  $T \geq 2$  targets  $t = 1, \dots, T$ . The set of  $m_k$  measurements available at scan  $k$  is denoted  $Z(k) = \{z_i(k), i = 1, \dots, m_k\}$ . Each measurement  $z_i(k)$  of  $Z(k)$  either originates from a target or from a FA. Denote  $\hat{z}^t(k|k-1)$  as the predicted measurement for target  $t$ , and all the possible innovations that could be used in the Kalman Filter to update the target state estimate are denoted  $\tilde{z}_i^t(k) \triangleq z_i(k) - \hat{z}^t(k|k-1)$ . In JPDAF, instead of using a particular innovation  $\tilde{z}_i^t(k)$ , it uses the weighted innovation  $\tilde{z}^t(k) = \sum_{i=1}^{m_k} \beta_i^t(k) \tilde{z}_i^t(k)$ , where  $\beta_i^t(k)$  is the probability that the measurement  $z_i$  originates from target  $t$ .  $\beta_0^t(k)$  is the probability that none measurements originate from the target  $t$ . The core of JPDAF is the computation of the a posteriori association probabilities  $\beta_i^t(k)$ , ( $i = 0, 1, \dots, m_k$ ) based on all possible joint association events  $\Theta(k) = \cap_{i=1}^{m_k} \Theta_i^{t_i}(k)$ , where  $\Theta_i^{t_i}(k)$  is the event that measurement  $z_i(k)$  originates from target  $t_i$  (by convention and notation convenience,  $t_i = 0$  means that the origin of measurement  $z_i$  is a FA.),

$0 \leq t_i \leq N_T$ . More precisely, one has to compute for  $i = 1, \dots, m_k$ ,  $\beta_i^t(k) = \sum_{\Theta(k)} P(\Theta(k)|Z^k) \hat{\omega}_{it}(\Theta(k))$  and  $\beta_0^t(k) = 1 - \sum_{i=1}^{m_k} \beta_i^t(k)$ , where  $Z^k$  is the set of all measurements available up to time  $k$ , and  $\hat{\omega}_{it}(\Theta(k))$  are the corresponding components of the association matrix characterizing the possible joint association  $\Theta(k)$ .

JPDAF is well theoretically founded and it does not require high memory. It provides pretty good results on simple MTT scenarios (with non-persisting interferences) with moderate FA densities. However the number of feasible joint association matrices increases exponentially with problem dimensions ( $m_k$  and  $N_T$ ) which makes the JPDAF intractable for complex dense MTT scenarios. For more details about JPDAF, please refer to [1], [2], [10]–[12], and [15].

## VI. SIMULATION RESULTS

The Converted Measurement KF (CMKF) is used in our MTT algorithm. We assume constant velocity target model. The process noise covariance matrix is:  $Q = \sigma_\nu^2 Q_T$ , where  $T$  is the sampling period,  $\sigma_\nu$  is the standard deviation of the process noise, and  $Q_T$  is as given in [3]. Here are the results of KDA KF, QADA-GNN KF, QADA-PDA KF, and JPDAF for the MTT scenario with maneuvering targets.

The noise-free group of targets simulation scenario (Fig.1) consists of four air targets moving from left to right (or from West to East). For the clear explanation of the results, targets are numbered starting at the beginning with 1st target that has the greater  $y$ -coordinate and continuing to 4th target with the smallest  $y$ -coordinate. The stationary sensor is located at the origin with range 10000 m. The sampling period is  $T_{\text{scan}} = 5$  sec, and the measurement standard deviations are 0.2 deg and 40 m for azimuth and range respectively. The targets move with constant velocity  $V = 100$  m/s. The first target for the first 8 scans moves without maneuvering keeping azimuth 120 deg from North. The group of two targets in the middle i.e. 2nd and 3rd move without maneuvering keeping azimuth 90 deg from North that means, horizontally from West to East. It is the main direction of the group movement. The 4th target starts with azimuth 60 deg and moves towards the middle group of rectilinearly moving targets. When it approaches the group, it starts a turn to the right with 30 deg. Its initial azimuth of 120 deg is decreased by the angle of turn and becomes 90 deg, i.e. coincides with the main direction. From 15th scan, the four targets move rectilinearly in parallel. The distance between them is 150 m. The absolute value of the corresponding transversal acceleration for the two maneuvers is  $1.495 \text{ m/s}^2$ . The total number of scans for the simulations is 30. The figure 2 shows the noised scenario for yielding to 0.15 FA per gate on average.

Our results are based on Monte Carlo (MC) simulations with 200 independent runs in applying KDA based KF, QADAGNN KF, QADA-PDA KF, and JPDAF. We compare the performance of these methods with different criteria, and we use an idealized track initiation in order to prevent uncontrolled impact of this stage on the statistical parameters of the tracking process during MC simulations.

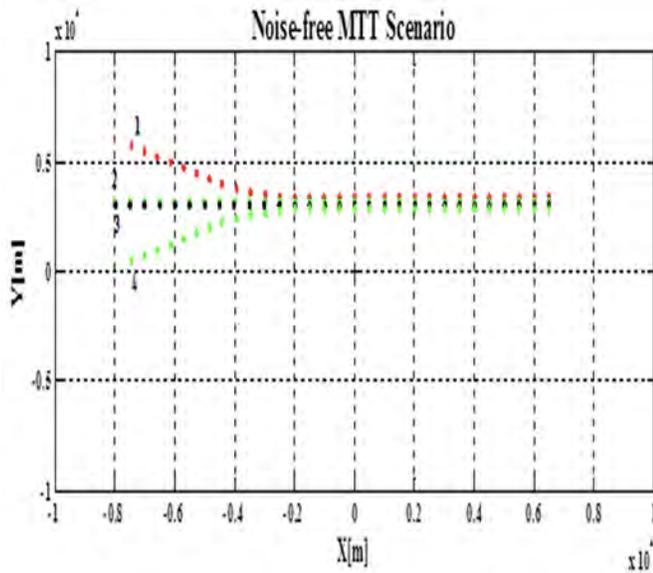


Fig. 1. Noise-free group of targets Scenario.

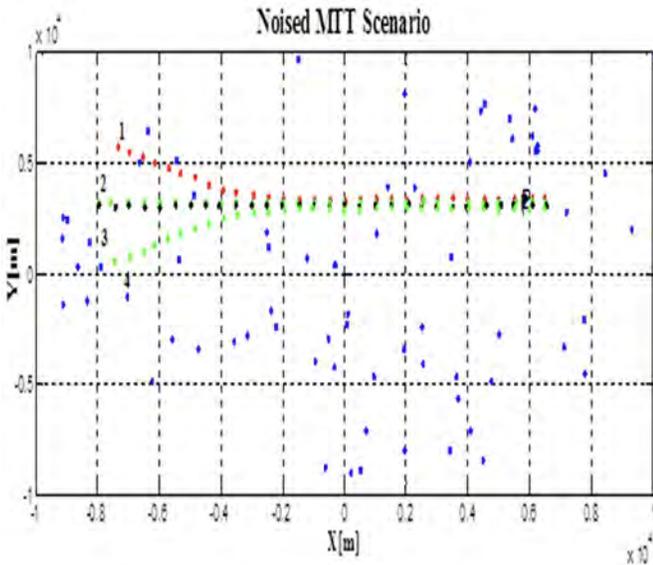


Fig. 2. Noised group of targets Scenario.

The true targets positions (known in our simulations) for the first two scans are used for track initiation. The evaluation of MTT performance is based on the criteria of Track Purity (TP), Track Life (TL), and percentage of miscorrelation (pMC):

- 1) TP criteria examines the ratio between the number of particular performed ( $j$ th observation -  $i$ th track) associations (in case of detected target) over the total number of all possible associations during the tracking scenario, but TP cannot be used with JPDAF because JPDAF is a soft assignment method. Instead of TP, we define the Probabilistic Purity Index (PPI). It considers the measurement that has the highest association probability computed by the JPDAF and check, (and count) if this

measurement originated from the target or not. PPI measures the ability of JPDAF to commit the highest probability to the correct target measurement in the soft assignment of all validated measurements.

- 2) TL is evaluated as an average number of scans before track's deletion. In our simulations, a track is canceled and deleted from the list of tracked tracks, when during 3 consecutive scans it cannot be updated with some measurement because there is no validated measurement in the validation gate. When using JPDAF, the track is canceled and deleted from the list of tracked tracks, when during 3 consecutive scans its own measurement does not fall in its gate. We call this, the "canceling/deletion condition". The status of the tracked tracks is denoted "alive".
- 3) pMC examines the relative number of incorrect observation-to-track associations during the scans. The MTT performance results for KDA only KF, QADAGNN KF, QADA-PDA KF, and JPDAF for average false alarms in gate  $FA = 0.15$  are given in Table 1. The MTT performance for QADA-PDA KF and QADA-GNN KF are estimated for both: Pignistic probabilities, and minimum Belief distance principles to compute the quality indicator.

TABLE I  
GROUP OF TARGETS SCENARIO: COMPARISON BETWEEN MTT PERFORMANCE RESULTS FOR 0.15 FA PER GATE.

(in %)	QADA-PDA		QADA-GNN		JPDAF	KDA
	BetP	Bld	BetP	Bld		
Average TL	88.12	<b>89.39</b>	84.31	<b>89.13</b>	78.42	70.02
Average pMC	2.67	<b>2.45</b>	3.28	<b>2.39</b>	5.92	5.71
Average TP	84.54	<b>86.14</b>	79.86	<b>85.92</b>	32.96 (PPI)	61.95

According to all criteria, QADA-PDA KF method shows the best performance, followed by QADA-GNN KF, and JPDAF. The KDA based KF approach, as one could expect, shows the worst performance. It is obvious that minimum Belief distance interval principle for computing the quality indicator leads to improved MTT performance (compared to the results based on Pignistic probabilities - BetP) for both QADA-PDA KF and QADA-GNN KF. Still QADA-PDA KF outperforms QADA-GNN KF based MTT.

In order to make a fair comparison between QADA KF and JPDAF, we will discuss also the root mean square errors (RMSE), associated with the filtered X and Y values, presented in Figs. 3–6. The results for QADA-GNN KF and QADA-PDA KF are obtained on the base of the improved QADA method using minimal Belief Interval distance criteria and with the new bba modeling, proposed in the paper.

Figs. 3 and 4 show the mean square X and Y error filtered, associated with target 1, and compared for KDA KF, QADA-GNN KF, QADA-PDA KF, and JPDAF. Figs. 5 and 6 consider the same errors for the middle track 3. All the results are compared to the sensor's errors along X and Y axis.

As a whole, one could see that rms errors, associated with QADA-PDA KF and QADA-GNN KF are a little bit less than the sensor's measurement errors, except around the scan

15th, where all the targets move in parallel. We see that the RMSE on Y filtered error for track 1 associated with KDA-JPDAF grows extremely after scan 12. This behavior could be explained by the fact, that from this scan on target 1 starts moving in parallel with the rest of targets, causing that way spatial persisting interferences and track coalescence effects in JPDAF. These effects degrade significantly the quality of JPDAF performance as already reported in [14]. The same effect of track coalescence could be observed for track 3, moving in parallel during all the scans. The RMSE on Y filtered associated with JPDAF performance is high during the whole tracking region.

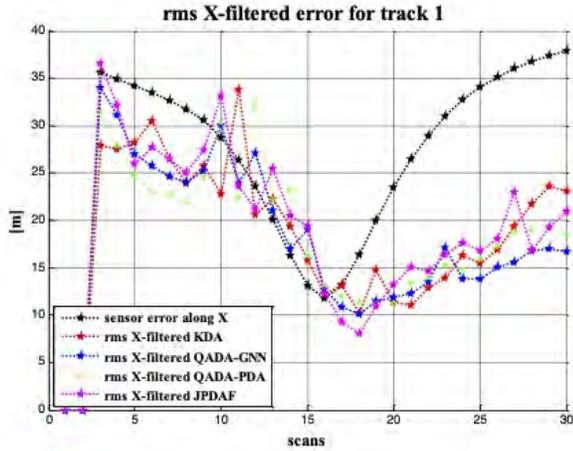


Fig. 3. RMSE on X for track 1 with the four tracking methods.

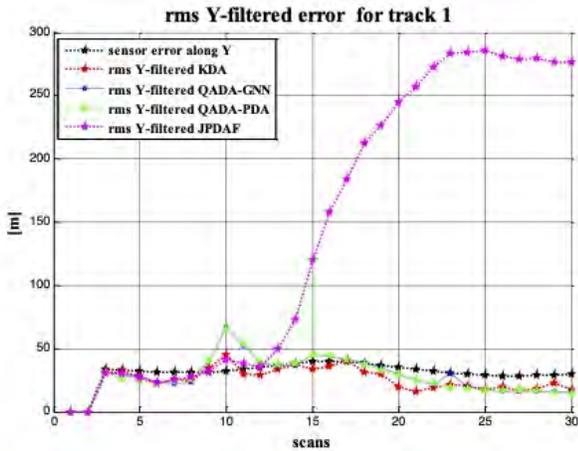


Fig. 4. RMSE on Y for track 1 with the four tracking methods.

## VII. CONCLUSIONS

This work evaluated with Monte Carlo simulations the efficiency of MTT performance in cluttered environment of four methods (a) classical MTT algorithm based on GNN approach for data association, utilizing Kinematic only Data; (b) QADA-GNN KF; (c) QADA-PDA KF; and (d) JPDAF.

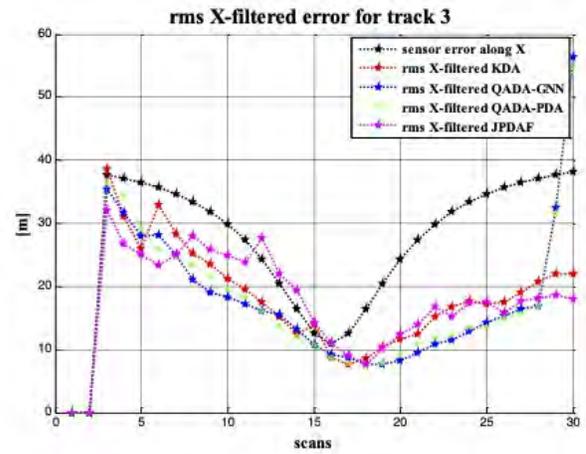


Fig. 5. RMSE on X for track 3 with the four tracking methods.

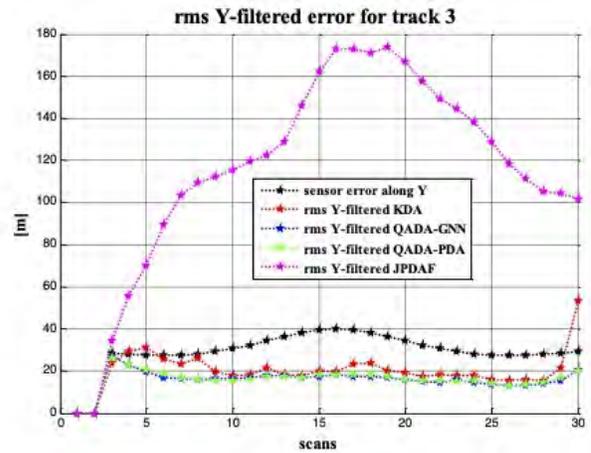


Fig. 6. RMSE on Y for track 3 with the four tracking methods.

QADA technique was improved by using new BBA modelling. It is also was modified by means of the new Belief interval distance applied for computing the quality indicator associated with the pairings in the optimal DA solution. The results were compared with those obtained by using Pignistic Probabilities. It was proved that this new approach leads to better MTT performance. The implemented groups of targets scenario show the advantages of applying QADA-KF. According to all performance criteria, the QADA-PDA KF gives the best performance, followed by QADA-GNN KF, and JPDAF. The KDA KF approach shows the worst performance (as expected). This scenario is particularly difficult for JPDAF because of several closely spaced and rectilinearly moving targets in clutter during many consecutive scans, and it leads to track coalescence effects due to persisting interferences. As a result, the tracking performance of JPDAF is degraded. Because the complexity of the calculation for joint association probabilities grows exponentially with the number of targets, JPDAF requires almost 3 times more computational time in comparison to other methods in the first (complex) scenario.

## REFERENCES

- [1] Y. Bar-Shalom, T.E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.
- [2] Y. Bar-Shalom (Ed.), *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, USA, 1990.
- [3] S. Blackman, R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.
- [4] J. Dezert, K. Benameur, *On the Quality of Optimal Assignment for Data Association*, in Proc. of Belief 2014 Int. Conf., Springer, L.N. in Compt. Sci., Vol.8764, pp. 374–382, 2014.
- [5] J. Dezert, K. Benameur, L. Rattou, J.-F. Grandin, *On the Quality Estimation of Optimal Multiple Criteria Data Association Solutions*, in Proc. of Fusion 2015 Int. Conf., Washington D.C., USA, July 6–9, 2015.
- [6] J. Dezert, A. Tchamova, P. Konstantinova, *Multitarget Tracking Performance based on the Quality Assessment of Data Association*, in Proc. of Fusion 2016 Int. Conf., Heidelberg, Germany, July 5–8, 2016.
- [7] Y. Bar-Shalom, T.E. Fortmann, M. Scheffe, *JPDA for Multiple Targets in Clutter*, in Proc. of Conf. on Inf. Sci. and Syst., Princeton, March 1980.
- [8] T.E. Fortmann et al., *Multitarget Tracking using Joint Probabilistic Data Association*, in Proc. of IEEE CDC, USA, Dec. 1980.
- [9] T.E. Fortmann, et al., *Sonar Tracking of Multiple Targets Using JPDAF*, IEEE Journal of Oceanic Eng., Vol. 8, No.3, pp. 173–184, July 1983.
- [10] Y. Bar-Shalom, X.-R. Li. *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, CT, USA, 1995.
- [11] Y. Bar-Shalom et al., *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, John Wiley & Sons, 2001.
- [12] Y. Bar-Shalom et al., *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, Storrs, CT, USA, 2011.
- [13] Y. Bar-Shalom, *Extension of the PDAF to Multitarget Environments*, in Proc. 5th Symp. Nonlinear Estimation, San Diego, USA, Sept. 1974.
- [14] R.J. Fitzgerald, *Track Biases and Coalescence with Probabilistic Data Association*, IEEE Trans. on AES, Vol. 21, No. 6, pp. 822–825, 1985.
- [15] J. Dezert, *Introduction au pistage multi-cibles multi-senseurs*, ENSTA Course (in French), Sept. 2003.
- [16] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, Volumes 1, 2, 3 & 4, ARP, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [17] J. Dezert, A. Tchamova, P. Konstantinova, *The Impact of the Quality Assessment of Optimal Assignment for Data Association in Multitarget Tracking Context*, Cybernetics and Inf. Techn. J., Vol.15, No.7, pp. 88–98, 2015.
- [18] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-Making with Belief Interval Distance*, in Proc. of Belief 2016 Int. Conf., Prague, CZ, Sept. 21–23, 2016.
- [19] J. Dezert, A. Tchamova, P. Konstantinova, E. Blasch, *A Comparative Analysis of QADA-KF with JPDAF for Multitarget Tracking in Clutter*, in Proc. of 20th International Conference on Information Fusion, Xi'an, China, July 10–13, 2017.
- [20] D. Han, J. Dezert, Y. Yang, *New Distance Measures of Evidence based on Belief Intervals*, in Proc. of Belief 2014 Int. Conf., Oxford, UK, Sept. 26–29, 2014.
- [21] D. Han, J. Dezert, Y. Yang, *Belief Interval Based Distance Measures in the Theory of Belief Functions*, IEEE Trans. on SMC, Vol. 48, No. 6, pp. 833–850, June 2018
- [22] A. Irpino, R. Verde, *Dynamic Clustering of Interval Data Using a Wasserstein-based Distance*, Pattern Recognition Letters, Vol. 29, pp. 1648–1658, 2008.
- [23] A.-L. Jousselme, P. Maupin, *Distances in evidence theory: Comprehensive survey and generalizations*, IJAR, Vol. 53, No. 2, pp. 118–145, 2012.
- [24] M. Bouchard M., A.-L. Jousselme, P.-E. Doré, *A proof for the positive definiteness of the Jaccard index matrix*, IJAR, Vol. 54, pp. 615–626, 2013.

# A Comparative Analysis of QADA-KF with JPDAF for Multitarget Tracking in Clutter

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Pavlina Konstantinova<sup>c</sup>, Erik Blasch<sup>d</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>c</sup>European Polytechnical University, Pernik, Bulgaria.

<sup>d</sup>AFRL, Rome, U.S.A.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, pavlina.konstantinova@gmail.com, erik.blasch.1@us.af.mil

Originally published as: J. Dezert, A. Tchamova, P. Konstantinova, E. Blasch, *A Comparative Analysis of QADA-KF with JPDAF for Multitarget Tracking in Clutter*, in Proc. of 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10–13, 2017, and reprinted with permission.

**Abstract**—This paper presents a comparative analysis of performances of two types of multi-target tracking algorithms: 1) the Joint Probabilistic Data Association Filter (JPDAF), and 2) classical Kalman Filter based algorithms for multi-target tracking improved with Quality Assessment of Data Association (QADA) method using optimal data association. The evaluation is based on Monte Carlo simulations for difficult maneuvering multiple-target tracking (MTT) problems in clutter.

**Keywords:** Data association, JPDAF, Belief Functions, QADA, PCR6 rule, Multitarget Tracking.

## I. INTRODUCTION

Multiple-target tracking (MTT) is a principle component of surveillance systems. The main objective of MTT is to estimate jointly, at each observation time moment, the number of targets continuously moving in a given region and their trajectories from the noisy sensor data. In a single-sensor case, the multitarget tracker receives a random number of measurements due to the uncertainty which results in low detection and false alarms, arising independently of the targets of interest. Because of the fact that detection probability is not perfect, some targets may go undetected at some sampling intervals. Additional complications appear, apart from the process and measurement noises, associated with a measurement origin uncertainty, missed detection, cancelling (death) of targets, etc.

Data association (DA) is a primary task of modern MTT systems [1]–[3]. It entails selecting the most trustable associations between uncertain sensor's measurements and existing targets at a given time. In the presence of a dense MTT environment, with false alarms and sensor detection probabilities less than unity, the problem of DA becomes more complex, because it should contend with many possibilities of pairings, some of which are in practice very imprecise, unreliable, and could lead to critical association mistakes in the overall tracking process.

In order to deal with these complex associations the most recent method to evaluate the Quality Assessment of Data Association (QADA) encountered in multiple target tracking applications in a mono-criterion context was proposed by Dezert and Benameur [4], and extended in [5] for the multi-criteria context. It is based on belief functions (BF) for achieving the quality of pairings belonging to the optimal data

assignment solution based on its consistency with respect to all the second best solutions, provided by a chosen algorithm. Most recently, in [6] the authors did discuss and propose the way in which Kalman filter (KF) could be enhanced in order to reflect the knowledge obtained based on the QADA method, called QADA-KF method.

Taking into account that QADA assumes the reward matrix is known, regardless of the manner in which it is obtained by the user, in this paper we propose and test the performance of two possible versions of QADA-KF. The first one utilizes the assignment matrix, provided by the Global Nearest Neighbour (GNN) method, called QADA-GNN KF approach. The second one utilizes the assignment matrix, provided by the Probabilistic Data Association (PDA) method, called QADA-PDA KF method.

These two QADA-KF methods are compared with the Joint Probabilistic Data Association Filter (JPDAF) [7]–[9] which is an extension of the Probabilistic Data Association Filter (PDAF) [1] to a fixed and known number of targets. JPDAF uses joint association events and joint association probabilities in order to avoid conflicting measurement-to-track assignments by making a soft (probabilistic) assignment of all validated measurements to multiple targets.

The main objective of this paper is to compare the performances of: (i) classical MTT algorithms based on the GNN approach for data association, utilizing Kinematic only Data (KDA) and Converted Measurement Kalman Filter (CMKF); (ii) QADA-GNN KF based MTT; (iii) QADA-PDA KF based MTT; (iiii) JPDAF based MTT. The evaluation is based on a Monte Carlo simulation for particular difficult maneuvering MTT problems in clutter.

This paper is organised as follows. In Section II the JPDAF is described and discussed. Section III is devoted to QADA based KF. Data association methods, providing an assignment matrix for QADA are discussed in Section IV. Two particular simulation MTT scenarios and results are presented for the KDA, QADA-GNN KF, QADA-PDA KF, and JPDAF in Section V. Conclusions are made in Section VI.

## II. JOINT PROBABILISTIC DATA ASSOCIATION FILTER

The Joint Probabilistic Data Association Filter (JPDAF) is an extension of the Probabilistic Data Association Filter (PDAF) for tracking multiple targets in clutter [1], [2], [10]–[12]. This Bayesian tracking filter uses the probabilistic assignment of all validated measurements belonging to the target gate to update its estimate. The preliminary version of JPDAF was proposed by Bar-Shalom in 1974 [13], then updated and finalized in [7]–[9]. The assumptions of JPDAF are the following:

- the number  $N_T$  of established targets in clutter is known;
- all the information available from the measurements  $\mathbf{Z}^k$  up to time  $k$  is summarized by the sufficient statistic  $\hat{\mathbf{x}}^t(k|k)$  (the approximate conditional mean), and covariance  $\mathbf{P}^t(k|k)$  for each target  $t$ ;
- the real state  $\mathbf{x}^t(k)$  of a target  $t$  at time  $k$  is modeled by a Gaussian pdf  $\mathcal{N}(\mathbf{x}^t(k); \hat{\mathbf{x}}^t(k|k), \mathbf{P}^t(k|k))$ ;
- each target  $t$  follows its own dynamic model;
- each target generates at most one measurement at each observation time and there are no merged measurements;
- each target is detected with some known detection probability  $P_d^t$ ;
- the false alarms (FA) are uniformly distributed in surveillance area and their number follows a Poisson pmf with FA density  $\lambda_{\text{FA}}$ .

In JPDAF, the measurement to target association probabilities are computed jointly across the targets and only for the latest set of measurements. This appealing theoretical (0-scan-back) approach however can give rise to very high combinatorics complexity if there are several persistent interferences, typically when several targets are crossing or if they move closely during several consecutive scans. Moreover some track coalescence effects may also appear which degrades substantially the JPDAF performances as it will shown in section V. These limitations of JPDAF have already been reported in [14]. Here we briefly recall the basics of JPDAF. For more details, please refer to [1], [2], [10]–[12], [15].

### A. JPDAF principle

Let's consider a cluster<sup>1</sup> of  $T \geq 2$  targets  $t = 1, \dots, T$ . The set of  $m_k$  measurements available at scan  $k$  is denoted  $\mathbf{Z}(k) = \{\mathbf{z}_i(k), i = 1, \dots, m_k\}$ . Each measurement  $\mathbf{z}_i(k)$  of  $\mathbf{Z}(k)$  either originates from a target or from a FA. Denote  $\hat{\mathbf{z}}^t(k|k-1)$  as the predicted measurement for target  $t$ , and all the possible innovations that could be used in the Kalman Filter to update the target state estimate are denoted  $\tilde{\mathbf{z}}_i^t(k) \triangleq \mathbf{z}_i(k) - \hat{\mathbf{z}}^t(k|k-1)$ ,  $i = 1, \dots, m_k$ . In JPDAF, instead of using a particular innovation  $\tilde{\mathbf{z}}_i^t(k)$ , it uses the weighted innovation  $\tilde{\mathbf{z}}^t(k) = \sum_{i=1}^{m_k} \beta_i^t(k) \tilde{\mathbf{z}}_i^t(k)$ , where  $\beta_i^t(k)$  is the probability that the measurement  $\mathbf{z}_i(k)$  originates from target  $t$ .  $\beta_0^t(k)$  is the probability that none measurements originate from the target  $t$ . The core of JPDAF is the computation of the a posteriori association probabilities  $\beta_i^t(k)$ ,  $i = 0, 1, \dots, m_k$  based on all

<sup>1</sup>A cluster is a group of targets which have some measurements in common in their validation gates (i.e. non-empty intersections).

possible joint association events  $\Theta(k) = \bigcap_{i=1}^{m_k} \Theta_i^{t_i}(k)$ , where  $\Theta_i^{t_i}(k)$  is the event that measurement  $\mathbf{z}_i(k)$  originates from target<sup>2</sup>  $t_i$ ,  $0 \leq t_i \leq N_T$ . More precisely, one has to compute for  $i = 1, \dots, m_k$ ,  $\beta_i^t(k) = \sum_{\Theta(k)} P\{\Theta(k)|\mathbf{Z}^k\} \hat{\omega}_{it}(\Theta(k))$  and  $\beta_0^t(k) = 1 - \sum_{i=1}^{m_k} \beta_i^t(k)$ , where  $\mathbf{Z}^k$  is the set of all measurements available up to time  $k$ , and  $\hat{\omega}_{it}(\Theta(k))$  are the corresponding components of the association matrix characterizing the possible joint association  $\Theta(k)$ .

### B. Feasible joint association events

Validation gates are used for finding the feasible joint events but not in the evaluation of their probabilities [12] (p. 388–389). To describe the observation situation, it uses the validation matrix  $\Omega = [\omega_{it}]$ ,  $i = 1, \dots, m_k$  and  $t = 0, \dots, N_T$  with elements  $\omega_{it} \in \{0, 1\}$  to indicate whether or not the measurement  $\mathbf{z}_i$  lies in the validation gate of target  $t$ . Because each measurement can potentially originate from a FA, all elements of the first column of  $\Omega$  corresponding to index  $t = 0$  (meaning FA, or none of the targets) are equal to one. From this validation matrix, all possible feasible joint association events  $\hat{\Omega}(\Theta(k)) = [\hat{\omega}_{it}(\Theta(k))]$  where  $\hat{\omega}_{it}(\Theta(k)) = 1$  if  $\Theta_i^{t_i}(k) \in \Theta(k)$ , and zero otherwise, are realized satisfying the following *feasibility conditions*:

- a measurement can have only one origin, that is for all  $i$

$$\sum_{t=0}^{N_T} \hat{\omega}_{it}(\Theta(k)) = 1. \quad (1)$$

- at most one measurement can originate from a target

$$\sum_{i=1}^{m_k} \hat{\omega}_{it}(\Theta(k)) \leq 1, \quad \text{for } t = 1, \dots, N_T. \quad (2)$$

The generation of all possible feasible joint association events is computationally expensive for complicated MTT scenarios, which is a serious limitation of JPDAF for real-world scenarios. A simple Matlab<sup>TM</sup> algorithm for the generation of matrices  $\hat{\Omega}(\Theta(k))$  is given in [15] (pp. 56–57), which is based on DFS (Depth First Search) detailed by Zhou in [16], [17], previously coded in FORTRAN in [18].

### C. Feasible joint association probabilities

Thanks to Bayes formula, the computation of the a posteriori joint association probabilities  $P\{\Theta(k)|\mathbf{Z}^k\}$  involved in the derivation of  $\beta_i^t(k)$  can be expressed as (see [1], [2], [10]–[12], [15] for full derivations)

$$\begin{aligned} P\{\Theta(k)|\mathbf{Z}^k\} &= \frac{1}{c} \cdot p[\mathbf{Z}(k)|\Theta(k), m_k, \mathbf{Z}^{k-1}] P\{\Theta(k)|m_k\} \\ &= \frac{1}{c} \cdot \frac{\phi(\Theta(k))!}{m_k!} \mu_F(\phi(\Theta(k))) V^{-\phi(\Theta(k))} \\ &\quad \times \prod_{i=1}^{m_k} [f_{t_i}(\mathbf{z}_i(k))]^{\tau_i(\Theta(k))} \\ &\quad \times \prod_{t=1}^T (P_d^t)^{\delta_t(\Theta(k))} (1 - P_d^t)^{1 - \delta_t(\Theta(k))}, \quad (3) \end{aligned}$$

<sup>2</sup>By convention and notation convenience,  $t_i = 0$  means that the origin of measurement  $\mathbf{z}_i$  is a FA.

where  $c$  is a normalization constant,  $V$  is the volume of the surveillance region, and the indicators  $\delta_t(\Theta(k))$  (target detection indicator),  $\tau_i(\Theta(k))$  (measurement association indicator),  $\phi(\Theta(k))$  (FA indicator) are defined by

$$\delta_t(\Theta(k)) \triangleq \sum_{i=1}^{m_k} \hat{\omega}_{it}(\Theta(k)) \leq 1 \quad t = 1, \dots, N_T, \quad (4)$$

$$\tau_i(\Theta(k)) \triangleq \sum_{t=1}^T \hat{\omega}_{it}(\Theta(k)), \quad (5)$$

$$\phi(\Theta(k)) \triangleq \sum_{i=1}^{m_k} [1 - \tau_i(\Theta(k))]. \quad (6)$$

$\mu_F(\phi(\Theta(k)))$  is the prior pmf of the number of false measurements (the clutter model) and

$$f_{t_i}(\mathbf{z}_i(k)) \triangleq \mathcal{N}[\mathbf{z}_i(k); \hat{\mathbf{z}}^{t_i}(k|k-1), \mathbf{S}^{t_i}(k)], \quad (7)$$

where  $\mathbf{S}^{t_i}(k)$  is the predicted covariance matrix of innovation  $\mathbf{z}_i(k) - \hat{\mathbf{z}}^{t_i}(k|k-1)$ .

Two versions of JPDAF have been proposed [1], [7]–[9]:

- **Parametric JPDAF:** Knowing the spatial density  $\lambda_{\text{FA}}$  of the false measurements, and using a Poisson pmf  $\mu_F(\phi(\Theta(k))) = \frac{(\lambda_{\text{FA}}V)^{\phi(k)}}{\phi(k)!} e^{-\lambda_{\text{FA}}V}$ , results in

$$P\{\Theta(k)|\mathbf{Z}^k\} = \frac{1}{c_1} \cdot \prod_{i=1}^{m_k} [\lambda_{\text{FA}}^{-1} \cdot f_{t_i}(\mathbf{z}_i(k))]^{\tau_i(\Theta(k))} \times \prod_{t=1}^{N_T} [P_d^t]^{\delta_t(\Theta(k))} [1 - P_d^t]^{1-\delta_t(\Theta(k))}, \quad (8)$$

where  $c_1$  is a normalization constant.

- **Non parametric JPDAF:** Using a diffuse prior pmf of number of FA  $\mu_F(\phi(k)) = \epsilon$ ,  $\forall \phi(k)$ , results in

$$P\{\Theta(k)|\mathbf{Z}^k\} = \frac{\phi(k)!}{c_2} \cdot \prod_{i=1}^{m_k} [V f_{t_i}(\mathbf{z}_i(k))]^{\tau_i(\Theta)} \times \prod_{t=1}^{N_T} [P_d^t]^{\delta_t(\Theta(k))} [1 - P_d^t]^{1-\delta_t(\Theta(k))}, \quad (9)$$

where  $c_2$  is a new normalization constant.

#### D. JPDAF state estimation

Once all feasible joint association events  $\Theta(k)$  have been generated and their a posteriori probabilities  $P\{\Theta(k)|\mathbf{Z}^k\}$  determined, all the marginal association probabilities  $\beta_i^t(k) = \sum_{\Theta(k)} P\{\Theta(k)|\mathbf{Z}^k\} \hat{\omega}_{it}(\Theta(k))$  and  $\beta_0^t(k) = 1 - \sum_{i=1}^{m_k} \beta_i^t(k)$  are computed. The state update and prediction are done with PDAF equations<sup>3</sup> given by

$$\hat{\mathbf{x}}^t(k|k) = \sum_{i=0}^{m_k} \beta_i^t(k) \hat{\mathbf{x}}_i^t(k|k), \quad (10)$$

<sup>3</sup>for the decoupled version of JPDAF. For the coupled version of JPDAF, see [10], [12].

with  $\hat{\mathbf{x}}_i^t(k|k)$  given by

$$\hat{\mathbf{x}}_{i>0}^t(k|k) = \hat{\mathbf{x}}^t(k|k-1) + \mathbf{K}^t(k) \tilde{\mathbf{z}}_i^t(k), \quad (11)$$

$$\hat{\mathbf{x}}_{i=0}^t(k|k) = \hat{\mathbf{x}}^t(k|k-1). \quad (12)$$

Using (11) and (12) in (10), then

$$\hat{\mathbf{x}}^t(k|k) = \hat{\mathbf{x}}^t(k|k-1) + \mathbf{K}^t(k) \sum_{i=1}^{m_k} \beta_i^t(k) \tilde{\mathbf{z}}_i^t(k), \quad (13)$$

$$\mathbf{P}^t(k|k) = \beta_0^t(k) \mathbf{P}^t(k|k-1) + (1 - \beta_0^t(k)) \mathbf{P}_c^t(k) + \tilde{\mathbf{P}}^t(k), \quad (14)$$

with

$$\mathbf{P}_c^t(k) = [\mathbf{I} - \mathbf{K}^t(k) \mathbf{H}(k)] \mathbf{P}^t(k|k-1), \quad (15)$$

$$\tilde{\mathbf{P}}^t(k) = \mathbf{K}^t(k) \left[ \sum_{i=1}^{m_k} \beta_i^t(k) \tilde{\mathbf{z}}_i^t(k) \tilde{\mathbf{z}}_i^{t'}(k) - \tilde{\mathbf{z}}(k) \tilde{\mathbf{z}}'(k) \right] \mathbf{K}'(k), \quad (16)$$

and

$$\mathbf{K}^t(k) \triangleq \mathbf{P}^t(k|k-1) \mathbf{H}'(k) \mathbf{S}^t(k)^{-1}, \quad (17)$$

$$\tilde{\mathbf{z}}_i^t(k) \triangleq \mathbf{z}_i(k) - \hat{\mathbf{z}}^t(k|k-1), \quad (18)$$

$$\tilde{\mathbf{z}}^t(k) \triangleq \sum_{i=1}^{m_k} \beta_i^t(k) \tilde{\mathbf{z}}_i^t(k). \quad (19)$$

It has been proved in [1] that  $\tilde{\mathbf{P}}^t(k)$  is always a semi-positive matrix. The target state prediction  $\hat{\mathbf{x}}^t(k+1|k)$  and  $\mathbf{P}^t(k+1|k)$  are obtained by the classical Kalman Filter (KF) equations [1] (assuming linear kinematic models), or by Extended KF equations. They will not be repeated here [2], [10].

In summary, JPDAF is well theoretically founded and it does not require high memory (0-scan-back). It provides pretty good results on simple MTT scenarios (with non persisting interferences) with moderate FA densities. However the number of feasible joint association matrices increases exponentially with problem dimensions ( $m_k$  and  $N_T$ ) which makes the JPDAF intractable for complex dense MTT scenarios.

### III. QADA BASED KALMAN FILTER

The aim of this paper is to compare the performance of the JPDAF based MTT algorithm with the classical<sup>4</sup> MTT algorithm, using the CMKF based on kinematics measurements, but improved by the QADA method.

The main idea behind the QADA method, proposed recently by Dezert and Benameur [4] is to compare the values  $a_1(i, j)$  in the first optimal DA solution  $\mathbf{A}_1$  with the corresponding values  $a_2(i, j)$  in second assignment solution  $\mathbf{A}_2$ , and to identify if there is a change of the optimal pairing  $(i, j)$ . In the MTT context  $(i, j)$  means an association between measurement  $\mathbf{z}_j$  and target  $T_i$ . QADA establishes a quality indicator associated with this pairing, depending on the stability of the pairing and also, on its relative impact in the global reward. The proposed method works also when the 1<sup>st</sup> and 2<sup>nd</sup> optimal assignments  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are not unique, i.e., there are

<sup>4</sup>Classical MTT algorithms are those based on hard assignment of a chosen measurement to a given target.

multiplicities available. In such a situation, the establishment of quality indicators could help in selecting one particular optimal assignment solution among multiple possible choices.

The construction of the quality indicator is based on belief functions (BF) and the Proportional Conflict Redistribution fusion rule no.6 (PCR6), defined within Dezert-Smarandache Theory (DSMT) [19]. It depends on the type of the pairing matching, and it is described in detail in [4].

In [6], the authors discuss and propose the way in which Kalman filter could be improved in order to reflect the knowledge obtained based on the QADA method.

Let's briefly recall what kind of information is obtained, having in hand the quality matrix, derived by QADA, in the MTT context. It gives knowledge about the confidence  $q(i, j)$  in all pairings  $(T_i, \mathbf{z}_j)$ ,  $i = 1, \dots, m$ ;  $j = 1, \dots, n$ , chosen in the first best assignment solution. The smaller quality (confidence) of hypothesis “ $\mathbf{z}_j$  belongs to  $T_i$ ” means, that the particular measurement error covariance  $\mathbf{R}$  was increased and the filter should not trust fully in the actual (true) measurement  $\mathbf{z}(k+1)$ .

Having this conclusion in mind, the authors propose, such a behaviour of the measurement error covariance to be modelled by  $\mathbf{R} = \frac{\mathbf{R}}{q(T_i, \mathbf{z}_j)}$ , for every pairing, chosen in the first best assignment and based on the corresponding quality value obtained. Then, when the Kalman filter gain decreases the true measurement  $\mathbf{z}_j(k+1)$  is trusted less in the updated state estimate  $\hat{\mathbf{x}}(k+1|k+1)$ .

#### IV. BUILDING ASSIGNMENT MATRIX FOR QADA

Data Association (DA) is a central problem in the modern MTT systems [1], [2]. It consists in finding the global optimal assignments of targets  $T_i$ ,  $i = 1, \dots, m$  to some measurements  $\mathbf{z}_j$ ,  $j = 1, \dots, n$  at a given time  $k$  by maximizing the overall gain in such a way, that no more than one target is assigned to a measurement, and reciprocally. The  $m \times n$  reward (gain/payoff) matrix  $\Omega = [\omega(i, j)]$  is defined by its elements  $\omega(i, j) > 0$ , representing the gain of the association of target  $T_i$  with the measurement  $\mathbf{z}_j$ .

These values are usually homogeneous to the likelihood ratios and could be established in different ways, described below. They provide the assignment matrix utilized by QADA in order to obtain the quality of pairings (interpreted as a confidence score) belonging to the optimal data assignment solution based on its consistency (stability) with respect to all the second best solutions, provided for a chosen algorithm.

QADA assumes the reward matrix is known, regardless of the manner in which it is obtained by the user. In this paper we propose two versions of QADA-KF. The first one utilizes the assignment matrix built from the single normalized distances, provided by the Global Nearest Neighbour method, called QADA-GNN KF method. The second one utilizes the assignment matrix, built from the posterior association probabilities, provided by the Probabilistic Data Association (PDA) method, called QADA-PDA KF method.

##### A. Assignment matrix based on GNN method

The GNN method finds and propagates the single most likely hypothesis during each scan to update KF. It is a

hard (i.e., binary) decision approach, as compared to the JPDAF which is a soft (i.e., probabilistic) decision approach using all validated measurements with their probabilities of association. GNN method was applied in [6] and [20] to obtain the assignment matrix, utilized in QADA. In this case the elements of assignment matrix  $\omega(i, j)$ ,  $i = 1, \dots, m$ ;  $j = 1, \dots, n$  represent the normalized distances  $d(i, j) \triangleq [(\mathbf{z}_j(k) - \hat{\mathbf{z}}_i(k|k-1))' \mathbf{S}^{-1}(k)(\mathbf{z}_j(k) - \hat{\mathbf{z}}_i(k|k-1))]^{1/2}$  between the validated measurement  $\mathbf{z}_j$  and target  $T_i$  satisfying the condition  $d^2(i, j) \leq \gamma$ . The distance  $d(i, j)$  is computed from the measurement  $\mathbf{z}_j(k)$  and its prediction  $\hat{\mathbf{z}}_i(k|k-1)$  (see [1] for details), and the inverse of the covariance matrix  $\mathbf{S}(k)$  of the innovation, computed by the tracking filter. The threshold  $\gamma$ , for which the probability of given observation to fall in the gate is 0.99, could be defined from the table of the Chi-square distribution with  $M$  degrees of freedom and allowable probability of a valid observation falling outside the gate. In this case the DA problem consists in finding the best assignment, that minimizes the overall cost.

##### B. Assignment matrix based on PDA method

The Probabilistic Data Association (PDA) method [1] calculates the association probabilities for validated measurements at a current time moment to the target of interest. PDA assumes the following hypotheses according to each validated measurement:

- $H_i(k)$ :  $\mathbf{z}_j(k)$  is a measurement, originated from the target of interest,  $i = 1, \dots, m$
- $H_0(k)$ : no one of the validated measurement originated from the target of interest

If  $N$  observations fall within the gate of track  $i$ ,  $N + 1$  hypotheses will be formed.

The probability of  $H_0$  is proportional to  $p_{i0} = \lambda_{\text{FA}}^N (1 - P_g P_d)$ , and the probability of  $H_j$  ( $j = 1, 2, \dots, N$ ) is proportional to

$$p_{ij} = \frac{\lambda_{\text{FA}}^{N-1} P_g P_d \cdot e^{-\frac{d_{ij}^2}{2}}}{(2\pi)^{M/2} \cdot \sqrt{|\mathbf{S}_{ij}|}}, \quad (20)$$

where  $P_g$  is the a priori probability that the correct measurement is in the validation gate<sup>5</sup> [1];  $P_d$  is the target detection probability;  $\lambda_{\text{FA}}$  is the spatial density of FA. The probabilities  $p_{ij}$  can be rewritten as [1]

$$p_{ij} = \begin{cases} \frac{b}{b + \sum_{i=1}^N \alpha_{i0}} & \text{for } j = 0 \text{ (no valid observ.)}, \\ \frac{\alpha_{ij}}{b + \sum_{i=1}^N \alpha_{i0}} & \text{for } 1 \leq j \leq N, \end{cases} \quad (21)$$

where

$$b \triangleq (1 - P_g P_d) \lambda_{\text{FA}} (2\pi)^{M/2} \sqrt{|\mathbf{S}_{ij}|}, \quad (22)$$

and

$$\alpha_{ij} \triangleq P_d \cdot e^{-\frac{d_{ij}^2}{2}}. \quad (23)$$

<sup>5</sup>In our simulations, we use  $P_g = 0.99$  and  $P_d = 0.99$ .

The assignment matrix used in QADA method is established from all  $p_{ij}$  given by (21) related with all association hypotheses. This matrix will have  $m$  rows (where  $m$  is the number of all targets of interest), and  $N + 1$  columns for the hypotheses generated. The  $(N + 1)$ th column will include the values  $p_{i0}$  associated with  $H_0(k)$ .

## V. SIMULATION SCENARIOS AND RESULTS

The Converted Measurement KF is used in our MTT algorithm. We assume constant velocity target model. The process noise covariance matrix is:  $Q = \sigma_v^2 Q_T$ , where  $T$  is the sampling period,  $\sigma_v$  is the standard deviation of the process noise and  $Q_T$  is as given in [3]. Here are the results of KDA KF, QADA-GNN KF, QADA-PDA KF, and JPDAF for two interesting MTT scenarios.

### A. Groups of targets simulation scenario

The noise-free groups of targets simulation scenario (Fig.1) consists of five air targets moving from North-West to South-East. For the clear explanation of the results, targets are numbered starting at the beginning with 1st target that has the greater y-coordinate and continuing to 5th target with the smallest y-coordinate. The three targets 2nd, 3rd, and 4th move together between them<sup>6</sup>. The stationary sensor is located at the origin with range 20000 m. The sampling period is  $T_{scan} = 5$  sec and the measurement standard deviations are 0.2 deg and 35 m for azimuth and range respectively. The targets move with constant velocity  $V = 100m/sec$ . The group of three targets in the middle i.e. 2nd to 4th move without maneuvering keeping azimuth 135 deg from North. It is the main direction of the group's movement. The first target starts with azimuth 165 deg and moves towards the middle group of rectilinearly moving targets. When it approaches the group, it starts a turn to the left with  $-30$  deg. Its initial azimuth of 165 deg is decreased by the angle of turn and becomes 135 deg, the main direction. The fifth target makes similar maneuver but in opposite direction - to the right. Its initial azimuth of 105 deg is increased by the turn of 30 deg and becomes 135 deg, and also coincides with the main direction. From 21th scan to 48th scan all the targets move rectilinearly in parallel. The distance between them is 150 m. From 48th scan, the first target makes a left turn to azimuth of 105 deg, that means  $-30$  degrees with respect to the main direction and starts to go away from the middle group. The fifth target makes right turn to azimuth of 165 deg that means  $+30$  deg from the main direction and also starts to go away. All maneuvers are with one and the same value of the angle (angle= 30 deg by absolute value), the same time duration and linear velocity. The absolute value of the corresponding transversal acceleration for all maneuvers is  $1.163m/s^2$ . The total number of scans for the simulations is 65. Fig. 2 shows the noised scenario for  $\lambda_{FA} = 16 \cdot 10^{-10}m^{-2}$  yielding to 0.2 FA per gate on average.

Our results are based on Monte Carlo (MC) simulations with 200 independent runs in applying KDA based KF, QADA-

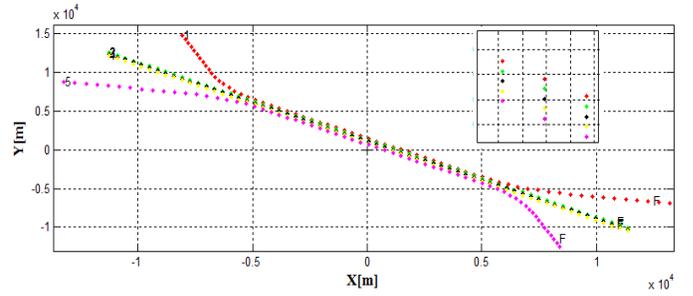


Figure 1. Noise-free groups of targets Scenario.

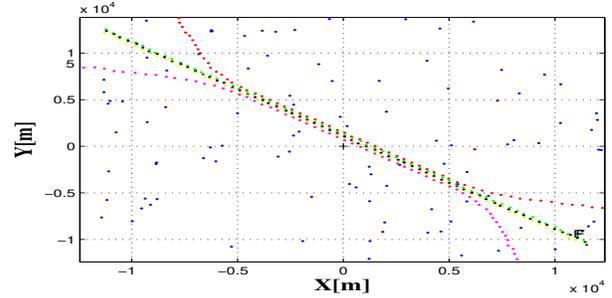


Figure 2. Noised groups of targets Scenario with  $\lambda_{FA} = 16 \cdot 10^{-10}m^{-2}$ .

GNN KF, QADA-PDA KF, and JPDAF<sup>7</sup>. We compare the performance of these methods with different criteria, and we use an idealized track initiation in order to prevent uncontrolled impact of this stage on the statistical parameters of the tracking process during MC simulations. The true targets positions (known in our simulations) for the first two scans are used for track initiation. The evaluation of MTT performance is based on the criteria of Track Purity (TP), Track Life (TL), and percentage of miscorrelation (pMC):

1) TP criteria examines the ratio between the number of particular performed ( $j$ th observation -  $i$ th track) associations (in case of detected target) over the total number of all possible associations during the tracking scenario, but TP cannot be used with JPDAF because JPDAF is a soft assignment method. Instead of TP, we define the Probabilistic Purity Index (PPI). It considers the measurement that has the highest association probability computed by the JPDAF and check, (and count) if this measurement originated from the target or not. PPI measures the ability of JPDAF to commit the highest probability to the correct target measurement in the soft assignment of all validated measurements.

2) TL is evaluated as an average number of scans before track's deletion. In our simulations, a track is cancelled and deleted from the list of tracked tracks, when during 3 consecutive scans it cannot be updated with some measurement because there is no validated measurement in the validation gate. When using JPDAF, the track is cancelled and deleted from the list of tracked tracks, when during 3 consecutive scans its own measurement does not fall in its gate. We call this,

<sup>6</sup>Note that three targets move together in the center.

<sup>7</sup>We have used the non parametric version of JPDAF in our simulations.

the “cancelling/deletion condition”. The status of the tracked tracks is denoted “alive”.

3) pMC examines the relative number of incorrect observation-to-track associations during the scans.

The MTT performance results for KDA only KF, QADA-GNN KF, QADA-PDA KF, and JPDAF for a low-noise case (0.2 FA per gate on average) are given in Table 1.

(in %)	KDA	JPDAF	QADA-GNN	QADA-PDA
Average TL	50.27	66.46	81.94	90.85
Average pMC	3.35	2.98	2.10	1.75
Average TP	45.61	PPI=29.14	79.32	87.61

Table 1

GROUPS OF TARGETS SCENARIO: COMPARISON BETWEEN MTT PERFORMANCE RESULTS FOR 0.2 FA PER GATE.

According to all criteria, the QADA-PDA KF method shows the best performance, followed by QADA-GNN KF, and JPDAF. The KDA based KF approach, as one could expect, shows the worst performance. Performance results for a more noisy scenario with 0.4 FA per gate on average are given in Table 2.

(in %)	KDA	QADA-GNN	JPDAF	QADA-PDA
Average TL	43.54	70.51	70.94	84.17
Average pMC	3.90	3.33	2.71	3.11
Average TP	38.22	66.43	PPI=25.65	78.51

Table 2

GROUPS OF TARGETS SCENARIO: COMPARISON BETWEEN MTT PERFORMANCE RESULTS FOR 0.4 FA PER GATE.

As we see, the results for 0.4 FA per gate scenario are degraded in comparison to the low-noise case. The average miscorrelation for QADA-PDA is slightly higher than for JPDAF, probably because QADA method is based on the 1<sup>st</sup> and 2<sup>nd</sup> best solutions only, and more information (i.e. the 3<sup>rd</sup> best assignment solution) should be used in such case to improve QADA performance, which is left for further research. According to TL and TP, still QADA-PDA KF based MTT shows stably better performance than JPDAF.

JPDAF based MTT outperforms QADA-GNN KF and KDA KF based MTT approaches according to the considered criteria. In order to make a fair comparison between QADA KF and JPDAF, we will discuss also the root mean square errors (RMSE), associated with the filtered X and Y values, presented in Figs. 3–7. Figs. 3 and 4 show the mean square X and Y error filtered, associated with target 1, and compared for KDA KF, QADA-GNN KF, QADA-PDA KF, and JPDAF. Figs. 5 and 6 consider the same errors for the middle of track 3. All the results are compared to the sensor’s errors along X and Y axis. We see that the RMSE on X filtered associated with KDA KF, QADA-GNN KF, and QADA-PDA KF are a little bit above from the sensor’s error in the region where target 1 makes maneuvers. For scans [20, 50], target 1 is moving in parallel to the group of other targets running rectilinearly and then these errors are less than respective sensors’s ones. The RMSE on X filtered associated with JPDAF performance is three times bigger in the region between scans 20th and 30th where target 1 starts moving in parallel to the rest of

rectilinearly moving targets. The RMSE on Y filtered is high during the whole region, where target 1 moves in parallel way.

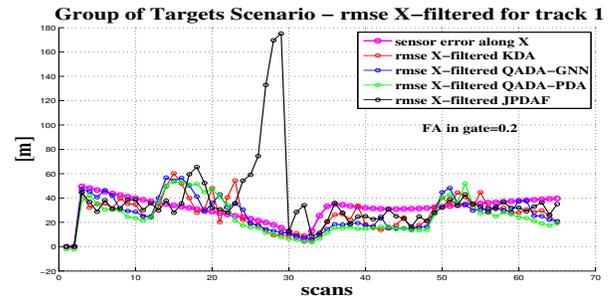


Figure 3. RMSE on X for track 1 with the four tracking methods.

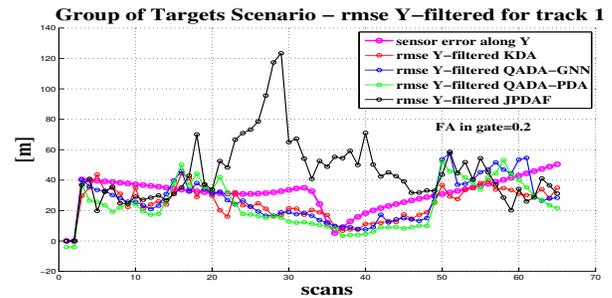


Figure 4. RMSE on Y for track 1 with the four tracking methods.

The RMSE on Y error filtered by JPDAF are especially critical for the middle track 3 which shows its poor performance in state estimation on Y direction. The RMSEs are more than 5 times bigger (in the region between scans 20th and 50th, where all five targets move in parallel) than the respective errors obtained by KDA KF, QADA-GNN KF, QADA-PDA KF, which are less than the sensor’s error. The RMSE on X filtered obtained with JPDAF is under the sensor’s error, beside KDA KF, QADA-GNN KF, QADA-PDA KF methods.

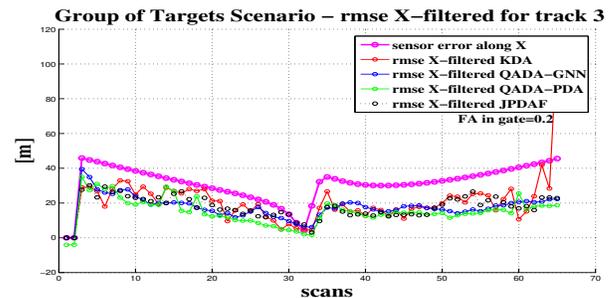


Figure 5. RMSE on X for track 3 with the four tracking methods.

The large value of RMSE on Y using the JPDAF can be explained by the specificity of the scenario because it has five targets moving closely during more than 30 consecutive scans with sensor’s measurement errors, and false alarms density, which yields to spatial persisting interferences and track coalescence effects in JPDAF, as shown in Fig. 7, where

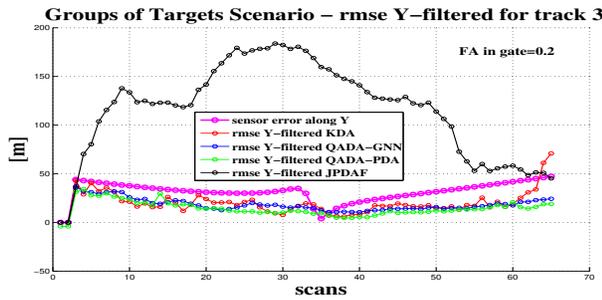
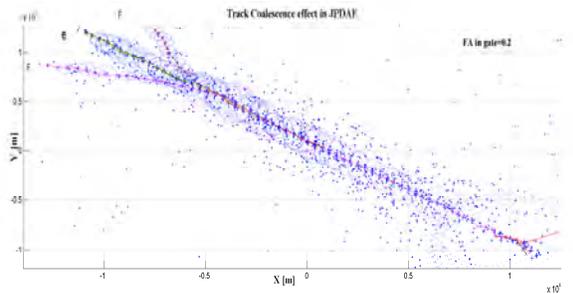


Figure 6. RMSE on Y for track 3 with the four tracking methods.

the red and green plots are the tracks estimates. These effects degrades significantly the quality of JPDAF performance as already reported in [14].


 Figure 7. JPDAF track coalescence (one run) with  $\lambda_{FA} = 16 \cdot 10^{-10} m^{-2}$ .

### B. Crossing targets simulation scenario

The second considered (crossing targets) scenario (Fig. 8) consists of two maneuvering targets moving with constant velocity  $38m/sec$ . At the beginning, both targets move from West to East. The stationary sensor is located at the origin with range 1200 m. The sampling period is  $T_{scan} = 1sec$  and the measurement standard deviations are 0.2 deg and 25 m for azimuth and range respectively.

The first target, having at the beginning greater y-coordinate, moves straightforward from West to East. Between the 8th and 12th scans it makes a 50 deg right turn, and then it moves straightforward during 8 scans. From the 20th scan to the 24th scan it makes a 50 deg left turn, and then it moves in East direction till the 41th scan. It makes a second 50 deg left turn between 41th and 45th scans, and then it moves straightforward during 8 scans. From 53th scan it makes a second 50 deg right turn till the 57th scan and then it moves in East direction. The trajectory of target 1 corresponds to the red plot of Fig. 8.

The second target makes a mirrored trajectory corresponding to the green plot of Fig. 8. From scan 1 to 8 it moves from West to East. During 8th to 12th scans it makes a 50 deg left turn. Then it moves straightforward during 8 scans. During 20th to 24th scans it makes a 50 deg right turn and then it moves in East direction till the 41th scan. It makes a second 50 deg right turn between the 41th and 45th scans, and then

it moves straightforward during 8 scans. From the 53th scan it makes a second 50 deg left turn till the 57th scan and then it moves in East direction. The total number of scans for the simulations is 65. Fig. 9 shows the respective noised scenario for  $\lambda_{FA} = 4 \cdot 10^{-7} m^{-2}$ .

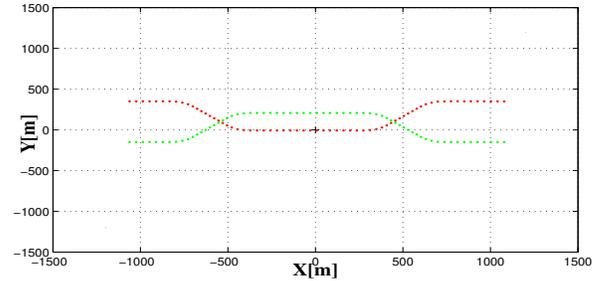
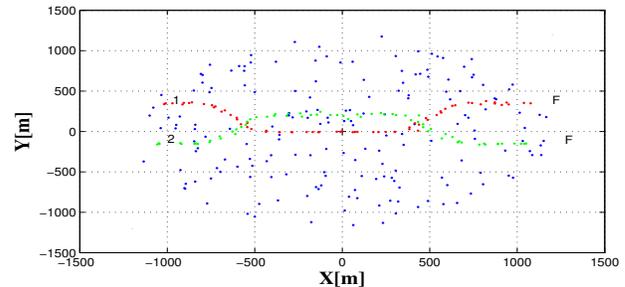


Figure 8. Noise-free Crossing targets Scenario.


 Figure 9. Noised Crossing targets Scenario with  $\lambda_{FA} = 4 \cdot 10^{-7} m^{-2}$ .

The MTT performance results obtained on the base of KDA only KF, QADA-GNN KF, QADA-PDA KF, and JPDAF for less noised case corresponding to 0.2 FA per gate are given in Table 3, and the performance results for a more noisy scenario with 0.4 FA per gate on average are given in Table 4.

(in %)	KDA	QADA-GNN	JPDAF	QADA-PDA
Average TL	77.06	88.93	91.25	93.47
Average pMC	2.40	2.24	2.08	2.11
Average TP	72.78	85.64	PPI=86.29	87.96

Table III

CROSSING TARGETS SCENARIO: COMPARISON BETWEEN MTT PERFORMANCE RESULTS FOR FA IN GATE = 0.2.

(in %)	KDA	QADA-GNN	JPDAF	QADA-PDA
Average TL	58.80	77.20	82.87	83.18
Average pMC	3.61	3.63	2.94	3.40
Average TP	52.90	72.01	PPI=76.94	77.15

Table IV

CROSSING TARGETS SCENARIO: COMPARISON BETWEEN MTT PERFORMANCE RESULTS FOR FA IN GATE = 0.4.

According to all criteria, the QADA-PDA KF shows again the best performance, but now JPDAF based MTT shows closed to QADA-PDA KF performance in comparison to the previous scenario, and exceeds the performance of QADA-GNN KF. Nevertheless the performances of all methods are

deteriorated in more noised case, when one has 0.4 FA in gate on average, this tendency is still kept. JPDAF has better (than in the previous scenario) performance, but still QADA-PDA KF exceeds its performance.

Figures 10-13 show that the RMS errors associated with X and Y filtered are below the sensor's error. They confirm the better performance of JPDAF in this particular scenario with only two maneuvering targets, which is simpler than the groups of targets scenario.

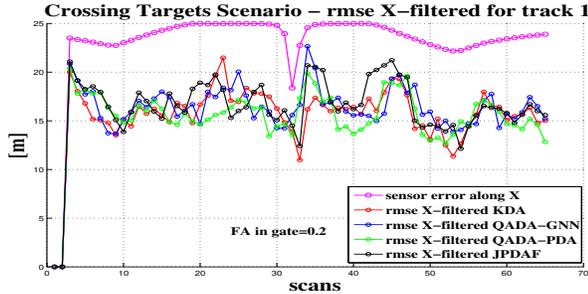


Figure 10. RMSE on X for track 1 with the four tracking methods.

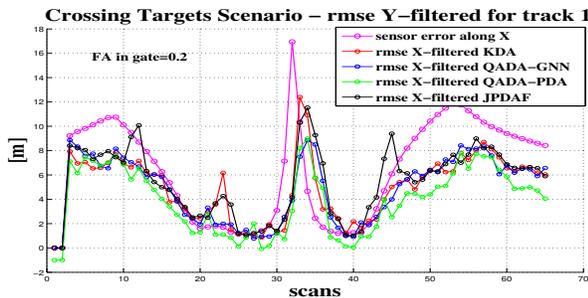


Figure 11. RMSE on Y for track 1 with the four tracking methods.

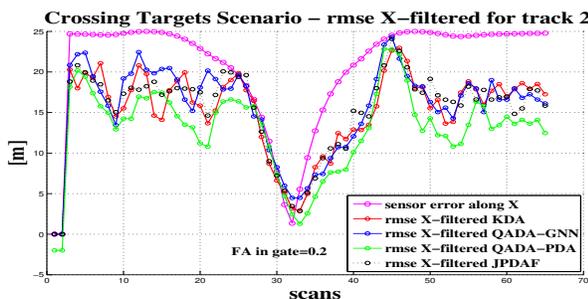


Figure 12. RMSE on X for track 2 with the four tracking methods.

## VI. CONCLUSIONS

This work evaluated with Monte Carlo simulations the efficiency of MTT performance in cluttered environment of four methods (i) classical MTT algorithm based on the GNN approach for data association, utilizing Kinematic only Data based Kalman Filter; (ii) QADA-GNN KF; (iii) QADA-PDA

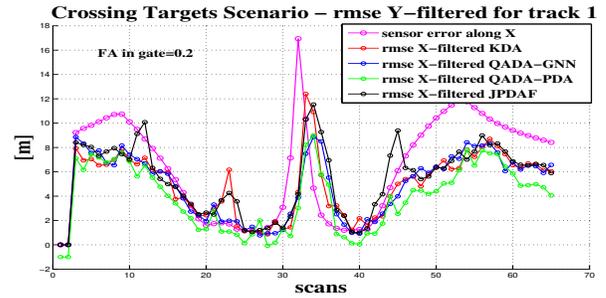


Figure 13. RMSE on Y for track 2 with the four tracking methods.

KF; and (iii) JPDAF. The first scenario (groups of targets) shows the advantages of applying QADA-KF. According to all performance criteria, the QADA-PDA KF gives the best performance, followed by QADA-GNN KF, and JPDAF. The KDA KF approach shows the worst performance (as expected). This scenario is particularly difficult for JPDAF because of several closely spaced and rectilinearly moving targets in clutter during many consecutive scans, and it leads to track coalescence effects due to persisting interferences. As a result, the tracking performance of JPDAF is degraded. Because the complexity of the calculation for joint association probabilities grows exponentially with the number of targets, JPDAF requires almost 3 times more computational time in comparison to other methods in the first (complex) scenario. In the second (only two crossing targets) MTT scenario, JPDAF shows better tracking performances in comparison to QADA-GNN KF. It is able to track more precisely these only two targets, because of non persisting interferences. Overall, our analysis shows that QADA-PDA KF method is the best of the four approaches to track multiple targets in clutter with a tractable complexity.

## REFERENCES

- [1] Y. Bar-Shalom, T.E. Fortmann, *Tracking and Data Association*, Academic Press, 1988.
- [2] Y. Bar-Shalom (Ed.), *Multitarget-Multisensor Tracking: Advanced Applications*, Artech House, Norwood, USA, 1990.
- [3] S. Blackman, R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.
- [4] J. Dezert, K. Benameur, *On the Quality of Optimal Assignment for Data Association*, Springer, L.N. in Compt. Sci., Vol.8764, pp. 374–382, 2014.
- [5] J. Dezert et al., *On the Quality Estimation of Optimal Multiple Criteria Data Association Solutions*, Proc. of Fusion 2015, July 2015.
- [6] J. Dezert, et al., *Multitarget Tracking Performance based on the Quality Assessment of Data Association*, Proc. of Fusion 2016, July 2016.
- [7] Y. Bar-Shalom, T.E. Fortmann, M. Scheffe, *JPDA for Multiple Targets in Clutter*, Proc. Conf. on Inf. Sci. and Syst., Princeton, March 1980.
- [8] T.E. Fortmann et al., *Multitarget Tracking using Joint Probabilistic Data Association*, in Proc. of IEEE CDC, Albuquerque, USA, Dec. 1980.
- [9] T.E. Fortmann, et al., *Sonar Tracking of Multiple Targets Using JPDAF*, IEEE Journal of Oceanic Eng., Vol. 8, No.3, pp. 173–184, July 1983.
- [10] Y. Bar-Shalom, X.-R. Li. *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, CT, USA, 1995.
- [11] Y. Bar-Shalom et al., *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, John Wiley & Sons, 2001.
- [12] Y. Bar-Shalom et al., *Tracking and Data Fusion: A Handbook of Algorithms*, YBS Publishing, Storrs, CT, USA, 2011.
- [13] Y. Bar-Shalom, *Extension of the PDAF to Multitarget Environments*, Proc. 5th Symp. Nonlinear Estimation, San Diego, USA, Sept. 1974.
- [14] R.J. Fitzgerald, *Track Biases and Coalescence with Probabilistic Data Association*, IEEE Trans. on AES, Vol. 21, No. 6, pp. 822–825, 1985.

- [15] J. Dezert, *Introduction au pistage multi-cibles multi-senseurs*, ENSTA Course (in French), Sept. 2003.
- [16] B. Zhou, *Multitarget Tracking in Clutter: Algorithms for Data Association and State Estimation*, Ph.D. Thesis, Penn. State Univ., 1992.
- [17] B. Zhou, N.K. Bose, *Multitarget Tracking in Clutter: Fast Algorithms for Data Association*, IEEE Trans. on AES, Vol. 29, No. 2, April 1993.
- [18] J. Dezert, *Poursuite Multi-Cibles Mono-Senseur*, ONERA Tech. Note 1988-10 (in French), France, 1988.
- [19] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, Volumes 1, 2, 3 & 4, ARP, 2004–2015.  
<http://www.onera.fr/staff/jean-dezert?page=2>
- [20] J. Dezert et al., *The Impact of the Quality Assessment of Optimal Assignment for Data Association in Multitarget Tracking Context*, Cybernetics and Inf. Techn. J., Vol.15, No.7, pp. 88–98, 2015.



# Multi-Criteria Decision-Making with Imprecise Scores and BF-TOPSIS

Jean Dezert<sup>a</sup>, Deqiang Han<sup>b</sup>, Jean-Marc Tacnet<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>CIESR, Xi'an Jiaotong University, Xi'an, China.

<sup>c</sup>UGA, Irstea, UR ETGR, St-Martin-d'Hères, France.

Emails: jean.dezert@onera.fr, deqhan@xjtu.edu.cn, tacnet@irstea.fr

Originally published as: J. Dezert, D. Han, J.-M. Tacnet, *Multi-Criteria Decision-Making with Imprecise Scores and BF-TOPSIS*, in Proc. of 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10–13, 2017, and reprinted with permission.

**Abstract**—In 2016 we developed a new approach for Multi-Criteria Decision-Making (MCDM) inspired by the technique for order preference by similarity to ideal solution (TOPSIS) and based on belief functions (BF). Our BF-TOPSIS (Belief Function based TOPSIS) approach assumes that the input score of each hypothesis for each criterion was a real precise number which is a quite restrictive assumption. In this paper we extend our BF-TOPSIS to deal with imprecise score values (intervals of real numbers) and we call it Imp-BF-TOPSIS. This new approach follows main ideas of BF-TOPSIS but extends its applicability for more realistic MCDM problems where the scores are given with a finite precision. Imp-BF-TOPSIS is based on Interval Arithmetic (IA), new probabilistic order relations between intervals and belief functions. We also present results of Imp-BF-TOPSIS for simple examples for illustrating its effectiveness.

**Keywords:** Information fusion, multi-criteria decision-making, MCDM, belief functions, TOPSIS.

## I. INTRODUCTION

The Multi-Criteria Decision-Making (MCDM) aims to choose an alternative among a known set of alternatives based on their quantitative or qualitative evaluations (scores) obtained with respect to different criteria. MCDM can be considered as a decision-level information fusion, and it has been widely used in many decision-making applications. In classical MCDM problem, all the criteria and all alternatives are known, and the score values are usually real numbers (precisely known). Depending on the context of the MCDM problem, the score can be interpreted either as a cost/expense or as a reward/benefit. In the sequel, by convention and without loss of generality we will interpret the score as a reward having monotonically increasing preference. Thus, the best alternative with respect to a given criteria will be the one providing the highest reward/benefit. The set of score values is represented by a quantitative benefit or payoff matrix. Each criterion can also have a relative importance weight. Many methods have been proposed in the literature to solve the classical MCDM [1]. When the score values are incomplete or imprecise (quantitative or qualitative), traditional approaches for classical MCDM problems do not work. In this paper, we focus on these unclassical MCDM problems. We propose to extend the BF-TOPSIS approach to deal with imprecise score values to cover a broader spectrum of real MCDM

applications. We use the theory of belief functions and the interval arithmetic. This extension of BF-TOPSIS method is referred as Imp-BF-TOPSIS method in the sequel, where *Imp* is an abbreviation standing for *Imprecise* to specify that the BF-TOPSIS will work with imprecise score values (or more generally with imprecise basic belief assignments (BBAs)). The rest of this paper is organized as follows. In section II, the formulation of classical MCDM problem is provided. In Section III, we introduce Interval Arithmetic and propose new (probabilistic) order relations for intervals as well as distances between intervals. In section IV, basics of belief functions are recalled. In section V we recall the principle of BF-TOPSIS for classical (precise scores) MCDM. The Imp-BF-TOPSIS for imprecise score values is presented in section VI, with simple examples in section VII. Section VIII concludes this paper.

## II. FORMULATION OF CLASSICAL MCDM

A classical MCDM problem has a given set of alternatives  $\mathbf{A} \triangleq \{A_1, A_2, \dots, A_M\}$  ( $M > 2$ ), and a given set of criteria  $\mathbf{C} \triangleq \{C_1, C_2, \dots, C_N\}$  ( $N \geq 1$ ). Each alternative  $A_i$  represents a possible choice (a possible decision to make). In a general context, each criterion is also characterized by a relative importance weighting factor  $w_j \in [0, 1]$ ,  $j = 1, \dots, N$  which are normalized by imposing the condition  $\sum_j w_j = 1$ . The set of normalized weighting factors is denoted by  $\mathbf{w} \triangleq \{w_1, w_2, \dots, w_N\}$ . The score of each alternative  $A_i$  with respect to each criteria  $C_j$  is expressed by a real number  $S_{ij}$  called the score value of  $A_i$  based on  $C_j$ . We denote  $\mathbf{S}$  the score  $M \times N$  matrix which is defined as  $\mathbf{S} \triangleq [S_{ij}]$ . The MCDM problem aims to select the best alternative  $A^* \in \mathbf{A}$  given  $\mathbf{S}$  and the weighting factors  $\mathbf{w}$  of criteria.

## III. CALCULUS WITH INTERVALS

A closed interval  $\mathbf{x}$  is denoted by  $\mathbf{x} = [\underline{x}, \bar{x}] = \{x | \underline{x} \leq x \leq \bar{x}\}$ .  $\underline{x} = \inf(\mathbf{x})$  is the infimum (lower endpoint) of  $\mathbf{x}$  and  $\bar{x} = \sup(\mathbf{x})$  is the supremum (upper endpoint) of  $\mathbf{x}$  taken values in  $\mathbb{R}$ . The set of intervals over  $\mathbb{R}$  is denoted by  $\mathbb{IR}$ . An interval in which one endpoint is included and the other is excluded is called a half-closed interval (or half-open interval) and it is called an open interval if its endpoints are excluded.

Any precise number  $x$  can be expressed with *imprecise number notation* as the degenerate interval  $\mathbf{x} = [x, x]$ . A non-degenerate interval is called a *proper interval*. The numbers  $wid(\mathbf{x}) \triangleq \bar{x} - \underline{x}$ ,  $rad(\mathbf{x}) = wid(\mathbf{x})/2$  and  $mid(\mathbf{x}) \triangleq \frac{1}{2}(\underline{x} + \bar{x})$  are respectively the width, the radius and the midpoint of  $\mathbf{x}$ . If  $\mathbf{x}$  is a precise number (i.e. a degenerate interval), then  $wid(\mathbf{x}) = 0$  and  $\mathbf{x} = [mid(\mathbf{x}), mid(\mathbf{x})]$ . The number  $mag(\mathbf{x}) \triangleq \max\{|x| \mid x \in \mathbf{x}\} = \max\{|\underline{x}|, |\bar{x}|\}$  is the magnitude of  $\mathbf{x}$ , and  $mig(\mathbf{x}) \triangleq \min\{|x| \mid x \in \mathbf{x}\} = \min\{|\underline{x}|, |\bar{x}|\}$  is the mignitude of  $\mathbf{x}$ . If  $\mathbf{x}$  and  $\mathbf{y}$  are overlapped intervals then  $\mathbf{x} \cap \mathbf{y}$  and  $\mathbf{x} \cup \mathbf{y}$  are also intervals defined by  $\mathbf{x} \cap \mathbf{y} = [\max\{\underline{x}, \underline{y}\}, \min\{\bar{x}, \bar{y}\}]$  and  $\mathbf{x} \cup \mathbf{y} = [\min\{\underline{x}, \underline{y}\}, \max\{\bar{x}, \bar{y}\}]$ . If  $\mathbf{x}$  and  $\mathbf{y}$  do not overlap, then  $\mathbf{x} \cap \mathbf{y}$  is empty, and  $\mathbf{x} \cup \mathbf{y}$  is not a proper interval but the union of two disjoint<sup>1</sup> intervals. In this case, the interval  $[\min\{\underline{x}, \underline{y}\}, \max\{\bar{x}, \bar{y}\}]$  is the tightest interval that includes  $\mathbf{x} \cup \mathbf{y}$  and it is called the *interval hull* of  $\mathbf{x}$  and  $\mathbf{y}$ . The interval  $\mathbf{x}$  is a subset of  $\mathbf{y}$  if  $(\underline{y} \leq \underline{x}) \wedge (\bar{x} \leq \bar{y})$ . The interval  $\mathbf{x}$  is equal to  $\mathbf{y}$  if  $(\underline{x} = \underline{y}) \wedge (\bar{x} = \bar{y})$ .

### A. Interval Arithmetic

Interval Arithmetic (IA) is an arithmetic defined on intervals of  $\mathbb{IR}$ . Its modern development started with Moore's works [4]–[7] and yielded recently to an IEEE Standard [8]. The INTLAB Matlab™ toolbox for IA has been developed and proposed by Rump in [9] with a tutorial in [10]. Other tools implementing IA are listed in [7] with more resources available on Kreinovich's interval computation web site [11]. The basic operations<sup>2</sup> on intervals are:

- **Addition:**  $\mathbf{x} + \mathbf{y} = [\underline{x} + \underline{y}, \bar{x} + \bar{y}]$
- **Subtraction:**  $\mathbf{x} - \mathbf{y} = [\underline{x} - \bar{y}, \bar{x} - \underline{y}]$ . In particular,  $-\mathbf{x} = [-\bar{x}, -\underline{x}]$ , because  $-\mathbf{x} = [0, 0] - [\underline{x}, \bar{x}]$ .
- **Multiplication:**  $\mathbf{x} \times \mathbf{y} = [\min\{S_{\times}(\mathbf{x}, \mathbf{y})\}, \max\{S_{\times}(\mathbf{x}, \mathbf{y})\}]$ , where  $S_{\times}(\mathbf{x}, \mathbf{y}) \triangleq \{\underline{x}\underline{y}, \underline{x}\bar{y}, \bar{x}\underline{y}, \bar{x}\bar{y}\}$  is the set of all possible products<sup>3</sup> of endpoints of  $\mathbf{x}$  and  $\mathbf{y}$ . In particular,  $-\mathbf{x} = [-\bar{x}, -\underline{x}]$  because  $-\mathbf{x} = [-1, -1] \times [\underline{x}, \bar{x}] = [\underline{x}, \bar{x}] \times [-1, -1]$ .
- **Division:**  $\mathbf{x}/\mathbf{y} = [\min\{S_{\div}(\mathbf{x}, \mathbf{y})\}, \max\{S_{\div}(\mathbf{x}, \mathbf{y})\}]$ , if  $0 \notin \mathbf{y}$  and where  $S_{\div}(\mathbf{x}, \mathbf{y}) \triangleq \{\underline{x}/\underline{y}, \underline{x}/\bar{y}, \bar{x}/\underline{y}, \bar{x}/\bar{y}\}$  is the set of all possible divisions of endpoints of  $\mathbf{x}$  and  $\mathbf{y}$ . If  $0 \in \mathbf{y}$  then the division by  $\mathbf{y}$  can be handled with more effort using extended interval arithmetic [7], [12] not detailed in this paper.
- **Inverse:** if  $\underline{x} > 0$  or  $\bar{x} < 0$ ,  $\frac{1}{\mathbf{x}} = [1/\bar{x}, 1/\underline{x}]$ .

The following algebraic properties hold for all  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{IR}$ :

- **Associativity:**  $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$  and  $(\mathbf{x}\mathbf{y})\mathbf{z} = \mathbf{x}(\mathbf{y}\mathbf{z})$ .
- **Commutativity:**  $(\mathbf{x} + \mathbf{y}) = (\mathbf{y} + \mathbf{x})$  and  $(\mathbf{x}\mathbf{y}) = (\mathbf{y}\mathbf{x})$ .
- **Neutral elements:**  $\mathbf{0} + \mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x}$  where  $\mathbf{0} \triangleq [0, 0]$ ,  $\mathbf{0} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{0} = \mathbf{0}$  and  $\mathbf{1} \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{1} = \mathbf{x}$  where  $\mathbf{1} \triangleq [1, 1]$ .

Proper intervals do not have additive or multiplicative inverses and the distributivity law does not hold for intervals. Instead, the following *sub-distributivity law* (weaker version of distributivity) holds:  $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{IR}, \mathbf{x}(\mathbf{y} + \mathbf{z}) \subseteq \mathbf{x}\mathbf{y} + \mathbf{x}\mathbf{z}$ .

Although the interval arithmetic is appealing and looks simple for basic operations with intervals, the so-called dependency problem is a major obstacle to its application when

complicate expressions have to be calculated to find tightest range enclosure. In fact, we must take care of dependencies of variables involved in formulas before applying IA in order to get tightest results. To reduce the dependency effect in the result, we need to replace (if possible) the original expression to compute by an equivalent simpler one having less (or none) redundant variables. For example, the derivation of  $\mathbf{x}/[\mathbf{x} + \mathbf{y}]$  for  $0 \notin \mathbf{x}$  must be computed with IA by  $1/[1 + \mathbf{y}/\mathbf{x}]$  to get tightest result. Also, the power 2 of  $\mathbf{x}$  must not be computed by  $[\underline{x}, \bar{x}] \times [\underline{x}, \bar{x}]$  because the unknown precise value of  $x$  in  $[\underline{x}, \bar{x}]$  must be exactly the same (strong dependency) in the multiplication operation in the derivation of  $\mathbf{x}^2$ . Hence,  $\mathbf{x}^2 = \{x^2 \mid -2 \leq x \leq 2\} = [0, 4]$  is different of  $[-2, 2] \times [-2, 2] = [-4, 4]$ .

### B. Basic interval functions

Here several functions that are used in the sequel. More interval functions can be found in [7], [13].

- **Absolute value** [13]:

$$|\mathbf{x}| = \begin{cases} [|\bar{x}|, |\underline{x}|], & \text{if } \bar{x} \leq 0 \\ [|\underline{x}|, |\bar{x}|], & \text{if } \underline{x} \geq 0 \\ [0, \max\{|\underline{x}|, |\bar{x}|\}], & \text{if } \underline{x} < 0 \text{ and } \bar{x} > 0 \end{cases} \quad (1)$$

- **Power** [7]:

- If  $n > 0$  is an odd number:  $\mathbf{x}^n = [\underline{x}^n, \bar{x}^n]$
- If  $n > 0$  is an even number

$$\mathbf{x}^n = \begin{cases} [\underline{x}^n, \bar{x}^n], & \text{if } \underline{x} > 0 \\ [\bar{x}^n, \underline{x}^n], & \text{if } \bar{x} < 0 \\ [0, \max\{\underline{x}^n, \bar{x}^n\}], & \text{if } 0 \in \mathbf{x}. \end{cases}$$

- If  $z > 0$  and  $\underline{x} > 0$ ,  $\mathbf{x}^z = [\underline{x}^z, \bar{x}^z]$ .

- **Square root** [7]:

$$\mathbf{x}^{\frac{1}{2}} = \sqrt{\mathbf{x}} = \begin{cases} [\sqrt{\underline{x}}, \sqrt{\bar{x}}], & \text{for } \underline{x} > 0 \\ [0, \sqrt{\bar{x}}], & \text{if } 0 \in \mathbf{x}. \end{cases}$$

### C. Order relations for intervals

The real numbers are ordered by the relation  $<$  (or  $>$ ) and comparing two real numbers is in general not a difficult task. In the methods developed in this paper, we need to compare imprecise numbers represented by intervals. We interpret an interval to mean "there is a point that lies between the bounds" and the relation between two intervals is a relation between the two points belonging to intervals (i.e. a *possibly* relation). Comparing intervals is not obvious in the general case when the intervals have a non-empty intersection. For this, we propose a method for comparing intervals and we then explain how to find the min (or max) element of a set of intervals. To make comparisons, we assume that the unknown precise value belonging to an imprecise number is uniformly distributed in the interval under concern. This assumption is motivated by the principle of insufficient reason. The comparative test that we propose does not provide a true or false answer (boolean function), but only a probability value that the test is satisfied or not. To implement the comparison between to intervals  $\mathbf{x}$

<sup>1</sup> $\mathbf{x} = [\underline{x}, \bar{x}]$  and  $\mathbf{y} = [\underline{y}, \bar{y}]$  are disjoint if  $(\bar{x} < \underline{y}) \vee (\bar{y} < \underline{x})$ .

<sup>2</sup>For simplicity, we use operations on closed intervals.

<sup>3</sup>The product of  $\mathbf{x}$  and  $\mathbf{y}$  will also be denoted  $\mathbf{x} \cdot \mathbf{y}$ , or  $\mathbf{x}\mathbf{y}$  for simplicity.

and  $\mathbf{y}$  of  $\mathbb{IR}$ , we define  $W \triangleq \text{wid}(\mathbf{x})\text{wid}(\mathbf{y})$  for notation convenience and we need to distinguish all possible situations as follows:

- Case 1:  $\underline{x} < \bar{x} < \underline{y} < \bar{y}$ . In this case,  $\mathbf{x} < \mathbf{y}$  with probability  $P(\mathbf{x} < \mathbf{y}) = 1$ .
- Case 2:  $\underline{y} < \bar{y} < \underline{x} < \bar{x}$ . In this case,  $\mathbf{x} < \mathbf{y}$  with probability  $P(\mathbf{x} < \mathbf{y}) = 0$ .
- Case 3:  $\underline{x} < \underline{y} < \bar{x} < \bar{y}$ . In this case,  $\mathbf{x} < \mathbf{y}$  with

$$P(\mathbf{x} < \mathbf{y}) = \frac{1}{W}[\text{wid}(\mathbf{a})\text{wid}(\mathbf{b}) + \text{wid}(\mathbf{a})\text{wid}(\mathbf{c}) + (\text{wid}(\mathbf{b})^2/2) + \text{wid}(\mathbf{b})\text{wid}(\mathbf{c})] \quad (2)$$

where  $\mathbf{a} \triangleq [\underline{x}, \underline{y}]$ ,  $\mathbf{b} \triangleq [\underline{y}, \bar{x}]$  and  $\mathbf{c} \triangleq [\bar{x}, \bar{y}]$ .

- Case 4:  $\underline{x} < \underline{y} < \bar{y} < \bar{x}$ . In this case,  $\mathbf{x} < \mathbf{y}$  with

$$P(\mathbf{x} < \mathbf{y}) = \frac{1}{W}[\text{wid}(\mathbf{a})\text{wid}(\mathbf{b}) + (\text{wid}(\mathbf{b})^2/2)] \quad (3)$$

where  $\mathbf{a} \triangleq [\underline{x}, \underline{y}]$  and  $\mathbf{b} \triangleq [\underline{y}, \bar{y}]$ .

- Case 5:  $\underline{y} < \underline{x} < \bar{y} < \bar{x}$ . In this case,  $\mathbf{x} < \mathbf{y}$  with

$$P(\mathbf{x} < \mathbf{y}) = \frac{1}{W}(\text{wid}(\mathbf{b})^2/2) \quad (4)$$

where  $\mathbf{b} \triangleq [\underline{x}, \bar{y}]$ .

- Case 6:  $\underline{y} < \underline{x} < \bar{x} < \bar{y}$ . In this case,  $\mathbf{x} < \mathbf{y}$  with

$$P(\mathbf{x} < \mathbf{y}) = \frac{1}{W}[\text{wid}(\mathbf{b})\text{wid}(\mathbf{c}) + (\text{wid}(\mathbf{b})^2/2)] \quad (5)$$

where  $\mathbf{b} \triangleq [\underline{x}, \bar{x}]$  and  $\mathbf{c} \triangleq [\bar{x}, \bar{y}]$ .

Formulae (2)-(5) are obtained by the probability calculus using uniform distributions over intervals and the total probability theorem. For case 3, one has  $P(\mathbf{x} < \mathbf{y}) = P(\mathbf{x} < \mathbf{y}, x \in \mathbf{a}, y \in \mathbf{b}) + P(\mathbf{x} < \mathbf{y}, x \in \mathbf{a}, y \in \mathbf{c}) + P(\mathbf{x} < \mathbf{y}, x \in \mathbf{b}, y \in \mathbf{b}) + P(\mathbf{x} < \mathbf{y}, x \in \mathbf{b}, y \in \mathbf{c})$  with  $P(\mathbf{x} < \mathbf{y}, x \in \mathbf{a}, y \in \mathbf{b}) = 1 \cdot \frac{\text{wid}(\mathbf{a}) \text{wid}(\mathbf{b})}{\text{wid}(\mathbf{x}) \text{wid}(\mathbf{y})}$ ,  $P(\mathbf{x} < \mathbf{y}, x \in \mathbf{a}, y \in \mathbf{c}) = 1 \cdot \frac{\text{wid}(\mathbf{a}) \text{wid}(\mathbf{c})}{\text{wid}(\mathbf{x}) \text{wid}(\mathbf{y})}$ ,  $P(\mathbf{x} < \mathbf{y}, x \in \mathbf{b}, y \in \mathbf{b}) = \frac{1}{2} \cdot \frac{\text{wid}(\mathbf{b}) \text{wid}(\mathbf{b})}{\text{wid}(\mathbf{x}) \text{wid}(\mathbf{y})}$  and  $P(\mathbf{x} < \mathbf{y}, x \in \mathbf{b}, y \in \mathbf{c}) = 1 \cdot \frac{\text{wid}(\mathbf{b}) \text{wid}(\mathbf{c})}{\text{wid}(\mathbf{x}) \text{wid}(\mathbf{y})}$ , which gives formula (2).

The value of  $P(\mathbf{x} > \mathbf{y})$  can be computed by a similar approach. Of course,  $P(\mathbf{x} \geq \mathbf{y}) = 1 - P(\mathbf{x} < \mathbf{y})$  and  $P(\mathbf{x} \leq \mathbf{y}) = 1 - P(\mathbf{x} > \mathbf{y})$ . Also, one has  $P(\mathbf{x} \neq \mathbf{y}) = P(\mathbf{x} < \mathbf{y}) + P(\mathbf{x} > \mathbf{y}) = 1 - P(\mathbf{x} = \mathbf{y})$ .

**Example 1:**  $\mathbf{x} = [-3, 0]$ ,  $\mathbf{y} = [-1, 4]$ , then  $P(\mathbf{x} < \mathbf{y}) = 0.9667$ .

Because we know how to compute the probability  $P(\mathbf{x} < \mathbf{y})$  for any two imprecise numbers  $\mathbf{x}$  and  $\mathbf{y}$ , we are able to find the min (or max) elements of a set of imprecise numbers  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M\}$  with a given associated probability. For instance, for finding the min element of  $X$  we proceed as follows:

- Calculate the  $M \times M$  square matrix<sup>4</sup>:

$$\mathbf{P} \triangleq [P_{ij} = P(\mathbf{x}_i < \mathbf{x}_j)] \quad (6)$$

<sup>4</sup>By construction all diagonal elements  $P_{ii}$  equal zero.

- Calculate the likelihood  $\lambda_i \triangleq \lambda(\mathbf{x}_i)$  of  $\mathbf{x}_i$  to be the min of  $X$  as the sum of  $P_{ij}$  for  $j \neq i$ , that is

$$\lambda_i = \sum_{j=1, \dots, M | j \neq i} P(\mathbf{x}_i < \mathbf{x}_j) = \sum_{j \neq i} P_{ij} \quad (7)$$

- The index of the most likely min element of  $X$  is

$$i_{\min} = \arg \max_{i=1, \dots, M} \lambda_i \quad (8)$$

The most likely min element of  $X$  is given by  $\mathbf{x}_{i_{\min}}$  with the probability  $P(\mathbf{x}_{i_{\min}} = \min\{X\}) = \lambda_i / (M - 1)$ .

An approach similar is applied to find the max element of  $X$  using the likelihood  $\lambda_i = \sum_{j \neq i} P(\mathbf{x}_i > \mathbf{x}_j)$  and  $i_{\max} = \arg \max_i \lambda_i$ . The max element of  $X$  will be given by  $\mathbf{x}_{i_{\max}}$  with the associated probability  $P(\mathbf{x}_{i_{\max}} = \max\{X\}) = \lambda_i / (M - 1)$ . Moreover and if needed, we can also sort (probabilistically) all the elements of  $X$  by decreasing (or increasing) order based on the likelihood values  $\lambda_i$ .

**Example 2:** Let's consider the set of intervals  $X = \{\mathbf{x}_1 = [-2, 2], \mathbf{x}_2 = [-3, 0], \mathbf{x}_3 = [0, 5], \mathbf{x}_4 = [-1, 3]\}$ . From formulas of  $P(\mathbf{x} < \mathbf{y})$  given for aforementioned cases 1–6, one obtains

$$\mathbf{P} \triangleq [P(\mathbf{x}_i < \mathbf{x}_j)] = \begin{bmatrix} 0 & 0.1667 & 0.9000 & 0.7188 \\ 0.8333 & 0 & 1.0000 & 0.9583 \\ 0.1000 & 0 & 0 & 0.2250 \\ 0.2812 & 0.0417 & 0.7750 & 0 \end{bmatrix}$$

with the likelihood values

$$\begin{bmatrix} \lambda_1 = \lambda(\mathbf{x}_1) \\ \lambda_2 = \lambda(\mathbf{x}_2) \\ \lambda_3 = \lambda(\mathbf{x}_3) \\ \lambda_4 = \lambda(\mathbf{x}_4) \end{bmatrix} = \begin{bmatrix} 1.7854 \\ 2.7917 \\ 0.3250 \\ 1.0979 \end{bmatrix}$$

The maximum likelihood is  $\lambda_2 = 2.7917$  and the corresponding index is  $i_{\min} = 2$ . This means that  $\mathbf{x}_2 = [-3, 0]$  is most likely the min element of  $X$  with the probability  $P(\mathbf{x}_2 = \min\{X\}) = 2.7917/3 = 0.9306$ . Using a similar approach, one will find that the max of  $X$  is  $\mathbf{x}_3 = [0, 5]$  with the probability  $P(\mathbf{x}_3 = \max\{X\}) = 2.6750/3 = 0.8917$ . Based on the likelihood values of the min element of  $X$  sorted in decreasing order, we obtain  $\mathbf{x}_2 < \mathbf{x}_1 < \mathbf{x}_4 < \mathbf{x}_3$ , which corresponds to what we intuitively expect in such example.

#### D. Distances between intervals

There are many ways to define strict distance metrics between two intervals. The simplest one is the Hausdorff distance between two intervals  $\mathbf{x}$  and  $\mathbf{y}$  of  $\mathbb{IR}$  which is the maximum distance  $d(x, y)$  of  $x \in \mathbf{x}$  to its nearest point  $y \in \mathbf{y}$ , where  $d(x, y)$  is any chosen metric [14], more precisely  $d_H(\mathbf{x}, \mathbf{y}) = \max_{x \in \mathbf{x}} \{\min_{y \in \mathbf{y}} d(x, y)\}$ . For simplicity, if we choose the  $L_1$  distance metric  $d_{L_1}(x, y) \triangleq |x - y|$ , then Hausdorff's distance is given by

$$d_H(\mathbf{x}, \mathbf{y}) = \max\{|\underline{x} - \underline{y}|, |\bar{x} - \bar{y}|\} = |\text{mid}(\mathbf{x}) - \text{mid}(\mathbf{y})| + |\text{rad}(\mathbf{x}) - \text{rad}(\mathbf{y})| \quad (9)$$

Another interesting distance successfully used for decision-making under uncertainty in the belief functions framework

[1], [15], [16], is Wassertein's distance metric [17], [18]  $d_W(\mathbf{x}, \mathbf{y})$  defined as

$$d_W(\mathbf{x}, \mathbf{y}) \triangleq \sqrt{[\text{mid}(\mathbf{x}) - \text{mid}(\mathbf{y})]^2 + \frac{1}{3}[\text{rad}(\mathbf{x}) - \text{rad}(\mathbf{y})]^2} \quad (10)$$

which corresponds to Mallows' distance [19] between two probability distributions when we assume that each interval is the support of a uniform distribution.

**Example 3:** If  $\mathbf{x} = [-3, 0]$  and  $\mathbf{y} = [-1, 4]$ , then  $d_H(\mathbf{x}, \mathbf{y}) = 4$  whereas  $d_W(\mathbf{x}, \mathbf{y}) \approx 3.0551$ .

#### IV. BASICS OF BELIEF FUNCTIONS

Belief functions have been introduced by Shafer in [20] to model epistemic uncertainty. We assume that the answer<sup>5</sup> of the problem under concern belongs to a known (or given) finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are exclusive<sup>6</sup>. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A basic belief assignment (BBA) associated with a given source of evidence is defined [20] as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are respectively defined by

$$\text{Bel}(A) = \sum_{\substack{B \subseteq A \\ B \in 2^\Theta}} m(B), \quad \text{and} \quad \text{Pl}(A) = 1 - \text{Bel}(\bar{A}). \quad (11)$$

If  $m(A) > 0$ ,  $A$  is called a focal element of  $m(\cdot)$ . When all focal elements are singletons then  $m(\cdot)$  is called a *Bayesian BBA* [20] and its corresponding  $\text{Bel}(\cdot)$  function is homogeneous to a (subjective) probability measure. The vacuous BBA, or VBBA for short, representing a totally ignorant source is defined as<sup>7</sup>  $m_v(\Theta) = 1$ .

Shafer [20] proposed to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD with Dempster's rule (i.e. the normalized conjunctive rule). The justification and behavior of Dempster's rule have been disputed over the years from many counter-examples involving high or low conflicting sources (from both theoretical and practical standpoints) as reported in [22]–[25]. Many rules of combination exist<sup>8</sup>, and we recommend the new interesting rules based on the proportional conflict redistribution (PCR) principle, see [21], Vol. 3 for details.

A true distance metric between two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined on the same FoD, has been defined in [15] as follows<sup>9</sup>

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))} \quad (12)$$

<sup>5</sup>i.e. the solution, or the decision to take.

<sup>6</sup>This is so-called Shafer's model of FoD [21].

<sup>7</sup>The complete ignorance is denoted  $\Theta$  in Shafer's book [20].

<sup>8</sup>see [21], Vol. 2 for a detailed list of fusion rules.

<sup>9</sup>Another well-known real distance metric  $d_J(m_1, m_2)$  had been proposed before by Josselme et al. in [26] which could also be used but we prefer to work with  $d_{BI}(m_1, m_2)$  distance for reasons explained in [27].

where the *Belief-Intervals* are defined by  $BI_1(X) \triangleq [\text{Bel}_1(X), \text{Pl}_1(X)]$  and  $BI_2(X) \triangleq [\text{Bel}_2(X), \text{Pl}_2(X)]$ , and where  $d_W(BI_1(X), BI_2(X))$  is Wassertein's distance between intervals calculated by (10).  $N_c = 1/2^{|\Theta|-1}$  is a normalization factor to get  $d_{BI}(m_1, m_2) \in [0, 1]$ .

Making decision on an element of FoD from a given BF ( $\text{Bel}(\cdot)$ ,  $\text{Pl}(\cdot)$ , or  $m(\cdot)$ ) can be done in many manners. For instance,

- in taking the argument of max of  $\{\text{Bel}(\theta_i), i = 1, \dots, n\}$ . This is a pessimistic decisional attitude.
- in taking the argument of max of  $\{\text{Pl}(\theta_i), i = 1, \dots, n\}$ . This is an optimistic decisional attitude.
- in approximating the BBA  $m(\cdot)$  by a subjective probability measure  $P(\cdot)$  and taking the argument of max of  $\{P(\theta_i), i = 1, \dots, n\}$ . This is a compromise decisional attitude.
- in taking the argument of min of  $\{d_{BI}(m(\cdot), m_{\theta_i}), i = 1, \dots, n\}$ , where  $m_{\theta_i}$  is the BBA entirely focused on  $\theta_i$  defined by  $m_{\theta_i}(X) = 1$ , if  $X = \theta_i$  and  $m_{\theta_i}(X) = 0$ , if  $X \neq \theta_i$ .

In the sequel, we will use the latter method which has been proved very effective in [16], [27].

#### V. BF-TOPSIS WITH PRECISE SCORES

Four BF-TOPSIS methods have been proposed in [1] with an increasing complexity and robustness to rank reversal phenomenon for MDCM support. In this section we briefly recall the main ideas of BF-TOPSIS. For further mathematical details, please refer to [1]. All these methods start with constructing BBAs from the precise score values of the score matrix  $\mathbf{S}$  as briefly explained. Only the way those BBAs are processed differs from one BF-TOPSIS method to another one.

##### A. From precise scores to precise BBAs

In [1], one has proved that BBAs can be consistently built from the precise score matrix  $\mathbf{S}$  as follows:

$$\text{Bel}_{ij}(A_i) \triangleq \begin{cases} \frac{\text{Sup}_j(A_i)}{A_{\max}^j} & \text{if } A_{\max}^j \neq 0 \\ 0 & \text{if } A_{\max}^j = 0 \end{cases} \quad (13)$$

$$\text{Bel}_{ij}(\bar{A}_i) \triangleq \begin{cases} \frac{\text{Inf}_j(A_i)}{A_{\min}^j} & \text{if } A_{\min}^j \neq 0 \\ 0 & \text{if } A_{\min}^j = 0 \end{cases} \quad (14)$$

where  $\bar{A}_i$  is the complement of  $A_i$  in the FoD  $\Theta \triangleq \{A_1, A_2, \dots, A_M\}$  ( $M > 2$ ), and

$$\text{Sup}_j(A_i) \triangleq \sum_{k \in \{1, \dots, M\} | S_{kj} \leq S_{ij}} |S_{ij} - S_{kj}| \quad (15)$$

$$\text{Inf}_j(A_i) \triangleq - \sum_{k \in \{1, \dots, M\} | S_{kj} \geq S_{ij}} |S_{ij} - S_{kj}| \quad (16)$$

The denominators involved in Eqs. (13)–(14), are defined by  $A_{\max}^j \triangleq \max_i \text{Sup}_j(A_i)$  and  $A_{\min}^j \triangleq \min_i \text{Inf}_j(A_i)$ , and they are supposed different from zero<sup>10</sup>. Therefore, the belief

<sup>10</sup>If  $A_{\max}^j = 0$  then  $\text{Bel}_{ij}(A_i) = 0$ , and if  $A_{\min}^j = 0$  then  $\text{Pl}_{ij}(A_i) = 1$ , so that  $\text{Bel}_{ij}(\bar{A}_i) = 0$ .

interval of choosing hypothesis  $A_i$  considering criterion  $C_j$  is given by:

$$[Bel_{ij}(A_i); Pl_{ij}(A_i)] \triangleq \left[ \frac{Sup_j(A_i)}{A_{\max}^j}, 1 - \frac{Inf_j(A_i)}{A_{\min}^j} \right] \quad (17)$$

From this belief interval, we deduce the BBA  $m_{ij}(\cdot)$  which is the triplet  $(m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i))$  defined by:

$$m_{ij}(A_i) \triangleq Bel_{ij}(A_i) \quad (18)$$

$$m_{ij}(\bar{A}_i) \triangleq Bel_{ij}(\bar{A}_i) = 1 - Pl_{ij}(A_i) \quad (19)$$

$$m_{ij}(A_i \cup \bar{A}_i) \triangleq m_{ij}(\Theta) = Pl_{ij}(A_i) - Bel_{ij}(A_i) \quad (20)$$

If a numerical value  $S_{ij}$  is missing in  $\mathbf{S}$ , one uses  $m_{ij}(\cdot) \triangleq (0, 0, 1)$ , i.e. one takes the vacuous belief assignment.

Using the formulae (13)–(20), we obtain from any  $M \times N$  precise score matrix<sup>11</sup>  $\mathbf{S}$  the general  $M \times N$  matrix  $\mathbf{M} \triangleq [m_{ij}(\cdot)]$  of BBAs that are involved in BF-TOPSIS methods. This construction of BBAs is very interesting for applications because it is invariant to the bias and scaling effects of score values [1]. Also, it allows us to model our lack of evidence (if any) with respect to an (or several) alternative(s) when their corresponding score values are missing for any reason.

### B. BF-TOPSIS1 method

From the BBA matrix  $\mathbf{M}$  and for each alternative  $A_i$ , one computes distances  $d_{BI}(m_{ij}, m_{ij}^{\text{best}})$  between  $m_{ij}(\cdot)$  and the ideal best BBA defined by  $m_{ij}^{\text{best}}(A_i) \triangleq 1$ , and the distances  $d_{BI}(m_{ij}, m_{ij}^{\text{worst}})$  between  $m_{ij}(\cdot)$  and the ideal worst BBA defined by  $m_{ij}^{\text{worst}}(\bar{A}_i) \triangleq 1$ . Then, one computes the weighted average distances with relative importance weighting factor  $w_j$  of criteria  $C_j$  as follows:

$$d^{\text{best}}(A_i) \triangleq \sum_{j=1}^N w_j \cdot d_{BI}(m_{ij}, m_{ij}^{\text{best}}) \quad (21)$$

$$d^{\text{worst}}(A_i) \triangleq \sum_{j=1}^N w_j \cdot d_{BI}(m_{ij}, m_{ij}^{\text{worst}}) \quad (22)$$

The relative closeness of the alternative  $A_i$  with respect to the ideal best solution  $A^{\text{best}}$  defined by

$$C(A_i, A^{\text{best}}) \triangleq \frac{d^{\text{worst}}(A_i)}{d^{\text{worst}}(A_i) + d^{\text{best}}(A_i)} \quad (23)$$

is used to make the preference ordering according to the descending order of  $C(A_i, A^{\text{best}}) \in [0, 1]$ , where a larger  $C(A_i, A^{\text{best}})$  value means a better alternative (higher preference).

<sup>11</sup>Note that each element  $m_{ij}(\cdot)$  is in fact a 3-uple of masses given by (18)–(20).

### C. BF-TOPSIS2 method

For each criteria  $C_j$ , one computes at first the relative closeness of each alternative  $A_i$  w.r.t. its ideal best solution  $A^{\text{best}}$  by

$$C_j(A_i, A^{\text{best}}) \triangleq \frac{d_{BI}(m_{ij}, m_{ij}^{\text{worst}})}{d_{BI}(m_{ij}, m_{ij}^{\text{worst}}) + d_{BI}(m_{ij}, m_{ij}^{\text{best}})} \quad (24)$$

The global relative closeness  $C(A_i, A^{\text{best}})$  of each alternative  $A_i$  with respect to its ideal best solution  $A^{\text{best}}$  used to make the final preference ordering is then obtained by the weighted average of  $C_j(A_i, A^{\text{best}})$ , that is

$$C(A_i, A^{\text{best}}) \triangleq \sum_{j=1}^N w_j \cdot C_j(A_i, A^{\text{best}}) \quad (25)$$

### D. BF-TOPSIS3 method

For each alternative  $A_i$ , one fuses the  $N$  precise BBAs  $m_{ij}(\cdot)$  discounted with importance factor  $w_j$  (see [28]) with PCR6 rule of combination [21] (Vol. 3) to get the precise fused BBA  $m_i^{\text{PCR6}}$ , from which one computes the distance  $d^{\text{best}}(A_i) = d_{BI}(m_i^{\text{PCR6}}, m_i^{\text{best}})$  between  $m_i^{\text{PCR6}}(\cdot)$  and its ideal best BBA  $m_i^{\text{best}}(A_i) \triangleq 1$ . Similarly, one computes the distance  $d^{\text{worst}}(A_i) = d_{BI}(m_i^{\text{PCR6}}, m_i^{\text{worst}})$  between  $m_i^{\text{PCR6}}(\cdot)$  and  $m_i^{\text{worst}}(\bar{A}_i) \triangleq 1$ . The relative closeness of each  $A_i$  with respect to ideal best solution  $C(A_i, A^{\text{best}})$  is computed by (23), and is used to make the preference ordering according to the descending order of  $C(A_i, A^{\text{best}})$ .

### E. BF-TOPSIS4 method

This method is similar to BF-TOPSIS3 except that we use the more complicate ZPCR6 fusion rule taking into account Zhang's degree of intersection of focal elements in the conjunctive consensus operator, see [29] for details.

## VI. BF-TOPSIS WITH IMPRECISE SCORES

In this section we present the extension of BF-TOPSIS methods to deal with imprecise score values  $S_{ij} = [\underline{S}_{ij}, \bar{S}_{ij}]$ ,  $i = 1, \dots, M$  and  $j = 1, \dots, N$ . These extensions will be referred as Imp-BF-TOPSIS in the sequel. The basic idea is to follow principles of BF-TOPSIS using Interval Arithmetic (IA) instead of classical arithmetic on reals.

### A. From imprecise scores to imprecise BBAs

The application of formulae (13)–(20) using IA operations does not work directly because of potential division by intervals including zero, and because of comparison tests involving boolean  $\leq$  and  $\geq$  functions. To circumvent these problems, we need to avoid intervals including zero, and replace boolean  $\leq$  and  $\geq$  functions by their probabilistic counterpart presented in section III-C. This is done as follows:

- Step 1 (Offset correction): To work only with positive intervals, we apply at first an offset correction of imprecise score values  $S_{ij} = [\underline{S}_{ij}, \bar{S}_{ij}]$  for each column  $j$  of imprecise score matrix  $\mathbf{S} = [S_{ij}]$ . This is a preprocessing step. We are allowed to do this because, by construction, the BBAs based on formulae (13)–(20) are invariant to

bias and scaling effects [1]. Therefore, we can always replace the original imprecise score  $[\underline{S}_{ij}, \bar{S}_{ij}]$  by

$$[\underline{S}'_{ij}, \bar{S}'_{ij}] = [\underline{S}_{ij}, \bar{S}_{ij}] + [\delta_j + \epsilon, \delta_j + \epsilon] \quad (26)$$

where  $\epsilon > 0$  is an arbitrary positive number to ensure the strict positivity of intervals, and the offset correction value  $\delta_j$  is given for  $j = 1, \dots, N$  by

$$\delta_j = - \min_{i=1, \dots, M} \{ \underline{S}_{ij} \} \quad (27)$$

**Example 4:** Let's consider the FoD  $\Theta \triangleq \{A_1, A_2, A_3, A_4\}$  with four alternatives, a criterion  $C_1$  and the following associated imprecise scores  $S_{11} = [-2, 2]$ ,  $S_{21} = [-3, 0]$ ,  $S_{31} = [0, 5]$  and  $S_{41} = [-1, 3]$ . The offset correction is then  $\delta_1 = -\min\{-2, -3, 0, -1\} = 3$ . If we take,  $\epsilon = 1$  then we will get the corrected (positive) imprecise scores

$$S'_{11} = [-2, 2] + [3 + 1, 3 + 1] = [2, 6]$$

$$S'_{21} = [-3, 0] + [3 + 1, 3 + 1] = [1, 4]$$

$$S'_{31} = [0, 5] + [3 + 1, 3 + 1] = [4, 9]$$

$$S'_{41} = [-1, 3] + [3 + 1, 3 + 1] = [3, 7]$$

- Step 2: Replace the  $S_{kj} \leq S_{ij}$  and  $S_{kj} \geq S_{ij}$  tests involved in (15) and (16), by their probabilistic counterparts  $P(S_{kj} \leq S_{ij}) \geq 0.5$  and  $P(S_{kj} \geq S_{ij}) \geq 0.5$  because  $S_{kj}$  and  $S_{ij}$  are imprecise numbers (i.e. intervals), where  $P(S_{kj} \leq S_{ij})$  and  $P(S_{kj} \geq S_{ij})$  are computed as in section III-C.

In the sequel, we assume that the offset correction of score as been applied (step 1 done) and for notation simplicity we denote these (corrected) strictly positive imprecise scores  $S_{ij}$ . The imprecise BBAs can now be computed from the (offset-corrected) imprecise scores values as follows<sup>12</sup>:

$$[\underline{Bel}_{ij}(A_i), \overline{Bel}_{ij}(A_i)] \triangleq \begin{cases} \frac{Sup_j(A_i)}{A_{\max}^j} & \text{if } A_{\max}^j \neq [0, 0] \\ [0, 0] & \text{if } A_{\max}^j = [0, 0] \end{cases} \quad (28)$$

$$[\underline{Bel}_{ij}(\bar{A}_i), \overline{Bel}_{ij}(\bar{A}_i)] \triangleq \begin{cases} \frac{Inf_j(A_i)}{A_{\min}^j} & \text{if } A_{\min}^j \neq [0, 0] \\ [0, 0] & \text{if } A_{\min}^j = [0, 0] \end{cases} \quad (29)$$

where

$$\begin{aligned} Sup_j(A_i) &= [\underline{Sup}_j(A_i), \overline{Sup}_j(A_i)] \\ &\triangleq \sum_{k \in \{1, \dots, M\} | P(S_{kj} \leq S_{ij}) \geq 0.5} |S_{ij} - S_{kj}| \end{aligned} \quad (30)$$

$$\begin{aligned} Inf_j(A_i) &= [\underline{Inf}_j(A_i), \overline{Inf}_j(A_i)] \\ &\triangleq - \sum_{k \in \{1, \dots, M\} | P(S_{kj} \geq S_{ij}) \geq 0.5} |S_{ij} - S_{kj}| \end{aligned} \quad (31)$$

The denominators involved in Eqs. (28)-(29), are defined by  $A_{\max}^j = [\underline{A}_{\max}^j, \bar{A}_{\max}^j] \triangleq \max_i Sup_j(A_i)$  and  $A_{\min}^j =$

$[\underline{A}_{\min}^j, \bar{A}_{\min}^j] \triangleq \min_i Inf_j(A_i)$ , and they are supposed different from  $[0, 0]$ <sup>13</sup>. Therefore, in non-degenerate case (when  $A_{\max}^j \neq [0, 0]$  and  $A_{\min}^j \neq [0, 0]$ ) the belief interval of hypothesis  $A_i$  considering criterion  $C_j$  has now imprecise bounds given by

$$Bel_{ij}(A_i) = [\underline{Bel}_{ij}(A_i), \overline{Bel}_{ij}(A_i)] = \frac{Sup_j(A_i)}{A_{\max}^j} \quad (32)$$

$$Pl_{ij}(A_i) = [\underline{Pl}_{ij}(A_i), \overline{Pl}_{ij}(A_i)] = [1, 1] - \frac{Inf_j(A_i)}{A_{\min}^j} \quad (33)$$

From these imprecise bounds, we calculate the imprecise BBAs  $m_{ij}(\cdot) = [\underline{m}_{ij}(\cdot), \bar{m}_{ij}(\cdot)]$  which is the triplet of intervals ( $m_{ij}(A_i) = [\underline{m}_{ij}(A_i), \bar{m}_{ij}(A_i)]$ ,  $m_{ij}(\bar{A}_i) = [\underline{m}_{ij}(\bar{A}_i), \bar{m}_{ij}(\bar{A}_i)]$ ,  $m_{ij}(A_i \cup \bar{A}_i) = [\underline{m}_{ij}(A_i \cup \bar{A}_i), \bar{m}_{ij}(A_i \cup \bar{A}_i)]$ ) defined by:

$$m_{ij}(A_i) \triangleq Bel_{ij}(A_i) \quad (34)$$

$$m_{ij}(\bar{A}_i) \triangleq Bel_{ij}(\bar{A}_i) = [1, 1] - Pl_{ij}(A_i) \quad (35)$$

$$m_{ij}(A_i \cup \bar{A}_i) \triangleq m_{ij}(\Theta) = Pl_{ij}(A_i) - Bel_{ij}(A_i) \quad (36)$$

If a numerical (imprecise) value  $S_{ij}$  is missing in  $\mathbf{S}$ , one uses  $m_{ij}(\cdot) \triangleq ([0, 0], [0, 0], [1, 1])$ , i.e. one takes the vacuous belief assignment expressed in its degenerate interval form.

Using the formulae (28)-(36), we obtain from any  $M \times N$  imprecise score matrix  $\mathbf{S}$  the general  $M \times N$  matrix  $\mathbf{M} \triangleq [m_{ij}(\cdot) = [\underline{m}_{ij}(\cdot), \bar{m}_{ij}(\cdot)]]$  of imprecise BBAs that are necessary in Imp-BF-TOPSIS methods.

It is worth to note that formulae (28)–(36) are fully consistent with (13)–(20) when all the elements  $S_{ij}$  of imprecise score matrix are degenerate (are precise numbers), that is when  $\underline{S}_{ij} = \bar{S}_{ij}$ . By choosing the midpoints of imprecise score values, we can always build a precise BBA that satisfies Shafer's BBA definition [1]. This midpoint-based BBA is always included in imprecise BBA bounds because of IA. Therefore, imprecise BBAs are always admissible and they can be combined by Dempster's or PCR6 rules thanks to IA operations. This has to be implemented with caution to avoid dependency effect [7].

### B. Imp-BF-TOPSIS1 and Imp-BF-TOPSIS2 methods

These methods are similar to BF-TOPSIS1 and BF-TOPSIS2 except that we use IA operations. The distances  $d_{BI}(m_{ij}, m_{ij}^{\text{best}})$  and  $d_{BI}(m_{ij}, m_{ij}^{\text{worst}})$  become imprecise numbers computed with formula (12) adapted for interval calculus, where  $m_{ij}^{\text{best}}(A_i) \triangleq [1, 1]$  and  $m_{ij}^{\text{worst}}(\bar{A}_i) \triangleq [1, 1]$ . Of course, all scalars involved in the formulae (12), (21), (22), and (25) must be expressed in their degenerate interval form in order to apply IA operations, for instance  $N_c$  is replaced by  $[N_c, N_c]$ ,  $w_j$  by  $[w_j, w_j]$ , etc. The final preference ordering is found according to the descending order of imprecise  $C(A_i, A^{\text{best}})$  obtained by the method explained at the end of section II-C, where a larger  $C(A_i, A^{\text{best}})$  means a better alternative (higher preference).

<sup>12</sup>Remember that operations involved in the formulas of this section are IA operations defined in section III.

<sup>13</sup>If  $A_{\max}^j = [0, 0]$  then  $Bel_{ij}(A_i) = [0, 0]$ , and if  $A_{\min}^j = [0, 0]$  then  $Pl_{ij}(A_i) = [1, 1]$ , so that  $Bel_{ij}(\bar{A}_i) = [0, 0]$ .

C. Imp-BF-TOPSIS3 and Imp-BF-TOPSIS4 methods

These methods are similar to BF-TOPSIS3 and BF-TOPSIS4<sup>14</sup> but with special adaptation of PCR6 and ZPCR6 formulae to reduce dependency effects with IA operations. For example, the expression  $\frac{m_1^2(X)m_2(X)}{m_1(X)+m_2(Y)}$  involved in PCR6 formula [21] (Vol. 3) for the fusion of two BBAs must be computed as  $[\frac{1}{m_1(X)m_2(Y)} + \frac{1}{m_1^2(X)}]^{-1}$  with IA to get the tightest range enclosure. The implementation of conjunctive rule must also be done with precaution when using IA to reduce the dependency effect in the derivation.

VII. EXAMPLES

A. Example 5 (mono-criterion)

• **Precise scores case** [1]: Let's consider a criterion  $C_1$  and seven alternatives  $A_i$ , ( $i = 1, \dots, 7$ ) with the precise score values  $S_{11} = 10$ ,  $S_{21} = 20$ ,  $S_{31} = -5$ ,  $S_{41} = 0$ ,  $S_{51} = 100$ ,  $S_{61} = -11$ , and  $S_{71} = 0$ . The direct ranking with the preference "greater is better" yields<sup>15</sup>  $A_5 \succ A_2 \succ A_1 \succ (A_4 \sim A_7) \succ A_3 \succ A_6$ . In applying formulas (13)–(20), we get the BBAs listed in Table I. Using BF-TOPSIS methods<sup>16</sup>, we get the distances, and the relative closeness measures of Table II. In sorting  $C(A_i, A^{best})$  by the descending order, we get the correct preferences order  $A_5 \succ A_2 \succ A_1 \succ (A_4 \sim A_7) \succ A_3 \succ A_6$  which is consistent with the direct ranking result.

Table I  
BBAS CONSTRUCTED FROM PRECISE SCORE VALUES.

	$m_{ij}(A_i)$	$m_{ij}(\bar{A}_i)$	$m_{ij}(A_i \cup \bar{A}_i)$
$A_1$	0.0955	0.5236	0.3809
$A_2$	0.1809	0.4188	0.4003
$A_3$	0.0102	0.8115	0.1783
$A_4$	0.0273	0.6806	0.2921
$A_5$	1.0000	0	0
$A_6$	0	1.0000	0
$A_7$	0.0273	0.6806	0.2921

Table II  
DISTANCES AND RELATIVE CLOSENESS MEASURES.

	$d_{BI}(m_{ij}, m_{ij}^{best})$	$d_{BI}(m_{ij}, m_{ij}^{worst})$	$C(A_i, A^{best})$
$A_1$	0.7380	0.0940	0.1130
$A_2$	0.6676	0.1615	0.1948
$A_3$	0.8112	0.0214	0.0257
$A_4$	0.7954	0.0405	0.0485
$A_5$	0	0.8229	1.0000
$A_6$	0.8229	0	0
$A_7$	0.7954	0.0405	0.0485

• **Imprecise scores case:** For simplicity, consider now the imprecise score values with midpoints consistent with previous example. For instance suppose  $S_{11} = [8, 12]$ ,  $S_{21} = [18, 22]$ ,  $S_{31} = [-7, -3]$ ,  $S_{41} = [-1, 1]$ ,  $S_{51} = [97, 103]$ ,  $S_{61} =$

$[-12, -10]$ , and  $S_{71} = [-1, 1]$ . The offset factor is equal to  $\delta_1 = 12$ . After offset corrections with  $\epsilon = 1$ , we get the corrected positive imprecise scores  $S_{11} = [21, 25]$ ,  $S_{21} = [31, 35]$ ,  $S_{31} = [6, 10]$ ,  $S_{41} = [12, 14]$ ,  $S_{51} = [110, 116]$ ,  $S_{61} = [1, 3]$ , and  $S_{71} = [12, 14]$ . In applying formulas (28)–(36), we get the imprecise BBAs listed in Table III.

Table III  
IMPRECISE BBAS CONSTRUCTED FROM IMPRECISE SCORE VALUES.

	$m_{ij}(A_i)$	$m_{ij}(\bar{A}_i)$	$m_{ij}(A_i \cup \bar{A}_i)$
$A_1$	[0.0701,0.1234]	[0.4375,0.6264]	[0.2501,0.4924]
$A_2$	[0.1452,0.2200]	[0.3606,0.4885]	[0.2915,0.4942]
$A_3$	[0.0049,0.0161]	[0.6538,1.0000]	[0,0.3413]
$A_4$	[0.0179,0.0376]	[0.5769,0.8046]	[0.1578,0.4051]
$A_5$	[0.9119,1.0000]	[0,0]	[0,0.0881]
$A_6$	[0,0]	[0.8365,1.0000]	[0,0.1635]
$A_7$	[0.0179,0.0376]	[0.5769,0.8046]	[0.1578,0.4051]

As we see all imprecise BBAs values of Table III include precise BBAs values of Table I. Note that all negative bounds encountered in derivations (if any) are set to zero, and all bounds greater than one in derivations (if any) are set to one because masses values must belong to  $[0, 1]$ . Each imprecise BBA represented by a row of Table III is said *admissible* because for a given hypothesis  $A_i$  one can find at least a point (a precise mass value) in each interval  $m_{ij}(A_i)$ ,  $m_{ij}(\bar{A}_i)$  and  $m_{ij}(A_i \cup \bar{A}_i)$  such that the sum of the masses equals one. If all imprecisions of scores reduce to zero, the results of Table III will coincide with results of Table I. Using Imp-BF-TOPSIS methods, we get the imprecise distances, and the imprecise relative closeness measures listed in Table IV.

Table IV  
IMPRECISE DISTANCES AND RELATIVE CLOSENESS MEASURES.

	$d_{BI}(m_{ij}, m_{ij}^{best})$	$d_{BI}(m_{ij}, m_{ij}^{worst})$	$C(A_i, A^{best})$
$A_1$	[0.5421,0.9338]	[0.0301,0.2864]	[0.0312,0.3457]
$A_2$	[0.5034,0.8313]	[0.0511,0.3253]	[0.0579,0.3926]
$A_3$	[0.5324,1.0000]	[0.0006,0.2977]	[0.0006,0.3586]
$A_4$	[0.5960,0.9956]	[0.0131,0.2317]	[0.0130,0.2800]
$A_5$	[0,0.1135]	[0.7176,0.8591]	[0.8634,1.0000]
$A_6$	[0.6900,0.9533]	[0,0.1360]	[0,0.1647]
$A_7$	[0.5960,0.9956]	[0.0131,0.2317]	[0.0130,0.2800]

For each element of  $\mathcal{C} = \{C(A_i, A^{best}), i = 1, \dots, M\}$ , we compute its likelihood  $\lambda_i \triangleq \lambda(C(A_i, A^{best}))$  to be the max of  $\mathcal{C}$  by the method explained in section III-C. Here, one gets

$$[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7] \approx [3.00, 3.57, 2.80, 2.29, 6.00, 1.02, 2.29]$$

In sorting  $\lambda_i$  by the descending order, we get  $A_5 \succ A_2 \succ A_1 \succ A_3 \succ A_6 \succ (A_4 \sim A_7)$ . This result is of course a bit different of what we obtain with precise midpoints of scores because of imprecision degree in the input scores, which is normal. However when the imprecision degree (i.e. the width of each score interval) of the input scores reduces to zero, we always obtain the same result as with (precise) midpoint of

<sup>14</sup>See their mathematical derivations in [1].

<sup>15</sup>where the symbol  $\succ$  means *better than* (or *is preferred to*).

<sup>16</sup>in mono-criterion case, all BF-TOPSIS methods are equivalent because there is no need of making fusion.

score intervals because of the consistency of interval arithmetic operators with arithmetic on real numbers.

### B. Example 6 (Multi-criteria)

An investor wants to invest some money in a company to get the highest profit. He considers four companies  $\mathcal{A} = \{A_1, A_2, A_3, A_4\}$  and must take a decision according to the following two criteria:  $C_1$  is the risk analysis (the min is better) with weight  $w_1 = 0.6$  and  $C_2$  is the growth analysis (the max is better) with  $w_2 = 0.4$ . Assume the imprecise scores are

$$S = \begin{array}{c} C_1(\text{in } \%) \quad C_2(\text{in } \%) \\ \begin{array}{c} A_1 \\ A_2 \\ A_3 \\ A_4 \end{array} \left[ \begin{array}{cc} [8, 12] & [8, 10] \\ [17, 19] & [15, 19] \\ [3, 5] & [8, 12] \\ [4, 8] & [5, 7] \end{array} \right] \end{array}$$

We get the final preference orderings with:

- Imp-BF-TOPSIS1: we get  $A_3 \succ A_4 \succ A_1 \succ A_2$  because  $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] \approx [1.4939, 1.0013, 1.9396, 1.5652]$
- Imp-BF-TOPSIS2: we get  $A_3 \succ A_4 \succ A_1 \succ A_2$  because  $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] \approx [1.4968, 0.9323, 2.0535, 1.5175]$
- Imp-BF-TOPSIS3: we get  $A_3 \succ A_4 \succ A_1 \succ A_2$  because  $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] \approx [1.5300, 1.3487, 1.5639, 1.5574]$
- Imp-BF-TOPSIS4: we get  $A_4 \succ A_3 \succ A_1 \succ A_2$  because  $[\lambda_1, \lambda_2, \lambda_3, \lambda_4] \approx [1.5093, 1.4686, 1.5101, 1.5120]$

Imp-BF-TOPSIS1–3 methods provide here the same preference order  $A_3 \succ A_4 \succ A_2 \succ A_1$ , hence  $A_3$  is the preferred choice. One sees that Imp-BF-TOPSIS3 and Imp-BF-TOPSIS4 have difficulties to provide very distinct likelihood values because unsurmountable dependency effects arise in IA operations when applying PCR6 and ZPCR6 rules which degrade substantially the final precision of the result. Based on this analysis, we recommend to use either Imp-BF-TOPSIS1 or Imp-BF-TOPSIS2 because they provide tightest enclosure results and they are much simpler to implement than Imp-BF-TOPSIS3 and Imp-BF-TOPSIS4.

## VIII. CONCLUSIONS

Four new methods (Imp-BF-TOPSIS1–Imp-BF-TOPSIS4) for MCDM have been proposed. We have shown how to calculate imprecise BBAs from imprecise scores, and how to evaluate the relative imprecise closeness of each alternative to the ideal best and worst solutions for making the preference ordering. These methods avoid scores normalization, and they can deal with imprecise scores, with missing scores, with the reliability of the sources as well, and they could also work with imprecise weightings of criteria. They are more complicated to implement (and slower) than their precise counterparts because of IA. They are consistent with BF-TOPSIS1–4 when the imprecision of scores reduces to zero. However because IA suffers of dependency effects, IA is not the *universal panacea* to work with imprecise values to get best results, specially

for combining imprecise BBAs. More research efforts need to be done to circumvent these problems (if possible) by better implementations (or by Monte-Carlo approach) in order to improve the performance of Imp-BF-TOPSIS3 and Imp-BF-TOPSIS4 methods. Application of these methods for natural risk assessment in mountains is under development and will be reported in future publications.

## REFERENCES

- [1] J. Dezert, D. Han, H. Yin, A New Belief Function Based Approach for Multi-Criteria Decision-Making Support, in Proc. of Fusion 2016.
- [2] C.L. Hwang, K. Yoon, Multiple Attribute Decision Making, Lecture Notes in Economics and Math. Syst. 186, Springer, Berlin, 1981.
- [3] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, H. Yin, The BF-TOPSIS Approach for Solving Non-Classical MCDM Problems, in Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21-23, 2016.
- [4] R. Moore, Interval Arithmetic and Automatic Error Analysis in Digital Computing, Ph.D. Thesis, Stanford Univ., U.S.A., October 1962.
- [5] R. Moore, Interval Analysis, Prentice-Hall, U.S.A., 1966.
- [6] R. Moore, Methods and Applications of Interval Analysis, Society for Industrial and Applied Mathematics, Philadelphia, U.S.A., 1979.
- [7] R. Moore, R. Kearfott, M. Cloud, Introduction to interval analysis, Society for Industrial and Applied Mathematics Press, 2009.
- [8] IEEE Standard for Interval Arithmetic, IEEE Std 1788, June 2015.
- [9] S.M. Rump, Interval library for Matlab (Intlab) <http://www.ti3.tu-harburg.de/rump/intlab/>
- [10] G.I. Hargreaves, Interval Analysis in MATLAB, Master's thesis, Dept. of Mathematics, Univ. of Manchester, 2002.
- [11] <http://www.cs.utep.edu/interval-comp/>
- [12] W.M. Kahan, A More Complete Interval Arithmetic: Lecture Notes for an Engineering Summer Course in Numerical Analysis at the University of Michigan, Technical report, University of Michigan, 1968.
- [13] V. Stahl, Interval Methods for Bounding the Range of Polynomials and Solving Systems of Nonlinear Equations, Ph.D. Thesis, Johannes Kepler Universität, Linz, Austria, September 1995.
- [14] G. Rote, Computing the minimum Hausdorff distance between two point sets on a line under translation, Inf. Proc. Letters (38), 123–127, 1991.
- [15] D. Han, J. Dezert, Y. Yang, New Distance Measures of Evidence based on Belief Intervals, in Proc. of Belief 2014, Oxford, UK, Sept. 2014.
- [16] D. Han, J. Dezert, Y. Yang, Belief interval Based Distances Measures in the Theory of Belief Functions, to appear in IEEE Tr. on SMC, 2017.
- [17] R. Verde, A. Irpino, A New Interval Data Distance Based on the Wasserstein Metric, in Proc. of the 31st Annual Conf. of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität, Freiburg, March 7–9, 2007.
- [18] A. Irpino, R. Verde, Dynamic Clustering of Interval Data Using a Wasserstein-based Distance, Pattern Reco. Letters 29, 1648–1658, 2008.
- [19] C.L. Mallows, A note on asymptotic joint normality, Annals of Math. Stat., Vol. 43, pp. 508–515, 1972.
- [20] G. Shafer, A Mathematical Theory of Evidence, Princeton Univ., 1976.
- [21] F. Smarandache, J. Dezert (Editors), Advances and applications of DSMT for information fusion, American Research Press, Vol. 1–4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [22] J. Dezert, P. Wang, A. Tchamova, On the validity of Dempster-Shafer theory, Proc. of Fusion 2012, Singapore, July 9–12, 2012.
- [23] A. Tchamova, J. Dezert, On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory, 6th IEEE Int. Conf. on Int. Syst., Sofia, Bulgaria, Sept. 6–8, 2012.
- [24] J. Dezert, A. Tchamova, On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule, Int. J. of Intell. Syst., Vol. 29, No. 3, pp. 223–252, March 2014.
- [25] F. Smarandache, J. Dezert, On the consistency of PCR6 with the averaging rule and its application to probability estimation, Proc. of Fusion 2013, Istanbul, Turkey, July 2013.
- [26] A.-L. Jousselme, D. Grenier, É. Bossé, A new distance between two bodies of evidence, Information Fusion, Vol. 2, No. 2, pp. 91–101, 2001.
- [27] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, Decision-Making with Belief Interval Distance, Proc. of Belief 2016, Prague, Sept. 2016.
- [28] F. Smarandache, J. Dezert, J.-M. Tacnet, Fusion of sources of evidence with different importances and reliabilities, Proc. of Fusion 2010, UK.
- [29] F. Smarandache, J. Dezert, Modified PCR rules of combination with degrees of intersections, Proc. of Fusion 2015, USA.

# Enhancing Multimodal Image Fusion: A Novel DSMT and SVM Classification Approach with Integrated Saliency Model

Hanan Anzid<sup>a,b</sup>, Gaétan Le-Goic<sup>b</sup>, Aissam Bekkari<sup>a</sup>, Alamin Mansouri<sup>b</sup>, and Driss Mammass<sup>a</sup>

<sup>a</sup>IRF-SIC Laboratory, Faculty of science, Agadir, Morocco.

<sup>b</sup>LE2I Laboratory, University of Burgundy-Franche-Comté, France.

hanananzid@gmail.com, Gaetan.Le-Goic@u-bourgogne.fr

Originally published as: H. Anzid, G. Le Goic, A. Bekkari, A. Mansouri, D. Mammass, *Improvement of Multimodal Images Classification Based on DSMT Using Visual Saliency Model Fusion with SVM*, Int. Journal of Computers & Technology, Vol. 18, 2018, and reprinted with permission.

**Abstract**—Multimodal images encompass diverse information that can be both complementary and redundant, thereby addressing the challenges associated with unimodal classification. By modeling and integrating these different pieces of information, multimodal approaches offer improved solutions. However, despite yielding acceptable results, this classification approach still falls short of the level achieved by the human brain’s powerful mechanism for effortlessly classifying visually observed scenes. To enhance the classification task within the realm of multimodal images, we present a methodology that leverages the Dezert-Smarandache Theory (DSMT). This approach involves fusing spectral and dense SURF features extracted from each modality, which are then pre-classified by the SVM classifier. Additionally, we incorporate the visual perception model into the fusion process. In order to demonstrate the effectiveness of incorporating salient features into the fusion process using DSMT, we conducted tests and validation on extensive datasets obtained from cultural heritage wall paintings. Each dataset consists of four imaging modalities, namely UV, IR, Visible, and fluorescence. The results obtained from these experiments show great promise.

**Keywords:** Visual saliency model, fusion, DSMT, SVM, Dense SURF features, Spectral features, Multimodal images, Classification.

## I. INTRODUCTION

Nowadays, multimodal imaging has gained increasing importance in computer vision application, and significant efforts have been put into developing methods of different tasks, such as Registration [1]–[4], Data fusion [5], Representation learning [6], Classification [7] and so on. In classification task, the unimodal image presents various problems as noisy data, incomplete information and distorted ones, etc. This often led to a misclassification. These limitations are overcome by using multimodal images, which are acquired from multiple sensors, and taken for the same object or scene. Each image or modality allows to provide different information that can sometimes be redundant, because the same area/scene is presented in a different sensor, and complementary for another modality, regarding the diversity of sensor technologies and their physical interaction mechanism. The use of this set of images together presents a real-world benefit to resolve a given problem with some various available information. The fusion of these data form a better quality classification.

However, these data are crippled with some imperfections such as conflict, ignorance, uncertainty and so on, which must be handled and taken into account by dedicated formalism as long as they present an aspect of reality. To fix such problem, several formalism exist as probability theory [8], Fuzzy theory [9], belief function formalism [10] and Dezert-Smarandache (DSMT) formalism [11], [12]. In this work, we benefit from the latest theory which is the most recent one, and it was introduced in order to deal with the high conflicted and uncertainty data thanks to its rich modelization and the combination operators (PCR5 and PCR6) that it integrates.

In classification task, belief function theory is widely exploited in many works [13]–[16]. Whereas DSMT or so-called plausible and paradoxical reasoning shows its efficiency in many applications, it was performed for multi-source remote sensing application [17] for supervised classification purpose by integrating contextual information obtained from ICM classifier with constraint and temporal information in hybrid DSMT process with adaptive decision rule, the authors also proposed a new decision rule based on DSMT transformation for change detection purpose [18]. In [19], the authors present an effective use of DSMT for multi-class classification by combining two SVM OAA (One-Against-All) implementation using PCR6 combination rule. A new method, based on fusing the attribute type information obtained from Ground Moving Target Indicator and imagery sensor using DSMT for tracking and classification, has been presented in [20]. Multidate fusion has been proposed in [21], [22] for the short-term prediction of the winter land cover. DSMT is also used in the medical case retrieval by [23], the authors used DSMT to fuse heterogeneous features of several sensors which will be included in CBR systems.

According to our study of the state of the art, all studied research works disregard the power of perceptual attention to well classify any scene thanks to the high human brain capacities. We benefit from this ability in our approach by integrating the visual perception model, using DSMT, with spectral and dense SURF features obtained from SVM classification for significant classification improvement.

The paper is organized as follows. After a brief presentation of mathematical background of DSMT formalism in section II, we present the overall system of the proposed method in

section III. Data and experiments are then given in section IV in order to evaluate the performance of our approach on real image datasets. A conclusion is given in section V.

## II. MATHEMATICAL BACKGROUND OF DSMT

Dezert-Smarandache theory was proposed jointly by Jean Dezert and Florentin Smarandache [24], and was an attempt to overcome belief function limitations by handling a high uncertainty and conflicting information. This theory can be described as follows:

We denote  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  the discernment space of the  $N$  class classification problem, and  $D^\Theta$  the hyper-power-set [25] that is the set of subsets of  $\Theta$ , with the union of classes and also their intersection, so that if  $X, Y \in D^\Theta$ , then  $X \cup Y \in D^\Theta$  and  $X \cap Y \in D^\Theta$ . Each source  $S_i$  contributes its belief mass  $m_i$  to  $X$ , known by the generalized basic belief assignment (gbba) step, and satisfying following properties:

$$m_i(X) : D^\Theta \rightarrow [0, 1], \quad (1)$$

$$m_i(\emptyset) = 0, \quad (2)$$

where  $\emptyset$  is the null set, and

$$\sum_{X \in D^\Theta} m_i(X) = 1. \quad (3)$$

The size of hyper-power-set presents a real limit in DSMT when  $N > 6$  ( $N$  number of classes) in Free model [26] which corresponds to the full hyper-power-set without any constraints, in contrary to hybrid model [26] which allows integrating constraints that can be exclusive and refined, and therefore minimizing  $D^\Theta$  size. The assigned generalist mass obtained from different sources are then combined and a new mass distribution is provided to  $D^\Theta$  elements. Combination step presents the kernel of the fusion process and each formalism proposed several combination operators. In DSMT formalism, all combination operators can be found in detail in [27], we quote the most used as Smets rule, Dempster-Shafer (normalized) operator, Yager operator, Zhang operator, DSMT rule, Dubois and Prade rule, PCR5 operator for  $N = 2$  and PCR6 operator for  $N > 2$ . To deal with a large number of the sources used in this work and the high uncertainty and conflicting information provided, we benefit from the performance of PCR6 combination rule in handling such problem.

The generalized belief functions Credibility noted  $Bel(\cdot)$  or  $Cr(\cdot)$ , Plausibility noted  $Pl(\cdot)$ , and DSMT transformation are derived from the function of basic mass and respectively defined for  $D^\Theta$  in  $[0, 1]$ :

$$Cr(X) = \sum_{\substack{x \in D^\Theta \\ x \subseteq X}} m(x) \quad (4)$$

$$Pl(X) = \sum_{\substack{x \in D^\Theta \\ x \cap X \neq \emptyset}} m(x) \quad (5)$$

$$DSmP_\epsilon(\emptyset) = 0, \text{ and } \forall X \in G^\Theta \setminus \{\emptyset\}$$

$$DSmP_\epsilon(X) = \sum_{Y \in G^\Theta} \frac{\sum_{\substack{Z \subseteq X \cap Y \\ C(Z)=1}} m(Z) + \epsilon \cdot C(X \cap Y)}{\sum_{\substack{Z \subseteq Y \\ C(Z)=1}} m(Z) + \epsilon \cdot C(Y)} \cdot m(Y) \quad (6)$$

where  $G^\Theta$  can present full  $D^\Theta$ , or reduced  $D^\Theta$  with constraint, depending on the model used (free, or hybrid).  $\epsilon$  is an adjustment parameter,  $C(X \cap Y)$  and  $C(Y)$  are respectively the cardinality of  $X \cap Y$  and  $Y$ .

The last step in DSMT process is making a final decision, which presents a real challenge in many applications. In this work, we are interested in improving classification, we have to take a decision about pixels' belonging to a simple class also called Singleton class, and in this case there are two ways: taking decision based on maximum of generalized basic belief mass (gbba), or based on generalized belief function already computed as follows:

- Maximum of credibility  $Cr(\cdot)$  is widely used in many applications [28], and it is considered as a pessimistic decision.
- Maximum of plausibility  $Pl(\cdot)$  which is considered as an optimistic decision.
- Maximum of DSMT that is a compromise decision between the above decisions which are based on using probabilistic transformation  $P(\cdot)$  in the interval of  $[Cr(\cdot), Pl(\cdot)]$ .

## III. OVERALL SYSTEM

### A. Pre-processing

Generally, the pre-processing that precedes classification aims to eliminate imperfections that taint information by a set of actions as filtering, gradient operations, etc. However, in the classification based on the theories of the uncertain, these imperfections are protected, modeled and combined to help to make a decision.

The registration is the usually used pre-processing in the fusion process, it aims at setting correspondence between two or more images of a scene obtained from one or various sensors potentially at different spatial positions and scales, by using an optimal spatial and radiometric transformations between the images.

In the case of multimodal images, registration is an issue because of the significant difference between images [29], [30]. An original methodology was proposed in a previous work to answer the particular issue of the registration with multimodal imaging inputs in which we exploit the SURF scale- and rotation-invariant descriptors for the identification and the description of the interest points and we introduce a relevance filtering based on both SURF distance and orientation features in matching step [1].

### B. Feature Extraction

Feature extraction is a pivotal step in the classification process. It aims to underline the relevant features that are correspondent to various classes. It is worth stating that the

appropriate choice of extracted features improves the performance of classification step. Spectral, Spatial and perceptual features are extracted in this work.

1) *Spectral Information*: The spectral information is widely used on large classification methods. In this work, we have extracted the spectral values of each pixel as a vector of attributes and then converted them to Cielab space model for a better correlation with human color processing.

2) *Dense SURF Description*: Speeded up robust feature (SURF) proposed by Herbert Bay [31] is a spatial descriptor which consists originally of two phases, Detection and description of keypoints. We proposed in a previous work [32] to skip the detection phase and to perform description one to each pixel in the image. This is done, at the first by assigning to each pixel the dominant orientation calculated by combining the Haar wavelets results within a circular neighborhood around each pixel, and then creating  $4 \times 4$  sub regions around the pixel. In each sub-region, a pixel wise Haar wavelets responses are computed, which in turn are summed up to form 64-elements descriptor.

3) *Saliency Information*: Based on a performed comparative analysis of saliency detection in our multimodal data [33], we extract the saliency features by using the method proposed by Rahtu et al [34]. This method used local features contrast in illuminance, color mapped to feature space  $F(x)$  that is divided into disjoint bins. A saliency measure is calculated by applying a sliding windows  $w$  divided into inner windows  $K$  and border  $B$  in which a hypothesis that points in  $K$  are salient and points in  $B$  are not, the measure can be defined as probability conditional and computed through the Bayes Formula as

$$S_0(x) = \frac{h_K(x)p_0}{h_K(x)p_0 + h_B(x)(1-p_0)}, \quad (7)$$

with  $0 < p_0 < 1$ , and  $h_B(x) = P(F(x)|H1)$ . A regularized saliency measure is then introduced to make it more robust to the noise.

The motivation of integrating saliency information in the fusion process is the fact that usually visual perception succeeds easily to classify any objet or scene.

### C. SVM Pre-Classification

Support vector machine is a supervised classification method introduced by Vapnik [35], [36], widely used in classification applications thanks to its performance to deal with high-dimensional data. Basically, it is designed for binary class by finding an optimal hyperplan that separates the two classes linearly-separated. In non-linear separable class, the feature space is mapped to some higher dimensional feature space where the classes are separable using a Kernel function  $K$  that should fulfill Mercers conditions, the most kernels used are Radial Basis Function RBF, in which the decision function is expressed as a flow

$$h(x) = \text{Sign}\left(\sum_{i=1}^N \alpha_i \exp\{-|x - x_i|^2 / (2 \cdot \sigma^2)\}\right), \quad (8)$$

where  $\alpha_i$  are Lagrange multipliers, and the associated Kernel function is

$$k(x, x') = e^{-\frac{|x-x'|^2}{2\sigma^2}} \quad (9)$$

In case of multiclass problem, two main approaches were proposed: One-Versus-Rest approach in which  $k$  binary classifiers  $SVM_k$  are constructed for  $k$ -class classification, and One-versus-One in which  $(k(k-1))/2$  binary classifiers are applied on each pair of classes.

In order to generate the probabilities for DSMT, we have performed a pre-classification [32] based on combining spectral information (see III-B1) and Dense SURF information (see III-B2) using SVM classifier with RBF kernel to handle non-linear high-dimensional data in our multimodal dataset, and One-Versus-Rest approach to deal with incomplete information provided from diverse modalities.

### D. DSMT Classification

1) *Mass function estimation*: Mass estimation function step is very crucial in fusion process, because the imperfections such as uncertainty, imprecision, paradox will be introduced. The most generation used for these masses is the probabilities from pre-classification. The SVM classification of  $k$  images generates the matrices of the probabilities  $P(x_S|\theta_i)$  of pixels belonging to the singleton class of the frame of discernment  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , the same for  $k$  saliency map generated using the proposed method in [34]. Each source (modality/saliency map) noted  $S_i^b$  ( $i = 1, \dots, K$ ) gives the probability of belonging to one, or two classes, and their complementary classes which presents the mass of the partial ignorance. Based on [19], we denote  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , and the gbba mass of each source is given by:

$$\begin{cases} m_S(\theta_i) = \frac{P(x|\theta_i)}{z}, \forall \theta_i \in \Theta, \\ m_S(\bar{\theta}_i) = \frac{P(x|\cup_{0 < j < n} \theta_j)}{z}, \forall \theta_j \in \Theta, \\ m_S(\emptyset) = 0. \end{cases} \quad (10)$$

where  $z = \sum_{j=0}^n P(x|\theta_j)$  is a normalization term used in order to make sure that the sum of masses is equal to one.

2) *Combination of masses and decision*: The estimated masses must be combined with appropriate rules that handles the conflict generated from different sources  $S_i^b$ . In this work, we have used PCR6 [37] rule in combination step because it shows a better performance compared with all combination rules cited in the previous section and tested on our datasets. The PCR6 is computed as follows:

Considering  $S$  independent sources, the combined  $m_{PCR6}(\cdot)$  masses acquired from  $S > 2$  sources are computed

as  $m_{PCR6}(\cdot) = 0$ , and  $\forall X \in D^\Theta \setminus \{\emptyset\}$  as

$$m_{PCR6}(X) = m_{12\dots S}(X) + \sum_{\substack{X_1, X_2, \dots, X_S \in D^\Theta \setminus \{\emptyset\} \\ X_1 \cap X_2 \cap \dots \cap X_S = \emptyset}} \left[ \sum_{r=1}^S \delta_{X_r}^X m_r(X_r) \right] \frac{m_1(X_1) m_2(X_2) \dots m_S(X_S)}{m_1(X_1) + m_2(X_2) + \dots + m_S(X_S)} \quad (11)$$

where

$$\delta_{X_r}^X \triangleq \begin{cases} 1, & \text{if } X = X_r, \\ 0, & \text{if } X \neq X_r, \end{cases} \quad (12)$$

and where the mass  $m_{12\dots S}(X)$  corresponds to the conjunctive consensus on  $X$  between  $N > 2$  sources, which is given by

$$m_{12\dots S}(X) \triangleq \sum_{\substack{X_1, X_2, \dots, X_S \in D^\Theta \setminus \{\emptyset\} \\ X_1 \cap X_2 \cap \dots \cap X_S = X}} m_1(X_1) m_2(X_2) \dots m_S(X_S)$$

Once the combination step is achieved, we calculate the generalized belief function and we use a probabilistic transformation DSMP that converts the combined masses measure to a probability measure using (6) to make a final decision.

#### IV. DATA AND EXPERIMENTS

##### A. Data

Large sets of multimodal images acquired on wall paintings from the Germolles palace are used to demonstrate our proposed method. This palace was offered by Dukes of Burgundy Philip to his wife Margaret Flanders in 1380, and it was the only remaining castle of the Dukes of Burgundy so well preserved, its wall painting was restored between 1989 and 1991. However, there were no conservation reports of the applied restoration. In order to detect the original from restored area, the conservator of Germolles used the multimodal images that have the advantage of being fast and relatively inexpensive solution for the examination of large areas of wall paintings. This technical photography consists of recording a set of images with a commercial digital photographic camera which has been modified by removing the thermal filter regularly positioned in front of the CCD. In this way it is possible to record images of reflected visible light (Vis), reflected infrared light (IRr), reflected ultraviolet light (UVr) and UV-fluorescence (UVf). This set of images provides information about the optical behaviour of the surface when reached by the different types of light and therefore provides information about the original portions of wall paintings from recent repainting.

For illustration purpose, we select an area of a south wall of the dressing room of Margaret represented in Figure 1. This area presents a large white P (for Philip) that covers the walls and painted in green, which is presented by four modalities VIS, UVF, UVR and IRR. Each modality measures  $3744 \times 5616$  pixels. IRR modality shows very well the parts over non-original green surface. The image of the UV-induced

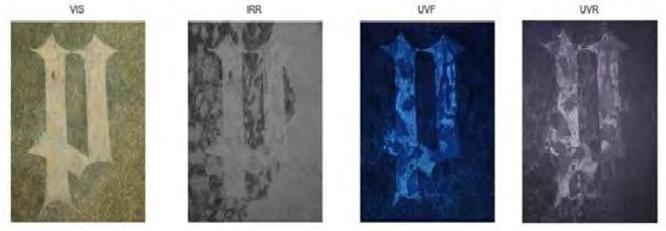


Figure 1. Multi-modal images of the same area.

fluorescence modality shows a relatively strong fluorescence corresponding to remains of an old/original paint layer over the white. The UVR image helps to identify the repainting original over the white of the letter P.

##### B. Experiments

The adopted methodology can be divided into four steps as illustrated in figure 2, which is started with the preprocessing by aligning each image with the VIS image that is used as a reference image.

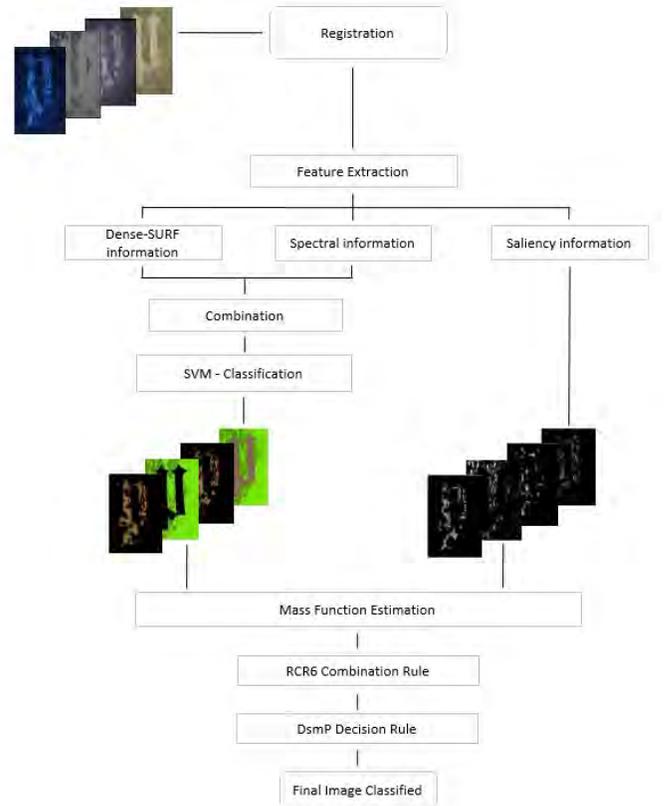


Figure 2. A representative illustration of the workflow.

In the second step, four topics have been identified: White original (WO), White repainted (WR), Green original (GO) and Green repainted (GR). Then spectral and Dense-SURF information is extracted and used jointly as the entry of the

SVM classifier using the RBF kernel. In parallel, Saliency information is extracted using the proposed method in [34], the provided maps are shown in figure 3.

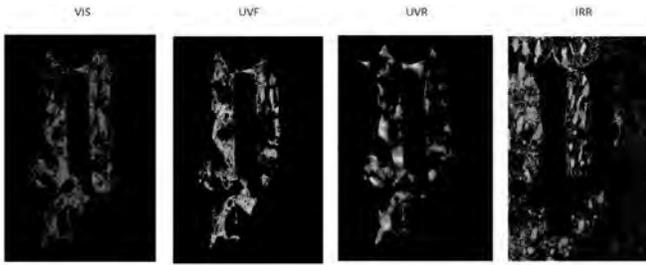


Figure 3. Saliency maps.

The third step is pre-classification using the SVM classifier that is applied to the images, in order to recover the probability matrixes of pixels belonging to classes. Each used modality highlights the presence of one or two classes. The UV-induced fluorescence modality shows a relatively strong fluorescence corresponding to the remains of an old painted layer of the white (WO) that reaches an accuracy of 92% using SVM, also UVR modality emphasizes WO class with a classification accuracy of 98%. Infrared light shows very well the parts over the original and repainted surface of the green and gets accuracy of 94% [32]. The provided maps are presented in figure 4.

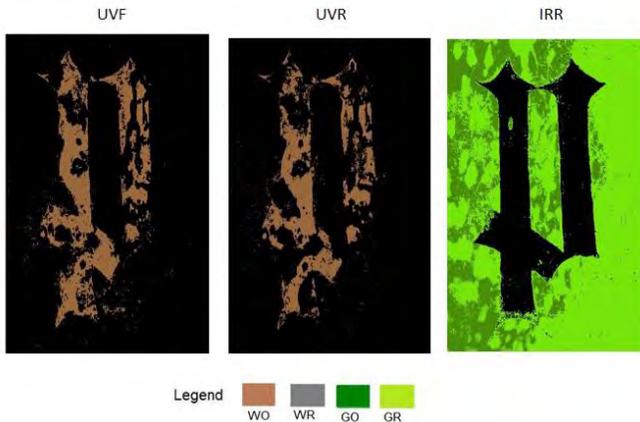


Figure 4. Multimodal SVM Classification.

The VIS modality reaches an accuracy of 98% with the classification of the two classes GO and GR, whereas this precision is reduced when classifying four classes because of the increase of the conflict. The classified image is presented in figure 5.

The last step presents the fusion process that is started with defining the frame of discernment  $\Theta = \{WO, WR, GO, GR\}$ . Due to the obtained information by SVM classification and saliency maps, there are some constraints that can be taken into account to deal with the real situation and to reduce the hyper power set  $D^\Theta$ , for example  $WO \cap GR = \emptyset$ .

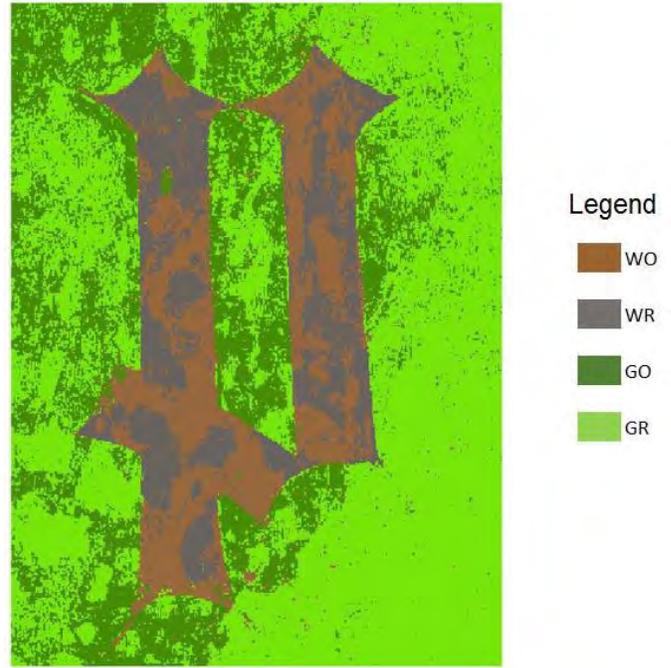


Figure 5. SVM classification of VIS modality.

Then the mass function that is associated with the emphasized class and its complementary in each modality are computed using equation (10). The PCR6 combination rule is used for combining the calculated masses basing on the equation (11), and as a final task, the decision is taken using maximum DsmP.

The final classified map, provided by DSMT only, is given in figure 6, and the final classified map obtained using DSMT-Saliency is shown in figure 7.

The results have progressed with the integration of the perceptual model in DSMT process, the visual analysis of the classification maps shows that the result of the proposed method much better with the ground truth over the WR and WO classes and appears to be closer to the reality, rather than the result obtained using DSMT only for the same classes, while the obtained map using unimodal image present a degraded result in terms of smoothness and connectivity between classes.

In this work, in order to evaluate the performance of the used methods and to compare the results, we have used the Overall accuracy (OA) that presents a percentage of correctly classified pixels, and Mean Error Rate (MER) that presents the percentage of misclassified pixels. Table I summarizes the obtained results using the different methods, from the results, we can note that the proposed method produces a better overall accuracy of 95.39% compared with the DSMT classification which provides an overall accuracy of 91.46% and the SVM classification that gives an overall accuracy of 86.43%, in terms of the error rate, the proposed method gives the low MER score of 4.61% compared with DSMT-Classification

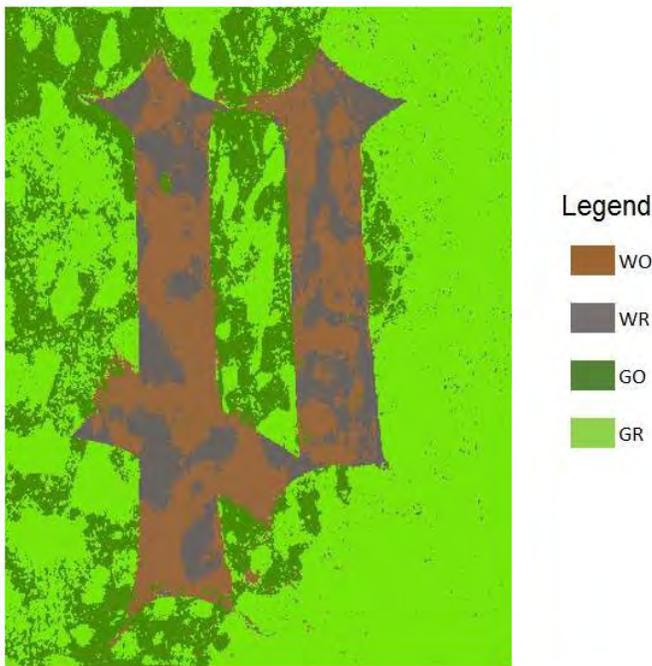


Figure 6. DSMT classification of multimodal images.

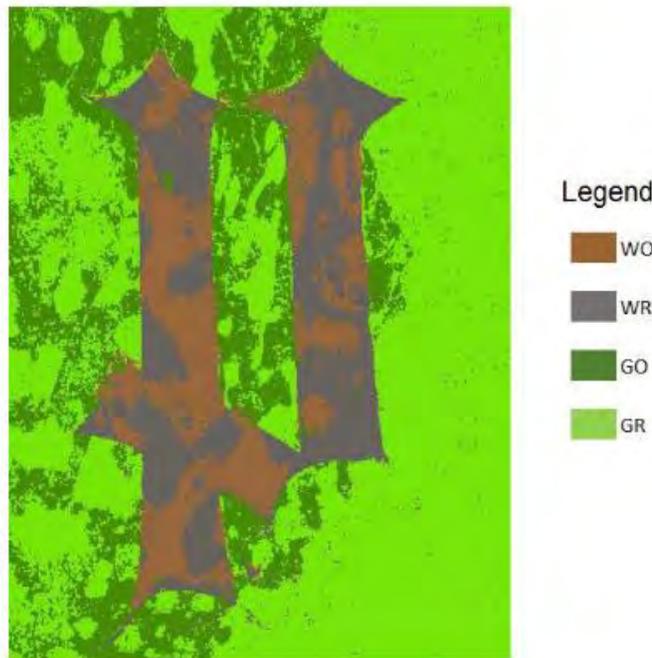


Figure 7. DSMT-Saliency classification of multimodal images.

and SVM-Classification that provides a MER of 8.53% and 12.60% respectively.

In conclusion, the use of DSMT theory with PCR6 combination rule provides a better result thanks to its effectiveness in managing correctly the conflict information that is provided from the different sources, and shows a significant

classification improvement compared with the unimodal SVM classification. Thus, the integration of saliency information in the fusion process presents a real benefit due to the powerful mechanism of the human brain in classification tasks.

Methods	OA	MER
SVM-Classification	86.43 %	12.60 %
DSMT-Classification	91.46 %	8.53 %
DSMT-Saliency-Classification	<b>95.39 %</b>	<b>4.61 %</b>

Table I  
ACCURACY AND ERRORS OF CLASSIFICATION RESULTS FROM DIFFERENT METHODS.

## V. CONCLUSION

In this paper, we have proposed a new method for multimodal image classification. As a first step, we have extracted spatial (Dense-SURF), spectral and saliency information. The extracted spatial and spectral information are combined and passing to the classifier SVM for pre-classification step. The SVM-classification results that are obtained from each modality is then fused using DSMT theory, the use of DSMT and SVM jointly provides better performance compared with the unimodal SVM classification. In the second step, the extracted saliency information is then modeled and combined with SVM classification results using DSMT process based on PCR6 combination rule and DSMT decision rule, the proposed method yields the best performance in terms of accuracy and error rate compared with DSMT-SVM classification and unimodal SVM classification.

## VI. ACKNOWLEDGEMENT

The authors thank the Château de Germolles managers for providing data and expertise and the COST Action TD1201 “Colour and Space in Cultural Heritage (COSCH)” ([www.cosch.info](http://www.cosch.info)) for supporting this case study. The authors also thank the PHC Toubkal/16/31: 34676YA program for the financial support.

## REFERENCES

- [1] H. Anzid, G. Le Goic, A. Bekkarri, A. Mansouri, D. Mammass, *Improving point matching on multimodal images using distance and orientation automatic filtering*, IEEE/ACS 13th Int. Conf. of Computer Systems and Applications (AICCSA), pp.1–8, Agadir, Morocco, 2016.
- [2] J. Huang, S. You, J. Zhao, *Multimodal image matching using self similarity*, 2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1–6, Washington, DC, USA, Oct 13–11, 2011.
- [3] Y. Ye, *Fast and robust registration of multimodal remote sensing images via dense orientated gradient feature*, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W7, pp. 1009–1015, 2017.
- [4] C. Samson, *Contribution à la classification des images satellitaires par approche variationnelle et équations aux dérivées partielles*, PhD. Thesis, University of Nice-Sophia Antipolis, France, Sept. 26, 2000.
- [5] L. Gómez-Chova, D. Tuia, G. Moser, G. Camps-Valls, *Multimodal classification of remote sensing images: A review and future directions*, Proceedings of the IEEE, Vol. 103, No. 9, pp. 1560–1584, Sept. 2015.
- [6] P. Kamavisdar, S. Saluja, S. Agrawal, *A survey on image classification approaches and techniques*, International Journal of Advanced Research in Computer and Communication Engineering, Vol.2, No. 1, pp. 1005–1009, Jan. 2013.

- [7] S. Majumdar, J. Bharadwaj, *Feature Level Fusion of Multimodal Images Using Haar Lifting Wavelet Transform*, World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol. 8, No. 6, pp. 1023–1027, 2014.
- [8] J.J. Clark, A.L. Yuille, *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic Publishers, 1990.
- [9] L.A. Zadeh, *Fuzzy sets*, Information and control, Vol. 8, No. 3, pp. 338–353, June 1965.
- [10] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, in Classic Works of the Dempster-Shafer Theory of Belief Functions, R. Yager and L. Liu (Eds.), pp. 57–72, Springer, 2008.
- [11] F. Smarandache, J. Dezert, *Proportional conflict redistribution rules for information fusion*, Chap. 1 of Advances and applications of DSMT for Information Fusion - Collected works, Vol. 2, F. Smarandache and J. Dezert (Eds.), American Research Press, Rehoboth, 2006.
- [12] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion, Vol. IV: Collected Works*, American Research Press, Rehoboth, 2015.
- [13] R. Khedam, A. Bouakache, G. Mercier, A. Belhadj-Aissa, *Fusion multilatée à l'aide de la théorie de Dempster-Shafer pour la détection et la cartographie des changements: application aux milieux urbain et périurbain de la région d'Alger*, Télédétection, Vol. 6, No. 4, pp. 359–404, 2006.
- [14] S. Foucher, M. Germain, J.-M. Boucher, G.B. Benie, *Multisource classification using ICM and Dempster-Shafer theory*, IEEE Trans. on Instrumentation and Measurement, Vol. 51, No. 2, pp. 277–281, April 2002.
- [15] M. Germain, J.-M. Boucher, G.B. Benie, *Fusion évidentielle multi source basée sur une nouvelle approche statistique floue*, in Proc. of ISIVC04 Conf, Brest, France, Jan. 2004.
- [16] T. Dencœur, *A k-nearest neighbor classification rule based on Dempster-Shafer theory*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 25, No. 5, pp. 804–813, May 1995.
- [17] A. Elhassouny, S. Idbraim, A. Bekkari, D. Mammass, D. Ducrot, *Application of DSMT-ICM with Adaptive decision rule to supervised classification in multisource remote sensing*, Journal of Computing, Vol. 5, No. 1, Jan. 2013.
- [18] A. Elhassouny, *Change detection by Fusion/Contextual Classification based on a Hybrid DSMT model and ICM with constraints*, International Journal Of computer Application, Vol. 35, No. 8, pp. 28–40, 2011.
- [19] N. Abbas, Y. Chibani, A. Martin, F. Smarandache, *The effective use of the DSMT for multi-class classification*, in Advances and Applications of DSMT for Information Fusion (collected works) - Vol. 4, F. Smarandache & J. Dezert Editors, pp. 359–380, American Research Press (ARP), 2015.
- [20] B. Pannetier, J. Dezert, *Multiple ground target tracking and classification with DSMT*, in Advances and Applications of DSMT for Information Fusion (collected works) - Vol. 4, F. Smarandache & J. Dezert Editors, pp. 337–342, American Research Press (ARP), 2015.
- [21] A.-L. Jousselme, A. Martin, P. Maupin, *Gestion de l'information paradoxale contrainte par des requêtes pour la classification de cibles dans un réseau de capteurs multi-modalités*, Colloque Systèmes Complexes d'Information et de Gestion des Risques pour l'Aide à la Décision (SCIGRAD'08), Brest, France, Nov. 24–25, 2008.
- [22] A. Bouakache, A. Belhadj-Aissa, G. Mercier, *Satellite image fusion using Dezert-Smarandache theory*, in Advances and Applications of DSMT for Information Fusion (collected works) - Vol. 3, F. Smarandache & J. Dezert Editors, American Research Press (ARP), 2009.
- [23] G. Quellec, M. Lamard, G. Cazuguel, C. Roux, B. Cochener, *Multimodal medical case retrieval using the Dezert-Smarandache theory*, in Proc. of Int. Conf. of the IEEE Engineering in Medicine and Biology Society, pp. 394–397, 2008.
- [24] J. Dezert, *Foundations for a new theory of plausible and paradoxical reasoning*, Information & Security Journal, Tzv. Semerdjiev Editor, Bulgarian Academy of Sciences, Vol. 9, 2002.
- [25] J. Dezert, F. Smarandache, *On the generation of hyper-powersets for the DSMT*, in Proc. of Fusion 2003 Int. Conf, Cairns, Australia, July 2003.
- [26] J. Dezert, F. Smarandache, *Presentation of DSMT*, Chap. 1 of Advances and Applications of DSMT for Information Fusion (collected works) - Vol. 1, F. Smarandache & J. Dezert Editors, American Research Press (ARP), 2004.
- [27] J. Dezert, F. Smarandache, *An introduction to DSMT*, in Advances and Applications of DSMT for Information Fusion (collected works) - Vol. 3, F. Smarandache & J. Dezert Editors, pp. 3-74, American Research Press (ARP), 2009.
- [28] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion (Collected works) - Vol. 2*, American Research Press (ARP), Rehoboth, U.S.A., 2006.
- [29] X. Shen, L. Xu, Q. Zhang, J. Jia, *Multi-modal and multi-spectral registration for natural images*, in Proc. of European Conf. on Computer Vision (CCV 2014), Part IV, Springer LNCS 8692, D. Fleet et al. (Eds.), pp. 309–324, 2014.
- [30] K. Karantza, A. Sotiras, N. Paragios, *Efficient and automated multi-modal satellite data registration through MRFs and linear programming*, in Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition Workshops, pp. 329–336, Columbus, OH, USA, June 23–28, 2014 .
- [31] H. Bay, T. Tuytelaars, L. Van Gool, *SURF: Speeded Up Robust Features*, in Proc. of 9th European Conference on Computer Vision (ECCV 2006), pp. 7–13, Springer Lecture Notes in Computer Science book series (LNIP, Volume 3951), 2006.
- [32] H. Anzid, G. Le Goic, A. Bekkari, A. Mansouri, D. Mammass, *Multi-modal images classification using SURF dense, Spectral information and Support Vector Machine*, in Proc. of the 2nd Int. Conf. on Intelligent Computing in Data Sciences ICDS, 2018 (also published in Procedia Computer Science, Vol. 148, pp. 107-115, 2019).
- [33] H. Anzid, G. Le Goic, A. Bekkari, A. Mansouri, D. Mammass, *Benchmarking Saliency Detection Methods on Multimodal Image Data*, Springer Lecture Notes in Computer Science, Vol. 10884, pp. 11–18, 2018.
- [34] E. Rahtu, J. Kannala, M. Salo, J. Heikkilä, *Segmenting salient objects from images and videos*, in Proc. of European Conference on Computer Vision, pp. 366–379, Springer Lecture Notes in Computer Science book series (LNIP, Vol. 6315), 2010.
- [35] L. Chapel, *Maintenir la viabilité ou la résilience d'un système : les machines à vecteurs de support pour rompre la malédiction de la dimensionnalité?*, PhD Thesis, Blaise Pascal University, Clermont II, France, 2007.
- [36] S. Aseervatham, *Apprentissage à base de noyaux sémantiques pour le traitement de données textuelles*, PhD. Thesis, Paris 13 University, Galilée Institut Laboratory of Data processing, Paris Nord, France, 217 pages, 2007.
- [37] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, Chap. 2 of Advances and Applications of DSMT for Information Fusion (collected works) - Vol. 2, F. Smarandache & J. Dezert Editors, pp. 69-88, American Research Press (ARP), Rehoboth, 2006.



# A New Probabilistic Transformation Based on Evolutionary Algorithm for Decision Making

Yilin Dong<sup>a</sup>, Xinde Li<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing, Jiangsu Province, China.

<sup>b</sup>The French Aerospace Lab (ONERA), Palaiseau, France.

Emails: dyl@seu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr

Originally published as: Y. Dong, X. Li, J. Dezert, *A New Probabilistic Transformation Based on Evolutionary Algorithm for Decision Making*, in Proc. of the 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10–13, 2017, and reprinted with permission.

**Abstract**—The study of alternative probabilistic transformation (PT) in DS theory has emerged recently as an interesting topic, especially in decision making applications. These recent studies have mainly focused on investigating various schemes for assigning both the mass of compound focal elements to each singleton in order to obtain Bayesian belief function for real-world decision making problems. In this paper, work by us also takes inspiration from both Bayesian transformation camps, with a novel evolutionary-based probabilistic transformation (EPT) to select the qualified Bayesian belief function with the maximum value of probabilistic information content (PIC) benefiting from the global optimizing capabilities of evolutionary algorithms. Verification of EPT is carried out by testing it on a set of numerical examples on 4D frames. On each problem instance, comparisons are made between the novel method and those existing approaches, which illustrate the superiority of the proposed method in this paper. Moreover, a simple constraint-handling strategy with EPT is proposed to tackle target type tracking (TTT) problem, simulation results of the constrained EPT on TTT problem prove the rationality of this modification.

**Keywords:** Evidence Reasoning, Probabilistic Transformation, Evolutionary Algorithm, Target Type Tracking problems, Decision Making.

## I. INTRODUCTION

Since the pioneering work of Dempster and Shafer [1], [2], which is known as Dempster-Shafer evidence theory (DST), in the late 70's regarding the possibility of distinguishing “unknown” and “imprecision” and fusing different evidences based on associative and commutative Dempster's combination rule, this new area of research (now known as evidence reasoning) has grown considerably as indicated by the notable increment of technical papers in peer-reviewed journals, conference and special sessions. However, the computational complexity of reasoning with DST is one of the major points of criticism this formalism has to face.

To overcome this difficulty, various approximating methods have been suggested that aim at reducing the number of focal elements in the frame of discernment (FoD) so as to maintain the tractability of computation computation. One common strategy is to simplify FoD by removing and/or aggregating focal elements for approximating original belief function [3]. Among these methods, probability transformations (PTs) seem particularly desirable for reducing such computation complexity by means of assigning the mass of non-singleton elements to each singleton [4], [5]. The research on this probabilistic

measure has received a lot of attentions and accordingly many efficient PTs have been pointed out by scholars in recent years. In them, a classical transformation, denoted as BetP [4], which offers a good compromise between the maximum of credibility (Bel) and the maximum of plausibility (Pl) for decision making. Unfortunately, BetP does not provide the highest probabilistic information content (PIC) [7]; Sudano [8] also proposed series of alternatives and principles of these similar to BetP, which were called PrPl, PrBel and PrHyb; CuzzP [9], which was proposed by Cuzzolin in the framework of DST in 2009, showed its ability of probabilistic transformation; Another novel transformation was proposed by Dezert and Smarandache in the framework of DSMT (free DSMT model, hybrid DSMT model or Shafer's model), which was called DSMT [7] and comprehensive comparisons have been made in [7] to prove the capabilities of DSMT in probabilistic transformation.

However, most mentioned aforementioned PTs have been always concentrated mainly on two crucial issues: (1) How to implement this operation (or assignment)? (2) How to evaluate the quality of this transformation? In this paper, we suggest a novel PT method based on evolutionary algorithm, namely, evolutionary-based probabilistic transformation (EPT), which alleviates the above two difficulties together based on optimization using a reasonable criteria. A similar idea was proposed by Han et.al [10] and the difference lies in the optimization approaches and objective functions. In the EPT method, the global search replaces the assigning operator used in the classical PTs and the evaluation criteria is embedded into EPT to provide important guidance for the searching procedure. Specifically, the mass of the singletons are randomly generated in evolutionary-based framework, which need to satisfy the constraints of probability distributions in evidence reasoning. Also, a selection operator is presented to assess the best individual in all populations by a special objective function (desirable evaluation criteria). Referring to the previous works [7], the PIC is used in this paper to select the best<sup>1</sup> solution as an objective function in EPT. Simulation results on 4D frames test cases show that the proposed EPT, in these problems, is able to outperform other PTs that pay special attention to some ratio created from the available information

<sup>1</sup>based on the highest PIC value.

(i.e. Bel or Pl). Moreover, we suggest a simple constraint-handling strategy with EPT that suits well for two target type tracking (TTT) problems. These first appealing results of EPT method encourage its use for more complex and real-world decision making problems.

The rest of this paper is organized as follows. In Section II we briefly summarize the basis of DST and several classical PT formulas. A novel EPT approach is presented in details in Section III. In Section IV several cases and comprehensive comparisons borrowed from previous papers are carried out to demonstrate the superiority of proposed method. Also, target type tracking problem and the pertinent analysis of EPT in TTT are described in detail in this section. Moreover, the limitation of EPT are also discussed in Section. V. The conclusion is drawn in Section. VI.

## II. BASIS OF BELIEF FUNCTIONS

In this section, we introduce the belief functions terminology of DST and the notations used in the sequel of this paper.

### A. DST basis

In DST [2], the elements  $\theta_i$  ( $i = 1, \dots, N$ ) of the frame of discernment (FoD)  $\Theta \triangleq \{\theta_1, \dots, \theta_N\}$  must be mutually exhaustive and exclusive. The power set of the FoD is denoted  $2^\Theta$  and a basic belief assignment (BBA), also called a mass function, is defined by the mapping:  $2^\Theta \rightarrow [0, 1]$ , which satisfies  $m(\emptyset) = 0$  and

$$\sum_{A \subseteq 2^\Theta} m(A) = 1 \quad (1)$$

where  $m(A)$  is defined as the BBA of  $A$ . The element  $A$  is called a focal element of  $m(\cdot)$  if  $m(A) > 0$ . The belief and plausibility functions, which are in one-to-one mapping with the BBA  $m(\cdot)$ , are defined for all  $A \subseteq \Theta$  by

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

$$Pl(A) = 1 - Bel(\bar{A}) = \sum_{A \cap B \neq \emptyset} m(B), \forall A \subseteq \Theta \quad (3)$$

where  $\bar{A} \triangleq \Theta \setminus A$  is the complement of  $A$  in  $\Theta$ . The belief interval  $[Bel(A), Pl(A)]$  represents the uncertainty committed to  $A$  and the bounds of this interval are usually interpreted as lower and upper bounds of the unknown (possibly subjective) probability of  $A$ . This interval plays an important role in the implementation of EPT as shown in details in Section III.

### B. DSMT basis

In the framework of Dezert-Smarandache Theory (DSMT) [5], the FoD  $\Theta$  is considered as a finite set of  $N$  exhaustive elements only (without the requirement of exclusivity of the elements). The BBA  $m(\cdot)$  is then defined on the hyper-power set of the FoD (i.e. the free Dedekind's lattice  $D^\Theta$ ), taking eventually into account some integrity constraints (if any). The main differences between DST and DSMT frameworks are: (1) the model on which one works with, and (2) the combination

rule. In the sequel, we will work with BBA defined only on the classical power-set for simplicity. Instead of distributing equally total conflicting mass onto elements of  $2^\Theta$  as within Dempster's rule through the normalization step, or transferring the partial conflicts onto partial uncertainties as within DSMT rule [4], we use the Proportional Conflict Redistribution rules (PCRs) [5] based on the transfer of conflicting masses (total or partial) proportionally to non-empty sets involved in the model according to all integrity constraints. In DSMT, the most effective rule is the PCR6 rule which is defined<sup>2</sup> for the fusion of two BBA's  $m_1(\cdot)$  and  $m_2(\cdot)$  as  $m_{PCR6}(\emptyset) = 0$  and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$

$$m_{PCR6}(A) = m_{12}(A) + \sum_{B \in 2^\Theta \setminus \{A\} | A \cap B = \emptyset} \left[ \frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right] \quad (4)$$

where  $m_{12}(A)$  is the conjunctive operator, and each element  $A$  and  $B$  are expressed in their disjunctive normal form.

### C. Classical Probabilistic Transformations

The efficiency of probabilistic transformation (PT) in the field of decision making has been analyzed in deep by Smets [4]. Various PTs have been proposed in the open literature and the main transformations are briefly recalled in this section.

1) *BetP*: Smets in [4], [6] first proposed pignistic probability to make decision which aims to transfer the mass of belief of each non-specific element onto the singletons. The classical pignistic probability is defined as  $BetP(\emptyset) = 0$ , and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ :

$$BetP(\theta_i) \triangleq \sum_{A \subseteq 2^\Theta, A \neq \emptyset} \frac{|\theta_i \cap A|}{|A|} \frac{m(A)}{1 - m(\emptyset)} \quad (5)$$

Because in Shafer's framework  $m(\emptyset) = 0$ , the formula (5) can simply be rewritten for any singleton  $\theta_i \in \Theta$  as

$$BetP(\theta_i) = \sum_{B \in 2^\Theta, \theta_i \subseteq B} \frac{1}{|B|} m(B) = m(\theta_i) + \sum_{B \in 2^\Theta, \theta_i \subset B} \frac{1}{|B|} m(B) \quad (6)$$

2) *CuzzP*: An intersection probability denoted as CuzzP [9] was proposed using the proportional repartition of the total non-specific mass (total non-specific mass ( $TNSM$ ) =  $\sum_{A \in 2^\Theta, |A|} m(A)$ ) for each contribution of the non-specific masses involved. CuzzP is defined by  $CuzzP(\emptyset) = 0$ , and for any singleton  $\theta_i \in \Theta$  by

$$CuzzP(\theta_i) \triangleq m(\theta_i) + \frac{Pl(\theta_i) - m(\theta_i)}{\sum_j (Pl(\theta_i) - m(\theta_j))} \cdot TNSM \quad (7)$$

<sup>2</sup>PCR6 rule coincides with PCR5 rule when combining only two BBA's [5].

3) *DSmP*: In 2008, Dezert and Smarandache [7] have proposed a new generalized pignistic transformation defined by  $DSmP_\varepsilon(\emptyset) = 0$  and for any singleton  $\theta_i \in \Theta$  by

$$DSmP_\varepsilon(\theta_i) \triangleq m(\theta_i) + (m(\theta_i) + \varepsilon) \times \left\{ \sum_{A \in 2^\Theta, \theta_N \subset A, |A| \geq 2} \frac{m(A)}{\sum_{B \in 2^\Theta, B \subset A, |B|=1} m(B) + \varepsilon \cdot |A|} \right\} \quad (8)$$

As shown in [7], DSmP makes a remarkable improvement compared with BetP, and CuzzP, since a more judicious redistribution of the ignorance masses to the singletons have been adopted by DSmP.

4) *PrBP1 and PrBP2*: Two novel pignistic probabilistic transformations were proposed by Pan in [11], which assume that the BBA is proportional to the product of  $Bel(\theta_i)$  and  $Pl(\theta_i)$  among each singleton element  $\theta_i$  of  $A \subseteq \Theta$ .

$$PrBP1(\theta_i) = \sum_{\theta_j \subseteq A} \left( \frac{Bel(\theta_j)Pl(\theta_j)}{\sum_{\theta_j \subseteq A} Bel(\theta_j)Pl(\theta_j)} \right) \cdot m(A) \quad (9)$$

Also, Pan et.al. assume that the masses are distributed proportionally to some given parameters  $s_i = Bel(\theta_i)/(1 - Pl(\theta_i))$  or  $s_i = Pl(\theta_i)/(1 - Bel(\theta_i))$ :

$$PrBP2(\theta_i) = \sum_{A, \theta_i \subseteq A} \left( \frac{s_i}{\sum_{j, \theta_j \subseteq A} s_j} \right) \cdot m(A) \quad (10)$$

As we can see, a Bayesian mass function which has only singleton focal elements can be obtained by any of these PTs.

#### D. Probabilistic Information Content (PIC)

The PIC criterion [12] is classically adopted to evaluate the performances of a probabilistic transformation of a BBA. If  $m(\cdot)$  is a Bayesian BBA defined on a discrete finite FoD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ , its PIC value is defined as<sup>3</sup>

$$PIC(m) \triangleq 1 + \frac{1}{\log_2 N} \sum_{i=1}^N m(\theta_i) \log_2 m(\theta_i) \quad (11)$$

The PIC metric actually measures the information content of a (probabilistic) source characterized by a Bayesian BBA  $m(\cdot)$ , which the value of this metric always belong to  $[0; 1]$ . It corresponds to the (normalized) dual of Shannon's entropy measure. When the Bayesian BBA is uniform over the FoD  $\Theta$ , one has  $m(\theta_i) = 1/N$  for  $i = 1, 2, \dots, N$  and the PIC metric is minimum, i.e.  $PIC(m) = PIC_{\min} = 0$ . The PIC metric is maximum, i.e.  $PIC(m) = PIC_{\max} = 1$  if the Bayesian BBA  $m(\cdot)$  is deterministic, that is if there exists an element  $\theta_i$  of  $\Theta$  such that  $m(\theta_i) = 1$ . While simple, appealing and generally adopted by the community, the PIC criteria is however not always sufficient to evaluate the efficiency of a PT as discussed in [14]. This point will be addressed in Section V.

<sup>3</sup>where  $0 \log_2(0) = 0$  by convention.

### III. EVOLUTIONARY-BASED PROBABILISTIC TRANSFORMATION (EPT)

The idea to approximate any BBA into a Bayesian BBA (i.e. a subjective probability measure) using the minimization of the Shannon entropy under compatibility constraints has been proposed recently by Han et al. in [10], [14] using "on-the-shelf" optimization techniques. In this paper, we present in details a new optimization method to achieve this PT based on a random evolutionary algorithm to acquire maximization of the PIC value. That is why we call it the Evolutionary-based Probabilistic Transformation (EPT) method.

Let's assume that the FoD of the original BBA  $m(\cdot)$  to approximate by a Bayesian BBA is  $\Theta \triangleq \{\theta_1, \theta_2, \dots, \theta_N\}$ . The EPT method consists of the following steps:

- Step 0 (setting parameters):  $t_{\max}$  is the max number of iterations;  $n_{\max}$  is the population size in each iteration;  $P_c$  is the crossover probability, and  $P_m$  is the mutation probability.
- Step 1 (population generation): A set  $\mathbf{P}_t$  of  $j = 1, 2, \dots, n_{\max}$  random probability values  $P_t^j = \{P^j(\theta_1), \dots, P^j(\theta_N)\}$  is generated such that the constraints (12)–(14) for  $j = 1, 2, \dots, n_{\max}$  are satisfied in order to make each random set of probabilities  $P_t^j$  compatible with the original BBA  $m(\cdot)$

$$P^j(\theta_i) \in [0; 1], \quad i = 1, 2, \dots, N \quad (12)$$

$$\sum_{i=1}^N P^j(\theta_i) = 1 \quad (13)$$

$$Bel(\theta_i) \leq P^j(\theta_i) \leq Pl(\theta_i), \quad i = 1, 2, \dots, N \quad (14)$$

- Step 2 (fitness assignment): To each probability set  $P_t^j$ , ( $j = 1, 2, \dots, n_{\max}$ ), we compute its PIC value and use it as its fitness factor  $F$ . More precisely, one takes  $F(P_t^j) = PIC(P_t^j)$ .
- Step 3 (best approximation of  $m(\cdot)$ ): the best set of probability  $P_t^{j_{\text{best}}}$  with highest PIC value is sought, and its associated index  $j_{\text{best}}$  are stored respectively in "Best-Individual" and "Index-of-BestIndividual".
- Step 4 (selection, crossover and mutation): The tournament selection, crossover and mutation operators drawn from evolutionary theory framework [13] are implemented to create the associated offspring population  $\mathbf{P}'_t$  based on the parent population  $\mathbf{P}_t$ . If  $F(P_t^{j_{\text{best}}}) \geq F(P_t'^{j_{\text{best}}})$ , then the "Best-Individual" remains unchanged; otherwise, Best-Individual =  $P_t'^{j_{\text{best}}}$ .
- Step 5 (Stopping EPT): The steps 1–4 illustrate the  $t$ -th iteration of EPT method. If  $t \geq t_{\max}$  then EPT method is completed, otherwise another iteration must be done by taking  $t + 1 = t$  and going back to step 1.

The scheme of EPT method is shown in Fig.1 and its pseudo-code is given in Algorithm 1.

### IV. SIMULATION RESULTS

In this section we compare EPT's results to other mentioned PTs. In particular, we show the important gain of PIC we

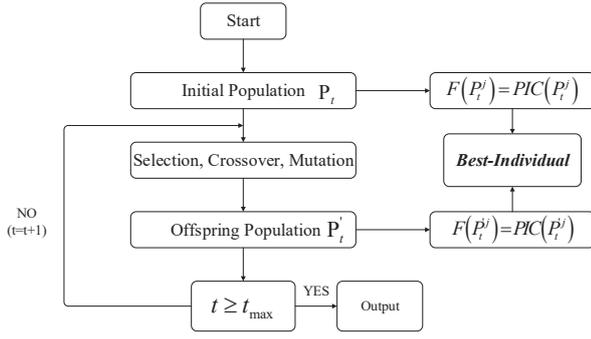


Figure 1: Scheme of EPT algorithm.

**Algorithm 1** Evolutionary-Based PT (EPT)

- 1: Define Stopping Criteria, ( $t \leq t_{\max}$ ); population Size  $n_{\max}$  for each iteration; crossover probability  $P_c$ , and mutation probability  $P_m$ .
- 2: Generate an initial random population  $\mathbf{P}_t$  of consistent probabilities  $P_t^j$  with  $m(\cdot)$ .
- 3: For each individual  $P_t^j$  in  $\mathbf{P}_t$  do
- 4: Calculate Fitness  $F(P_t^j) = PIC(P_t^j)$  of  $P_t^j$
- 5: Store the best individual  $P_t^{j_{\text{best}}}$
- 6: End
- 7: Repeat:
- 8: Selection: Select 2 individuals based on fitness
- 9: Crossover: exchange parts of 2 individuals with probability  $P_c$
- 10: Mutation: mutate the child individuals with probability  $P_m$
- 11: After these three sub-steps, the updated population  $\mathbf{P}'_t$  is obtained
- 12: Calculate the fitness of individuals of  $\mathbf{P}'_t$ , and store the best individual  $P_t^{j_{\text{best}}}$
- 13: If  $F(P_t^{j_{\text{best}}}) \geq F(P_t^{j_{\text{best}}})$
- 14: Best-Individual remains unchanged
- 15: else
- 16: Best-Individual =  $P_t^{j_{\text{best}}}$
- 17: If  $t \geq t_{\max}$  then stops, otherwise  $t + 1 \rightarrow t$  and go back to line 7

can obtain, and the capability of EPT to improve target type tracking.

### A. Examples and comparisons

In order to compare different PTs with EPT, two cases borrowed from [11] and [12] are considered, where PIC is used for evaluation. In all the following cases, the parameters  $t_{\max}$ ,  $n_{\max}$ ,  $P_c$  and  $P_m$  necessary to EPT method have been set to  $t_{\max} = 50$ ,  $n_{\max} = 1000$ ,  $P_c = 0.9$  and  $P_m = 0.1$  respectively.

**Example 1:**  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$

The BBA  $m(\cdot)$  to approximate by a Bayesian BBA (probability measure) is

$$\begin{aligned}
 m(\theta_1) &= 0.16, m(\theta_2) = 0.14, m(\theta_3) = 0.01, m(\theta_4) = 0.02 \\
 m(\theta_1 \cup \theta_2) &= 0.20, m(\theta_1 \cup \theta_3) = 0.09, m(\theta_1 \cup \theta_4) = 0.04 \\
 m(\theta_2 \cup \theta_3) &= 0.04, m(\theta_2 \cup \theta_4) = 0.02, m(\theta_3 \cup \theta_4) = 0.01 \\
 m(\theta_1 \cup \theta_2 \cup \theta_3) &= 0.10, m(\theta_1 \cup \theta_2 \cup \theta_4) = 0.03 \\
 m(\theta_1 \cup \theta_3 \cup \theta_4) &= 0.03, m(\theta_2 \cup \theta_3 \cup \theta_4) = 0.03 \\
 m(\Theta) &= 0.08
 \end{aligned}$$

The Bayesian BBA obtained by classical PT (5)–(10) and EPT with their corresponding PIC values calculated by (11) are given in Table I. As expected, the EPT provides the maximum PIC.

Table I: Results of Different PTs in Example 1.

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	PIC
<i>CuzzP</i>	0.3860	0.3382	0.1607	0.1151	0.0790
<i>BetP</i>	0.3983	0.3433	0.1533	0.1051	0.0926
<i>DSmP<sub>0</sub></i>	0.5176	0.4051	0.0303	0.0470	0.3100
<i>DSmP<sub>0,001</sub></i>	0.5162	0.4043	0.0319	0.0476	0.3058
<i>PrBP1</i>	0.5419	0.3998	0.0243	0.0340	0.3480
<i>PrBP2</i>	0.5578	0.3842	0.0226	0.0354	0.3529
<b>EPT</b>	<b>0.7246</b>	<b>0.2218</b>	<b>0.0266</b>	<b>0.0270</b>	<b>0.4508</b>

**Example 2:**  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$

In this case, we randomly generate BBAs and compare EPT with classical PTs (*CuzzP*, *BetP*, *DSmP*, *PrBP1* and *PrBP2* given by (5)–(10)). The original BBAs to approximate are generated according to Algorithm 2 of [15].

**Algorithm 2** Random generation of BBA

- 1: Input: Frame of Discernment  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$
- 2:  $N_{\max}$ : Maximum number of focal element
- 3: Output: BBA-m
- 4: Generate  $K(\Theta)$ , which is the power set of  $\Theta$
- 5: Generate a random permutation of  $K(\Theta) \rightarrow R(\Theta)$
- 6: Generate an integer between 1 and  $N_{\max} \rightarrow l$
- 7: For each First  $k$  elements of  $R(\Theta)$  do
- 8: Generate a value within  $[0, 1] \rightarrow m_i, i = 1, \dots, l$
- 9: End
- 10: Normalize the vector  $m = [m_1, m_2, \dots, m_l] \rightarrow m'$
- 11:  $m(\theta_i) = m'_i$

In our test, we have set the cardinality of the FoD to 4 and fixed the number of focal elements to  $l = N_{\max} = 15$ . We randomly generate  $L = 30$  BBA's. Six PT methods are tested and PIC is used to evaluate the quality of their corresponding results are shown in Fig.2. As we can see, EPT outperforms significantly other methods based on maximum of PIC criterion, which is normal.

### B. Evaluation of EPT in Target Type Tracking problem

Target Type Tracking (TTT) problem can be briefly described below [16]:

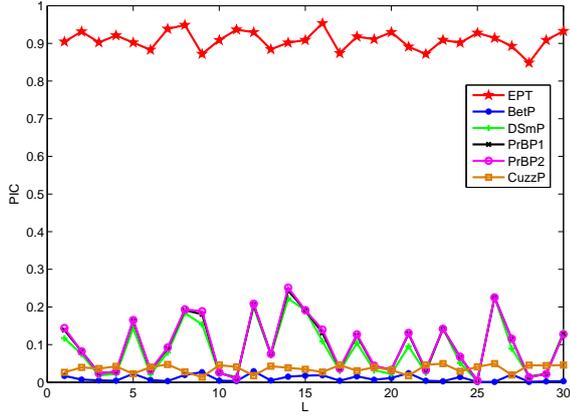


Figure 2: PIC values obtained with EPTs and classical PTs.

### 1) Target Type Tracking Problem (TTT):

1. Considering  $\zeta = 1, 2, \dots, \zeta_{max}$  be the time index and let  $N$  possible target types  $Tar_{\zeta} \in \Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  in the surveillance area; For instance, in the normal air target surveillance systems the FoD could be  $\Theta = \{Fighter, Cargo\}$ . That is,  $Tar_1 = \theta_1 \triangleq Fighter$ ,  $Tar_2 = \theta_2 \triangleq Cargo$ . Similarly, the FoD in a ground target surveillance systems could be  $\Theta_{ground} = \{Tank, Truck, Car, Bus\}$ . In this paper, we just consider the air target surveillance systems to prove the practicability of EPT.
2. At every time  $\zeta$ , the true type of the target  $Tar(\zeta) \in \Theta$  is immediately observed by an attribute-sensor (here, we assume a possible target probability).
3. A defined classifier is applied to process the attribute measurement of the sensor which provides the probability  $Tar_d(\zeta)$  on the type of the observed target at each instant  $\zeta$ .
4. The sensor is in general not totally reliable and is characterized by a  $N \times N$  confusion matrix:

$$\mathbf{M} = [M_{ij} = P(Tar_d = Tar_j | TrueType = Tar_i)] \quad (15)$$

where  $0 \leq i \leq N; 0 \leq j \leq N$ .

Here, we briefly summarize the main steps of TTT using EPT.

1. Initialization. Determine the target type frame  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  and set the initial BBA  $m^{initial}(\theta_1 \cup \theta_2 \cup \dots \cup \theta_N) = 1$  since there is no information about the first target type that will be observed;
2. Updating BBA. An observed BBA  $m_{obs}(\cdot)$  on types of unknown observed target is defined from current target type declaration and confusion matrix  $\mathbf{M}$ ;
3. Combination. We combine the current BBA  $m_{obs}(\cdot)$  with initial BBA  $m^{initial}(\cdot)$  according to PCR6 combination rule:  $m_{PCR6}(\cdot) = m_{obs}(\cdot) \oplus m^{initial}(\cdot)$ ;

4. Approximation. Using  $EPT(\cdot)$  to approximate  $m_{PCR6}(\cdot)$  into a Bayesian BBA;
5. Decision Making. Taking a final decision about the type of the target at current observation time based on the obtained Bayesian BBA;
6. Updating BBA. Set  $m^{initial}(\cdot) = m_{PCR6}(\cdot)$ , and increase time index  $\zeta = \zeta + 1$  and go back to step 2.

2) Raw Dataset of TTT: We have tested our EPT-based TTT on a very simple scenario for a 2D TTT, namely  $\Theta = \{Fighter, Cargo\}$  for two types of classifiers. The matrix  $\mathbf{M}_1$  corresponds to the confusion matrix of the good classifier, and  $\mathbf{M}_2$  corresponds to the confusion matrix of the poor classifier.

$$\mathbf{M}_1 = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}; \mathbf{M}_2 = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix} \quad (16)$$

In our scenario, a true Target Type sequence over 120 scans is generated according to Fig. 3. We can observe clearly from Fig. 3 that Cargo (which is denoted as Type 2) appears at first in the sequence, and then the observation of the Target Type switches three times onto Fighter Type (Type 1) during different time duration (namely, 20s, 10s, 5s).

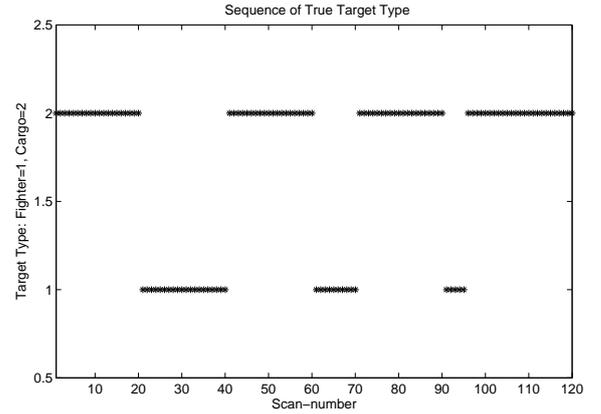


Figure 3: Raw Sequence of True Target Type.

**A pathological case for TTT:** Our analysis has shown that EPT can nevertheless be in troubles for tracking two target types as proved in this simple particular example (when  $0 \leq m(\theta_1 \cup \theta_2) \leq 0.1$ ). Let's consider the following BBA

$$m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0, 1, 0]$$

According to the compatibility constraints (12)–(14), the population  $\mathbf{P}'_t$  is obtained from  $\mathbf{P}_t$  through a selection procedure. Next, individual  $P_t^j$  in  $\mathbf{P}'_t$  which is denoted as  $P_t^j = [m'(\theta_1), m'(\theta_2)]$  is subject to initial constraint (1) and (17):

$$\begin{aligned} m'(\theta_1) &\geq (Bel(\theta_1) = m(\theta_1) = 0) \\ m'(\theta_1) &\leq (Pl(\theta_1) = m(\theta_1) + m(\theta_1 \cup \theta_2) = 0 + 0 = 0); \\ m'(\theta_2) &\geq (Bel(\theta_2) = m(\theta_2) = 1) \\ m'(\theta_2) &\leq (Pl(\theta_2) = m(\theta_2) + m(\theta_2 \cup \theta_1) = 1 + 0 = 1); \end{aligned} \quad (17)$$

From the above inequalities, one sees that only one probability measure  $P_t^S = [m(\theta_1), m(\theta_2)] = [0, 1]$  (where the superscript index  $S$  means *Single*) satisfies this constraint. However because of mechanism of EPT and real-coded generic algorithm (RCGA), the probabilities  $P_t^j$  in population  $\mathbf{P}_t$  which are randomly generated in the interval  $[0, 1]$ , will have a very little chance to be equal to the suitable measure  $[0, 1]$  satisfying the constraints. That is why EPT becomes inefficient in this case which occurs with a probability of  $1/n_{\max}$ , where  $n_{\max}$  denotes the size of population<sup>4</sup>  $\mathbf{P}_t$ . Unfortunately, in TTT decision making problems, such case cannot be avoided because it can really happens.

To circumvent this problem and make EPT approach working in all circumstances, we need to modify a bit the EPT method to generate enough individuals for making selection step efficiently when the bounds of belief interval  $[Bel, Pl]$  take their min and max values  $([0.9, 0.05, 0.05], [0.05, 0.9, 0.05])$ . For achieving this, we propose to enlarge the interval through a parameter  $\lambda$ , and maintain the property of original interval in some degree at the same time. More precisely, the modified belief interval, denoted  $[Bel', Pl']$ , is heuristically computed by a simple thresholding technique as follows:

First, we assume that the original BBA we consider here for FoD  $\Theta = \{\theta_1, \theta_2\}$  is  $[\theta_1, \theta_2, \theta_1 \cup \theta_2] = [a, b, c], (a + b + c) = 1; 0 \leq c \leq 0.1$

Step 1: Let  $m'(\theta_1 \cup \theta_2) = c + \lambda$ ;

Step 2: if  $a > b$

$$m'(\theta_1) = a - \lambda; m'(\theta_2) = b; m'(\theta_1 \cup \theta_2) = c + \lambda; \quad (18)$$

Step 3: if  $a \leq b$

$$m'(\theta_1) = a; m'(\theta_2) = b - \lambda; m'(\theta_1 \cup \theta_2) = c + \lambda; \quad (19)$$

So the value of  $[Bel'(\theta_1), Pl'(\theta_1)]$  and  $[Bel'(\theta_2), Pl'(\theta_2)]$  can be calculated based on Eq.(18),Eq.(19), which are presented as follows:

When  $a > b$ :

$$\begin{cases} Pl'(\theta_1) = m(\theta_1) + m'(\theta_1 \cup \theta_2) = a - \lambda + c + \lambda = a + c; \\ Bel'(\theta_1) = 1 - Pl'(\bar{\theta}_1) = 1 - (b + c + \lambda) = a - \lambda. \end{cases} \quad (20)$$

$$\begin{cases} Pl'(\theta_2) = m(\theta_2) + m'(\theta_1 \cup \theta_2) = b + c + \lambda = b + c + \lambda; \\ Bel'(\theta_2) = 1 - Pl'(\bar{\theta}_2) \\ = 1 - (a - \lambda + c + \lambda) = 1 - (a + c) = b. \end{cases} \quad (21)$$

When  $a \leq b$ :

$$\begin{cases} Pl'(\theta_1) = m(\theta_1) + m'(\theta_1 \cup \theta_2) = a + c + \lambda; \\ Bel'(\theta_1) = 1 - Pl'(\bar{\theta}_1) \\ = 1 - (b - \lambda + c + \lambda) = 1 - (b + c) = a. \end{cases} \quad (22)$$

<sup>4</sup>In our simulation, we did take  $n_{\max} = 1000$ .

$$\begin{cases} Pl'(\theta_2) = m(\theta_2) + m'(\theta_1 \cup \theta_2) = b - \lambda + c + \lambda = b + c; \\ Bel'(\theta_2) = 1 - Pl'(\bar{\theta}_2) = 1 - (a + c + \lambda) = b - \lambda. \end{cases} \quad (23)$$

**Explanation:** Through step 1, one computes the total singleton mass one has in the entire BBA and the threshold value 0.9 allows to evaluate if the percentage of singleton mass is big enough or not. Here, we not only consider the unique extreme case  $m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0, 1, 0]$ , but also other possible cases such as  $m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0.0001, 0.9998, 0.0001]$ . Why do we consider the concept of percentage? Actually, the higher percentage of singleton mass, the smaller interval for  $P_t^j$ , in other words, the higher value of  $m(\theta_1 \cup \theta_2)$ , the bigger interval for  $P_t^j$  which can be shown in Eq.(17); The step 2 and step 3 give the way of calculating the updated upper bound of belief interval  $[Bel', Pl']$  and Eq.(20)–Eq.(23) prove that the parameter  $\lambda$  determines the range of the interval; Next, we give two examples to show how the above method works:

#### The pathological case 1 for TTT (using modified EPT)

$$m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0.0001, 0.9998, 0.0001].$$

Here, the parameter  $\lambda$  is arbitrarily<sup>5</sup> set to 0.4. Then one computes in step 2 the modified plausibility bounds  $Bel'(\theta_1) = 0.0001, Pl'(\theta_1) = 0.0001 + 0.0001 + \lambda = 0.4002$  and  $Bel'(\theta_2) = 0.9998 - 0.4 = 0.5998, Pl'(\theta_2) = 0.9999$ . So we get  $[Bel'(\theta_1), Pl'(\theta_1)] = [0.0001, 0.4002]$  and  $[Bel'(\theta_2), Pl'(\theta_2)] = [0.5998, 0.9999]$ .

Consequently, any Bayesian BBA  $P_t^j = [m'(\theta_1), m'(\theta_2)]$  must be generated according the (modified) compatibility constraints

$$\begin{aligned} m'(\theta_1) &\in [Bel'(\theta_1), Pl'(\theta_1)] = [0.0001, 0.4002] \\ m'(\theta_2) &\in [Bel'(\theta_2), Pl'(\theta_2)] = [0.5998, 0.9999] \end{aligned}$$

#### The pathological case 2 for TTT (using modified EPT)

$$m_{target}(\cdot) = [\theta_1, \theta_2, \theta_1 \cup \theta_2] = [0.45, 0.48, 0.07].$$

Here, the parameter  $\lambda$  is set to 0.2. Then any Bayesian BBA  $P_t^j = [m'(\theta_1), m'(\theta_2)]$  must be generated according the (modified) compatibility constraints

$$\begin{aligned} m'(\theta_1) &\in [Bel'(\theta_1), Pl'(\theta_1)] = [0.45, 0.72] \\ m'(\theta_2) &\in [Bel'(\theta_2), Pl'(\theta_2)] = [0.28, 0.55] \end{aligned}$$

In order to evaluate the influence of the parameter  $\lambda$ , we have reexamined all the pathological cases based on the following procedure:

- 1) The value of parameter  $\lambda$  is taken to five possible values: 0, 0.1, 0.2, 0.3, 0.4, 0.5;
- 2) We randomly generate initial population  $\mathbf{P}_t$  based on  $\lambda$ , which is also subjected to the constraints (12)–(14).

<sup>5</sup>The value of the parameter  $\lambda$  can be chosen to any value in  $[0, 1]$  by the designer for his/her own reason to ensure the alternative interval effectively in modified EPT version.

With this simulation, we can observe in Fig.4 the impact of  $\lambda$  value on the number of  $P_t^j$  in  $\mathbf{P}_t$ . When  $\lambda = 0$  happens <sup>6</sup>, there exists no suitable  $P_t^j$  for case one which demonstrates the necessity to circumvent the pathological case problem. Obviously, the number of  $P_t^j$  increases with the increase of  $\lambda$  value, which efficiently proves the advantage of using the modified EPT approach to make selection step of the evolutionary algorithm more efficient. One point we need to clarify is that the intervals i.e.  $[Bel'(\theta_1), Pl'(\theta_1)]$ ,  $[Bel'(\theta_2), Pl'(\theta_2)]$  induced from parameter  $\lambda$  above aims at guaranteeing enough number of  $P_t^j$  in  $\mathbf{P}_t$  in the implementation of EPT.

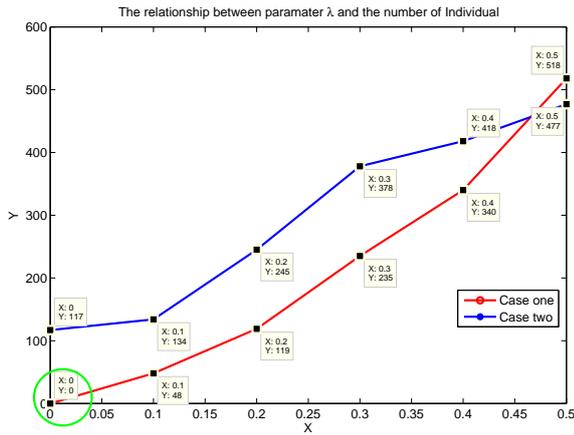


Figure 4: Impact of  $\lambda$  (x-axis) on individuals in  $\mathbf{P}_t$  (y-axis).

3) *Simulation Results of TTT Based on Modified EPT*: Our simulation consists in 100 Monte-Carlo runs and we show in the sequel the averaged performances of EPT and DSMP. The figures 5–8 illustrate the Bayesian BBA’s obtained by DSMP [7] -(part a) and our new EPT method-(part b) based on TTT using PCR6 fusion rule. One sees that regardless of the good classifier  $M_1$  and poor classifier  $M_2$ , EPT is able to track properly the quick changes of target type.

### V. LIMITATION OF EPT

As pointed out by Han et al. in [14], in general it is not enough, nor comprehensive to evaluate the quality of probabilistic transformation of a BBA from only the PIC criterion, even if the chosen PT provides highest PIC value by optimization. Our EPT approach, is not exempt of this problem of course as we can see in the simple example below, where no optimization technique provides useful (robust) solution.

Let’s consider the FoD  $\Theta = \{\theta_1, \theta_2\}$  with the BBA to approximate chosen as follows:

$$m(\theta_1) = 0.10001, m(\theta_2) = 0.10000, m(\theta_1 \cup \theta_2) = 0.79999$$

Based on PIC value optimization using EPT (or any other efficient optimization techniques), we will obtain the Bayesian

<sup>6</sup>which actually the original EPT is applied

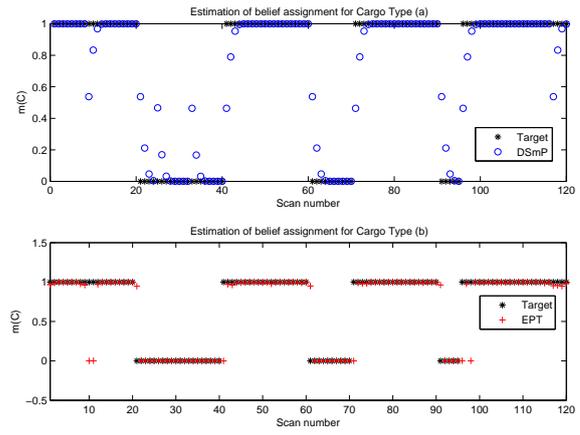


Figure 5: Belief Mass for Cargo Type for  $M_1$ .

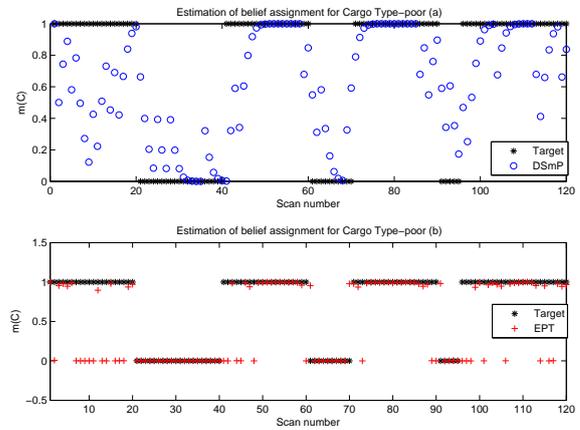


Figure 6: Belief Mass for Cargo Type for  $M_2$ .

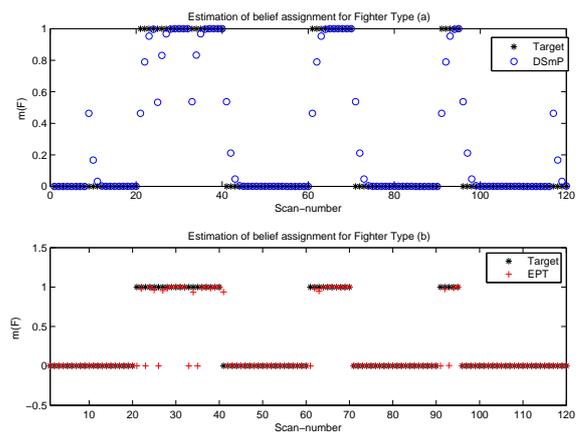


Figure 7: Belief Mass for Fighter Type for  $M_1$ .

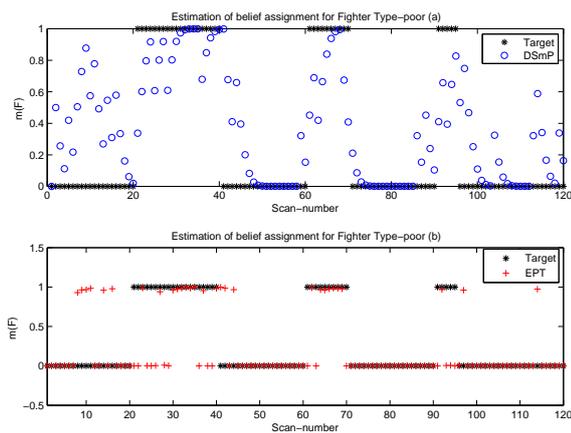


Figure 8: Belief Mass for Fighter Type for  $M_2$ .

BBA  $m(\theta_1, \theta_2) = [0.0001605, 0.9998394]$  with  $PIC = 0.9977$ . This simple example shows that in the original BBA  $m(\theta_1)$  is almost the same as  $m(\theta_2)$  and there is no solid reason to get a very high probability for  $\theta_2$  and a small one for  $\theta_1$  in the Bayesian BBA, even if a highest PIC is reached. Exaggerated high PIC is not always preferred (unreasonable or directly make wrong decisions), which can be seen in Fig.6 and Fig.8, although the PIC should be as high as possible for decision making problems. Therefore, a reasonable compromise must be found between PIC level and also fidelity level of the transformations to the original BBA, which is a theoretical open challenging problem left for further research works.

## VI. CONCLUSION

An evolutionary algorithm for probabilistic transformation (EPT) has been proposed in this paper. It uses the genetic algorithm to obtain Bayesian belief function with highest PIC value. The utility of EPT was verified on a set of three probabilistic transformation cases borrowed from the literature. On these cases, the performance of EPT has been compared to other existing probabilistic transformations. Our results indicate that EPT performs better than others on all problems from PIC increasing standpoint. The shortcomings of original EPT version have been clearly identified on two type tracking problems, and they have been overcome thanks to a modification of belief interval constraints. As future works, we would like to establish more appropriate evaluation criteria and make more comparisons between performances of this EPT approach with other recent proposed evolutionary algorithms. We would also make more investigations on EPT to extend it to work with more than two targets.

## ACKNOWLEDGMENT

This work was supported by NNSF of China (No. 61175091), Qing Lan Project of Jiangsu Province, Aeronautical Science Foundation of China (20140169002), and Six Major Top-talent Plan of Jiangsu Province.

## REFERENCES

- [1] A.P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Stat.*, Vol. 38, No. 4, pp. 325–339, 1967.
- [2] G. Shafer, *A mathematical theory of evidence*, Princeton Univ. Press, 1976.
- [3] T. Denœux, Inner and outer approximation of belief structures using a hierarchical clustering approach, *International Journal of Uncertainty, Fuzziness, and Knowledge-based Systems*, Vol.9, No.4, pp. 437-460, 2001.
- [4] P. Smets, The Combination of Evidence in the Transferable Belief Model, *IEEE Trans. on PAMI*, Vol. 5, pp. 29–39, 1990.
- [5] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, American Research Press, Rehoboth, NM, U.S.A., Vol. 1–4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [6] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, *Int. J. of Approximate Reasoning*, Vol. 38, 2005.
- [7] J. Dezert, F. Smarandache, A new probabilistic transformation of belief mass assignment, In *Proc. of 11th Int. Conf. on Information Fusion*, Cologne, Germany, pp. 1–8, June-July 2008.
- [8] J. Sudano, Equivalence Between Belief Theories and Naïve Bayesian Fusion for Systems with Independent Evidential Data-Part I, *The theory*, In *Proc. of Fusion 2003*, Cairns, Australia, July 2003.
- [9] F. Cuzzolin, The Intersection probability and its properties, in *Proc. of ECSQARU 2009*, Verona, Italy, July 1–3, 2009.
- [10] D. Han, J. Dezert, C.Z. Han. New Basic belief assignment approximations based on optimization, In *Proc. of 15th Int. Conf. on Information Fusion*, pp. 286–293, July 2012.
- [11] W. Pan, J.Y. Hong, New methods of transforming belief functions to pignistic probability functions in evidence theory, 2009 Int. Workshop on Intelligent System and Applications, 2009 (In Chinese).
- [12] J. Sudano, The system probability information content (PIC) relationship to contributing components, combining independent multisource beliefs, hybrid and pedigree pignistic probabilities, in *Proc. of Int. Conf. on Information Fusion (Fusion 2002)*, Annapolis, Maryland, U.S.A, Vol. 2, pp. 1277–1283, July 2002.
- [13] N. Srinivas, K. Deb, Multiobjective function optimization using non-dominated sorting genetic algorithms, *Evolutionary Computation*, Vol. 2, No. 3, pp. 221-248, Fall 1995.
- [14] D.Q. Han, J. Dezert, Z.S. Duan, Evaluation of Probability Transformations of Belief Functions for Decision Making. *IEEE Transactions on Systems, Man, And Cybernetics: Systems*, Vol. PP, No. 99, 2015.
- [15] A.-L. Josselme, P. Maupin, On some properties of distances in evidence theory, in *Workshop on Theory of Belief Functions*, pp.1–6, Brest, France, March 31st–April 2nd, 2010.
- [16] J. Dezert, A. Tchamova, F. Smarandache, P. Konstantinova, Target Type Tracking with PCR5 and Dempster’s rules: A Comparative Analysis, in *Proc. of Int. Conf. on Information Fusion (Fusion 2006)*, Florence, Italy, July 2006.

# A Hierarchical Flexible Coarsening Method to Combine BBAs in Probabilities

Yilin Dong<sup>a</sup>, Xinde Li<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing, Jiangsu Province, China.

<sup>b</sup>The French Aerospace Lab (ONERA), Palaiseau, France.

Emails: dyl@seu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr

Originally published as: Y. Dong, X. Li, J. Dezert, *A Hierarchical Flexible Coarsening Method to Combine BBAs in Probabilities*, in Proc. of the 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10–13, 2017, and reprinted with permission.

**Abstract**—In many applications involving epistemic uncertainties usually modeled by belief functions, it is often necessary to approximate general (non-Bayesian) basic belief assignments (BBAs) to subjective probabilities (called Bayesian BBAs). This necessity occurs if one needs to embed the fusion result in a system based on the probabilistic framework and Bayesian inference (e.g. tracking systems), or if one wants to use classical decision theory to make a decision. There exists already several methods (probabilistic transforms) to approximate any general BBA to a Bayesian BBA. From a fusion standpoint, two approaches are usually adopted: 1) one can approximate at first each BBA in subjective probabilities and use Bayes fusion rule to get the final Bayesian BBA, or 2) one can fuse all the BBAs with a fusion rule, typically Dempster-Shafer's, or PCR6 rules (which is very costly in computations), and convert the combined BBA in a subjective probability measure. The former method is the simplest method but it generates a high loss of information included in original BBAs, whereas the latter is intractable for high dimension problems. This paper presents a new method to achieve this task based on hierarchical decomposition (coarsening) of the frame of discernment, which can be seen as an intermediary approach between the two aforementioned methods. After the presentation of this new method, we show through simulations how it performs with respect to other methods.

**Keywords:** Information fusion, belief functions, DST, DSMT, PCR6 rule, coarsening.

## I. INTRODUCTION

The theory of belief functions, known as Dempster-Shafer Theory (DST) has been developed by Shafer [1] in 1976 from Dempster's works [2]. Belief functions allow to model epistemic uncertainty and they have been already used in many applications since the 1990's [3], mainly those related to expert systems, decision-making support and information fusion. To palliate some limitations of DST, Dezert and Smarandache have proposed an extended mathematical framework of belief functions with new efficient quantitative and qualitative rules of combinations, which is called DSMT (Dezert and Smarandache Theory) in the literature [4], [5] with applications listed in [6]. One of the major drawbacks of DST and DSMT is their high computational complexities, as soon as the fusion space (i.e. frame of discernment - FoD) and the number of sources to combine are large<sup>1</sup>.

<sup>1</sup>DSMT is more complex than DST, and the Proportional Conflict Redistribution rule #6 (PCR6 rule) becomes computationally intractable in the worst case as soon as the cardinality of the Frame of Discernment (FoD) is greater than six.

To reduce the computational cost of operations with belief functions when the number of focal elements is very large, several approaches have been proposed by different authors. Basically, the existing approaches rely either on efficient implementations of computations as proposed for instance in [7], [8], or on approximation techniques of original Basic Belief Assignment (BBA) to combine [9]–[12], or both. In many applications involving epistemic uncertainties usually modeled by belief functions, it is often necessary to approximate general (non-Bayesian) basic belief assignments (BBAs) to subjective probabilities (called Bayesian BBAs). This necessity occurs if one needs to embed the fusion result in a system based on the probabilistic framework and Bayesian inference (e.g. tracking systems), or if one wants to use classical decision theory to make a decision. From a fusion standpoint, two approaches are usually adopted: 1) one can approximate at first each BBA in subjective probabilities and use Bayes fusion rule to get the final Bayesian BBA, or 2) one can fuse all the BBAs with a fusion rule, typically Dempster-Shafer's, or PCR6 rules (which is very costly in computations), and convert the combined BBA in a subjective probability measure. The former method is the simplest method but it generates a high loss of information included in original BBAs, whereas the latter direct method is intractable for high dimension problems. This paper presents a new method to achieve this task based on hierarchical decomposition (coarsening) of the frame of discernment, which can be seen as an intermediary approach between the two aforementioned methods.

This paper presents a new approach to fuse BBAs into a Bayesian BBA in order to reduce computational burden and keep the fusion tractable even for large dimension problems. This method is based on a hierarchical decomposition (coarsening) framework which allows to keep as much as possible information of original BBAs in preserving lower complexity. The main contributions of this paper are:

- 1) the presentation of the FoD bintree decomposition on which will be done the BBAs approximations;
- 2) the presentation of the fusion of approximate BBAs from bintree representation.

This hierarchical structure allows to encompass bintree decomposition and BBAs approximations on it to obtain the final approximate fusionned Bayesian BBA.

This paper is organized as follows. In section II, we recall some basics of DST and DSMT that are relevant to the new method presented in this paper. More details with examples can easily be found in [1], [5]. We will also briefly recall our preliminary works about hierarchical coarsening of FoD. Section III presents the novel hierarchical flexible (adaptive) coarsening method which can be regarded as the extension of our previous works. Two simple examples are given in section IV to illustrate the detailed calculation steps. Simulation experiments are presented in section V to show the rationality of this new approach. Finally, Sect.VI concludes the paper with future works perspectives.

## II. MATHEMATICAL BACKGROUND

This section provides a brief reminder of basics of DST and DSMT, and of original hierarchical coarsening method which are necessary for the presentation and the understanding of the more general flexible coarsening approximate method of section III.

### A. Basics of DST and DSMT

In DST framework, the frame of discernment<sup>2</sup>  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$  ( $n \geq 2$ ) is a set of exhaustive and exclusive elements (hypotheses) which represent the possible solutions of the problem under consideration and thus Shafer's model assumes  $\theta_i \cap \theta_j = \emptyset$  for  $i \neq j$  in  $\{1, \dots, n\}$ . A basic belief assignment (BBA)  $m(\cdot)$  is defined by the mapping:  $2^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . In DSMT, one can abandon Shafer's model (if Shafer's model doesn't fit with the problem) and refute the principle of the third excluded middle<sup>3</sup>. Instead of defining the BBAs on the power set  $2^\Theta \triangleq (\Theta, \cup)$  of the FoD, the BBAs are defined on the so-called *hyper-power set* (or Dedekind's lattice) denoted  $D^\Theta \triangleq (\Theta, \cup, \cap)$  whose cardinalities follows Dedekind's numbers sequence, see [5], Vol.1 for details and examples. A (generalized) BBA, called a mass function,  $m(\cdot)$  is defined by the mapping:  $D^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in D^\Theta} m(A) = 1$ . DSMT framework encompasses DST framework because  $2^\Theta \subset D^\Theta$ . In DSMT we can take into account also a set of *integrity constraints* on the FoD (if known), by specifying all the pairs of elements which are really disjoint. Stated otherwise, Shafer's model is a specific DSMT model where all elements are known to be disjoint.  $A \in D^\Theta$  is called a focal element of  $m(\cdot)$  if  $m(A) > 0$ . A BBA is called a Bayesian BBA if all of its focal elements are singletons and Shafer's model is assumed, otherwise it is called non-Bayesian [1]. A full ignorance source is represented by the vacuous BBA  $m_v(\Theta) = 1$ . The belief (or credibility) and plausibility functions are respectively defined by  $Bel(X) \triangleq \sum_{Y \in D^\Theta | Y \subseteq X} m(Y)$  and  $Pl(X) \triangleq \sum_{Y \in D^\Theta | Y \cap X \neq \emptyset} m(Y)$ .  $BI(X) \triangleq [Bel(X), Pl(X)]$  is called the belief interval of  $X$ . Its length  $U(X) \triangleq Pl(X) - Bel(X)$  measures the degree of uncertainty of  $X$ .

<sup>2</sup>We use the symbol  $\triangleq$  to mean equals by definition.

<sup>3</sup>The third excluded middle principle assumes the existence of the complement for any elements/propositions belonging to the power set  $2^\Theta$ .

In 1976, Shafer did propose Dempster's rule<sup>4</sup> to combine BBAs in DST framework. DS rule is defined by  $m_{DS}(\emptyset) = 0$  and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ ,

$$m_{DS}(A) = \frac{\sum_{B, C \in 2^\Theta | B \cap C = A} m_1(B)m_2(C)}{1 - \sum_{B, C \in 2^\Theta | B \cap C = \emptyset} m_1(B)m_2(C)} \quad (1)$$

DS rule formula is commutative and associative and can be easily extended to the fusion of  $S > 2$  BBAs. Unfortunately, DS rule has been highly disputed during the last decades by many authors because of its counter-intuitive behavior in high or even low conflict situations, and that is why many rules of combination have been proposed in the literature to combine BBAs [13]. To palliate DS rule drawbacks, the very interesting PCR6 (Proportional Conflict redistribution rule #6) has been proposed in DSMT and it is usually adopted<sup>5</sup> in recent applications of DSMT. The fusion of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  by the PCR6 rule is obtained by  $m_{PCR6}(\emptyset) = 0$  and  $\forall A \in D^\Theta \setminus \{\emptyset\}$

$$m_{PCR6}(A) = m_{12}(A) + \sum_{B \in D^\Theta \setminus \{A\} | A \cap B = \emptyset} \left[ \frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right] \quad (2)$$

where  $m_{12}(A) = \sum_{B, C \in D^\Theta | B \cap C = A} m_1(B)m_2(C)$  is the conjunctive operator, and each element  $A$  and  $B$  are expressed in their disjunctive normal form. If the denominator involved in the fraction is zero, then this fraction is discarded. The general PCR6 formula for combining more than two BBAs altogether is given in [5], Vol. 3. We adopt the generic notation  $m_{12}^{PCR6}(\cdot) = PCR6(m_1(\cdot), m_2(\cdot))$  to denote the fusion of  $m_1(\cdot)$  and  $m_2(\cdot)$  by PCR6 rule. PCR6 is not associative and PCR6 rule can also be applied in DST framework (with Shafer's model of FoD) by replacing  $D^\Theta$  by  $2^\Theta$  in Eq. (2).

### B. Hierarchical coarsening for fusion of Bayesian BBAs

Here, we briefly recall the principle of hierarchical coarsening of FoD to reduce the computational complexity of PCR6 combination of original Bayesian BBAs. The fusion of original non-Bayesian BBAs will be presented in the next section.

This principle was called *rigid grouping* in our previous works [17]–[19]. The goal of this coarsening is to replace the original (refined) Frame of Discernment (FoD)  $\Theta$  by a set of coarsened ones to make the computation of PCR6 rule tractable. Because we consider here only Bayesian BBA to combine, their focal elements are only singletons of the FoD  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$ , with  $n \geq 2$ , and we assume Shafer's model of the FoD  $\Theta$ .

A coarsening of the FoD  $\Theta$  means to replace it with another FoD less specific of smaller dimension  $\Omega = \{\omega_1, \dots, \omega_k\}$  with  $k < n$  from the elements of  $\Theta$ . This can be done in many ways depending the problem under consideration. Generally, the elements of  $\Omega$  are singletons of  $\Theta$ , and disjunctions of

<sup>4</sup>We use DS index to refer to Dempster-Shafer's rule (DS rule) because Shafer did really promote Dempster's rule in his milestone book [1].

<sup>5</sup>PCR6 rule coincides with PCR5 when combining only two BBAs [5].

elements of  $\Theta$ . For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , then the possible coarsened frames built from  $\Theta$  could be, for instance,  $\Omega = \{\omega_1 = \theta_1, \omega_2 = \theta_2, \omega_3 = \theta_3 \cup \theta_4\}$ , or  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2, \omega_2 = \theta_3 \cup \theta_4\}$ , etc. When dealing with Bayesian BBAs, the projection<sup>6</sup>  $m^\Omega(\cdot)$  of the original BBA  $m^\Theta(\cdot)$  is simply obtained by taking

$$m^\Omega(\omega_i) = \sum_{\theta_j \subseteq \omega_i} m^\Theta(\theta_j) \quad (3)$$

The hierarchical coarsening process (or rigid grouping) is a simple dichotomous approach of coarsening obtained as follows:

- If  $n = |\Theta|$  is an even number:

The disjunction of the  $n/2$  first elements  $\theta_1$  to  $\theta_{n/2}$  of  $\Theta$  define the element  $\omega_1$  of  $\Omega$ , and the last  $n/2$  elements  $\theta_{n/2+1}$  to  $\theta_n$  of  $\Theta$  define the element  $\omega_2$  of  $\Omega$ , that is

$$\Omega \triangleq \{\omega_1 = \theta_1 \cup \dots \cup \theta_{n/2}, \omega_2 = \theta_{n/2+1} \cup \dots \cup \theta_n\}$$

and based on (3), one has

$$m^\Omega(\omega_1) = \sum_{j=1, \dots, n/2} m^\Theta(\theta_j) \quad (4)$$

$$m^\Omega(\omega_2) = \sum_{j=n/2+1, \dots, n} m^\Theta(\theta_j) \quad (5)$$

For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , and one considers the Bayesian BBA  $m^\Theta(\theta_1) = 0.1$ ,  $m^\Theta(\theta_2) = 0.2$ ,  $m^\Theta(\theta_3) = 0.3$  and  $m^\Theta(\theta_4) = 0.4$ , then  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2, \omega_2 = \theta_3 \cup \theta_4\}$  and  $m^\Omega(\omega_1) = 0.1 + 0.2 = 0.3$  and  $m^\Omega(\omega_2) = 0.3 + 0.4 = 0.7$ .

- If  $n = |\Theta|$  is an odd number:

In this case, the element  $\omega_1$  of the coarsened frame  $\Omega$  is the disjunction of the  $[n/2 + 1]$ <sup>7</sup> first elements of  $\Theta$ , and the element  $\omega_2$  is the disjunction of other elements of  $\Theta$ . That is

$$\Omega \triangleq \{\omega_1 = \theta_1 \cup \dots \cup \theta_{[n/2+1]}, \omega_2 = \theta_{[n/2+1]+1} \cup \dots \cup \theta_n\}$$

and based on (3), one has

$$m^\Omega(\omega_1) = \sum_{j=1, \dots, [n/2+1]} m^\Theta(\theta_j) \quad (6)$$

$$m^\Omega(\omega_2) = \sum_{j=[n/2+1]+1, \dots, n} m^\Theta(\theta_j) \quad (7)$$

For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and one considers the Bayesian BBA  $m^\Theta(\theta_1) = 0.1$ ,  $m^\Theta(\theta_2) = 0.2$ ,  $m^\Theta(\theta_3) = 0.3$ ,  $m^\Theta(\theta_4) = 0.3$  and  $m^\Theta(\theta_5) = 0.1$ , then  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2 \cup \theta_3, \omega_2 = \theta_4 \cup \theta_5\}$  and  $m^\Omega(\omega_1) = 0.1 + 0.2 + 0.3 = 0.6$  and  $m^\Omega(\omega_2) = 0.3 + 0.1 = 0.4$ .

Of course, the same coarsening applies to all original BBAs  $m_s^\Theta(\cdot)$ ,  $s = 1, \dots, S$  of the  $S > 1$  sources of evidence to work with less specific BBAs  $m_s^\Omega(\cdot)$ ,  $s = 1, \dots, S$ . The less specific

<sup>6</sup>For clarity and convenience, we put explicitly as upper index the FoD of which the belief mass refers.

<sup>7</sup>The notation  $[x]$  means the integer part of  $x$ .

BBAs (called coarsened BBAs by abuse of language) can then be combined with PCR6 rule of combination according to formula (2). This dichotomous coarsening method is repeated iteratively  $l$  times as schematically represented by a bintree<sup>8</sup>. The last step of this hierarchical process is to calculate the combined (Bayesian) BBA of all focal elements according to the connection weights of the bintree structure, where the number of iterations (or layers)  $l$  of the tree depends on the cardinality  $|\Theta|$  of the original FoD  $\Theta$ . Specifically, the assignment of each focal element is updated according to the connection weights of link paths from root to terminal nodes. This principle is illustrated in details in the following example.

**Example 1:** Let's consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and the following three Bayesian BBAs

Focal elem.	$m_1^\Theta(\cdot)$	$m_2^\Theta(\cdot)$	$m_3^\Theta(\cdot)$
$\theta_1$	0.1	0.4	0
$\theta_2$	0.2	0	0.1
$\theta_3$	0.3	0.1	0.5
$\theta_4$	0.3	0.1	0.4
$\theta_5$	0.1	0.4	0

The hierarchical coarsening and fusion of BBAs is obtained from the following steps:

**Step 1:** We define the bintree structure based on iterative half split of FoD as shown in Fig. 1.

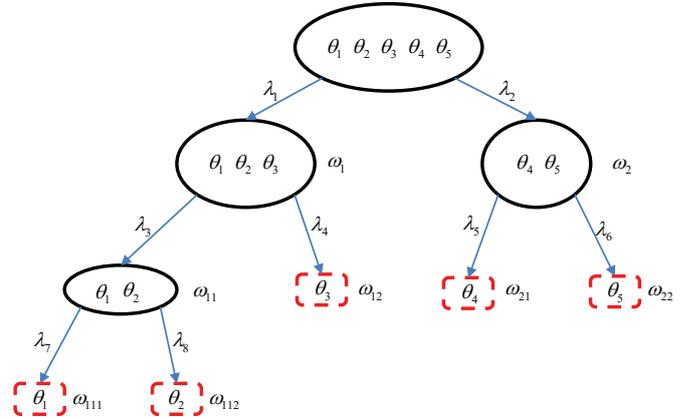


Figure 1: Fusion of Bayesian BBAs using bintree coarsening for Example 1.

The connecting weights are denoted as  $\lambda_1, \dots, \lambda_8$ . The elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$ :  $\Omega_1 = \{\omega_1 \triangleq \theta_1 \cup \theta_2 \cup \theta_3, \omega_2 \triangleq \theta_4 \cup \theta_5\}$
- At layer  $l = 2$ :

$$\Omega_2 = \{\omega_{11} \triangleq \theta_1 \cup \theta_2, \omega_{12} \triangleq \theta_3, \omega_{21} \triangleq \theta_4, \omega_{22} \triangleq \theta_5\}$$

- At layer  $l = 3$ :  $\Omega_3 = \{\omega_{111} \triangleq \theta_1, \omega_{112} \triangleq \theta_2\}$

<sup>8</sup>Here we consider bintree only for simplicity, which means that the coarsened frame  $\Omega$  consists of two elements only. Of course a similar method can be used with tri-tree, quad-tree, etc.

**Step 2:** The BBAs of elements of the (sub-)frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use (6)-(7) because  $|\Theta| = 5$  is an odd number. Therefore, we get

Focal elem.	$m_1^{\Omega_1}(\cdot)$	$m_2^{\Omega_1}(\cdot)$	$m_3^{\Omega_1}(\cdot)$
$\omega_1 \triangleq \theta_1 \cup \theta_2 \cup \theta_3$	0.6	0.5	0.6
$\omega_2 \triangleq \theta_4 \cup \theta_5$	0.4	0.5	0.4

- At layer  $l = 2$ : We work with the two subframes  $\Omega_{21} \triangleq \{\omega_{11}, \omega_{12}\}$  and  $\Omega_{22} \triangleq \{\omega_{21}, \omega_{22}\}$  of  $\Omega_2$  with the BBAs:

Focal elem.	$m_1^{\Omega_{21}}(\cdot)$	$m_2^{\Omega_{21}}(\cdot)$	$m_3^{\Omega_{21}}(\cdot)$
$\omega_{11} \triangleq \theta_1 \cup \theta_2$	$\frac{1}{2}$	$\frac{4}{5}$	$\frac{1}{6}$
$\omega_{12} \triangleq \theta_3$	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{5}{6}$
Focal elem.	$m_1^{\Omega_{22}}(\cdot)$	$m_2^{\Omega_{22}}(\cdot)$	$m_3^{\Omega_{22}}(\cdot)$
$\omega_{21} \triangleq \theta_4$	$\frac{3}{4}$	$\frac{1}{5}$	1
$\omega_{22} \triangleq \theta_5$	$\frac{1}{4}$	$\frac{4}{5}$	0

These mass values are obtained by the proportional redistribution of the mass of each focal element with respect to the mass of its parent focal element in the bin tree. For example, the value  $m_2^{\Omega_{21}}(\omega_{11}) = 4/5$  is derived by taking

$$m_2^{\Omega_{21}}(\omega_{11}) = \frac{m_2^\Theta(\theta_1) + m_2^\Theta(\theta_2)}{m_2^\Theta(\theta_1) + m_2^\Theta(\theta_2) + m_2^\Theta(\theta_3)} = \frac{0.4}{0.5} = \frac{4}{5}$$

Other mass values are computed similarly using this proportional redistribution method.

- At layer  $l = 3$ : We use again the proportional redistribution method which gives us

Focal elem.	$m_1^{\Omega_3}(\cdot)$	$m_2^{\Omega_3}(\cdot)$	$m_3^{\Omega_3}(\cdot)$
$\omega_{111} \triangleq \theta_1$	$\frac{1}{3}$	1	0
$\omega_{112} \triangleq \theta_2$	$\frac{2}{3}$	0	1

**Step 3:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening elements. In each layer  $l$ , we fuse sequentially<sup>9</sup> the three BBAs using PCR6 formula (2). More precisely, we compute at first  $m_{12}^{PCR6, \Omega_l}(\cdot) = PCR6(m_1^{\Omega_l}(\cdot), m_2^{\Omega_l}(\cdot))$  and then  $m_{(12)3}^{PCR6, \Omega_l}(\cdot) = PCR6(m_{12}^{PCR6, \Omega_l}(\cdot), m_3^{\Omega_l}(\cdot))$ . Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :

$$\lambda_1 = m_{(12)3}^{PCR6, \Omega_1}(\omega_1) = 0.6297$$

$$\lambda_2 = m_{(12)3}^{PCR6, \Omega_1}(\omega_2) = 0.3703$$

- At layer  $l = 2$ :

$$\lambda_3 = m_{(12)3}^{PCR6, \Omega_{21}}(\omega_{11}) = 0.4137$$

$$\lambda_4 = m_{(12)3}^{PCR6, \Omega_{21}}(\omega_{12}) = 0.5863$$

$$\lambda_5 = m_{(12)3}^{PCR6, \Omega_{22}}(\omega_{21}) = 0.8121$$

$$\lambda_6 = m_{(12)3}^{PCR6, \Omega_{22}}(\omega_{22}) = 0.1879$$

<sup>9</sup>Because PCR6 fusion is not associative, we should apply the general PCR6 formula to get best results. Here we use sequential fusion to reduce the computational complexity even if the fusion result is approximate.

- At layer  $l = 3$ :

$$\lambda_7 = m_{(12)3}^{PCR6, \Omega_3}(\omega_{111}) = 0.3103$$

$$\lambda_8 = m_{(12)3}^{PCR6, \Omega_3}(\omega_{112}) = 0.6897$$

**Step 4:** The final assignment of belief mass to the elements of original FoD  $\Theta$  are calculated using the product of the connection weights of link paths from root (top) node to terminal nodes (leaves). We finally get the following *resulting combined and normalized* Bayesian BBA

$$m^\Theta(\theta_1) = \lambda_1 \cdot \lambda_3 \cdot \lambda_7 = 0.6297 \cdot 0.4137 \cdot 0.3103 = 0.0808$$

$$m^\Theta(\theta_2) = \lambda_1 \cdot \lambda_3 \cdot \lambda_8 = 0.6297 \cdot 0.4137 \cdot 0.6897 = 0.1797$$

$$m^\Theta(\theta_3) = \lambda_1 \cdot \lambda_4 = 0.6297 \cdot 0.5863 = 0.3692$$

$$m^\Theta(\theta_4) = \lambda_2 \cdot \lambda_5 = 0.3703 \cdot 0.8121 = 0.3007$$

$$m^\Theta(\theta_5) = \lambda_2 \cdot \lambda_6 = 0.3703 \cdot 0.1879 = 0.0696$$

### III. NEW HIERARCHICAL FLEXIBLE COARSENING METHOD

Contrary to the (rigid) hierarchical coarsening method presented in section II, in our new flexible coarsening approach the elements  $\theta_i$ ,  $i = 1, \dots, n$  in FoD  $\Theta$  will not be half split to build coarsening focal elements  $\omega_j$ ,  $j = 1, \dots, k$  of the FoD  $\Omega_l$ . In the hierarchical flexible (adaptive) coarsening method, the elements  $\theta_i$  chosen to belong to the same group are determined using the consensus information drawn from the BBAs provided by the sources. Specifically, the degrees of disagreement between the provided sources on decisions  $(\theta_1, \theta_2, \dots, \theta_n)$  are first calculated using the belief-interval based distance  $d_{BI}$  [16], [20] to obtain *disagreement vector*. Then, the k-means algorithm is applied for clustering elements  $\theta_i$ ,  $i = 1, \dots, n$  based on the corresponding value in consensus vector. It is worth noting that values of disagreement reflect the preferences of independent sources of evidence for the same focal element. If they are small, it means that all sources have a consistent opinion and these elements should be clustered in the same group. Conversely, if disagreement values are large, it means that the sources have strong disagreement on these focal elements, and these focal elements need to be clustered in another group.

#### A. Calculating the disagreement vector

Let us consider several BBAs  $m_s^\Theta(\cdot)$ , ( $s = 1, \dots, S$ ) defined on same FoD  $\Theta$  of cardinality  $|\Theta| = n$ . The specific BBAs  $m_{\theta_i}(\cdot)$ ,  $i = 1, \dots, n$  entirely focused on  $\theta_i$  are defined by  $m_{\theta_i}(\theta_i) = 1$ , and for  $X \neq \theta_i$   $m_{\theta_i}(X) = 0$ . The disagreement of opinions of two sources about  $\theta_i$  is defined as the  $L_1$ -distance between the  $d_{BI}$  distances of the BBAs  $m_s^\Theta(\cdot)$ ,  $s = 1, 2$  to  $m_{\theta_i}(\cdot)$ , which is expressed by

$$D_{12}(\theta_i) \triangleq |d_{BI}(m_1^\Theta(\cdot), m_{\theta_i}(\cdot)) - d_{BI}(m_2^\Theta(\cdot), m_{\theta_i}(\cdot))| \quad (8)$$

The disagreement of opinions of  $S \geq 3$  sources about  $\theta_i$ , is defined as

$$D_{1-S}(\theta_i) \triangleq \frac{1}{2} \sum_{i=1}^S \sum_{j=1}^S |d_{BI}(m_i^\Theta(\cdot), m_{\theta_i}(\cdot)) - d_{BI}(m_j^\Theta(\cdot), m_{\theta_i}(\cdot))| \quad (9)$$

where  $d_{BI}$  distance is defined by<sup>10</sup> [20]

$$d_{BI}^E(m_1, m_2) \triangleq \sqrt{n_c \cdot \sum_{i=1}^{2^n-1} [d^I(BI_1(\theta_i), BI_2(\theta_i))]^2} \quad (10)$$

Here,  $n_c = 1/2^{n-1}$  is the normalization constant and  $d^I([a, b], [c, d])$  is the Wasserstein's distance defined by  $d^I([a, b], [c, d]) = \sqrt{[\frac{a+b}{2} - \frac{c+d}{2}]^2 + \frac{1}{3}[\frac{b-a}{2} - \frac{d-c}{2}]^2}$ . And  $BI(\theta_i) = [Bel(\theta_i), Pl(\theta_i)]$ .

The disagreement vector  $\mathbf{D}_{1-S}$  is defined by

$$\mathbf{D}_{1-S} \triangleq [D_{1-S}(\theta_1), \dots, D_{1-S}(\theta_n)] \quad (11)$$

### B. Clustering focal elements

Once  $\mathbf{D}_{1-S}$  is derived, a clustering algorithm is used to coarsen focal elements according to their corresponding values in  $\mathbf{D}_{1-S}$ . In this paper, we have used the k-means algorithm<sup>11</sup> to cluster focal elements. For each source  $s = 1, \dots, S$ , the mass assignments of focal elements in two<sup>12</sup> different clusters are added up according to formulas (12)–(13).

$$m_s^\Omega(\omega_1) = \sum_{\theta_i \in \omega_1} m^\Theta(\theta_i) \quad (12)$$

$$m_s^\Omega(\omega_2) = \sum_{\theta_j \in \omega_2} m^\Theta(\theta_j) \quad (13)$$

### C. Combination of the BBAs

Based on the disagreement vector and k-means algorithm, a new adaptive bintree structure based on this flexible coarsening decomposition is obtained (see example in the next section) and the elements in FoD  $\Theta$  are grouped more reasonably in each layer of the decomposition. Once the adaptive bintree structure is derived, other steps (multiplications of link weights) can be implemented which are identical to hierarchical (rigid) coarsening method presented in section II to get the final combined Bayesian BBA.

### D. Summary of the method

The fusion method of BBAs to get a combined Bayesian BBA based on hierarchical flexible decomposition of the FoD consists of the four steps below illustrated in Fig. 2.

- **Step 1 (pre-processing):** At first, all input BBAs to combine are approximated to Bayesian BBAs with DSMT transform.
- **Step 2 (disagreement vector):**  $\mathbf{D}_{1-S}(\cdot)$  is calculated using  $d_{BI}$  distances to estimate the degree of disagreement of BBAs  $m_1^\Theta, \dots, m_S^\Theta$  on potential decisions  $\theta_1, \dots, \theta_n$ .
- **Step 3 (adaptive bintree):** The adaptive bintree decomposition of the FoD  $\Theta$  is obtained using k-Means algorithm to get elements of subframes  $\Omega_l$ .
- **Step 4 (assignments and connection weights):** For each source  $m_s^\Theta(\cdot)$  to combine, the mass assignment of

<sup>10</sup>For simplicity, we assume Shafer's model so that  $|2^\Theta| = 2^n$ , otherwise the number of elements in the summation of (10) should be  $|D^\Theta| - 1$  with another normalization constant  $n_c$ .

<sup>11</sup>which is implemented in Matlab<sup>TM</sup>

<sup>12</sup>because we use here the bisection decomposition.

each element of subframe  $\Omega_l$  is computed by (12)–(13). The weight of links between two layers of the bintree decomposition are obtained with PCR6 rule<sup>13</sup>.

- **Step 5 (fusion):** The final result (combined Bayesian BBA) is computed by the product of weights of link paths from root to terminal nodes.

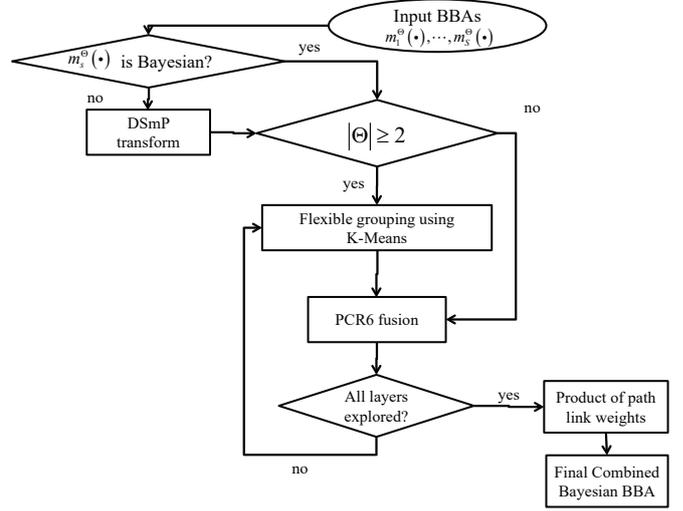


Figure 2: Hierarchical flexible decomposition of FoD for fusion.

## IV. TWO SIMPLE EXAMPLES

### A. Example 1 (fusion of Bayesian BBAs)

Let us revisit example 1 presented in section II-B. It can be verified in applying formula (9) that the disagreement vector  $\mathbf{D}_{1-3}$  for this example is equal to

$$\mathbf{D}_{1-3} = [0.4085, 0.2156, 0.3753, 0.2507, 0.4086]$$

The derivation of  $D_{1-3}(\theta_1)$  is given below for convenience.

$$\begin{aligned} D_{1-3}(\theta_1) &= |d_{BI}(m_1^\Theta(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_2^\Theta(\cdot), m_{\theta_1}(\theta_1))| \\ &\quad + |d_{BI}(m_2^\Theta(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_3^\Theta(\cdot), m_{\theta_1}(\theta_1))| \\ &\quad + |d_{BI}(m_1^\Theta(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_3^\Theta(\cdot), m_{\theta_1}(\theta_1))| \\ &= 0.4085. \end{aligned}$$

Based on the disagreement vector and k-means algorithm, a new adaptive bintree structure is obtained and shown in Fig. 3. Compared to Fig. 1, the elements in FoD  $\Theta$  are grouped more reasonably. In vector  $\mathbf{D}_{1-3}$ ,  $\theta_1$  and  $\theta_5$  lie in similar degree of disagreement so that they are put in the same group. Similarly for  $\theta_2$  and  $\theta_4$ . However, element  $\theta_3$  seems *weird*, which is put alone at the beginning of flexible coarsening. Once this adaptive bintree decomposition is obtained, other steps can be implemented which are identical to hierarchical coarsening method of section II to get the final combined BBA.

The flexible coarsening and fusion of BBAs is obtained from the following steps:

<sup>13</sup>general formula preferred, or applied sequentially to reduce complexity.

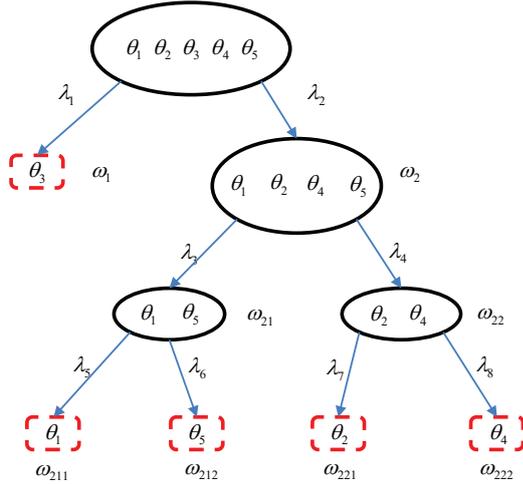


Figure 3: Example 1: Flexible bintree decomposition of FoD.

**Step 1:** According to Fig.3, the elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$ :  $\Omega_1 = \{\omega_1 \triangleq \theta_3, \omega_2 \triangleq \theta_1 \cup \theta_2 \cup \theta_4 \cup \theta_5\}$
- At layer  $l = 2$ :  $\Omega_2 = \{\omega_{21} \triangleq \theta_1 \cup \theta_5, \omega_{22} \triangleq \theta_2 \cup \theta_4\}$
- At layer  $l = 3$ :  $\Omega_3 = \{\omega_{211} \triangleq \theta_1, \omega_{212} \triangleq \theta_5, \omega_{221} \triangleq \theta_2, \omega_{222} \triangleq \theta_4\}$

**Step 2:** The BBAs of elements of the (sub-)frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use (12)-(13) and we get

Focal elem.	$m_1^{\Omega_1}(\cdot)$	$m_2^{\Omega_1}(\cdot)$	$m_3^{\Omega_1}(\cdot)$
$\omega_1 \triangleq \theta_3$	0.3	0.1	0.5
$\omega_2 \triangleq \theta_1 \cup \theta_2 \cup \theta_4 \cup \theta_5$	0.7	0.9	0.5

- At layer  $l = 2$ : We use again the proportional redistribution method which gives us:

Focal elem.	$m_1^{\Omega_2}(\cdot)$	$m_2^{\Omega_2}(\cdot)$	$m_3^{\Omega_2}(\cdot)$
$\omega_{21} \triangleq \theta_1 \cup \theta_5$	$\frac{3}{7}$	$\frac{4}{9}$	$\frac{1}{5}$
$\omega_{22} \triangleq \theta_2 \cup \theta_4$	$\frac{4}{7}$	$\frac{5}{9}$	$\frac{4}{5}$

- At layer  $l = 3$ : We work with the two subframes  $\Omega_{31} \triangleq \{\omega_{211}, \omega_{212}\}$  and  $\Omega_{32} \triangleq \{\omega_{221}, \omega_{222}\}$  of  $\Omega_3$  with the BBAs

Focal elem.	$m_1^{\Omega_{31}}(\cdot)$	$m_2^{\Omega_{31}}(\cdot)$	$m_3^{\Omega_{31}}(\cdot)$
$\omega_{211} \triangleq \theta_1$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\omega_{212} \triangleq \theta_5$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Focal elem.	$m_1^{\Omega_{32}}(\cdot)$	$m_2^{\Omega_{32}}(\cdot)$	$m_3^{\Omega_{32}}(\cdot)$
$\omega_{221} \triangleq \theta_2$	$\frac{2}{3}$	0	$\frac{1}{3}$
$\omega_{222} \triangleq \theta_4$	$\frac{3}{5}$	1	$\frac{4}{5}$

**Step 3:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening elements. Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :

$$\lambda_1 = 0.2226; \quad \lambda_2 = 0.7774.$$

- At layer  $l = 2$ :

$$\lambda_3 = 0.2200; \quad \lambda_4 = 0.7800.$$

- At layer  $l = 3$ :

$$\lambda_5 = 0.5; \quad \lambda_6 = 0.5; \quad \lambda_7 = 0.0669; \quad \lambda_8 = 0.9331.$$

**Step 4:** We finally get the following *resulting combined and normalized Bayesian BBA*

$$m^\Theta(\cdot) = \{0.0855, 0.0406, 0.2226, 0.5658, 0.0855\}.$$

**B. Example 2 (with non-Bayesian BBAs)**

**Example 1bis:** Let's consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and the following BBAs given by

Focal elem.	$m_1^\Theta(\cdot)$	$m_2^\Theta(\cdot)$	$m_3^\Theta(\cdot)$
$\theta_1$	0.1	0.4	0
$\theta_2$	0.2	0	0
$\theta_3$	0.3	0.05	0
$\theta_4$	0.03	0.05	0
$\theta_5$	0.1	0.04	0
$\theta_1 \cup \theta_2$	0.1	0.04	0
$\theta_2 \cup \theta_3 \cup \theta_5$	0	0.02	0.1
$\theta_3 \cup \theta_4$	0.02	0.1	0.2
$\theta_1 \cup \theta_5$	0.1	0.3	0.2
$\Theta$	0.05	0	0.5

**Step 1 (Pre-Processing):** All these three BBAs are transformed into Bayesian BBAs with DSMP transform and the generated BBAs are illustrated as

Focal elem.	$m_1^\Theta(\cdot)$	$m_2^\Theta(\cdot)$	$m_3^\Theta(\cdot)$
$\theta_1$	0.1908	0.7127	0.2000
$\theta_2$	0.2804	0	0.1334
$\theta_3$	0.3387	0.1111	0.2333
$\theta_4$	0.0339	0.1	0.2000
$\theta_5$	0.1562	0.0761	0.2333

It can be verified in applying formula (9) that the disagreement vector  $\mathbf{D}_{1-3}$  for this example is equal to

$$\mathbf{D}_{1-3} = [0.5385, 0.3632, 0.3453, 0.2305, 0.2827]$$

**Step 2:** According to the clustering algorithm, the elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$ :  $\Omega_1 = \{\omega_1 \triangleq \theta_1, \omega_2 \triangleq \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5\}$
- At layer  $l = 2$ :  $\Omega_2 = \{\omega_{21} \triangleq \theta_2 \cup \theta_3, \omega_{22} \triangleq \theta_4 \cup \theta_5\}$
- At layer  $l = 3$ :  $\Omega_3 = \{\omega_{211} \triangleq \theta_2, \omega_{212} \triangleq \theta_3, \omega_{221} \triangleq \theta_4, \omega_{222} \triangleq \theta_5\}$

**Step 3:** The BBAs of elements of the (sub-)frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use (12)-(13) and we get

Focal elem.	$m_1^{\Omega_1}(\cdot)$	$m_2^{\Omega_1}(\cdot)$	$m_3^{\Omega_1}(\cdot)$
$\omega_1 \triangleq \theta_1$	0.1908	0.7127	0.2000
$\omega_2 \triangleq \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5$	0.8092	0.2873	0.8000

- At layer  $l = 2$ : We use again the proportional redistribution method which gives us:

Focal elem.	$m_1^{\Omega_2}(\cdot)$	$m_2^{\Omega_2}(\cdot)$	$m_3^{\Omega_2}(\cdot)$
$\omega_{21} \triangleq \theta_2 \cup \theta_3$	0.7651	0.3867	0.4584
$\omega_{22} \triangleq \theta_4 \cup \theta_5$	0.2349	0.6133	0.5416

- At layer  $l = 3$ : We work with the two subframes  $\Omega_{31} \triangleq \{\omega_{211}, \omega_{212}\}$  and  $\Omega_{32} \triangleq \{\omega_{221}, \omega_{222}\}$  of  $\Omega_3$  with the BBAs:

Focal elem.	$m_1^{\Omega_{31}}(\cdot)$	$m_2^{\Omega_{31}}(\cdot)$	$m_3^{\Omega_{31}}(\cdot)$
$\omega_{211} \triangleq \theta_2$	0.4529	0	0.3638
$\omega_{212} \triangleq \theta_3$	0.5471	1	0.6362
Focal elem.	$m_1^{\Omega_{32}}(\cdot)$	$m_2^{\Omega_{32}}(\cdot)$	$m_3^{\Omega_{32}}(\cdot)$
$\omega_{221} \triangleq \theta_4$	0.1783	0.5679	0.4616
$\omega_{222} \triangleq \theta_5$	0.8217	0.4321	0.5384

**Step 4:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening elements. Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :

$$\lambda_1 = 0.2345; \quad \lambda_2 = 0.7655.$$

- At layer  $l = 2$ :

$$\lambda_3 = 0.5533; \quad \lambda_4 = 0.4467.$$

- At layer  $l = 3$ :

$$\lambda_5 = 0.1606; \quad \lambda_6 = 0.8394;$$

$$\lambda_7 = 0.3349; \quad \lambda_8 = 0.6651.$$

**Step 5:** We finally get the following *resulting combined and normalized Bayesian BBA*

$$m^\ominus(\cdot) = \{0.2345, 0.0681, 0.3555, 0.1145, 0.2274\}.$$

## V. SIMULATION RESULTS AND PERFORMANCES

### A. Flexible Grouping of Singletons

1) *Similarity*:<sup>14</sup> Assuming that  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}\}$  and first, we randomly generate 2 BBAs, denoted as  $m_1^\ominus(\cdot)$  and  $m_2^\ominus(\cdot)$ , which can be seen in Table I.

Table I: BBAs for Two Sources  $m_1^\ominus(\cdot)$  and  $m_2^\ominus(\cdot)$

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$
$m_1^\ominus(\cdot)$	0.1331	0.0766	0.0175	0.0448	0.0229
$m_2^\ominus(\cdot)$	0.1020	0.0497	0.1094	0.0612	0.0612
	$\theta_6$	$\theta_7$	$\theta_8$	$\theta_9$	$\theta_{10}$
$m_1^\ominus(\cdot)$	0.1142	0.0023	0.2254	0.1583	3.4959e-04
$m_2^\ominus(\cdot)$	0.0069	0.0070	0.0128	0.0833	0.0338
	$\theta_{11}$	$\theta_{12}$	$\theta_{13}$	$\theta_{14}$	$\theta_{15}$
$m_1^\ominus(\cdot)$	0.0075	0.0514	0.1121	0.0314	0.0021
$m_2^\ominus(\cdot)$	0.1180	0.1202	0.1351	0.0686	0.0309

In order to fully verify the similarity between hierarchical flexible coarsening method and PCR6 in DSMT, a new strict

<sup>14</sup>Similarity represents the approximate degree between fusion results using flexible coarsening and PCR6.

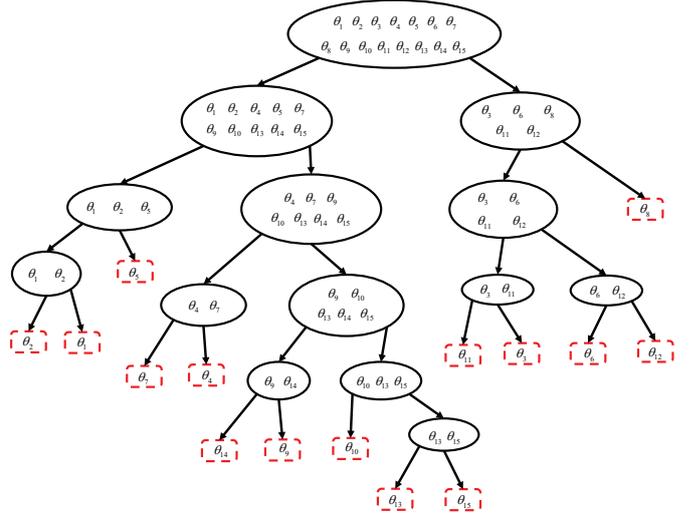


Figure 4: Structure of Hierarchical Flexible Coarsening.

distance metric between two BBAs, denoted  $d_{BI}^E$ , was recently proposed in [20], [16] and it will be used in this paper.

In this paper, we regard  $d_{BI}^E$  as one criteria for evaluating the degree of similarity between the fusion results obtained from flexible coarsening and PCR6.

Based on (8) and (10), the disagreement vector  $\mathbf{D}(\cdot)$  is obtained:

$$\mathbf{D}(\cdot) = (0.0032, 0.0020, 0.0290, 0.0092, 0.0147, 0.0228, \\ 0.0059, 0.0537, 0.0154, 0.0131, 0.0338, 0.0235, \\ 0.0118, 0.0145, 0.0120).$$

Thus, bintree structure of hierarchical flexible coarsening is illustrated in Fig. 4 and the similarity between fusion results of hierarchical flexible coarsening and PCR6 is 0.9783. And the similarity between hierarchical coarsening method and PCR6 is 0.9120. In particular, terminal nodes (the red small box in Fig. 4) of flexible grouping are not in accordance with the original order  $\theta_1, \theta_2, \dots, \theta_{15}$ . This is quite different compared to original hierarchical coarsening method.

From the point of view of statistics, 100 BBAs are randomly generated to be fused with three methods: hierarchical flexible coarsening, hierarchical coarsening and also PCR6. Comparisons are made in Fig. 5, which show the superiority of our new approach proposed in this paper (Average value of new method is 97% and the old method is 93.5%).

### B. Flexible Grouping of Conflicting Focal Elements

Assuming that there are five sources of evidence  $m_1^\ominus(\cdot), m_2^\ominus(\cdot), m_3^\ominus(\cdot), m_4^\ominus(\cdot), m_5^\ominus(\cdot)$ , and the restricted hype-power set  $D^\ominus = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_1 \cap \theta_2, \theta_5 \cap \theta_6 \cap \theta_7, \theta_1 \cap \theta_5 \cap \theta_9 \cap \theta_{10}\}$ . And then we randomly generate 1000 BBAs for each source to calculate the similarity using (10). From Fig. 6, we can find that hierarchical flexible coarsening method can also maintain high degree of similarity which performs better than hierarchical coarsening.

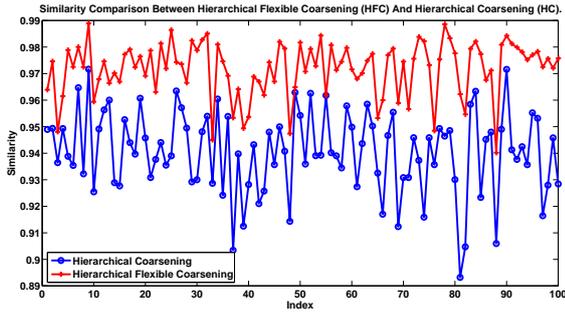


Figure 5: Comparisons Between HFC and HC (Only Singletons).

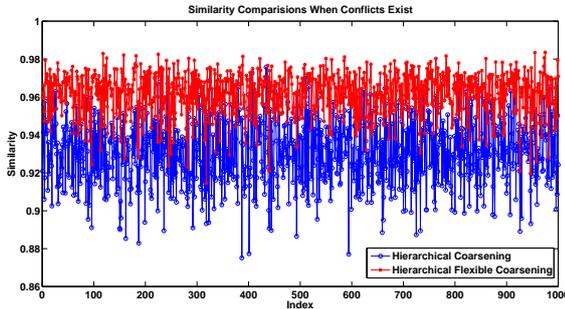


Figure 6: Comparisons Between HFC and HC (Singletons and Conflicting Focal Elements).

C. Flexible Grouping of Uncertain and Hybrid Focal Elements

We can also deal with uncertain and hybrid focal elements. Assuming that there are also five sources of evidence  $m_1^\ominus(\cdot), m_2^\ominus(\cdot), m_3^\ominus(\cdot), m_4^\ominus(\cdot), m_5^\ominus(\cdot)$  and  $D_1^\ominus = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_1 \cup \theta_2, \theta_5 \cup \theta_6 \cup \theta_7, \theta_1 \cup \theta_5 \cup \theta_9 \cup \theta_{10}\}$ ;  $D_2^\ominus = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_2 \cap \theta_4 \cup \theta_6, \theta_1 \cup \theta_3 \cap \theta_5 \cup \theta_7 \cap \theta_9\}$ <sup>15</sup>. And then we respectively and randomly generate 1000 BBAs for these two cases  $D_1^\ominus$  and  $D_2^\ominus$ . Finally, we calculate the average similarity degree of HFC and HC with PCR6 in Table II, which illustrates HFC performs better than old method. However, there exist the extra time cost of HFC compared to HC due to the clustering steps in coarsening process.

Table II: Similarity Comparisons

	Hierarchical Flexible Coarsening	Hierarchical Coarsening
$D_1^\ominus$	98%	91%
$D_2^\ominus$	97%	93%

VI. CONCLUSION AND PERSPECTIVES

A novel hierarchical flexible approximate method in DSMT is proposed here. Compared to original hierarchical coarsen-

<sup>15</sup>In this case,  $D_1^\ominus$  represents uncertain focal elements and  $D_2^\ominus$  represents hybrid focal elements.

ing, flexible strategy guarantees higher similarity with PCR6 rules in fusion process. Besides, whether focal elements in hyper power set are singletons, conflicting focal elements, uncertain or even hybrid focal elements, the new method works well. In the future work, we will focus on the general framework of hierarchical coarsening, which could generate final non-Bayesian BBAs in order to avoid loss of information. Furthermore, other advantages or disadvantages of our proposed methods such as computational efficiency and time consumption need to be further investigated.

ACKNOWLEDGMENT

This work was supported by NNSF of China (No.61175091), Qing Lan Project of Jiangsu Province, Aeronautical Science Foundation of China (20140169002), and Six Major Top-talent Plan of Jiangsu Province.

REFERENCES

- [1] G. Shafer, A mathematical theory of evidence, Princeton Univ. Press, 1976.
- [2] A. Dempster, Upper and lower probabilities induced by a multivalued mapping, in Annals of Mathematical Statistics, Vol. 38, pp. 325–339, 1967.
- [3] P. Smets, Practical uses of belief functions, in K.B. Laskey and H. Prade Editors, 15th Conf. on Uncertainty in Artificial Intelligence, pp. 612–621, Stockholm, Sweden, 1999.
- [4] J. Dezert, Foundations for a new theory of plausible and paradoxical reasoning, in Information & Security: An Int. Journal, Vol. 9, 2002.
- [5] F. Smarandache, J. Dezert (Editors), Advances and applications of DSMT for information fusion, American Research Press, Rehoboth, NM, U.S.A., Vol. 1–4, 2004–2015. Available at webpage 2 of [6].
- [6] <http://www.onera.fr/staff/jean-dezert>
- [7] R. Kennes, Computational aspects of the Möbius transform of graphs, in IEEE Trans. on SMC, Vol. 22, pp. 201–223, 1992.
- [8] G. Shafer, R. Logan, Implementing Dempster’s rule for hierarchical evidence, in Artificial Intelligence, Vol. 33, pp. 271–298, 1987.
- [9] Y. Yang, Y.L. Liu, Iterative approximation of basic belief assignment based on distance of evidence, in Plos One, Vol. 11, No. 2, 2016.
- [10] T. Denœux, Inter and outer approximation of belief structures using a hierarchical clustering approach, in Int. J. of Uncertainty, Fuzziness and Knowledge-Based System, Vol. 9, No 4, pp. 437–460, 2001.
- [11] Y. Yang, D. Han, C. Han, F. Cao, A novel approximation of basic probability assignment based on rank-level fusion, in Chinese Journal of Aeronautics, Vol. 26, No. 4, pp. 993–999, 2013.
- [12] D. Han, Y. Yang, J. Dezert, Two novel methods of BBA approximation based on focal element redundancy, in Proc. of Fusion 2015, Washington, D.C., USA, July 2015.
- [13] P. Smets, Analyzing the combination of conflicting belief functions, in Information Fusion, Vol. 8, pp. 387–412, 2006.
- [14] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, in Int. J. of Approx. reasoning, Vol. 38, 2005.
- [15] J. Dezert, F. Smarandache, A new probabilistic transformation of belief mass assignment, in Proc. of Fusion 2008.
- [16] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, Decision-making with belief interval distance, in Proc. of Belief 2016 Int. Conf., Prague, CZ, 2016.
- [17] X.D. Li, J. Dezert, X.H. Huang, Z.D. Meng, X.J. Wu, A fast approximate reasoning method in hierarchical DSMT (A), in Acta Electronica Sinica, Vol. 38, No. 11, pp. 2567–2572, 2010.
- [18] X.D. Li, W.D. Yang, X.J. W, J. Dezert, A fast approximate reasoning method in hierarchical DSMT (B), in Acta Electronica Sinica, Vol. 39, No. 3A, pp. 32–36, 2011.
- [19] X.D. Li, W.D. Yang, J. Dezert, A fast approximate reasoning method in hierarchical DSMT (C), in J. Huazhong Univ. of Sci. and Tech. (Natural Science Edition), Vol. 39, pp. 151–156, 2011.
- [20] D. Han, J., Dezert, Y., Yang, New distance measures of evidence based on belief intervals, in Proc. of Belief 2014, Oxford, UK, 2014.

# A Fast Combination Method in DSMT And Its Application to Recommender System

Yilin Dong<sup>a</sup>, Xinde Li<sup>a</sup>, Yihai Liu<sup>b</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE, Ministry of Education,  
School of Automation, Southeast University, Nanjing, Jiangsu Province, China.

<sup>b</sup>Jiangsu Automation Research Institute, Lianyungang, Jiangsu Province, China.

Emails: dyl@seu.edu.cn, xindeli@seu.edu.cn, liuyihai@126.com

Originally published as: Y. Dong, X. Li, Y. Liu, *A Fast Combination Method in DSMT and its Application to Recommender System*, PLoS ONE, Vol. 13(1): e0189703, 2018, and reprinted with permission.

**Abstract**—In many applications involving epistemic uncertainties usually modeled by belief functions, it is often necessary to approximate general (non-Bayesian) basic belief assignments (BBAs) to subjective probabilities (called Bayesian BBAs). This necessity occurs if one needs to embed the fusion result in a system based on the probabilistic framework and Bayesian inference (e.g. tracking systems), or if one needs to make a decision in the decision making problems. In this paper, we present a new fast combination method, called modified rigid coarsening (MRC), to obtain the final Bayesian BBAs based on hierarchical decomposition (coarsening) of the frame of discernment. Regarding this method, focal elements with probabilities are coarsened efficiently to reduce computational complexity in the process of combination by using disagreement vector and a simple dichotomous approach. In order to prove the practicality of our approach, this new approach is applied to combine users' soft preferences in recommender systems (RSs). Additionally, in order to make a comprehensive performance comparison, the proportional conflict redistribution rule #6 (PCR6) is regarded as a baseline in a range of experiments. According to the results of experiments, MRC is more effective in accuracy of recommendations compared to original Rigid Coarsening (RC) method and comparable in computational time.

**Keywords:** Recommender system, DSMT, PCR6.

## I. INTRODUCTION

The theory of belief functions, known as Dempster-Shafer Theory (DST) was developed by Shafer [1] in 1976 from Dempster's works [2]. Belief functions allow one to model epistemic uncertainty [3] and they have been already used in many applications since the 1990's [4], mainly those relevant to expert systems, decision-making support and information fusion. To palliate some limitations (such as high computational complexity) of DST, Dezert and Smarandache proposed an extended mathematical framework of belief functions with new efficient quantitative and qualitative rules of combinations, which was called DSMT (Dezert and Smarandache Theory) in literature [5], [6] with applications listed in [7]. One of the major drawbacks of DST and DSMT is their high computational complexities, on condition that the fusion space (i.e. frame of discernment – FoD) and the number of sources to combine are large. DSMT is more complex than DST, and the Proportional Conflict Redistribution rule #6 (PCR6 rule) becomes computationally intractable in the worst case as soon

as the cardinality of the Frame of Discernment (FoD) is greater than six.

To reduce the computational cost of operations with belief functions when the number of focal elements is very large, several approaches have been proposed by different authors. Basically, the existing approaches rely either on efficient implementations of computations as proposed for instance in [8], [9], or on approximation techniques of original Basic Belief Assignment (BBA) to combine [10]–[14], or both. From a fusion standpoint, two approaches are usually adopted: 1) one can approximate at first each BBA in subjective probabilities and use Bayes fusion rule to get the final Bayesian BBA [11], [12], or 2) one can fuse all the BBAs with a fusion rule, typically Dempster-Shafer's, or proportional conflict redistribution rule #6 (PCR6) rules (which is very costly in computations), and convert the combined BBA in a subjective probability measure [10], [14]. The former method is the simplest method but it generates a high loss of information included in the original BBAs, whereas the latter method is intractable for high dimension issues.

This paper presents a new combination method, called modified rigid coarsening (MRC), to get the final Bayesian BBAs based on hierarchical decomposition (coarsening) of the frame of discernment, which can be seen as an intermediary approach between the two aforementioned methods. This hierarchical structure allows to encompass bintree decomposition and mass of coarsening FoD on it. To prove the practicality of our proposed method, MRC is applied to combine users' preferences so as to provide the suitable recommendation for RSs. This paper is an extended version of our preliminary work on original rigid coarsening (RC) published in [15]. In this paper, more detailed analyses of this new combination method are provided. More importantly, this innovative method is also applied into the real application. These are all added values (contributions) of this paper.

The main contributions of this paper are:

- 1) the presentation of the FoD bintree decomposition on which will be done the BBAs approximations;
- 2) user preferences in Recommender Systems (RSs) are modeled by DSMT-Modeling Function.

In order to measure the efficiency and effectiveness of the MRC, it is integrated in the RSs based on DSMT and compared to traditional methods in the experiments. The results show that regarding the accuracy of recommendations, MRC is extremely close to classical PCR6; and the computational time of MRC can be obviously superior to that of PCR6.

The remainder of this paper is organized as follows. In section II, we review relevant prior work on DST and DSMT first. In section III, MRC is presented. In section IV, a recommendation system based on DSMT, that employs MRC to combine users' preferences, is shown. In section V, we evaluate our proposed algorithm based on two public datasets: Movielens and Flixster. Finally, we conclude and discuss future work.

## II. MATHEMATICAL BACKGROUND

This section provides a brief reminder of the basics of DST and DSMT, which is necessary for the presentation and understanding of the more general MRC of Section III.

In DST framework, the frame of discernment<sup>1</sup>  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$  ( $n \geq 2$ ) is a set of exhaustive and exclusive elements (hypotheses) which represents the possible solutions of the problem under consideration and thus Shafer's model assumes  $\theta_i \cap \theta_j = \emptyset$  for  $i \neq j$  in  $\{1, \dots, n\}$ . A basic belief assignment (BBA)  $m(\cdot)$  is defined by the mapping:  $2^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . In DSMT, one can abandon Shafer's model (if Shafer's model doesn't fit with the problem) and refute the principle of the third excluded middle. The third excluded middle principle assumes the existence of the complement for any elements/propositions belonging to the power set  $2^\Theta$ . Instead of defining the BBAs on the power set  $2^\Theta \triangleq (\Theta, \cup)$  of the FoD, the BBAs are defined on the so-called *hyper-power set* (or Dedekind's lattice) denoted  $D^\Theta \triangleq (\Theta, \cup, \cap)$  whose cardinalities follows Dedekind's numbers sequence, see [6], Vol.1 for details and examples. A (generalized) BBA, called a mass function,  $m(\cdot)$  is defined by the mapping:  $D^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in D^\Theta} m(A) = 1$ . The DSMT framework encompasses DST framework because  $2^\Theta \subset D^\Theta$ . In DSMT, we can take into account also a set of *integrity constraints* on the FoD (if known), by specifying all the pairs of elements which are really disjoint. Stated otherwise, Shafer's model is a specific DSMT model where all elements are deemed to be disjoint.  $A \in D^\Theta$  is called a focal element of  $m(\cdot)$  if  $m(A) > 0$ . A BBA is called a Bayesian BBA if all of its focal elements are singletons and Shafer's model is assumed, otherwise it is called non-Bayesian [1]. A full ignorance source is represented by the vacuous BBA  $m_v(\Theta) = 1$ . The belief (or credibility) and plausibility functions are respectively defined by  $Bel(X) \triangleq \sum_{Y \in D^\Theta | Y \subseteq X} m(Y)$  and  $Pl(X) \triangleq \sum_{Y \in D^\Theta | Y \cap X = \emptyset} m(Y)$ .  $BI(X) \triangleq [Bel(X), Pl(X)]$  is called the belief interval of  $X$ . Its length  $U(X) \triangleq Pl(X) - Bel(X)$  measures the degree of uncertainty of  $X$ .

<sup>1</sup>Here, we use the symbol  $\triangleq$  to mean equals by definition.

In 1976, Shafer did propose Dempster's rule and we use DS index to refer to Dempster-Shafer's rule (DS rule) because Shafer did really promote Dempster's rule in his milestone book [1]) to combine BBAs in DST framework. DS rule is defined by  $m_{DS}(\emptyset) = 0$  and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ ,

$$m_{DS}(A) = \frac{\sum_{B,C \in 2^\Theta | B \cap C = A} m_1(B)m_2(C)}{1 - \sum_{B,C \in 2^\Theta | B \cap C = \emptyset} m_1(B)m_2(C)}. \quad (1)$$

The DS rule formula is commutative and associative and can be easily extended to the fusion of  $S > 2$  BBAs. Unfortunately, DS rule has been highly disputed during the last decades by many authors because of its counter-intuitive behavior in high or even low conflict situations, and that is why many rules of combination were proposed in literature to combine BBAs [16]. To palliate DS rule drawbacks, the very interesting PCR6 was proposed in DSMT and it is usually adopted (PCR6 rule coincides with PCR5 when combining only two BBAs [6]) in recent applications of DSMT. The fusion of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  by the PCR6 rule is obtained by  $m_{PCR6}(\emptyset) = 0$  and  $\forall A \in D^\Theta \setminus \{\emptyset\}$

$$m_{PCR6}(A) = m_{12}(A) + \sum_{B \in D^\Theta \setminus \{A\} | A \cap B = \emptyset} \left[ \frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right]. \quad (2)$$

where  $m_{12}(A) = \sum_{B,C \in D^\Theta | B \cap C = A} m_1(B)m_2(C)$  is the conjunctive operator, and each element A and B are expressed in their disjunctive normal form. If the denominator involved in the fraction is zero, then this fraction is discarded. The general PCR6 formula for combining more than two BBAs altogether is given in [6], Vol. 3. We adopt the generic notation  $m_{12}^{PCR6}(\cdot) = PCR6(m_1(\cdot), m_2(\cdot))$  to denote the fusion of  $m_1(\cdot)$  and  $m_2(\cdot)$  by PCR6 rule. PCR6 is not associative and PCR6 rule can also be applied in DST framework (with Shafer's model of FoD) by replacing  $D^\Theta$  by  $2^\Theta$  in (2).

## III. MODIFIED RIGID COARSENING FOR FUSION OF BAYESIAN BBAS

Here, we introduce the principle of MRC of FoD to reduce the computational complexity of PCR6 combination of original Bayesian BBAs. Considering the case of non-Bayesian BBAs, it requires decoupling all non-singletons in these BBAs in advance. The fusion of original nonBayesian BBAs needs to be decoupled by using DSMT in advance, which will be explained in Section IV.

### A. Rigid coarsening

This proposal was initially called *rigid coarsening (RC)* in our previous works [17]–[19] and currently improved in our recent work [15]. The goal of this coarsening is to replace the original (refined) FoD  $\Theta$  by a set of coarsened ones to make computation of the PCR6 rule tractable. Because we consider here only Bayesian BBA to combine, their focal

elements are only singletons of the FoD  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$ , with  $n \geq 2$ , and we assume Shafer's model of the FoD  $\Theta$ . A coarsening of the FoD  $\Theta$  means to replace it with another FoD less specific of smaller dimension  $\Omega = \{\omega_1, \dots, \omega_k\}$  with  $k < n$  from the elements of  $\Theta$ . This can be done in many ways depending the problem under consideration. Generally, the elements of  $\Omega$  are singletons of  $\Theta$ , and disjunctions of elements of  $\Theta$ . For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , then a possible coarsened frame built from  $\Theta$  could be, for instance,  $\Omega = \{\omega_1 = \theta_1, \omega_2 = \theta_2, \omega_3 = \theta_3 \cup \theta_4\}$ , or  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2, \omega_2 = \theta_3 \cup \theta_4\}$ , etc.

**Definition 1:** When dealing with Bayesian BBAs, the projection<sup>2</sup>  $m^\Omega(\cdot)$  of the original BBA  $m^\Theta(\cdot)$  is simply obtained by taking

$$m^\Omega(\omega_i) = \sum_{\theta_j \in \omega_i} m^\Theta(\theta_j). \quad (3)$$

The rigid coarsening process is a simple dichotomous approach of coarsening obtained as follows:

- If  $n = |\Theta|$  is an even number:

The disjunction of the  $n/2$  first elements  $\theta_1$  to  $\theta_{\frac{n}{2}}$  of  $\Theta$  define the element  $\omega_1$  of  $\Omega$ , and the last  $n/2$  elements  $\theta_{\frac{n}{2}+1}$  to  $\theta_n$  of  $\Theta$  define the element  $\omega_2$  of  $\Omega$ , that is

$$\Omega \triangleq \{\omega_1 = \theta_1 \cup \dots \cup \theta_{\frac{n}{2}}, \omega_2 = \theta_{\frac{n}{2}+1} \cup \dots \cup \theta_n\}$$

and based on (3), one has

$$m^\Omega(\omega_1) = \sum_{j=1, \dots, \frac{n}{2}} m^\Theta(\theta_j), \quad (4)$$

$$m^\Omega(\omega_2) = \sum_{j=\frac{n}{2}+1, \dots, n} m^\Theta(\theta_j). \quad (5)$$

For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , and one considers the Bayesian BBA  $m^\Theta(\theta_1) = 0.1$ ,  $m^\Theta(\theta_2) = 0.2$ ,  $m^\Theta(\theta_3) = 0.3$  and  $m^\Theta(\theta_4) = 0.4$ , then  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2, \omega_2 = \theta_3 \cup \theta_4\}$  and  $m^\Omega(\omega_1) = 0.1 + 0.2 = 0.3$  and  $m^\Omega(\omega_2) = 0.3 + 0.4 = 0.7$ .

- If  $n = |\Theta|$  is an odd number:

In this case, the element  $\omega_1$  of the coarsened frame  $\Omega$  is the disjunction of the<sup>3</sup>  $[n/2 + 1]$  first elements of  $\Theta$ , and the element  $\omega_2$  is the disjunction of other elements of  $\Theta$ . That is

$$\Omega \triangleq \{\omega_1 = \theta_1 \cup \dots \cup \theta_{[\frac{n}{2}+1]}, \omega_2 = \theta_{[\frac{n}{2}+1]+1} \cup \dots \cup \theta_n\}$$

and based on (3), one has

$$m^\Omega(\omega_1) = \sum_{j=1, \dots, [\frac{n}{2}+1]} m^\Theta(\theta_j), \quad (6)$$

<sup>2</sup>For clarity and convenience, we put explicitly as upper index the FoD for which the belief mass refers.

<sup>3</sup>The notation  $[x]$  means the integer part of  $x$  first elements of  $\Theta$

$$m^\Omega(\omega_2) = \sum_{j=[\frac{n}{2}+1]+1, \dots, n} m^\Theta(\theta_j). \quad (7)$$

For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and one considers the Bayesian BBA  $m^\Theta(\theta_1) = 0.1$ ,  $m^\Theta(\theta_2) = 0.2$ ,  $m^\Theta(\theta_3) = 0.3$ ,  $m^\Theta(\theta_4) = 0.3$  and  $m^\Theta(\theta_5) = 0.1$ , then  $\Omega = \{\omega_1 = \theta_1 \cup \theta_2 \cup \theta_3, \omega_2 = \theta_4 \cup \theta_5\}$  and  $m^\Omega(\omega_1) = 0.1 + 0.2 + 0.3 = 0.6$  and  $m^\Omega(\omega_2) = 0.3 + 0.1 = 0.4$ .

Of course, the same coarsening strategy applies to all original BBAs  $m_s^\Theta = (\cdot)$ ,  $s = 1, \dots, S$  of the  $S > 1$  sources of evidence to work with less specific BBAs  $m_s^\Omega = (\cdot)$ ,  $s = 1, \dots, S$ . The less specific BBAs (called coarsened BBAs by abuse of language) can then be combined with the PCR6 rule of combination according to formula (2). This dichotomous coarsening method is repeated iteratively  $l$  times as schematically represented by a bintree. Here, we consider bintree only for simplicity, which means that the coarsened frame  $\Omega$  consists of two elements only. Of course a similar method can be used with tri-tree, quad-tree, etc. The last step of this hierarchical process is to calculate the combined (Bayesian) BBA of all focal elements according to the connection weights of the bintree structure, where the number of layers  $l$  of the tree depends on the cardinality  $|\Theta|$  of the original FoD  $\Theta$ . Specifically, the mass of each focal element is updated depending on the connection weights of link paths from root to terminal nodes. This principle is illustrated in details in the following example.

**Example 1:** Let's consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , and the following three Bayesian BBAs can be seen in Table I:

Table I  
THREE BAYESIAN BBAs FOR EXAMPLE 1.

Focal elements	$m_1^\Theta(\cdot)$	$m_2^\Theta(\cdot)$	$m_3^\Theta(\cdot)$
$\theta_1$	0.1	0.4	0
$\theta_2$	0.2	0	0.1
$\theta_3$	0.3	0.1	0.5
$\theta_4$	0.3	0.1	0.4
$\theta_5$	0.1	0.4	0

The rigid coarsening and fusion of BBAs is deduced from the following steps:

**Step 1:** We define the bintree structure based on iterative half split of FoD as shown in Fig 1.

The connecting weights are denoted as  $\lambda_1, \dots, \lambda_8$ . The elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$  :  $\Omega_1 = \{\omega_1 \triangleq \theta_1 \cup \theta_2 \cup \theta_3, \omega_2 \triangleq \theta_4 \cup \theta_5\}$
- At layer  $l = 2$  :  $\Omega_2 = \{\omega_{11} \triangleq \theta_1 \cup \theta_2, \omega_{12} \triangleq \theta_3, \omega_{21} = \theta_4, \omega_{22} = \theta_5\}$
- At layer  $l = 3$  :  $\Omega_3 = \{\omega_{111} \triangleq \theta_1, \omega_{112} \triangleq \theta_2\}$

**Step 2:** The BBAs of elements of the (sub-) frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use Eqs (6) and (7) because  $|\Theta| = 5$  is an odd number. Therefore, we get the BBAs in Table II:
- At layer  $l = 2$ : We work with the two subframes  $\Omega_{21} \triangleq \{\omega_{11}, \omega_{12}\}$  and  $\Omega_{22} \triangleq \{\omega_{21}, \omega_{22}\}$  of  $\Omega_2$  with the

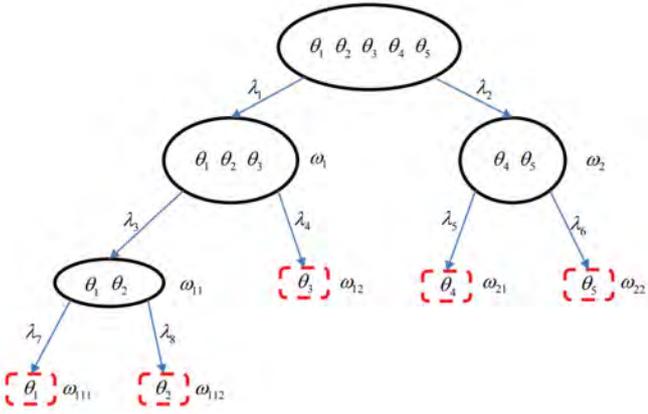


Figure 1. Fusion of Bayesian BBAs using bintree coarsening for Example 1.

Table II  
THE BBAs OF ELEMENTS OF THE SUB-FRAMES  $\Omega_1$  FOR EXAMPLE 1.

Focal elements	$m_1^\Theta(\cdot)$	$m_2^\Theta(\cdot)$	$m_3^\Theta(\cdot)$
$\omega_1 \triangleq \theta_1 \cup \theta_2 \cup \theta_3$	0.6	0.6	0.6
$\omega_2 \triangleq \theta_4 \cup \theta_5$	0.4	0.5	0.4

BBAs in Tables III and IV:

Table III  
THE BBAs OF ELEMENTS OF THE SUB-FRAMES  $\Omega_{21}$  FOR EXAMPLE 1.

Focal elements	$m_1^{\Omega_{21}}(\cdot)$	$m_2^{\Omega_{21}}(\cdot)$	$m_3^{\Omega_{21}}(\cdot)$
$\omega_{11} \triangleq \theta_1 \cup \theta_2$	$\frac{1}{2}$	$\frac{4}{5}$	$\frac{1}{6}$
$\omega_{12} \triangleq \theta_3$	$\frac{1}{2}$	$\frac{1}{5}$	$\frac{5}{6}$

Table IV  
THE BBAs OF ELEMENTS OF THE SUB-FRAMES  $\Omega_{22}$  FOR EXAMPLE 1.

Focal elements	$m_1^{\Omega_{22}}(\cdot)$	$m_2^{\Omega_{22}}(\cdot)$	$m_3^{\Omega_{22}}(\cdot)$
$\omega_{21} \triangleq \theta_4$	$\frac{3}{4}$	$\frac{1}{5}$	1
$\omega_{22} \triangleq \theta_5$	$\frac{1}{4}$	$\frac{4}{5}$	0

These mass values are obtained by the proportional redistribution of the mass of each focal element with respect to the mass of its parent focal element in the bin tree. For example,  $m_2^{\Omega_{21}}(\omega_{11}) = 4/5$  is derived by taking

$$m_2^{\Omega_{21}}(\omega_{11}) = \frac{m_2^\Theta(\theta_1) + m_2^\Theta(\theta_2)}{m_2^\Theta(\theta_1) + m_2^\Theta(\theta_2) + m_2^\Theta(\theta_3)} = \frac{0.4}{0.5} = \frac{4}{5}$$

Other masses of coarsening focal elements are computed similarly using this proportional redistribution method.

- At layer  $l = 3$ : We use again the proportional redistribution method which gives us the BBAs of the sub-frames  $\Omega_3$  in Table V:

Table V  
THE BBAs OF ELEMENTS OF THE SUB-FRAMES  $\Omega_3$  FOR EXAMPLE 1.

Focal elements	$m_1^{\Omega_3}(\cdot)$	$m_2^{\Omega_3}(\cdot)$	$m_3^{\Omega_3}(\cdot)$
$\omega_{111} \triangleq \theta_1$	$\frac{1}{3}$	1	0
$\omega_{112} \triangleq \theta_2$	$\frac{2}{3}$	0	1

**Step 3:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening. In each layer  $l$ , we fuse sequentially the three BBAs using PCR6 formula (2). Because PCR6 fusion is not associative, we should apply the general PCR6 formula to get best results. Here we use sequential fusion to reduce the computational complexity even if the fusion result is approximate. More precisely, we compute at first  $m_{12}^{PCR6, \Omega_l}(\cdot) = PCR6(m_1^{\Omega_l}(\cdot), m_2^{\Omega_l}(\cdot))$  and  $m_{(12)3}^{PCR6, \Omega_l}(\cdot) = PCR6(m_{12}^{\Omega_l}(\cdot), m_3^{\Omega_l}(\cdot))$ . Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :

$$\lambda_1 = m_{(12)3}^{PCR6, \Omega_1}(\omega_1) = 0.6297$$

$$\lambda_2 = m_{(12)3}^{PCR6, \Omega_1}(\omega_2) = 0.3703$$

- At layer  $l = 2$ :

$$\lambda_3 = m_{(12)3}^{PCR6, \Omega_{21}}(\omega_{11}) = 0.4137$$

$$\lambda_4 = m_{(12)3}^{PCR6, \Omega_{21}}(\omega_{12}) = 0.5863$$

$$\lambda_5 = m_{(12)3}^{PCR6, \Omega_{22}}(\omega_{21}) = 0.8121$$

$$\lambda_6 = m_{(12)3}^{PCR6, \Omega_{22}}(\omega_{22}) = 0.1879$$

- At layer  $l = 3$ :

$$\lambda_7 = m_{(12)3}^{PCR6, \Omega_3}(\omega_{111}) = 0.3103$$

$$\lambda_8 = m_{(12)3}^{PCR6, \Omega_3}(\omega_{112}) = 0.6897$$

**Step 4:** The final assignments of elements in original FoD  $\Theta$  are calculated using the product of the connection weights of link paths from root (top) node to terminal nodes (leaves). We eventually get the *combined and normalized* Bayesian BBA:

$$m^\Theta(\theta_1) = \lambda_1 \cdot \lambda_3 \cdot \lambda_7 = 0.6297 \cdot 0.4137 \cdot 0.3103 = 0.0808$$

$$m^\Theta(\theta_2) = \lambda_1 \cdot \lambda_3 \cdot \lambda_8 = 0.6297 \cdot 0.4137 \cdot 0.6897 = 0.1797$$

$$m^\Theta(\theta_3) = \lambda_1 \cdot \lambda_4 = 0.6297 \cdot 0.5863 = 0.3692$$

$$m^\Theta(\theta_4) = \lambda_2 \cdot \lambda_5 = 0.3703 \cdot 0.8121 = 0.3007$$

$$m^\Theta(\theta_5) = \lambda_2 \cdot \lambda_6 = 0.3703 \cdot 0.1879 = 0.0696$$

### B. Modified rigid coarsening

One of the issues with RC described in the previous section is that *no extra self-information of focal elements is embedded into the coarsening process*. In this paper, the elements  $\theta_i$  selected to belong to the same group are determined using the consensus information drawn from the BBAs provided by the sources. Specifically, the degrees of disagreement between the provided sources on decisions  $(\theta_1, \theta_2, \dots, \theta_n)$  are first calculated using the belief-interval based distance  $d_{BI}$  [20] to obtain *disagreement vector*. And then all focal elements in FoD are sorted in an ascending order. Finally, the simple dichotomous approach is utilized to hierarchical coarsen those **Re-sorted** focal elements.

**Calculating the disagreement vector.** Let us consider several BBAs  $m_s^\ominus(\cdot), (s = 1, \dots, S)$  defined on same FoD  $\ominus$  of cardinality  $|\ominus| = n$ . The specific BBAs  $m_{\theta_i}(\cdot), i = 1, \dots, n$  entirely focused on  $\theta_i$  are defined by  $m_{\theta_i}(\theta_i) = 1$ , and for  $X \neq \theta_i, m_{\theta_i}(X) = 0$ .

**Definition 2:** The disagreement of opinions of two sources about  $\theta_i$  is defined as the  $L_1$ -distance between the  $d_{BI}$  distances of the BBAs  $m_s^\ominus(\cdot), s = 1, 2$  to  $m_{\theta_i}(\cdot)$ , which is expressed by

$$D_{12}(\theta_i) \triangleq |d_{BI}(m_1^\ominus(\cdot), m_{\theta_i}(\cdot)) - d_{BI}(m_2^\ominus(\cdot), m_{\theta_i}(\cdot))|. \quad (8)$$

**Definition 3:** The disagreement of opinions of  $S \geq 3$  sources about  $\theta_i$ , is defined as

$$D_{1-S}(\theta_i) \triangleq \frac{1}{2} \sum_{i=1}^S \sum_{j=1}^S |d_{BI}(m_i^\ominus(\cdot), m_{\theta_i}(\cdot)) - d_{BI}(m_j^\ominus(\cdot), m_{\theta_i}(\cdot))|. \quad (9)$$

where  $d_{BI}$  distance is defined by [20]. For simplicity, we assume Shafer's model so that  $|2^\ominus| = 2^n$ , otherwise the number of elements in the summation of (10) should be  $|D^\ominus| - 1$  with another normalization constant  $n_c$ .

$$d_{BI}^E(m_1, m_2) \triangleq \sqrt{n_c \cdot \sum_{i=1}^{2^n-1} [d^I(BI_1(\theta_i), BI_2(\theta_i))]^2}. \quad (10)$$

Here,  $n_c = 1/2^{n-1}$  is the normalization constant and  $d^I([a, b], [c, d])$  is the Wasserstein's distance defined by  $d^I([a, b], [c, d]) = \sqrt{[\frac{a+b}{2} - \frac{c+d}{2}]^2 + \frac{1}{3}[\frac{b-a}{2} - \frac{d-c}{2}]^2}$ , and  $BI(\theta_i) = [Bel(\theta_i), Pl(\theta_i)]$ .

The disagreement vector  $D_{1-S}$  is defined by

$$D_{1-S} \triangleq [D_{1-S}(\theta_1), \dots, D_{1-S}(\theta_n)]. \quad (11)$$

**Modified rigid coarsening by using the disagreement vector.** Once  $D_{1-S}$  is derived, all focal elements  $\{\theta_1, \theta_2, \dots, \theta_n\}$  are sorted according to their corresponding values in  $D_{1-S}$ .

Let us revisit example 1 presented in the previous subsection. It can be verified in applying formula (9) that the disagreement vector  $D_{1-3}$  for this example is equal to

$$D_{1-3} = [0.4085, 0.2156, 0.3753, 0.2507, 0.4086]$$

The derivation of  $D_{1-3}(\theta_1)$  is given below for convenience:

$$\begin{aligned} D_{1-3} &= |d_{BI}(m_1^\ominus(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_2^\ominus(\cdot), m_{\theta_1}(\theta_1))| \\ &+ |d_{BI}(m_2^\ominus(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_3^\ominus(\cdot), m_{\theta_1}(\theta_1))| \\ &+ |d_{BI}(m_1^\ominus(\cdot), m_{\theta_1}(\theta_1)) - d_{BI}(m_3^\ominus(\cdot), m_{\theta_1}(\theta_1))| \\ &= 0.4085. \end{aligned}$$

Based on the disagreement vector, a new bintree structure is obtained and shown in Fig 2.

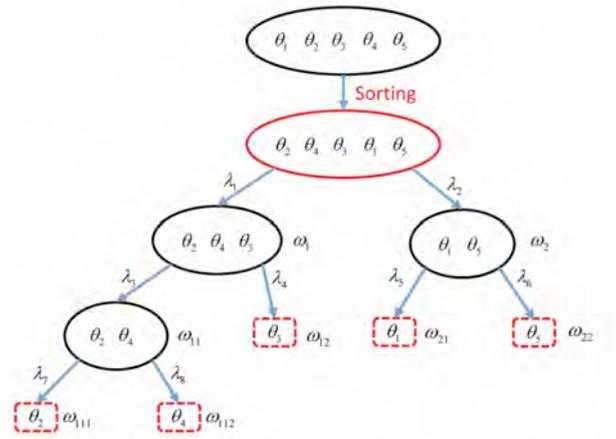


Figure 2. Fusion of Bayesian BBAs using MRC for Example 1.

Compared with Fig 1, the elements in FoD  $\ominus$  are grouped more reasonably. In vector  $D_{1-3}$ ,  $\theta_1$  and  $\theta_5$  lie in similar degree of disagreement so that they are put in the same group. Similarly for  $\theta_2$  and  $\theta_4$ . However, element  $\theta_3$  seems *weird*, which is put alone in the process of coarsening. Once this new bintree decomposition is obtained, other steps can be implemented which are identical to rigid coarsening in section to get the final combined BBA.

**Step 1:** According to Fig 2, the elements of the frames  $\Omega_l$  are defined as follows:

- At layer  $l = 1$ :

$$\Omega_1 = \{\omega_1 \triangleq \theta_2 \cup \theta_4 \cup \theta_3, \omega_2 \triangleq \theta_1 \cup \theta_5\}.$$

- At layer  $l = 2$ :

$$\Omega_2 = \{\omega_{11} \triangleq \theta_2 \cup \theta_4, \omega_{12} \triangleq \theta_3, \omega_{21} \triangleq \theta_1, \omega_{22} \triangleq \theta_5\}.$$

- At layer  $l = 3$ :

$$\Omega_3 = \{\omega_{111} \triangleq \theta_2, \omega_{112} \triangleq \theta_4\}.$$

**Step 2:** The BBAs of elements of the (sub-) frames  $\Omega_l$  are obtained as follows:

- At layer  $l = 1$ , we use (6) and (7) and we get Table VI

Table VI

 THE BBAS OF ELEMENTS OF THE SUB-FRAMES  $\Omega_1$  USING MRC FOR EXAMPLE 1.

Focal elements	$m_1^{\Omega_1}(\cdot)$	$m_2^{\Omega_1}(\cdot)$	$m_3^{\Omega_1}(\cdot)$
$\omega_1 \triangleq \theta_2 \cup \theta_4 \cup \theta_3$	0.8	0.2	1.0
$\omega_2 \triangleq \theta_1 \cup \theta_5$	0.2	0.8	0.0

- At layer  $l = 2$ , We use again the proportional redistribution method which gives us Tables VII and VIII. Here, masses of  $\omega_{21}, \omega_{22}$  in  $m_3^{\Omega_{22}}(\cdot)$  are not considered because the mass of their parent focal element ( $m_3^{\Omega_1}(\omega_2)$ ) in bintree is 0.

Table VII

 THE BBAS OF ELEMENTS OF THE SUB-FRAMES  $\Omega_{21}$  USING MRC FOR EXAMPLE 1.

Focal elements	$m_1^{\Omega_{21}}(\cdot)$	$m_2^{\Omega_{21}}(\cdot)$	$m_3^{\Omega_{21}}(\cdot)$
$\omega_{11} \triangleq \theta_2 \cup \theta_4$	$\frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{2}$
$\omega_{12} \triangleq \theta_3$	$\frac{3}{8}$	$\frac{1}{2}$	$\frac{1}{2}$

Table VIII

 THE BBAS OF ELEMENTS OF THE SUB-FRAMES  $\Omega_{22}$  USING MRC FOR EXAMPLE 1.

Focal elements	$m_1^{\Omega_{22}}(\cdot)$	$m_2^{\Omega_{22}}(\cdot)$	$m_3^{\Omega_{22}}(\cdot)$
$\omega_{21} \triangleq \theta_1$	$\frac{1}{2}$	$\frac{1}{2}$	-
$\omega_{22} \triangleq \theta_5$	$\frac{1}{2}$	$\frac{1}{2}$	-

- At layer  $l = 3$ , We work with the two subframes of  $\Omega_3$ , that is  $\Omega_3 \triangleq \{\omega_{111}, \omega_{112}\}$ , with the BBAs in Table IX.

Table IX

 THE BBAS OF ELEMENTS OF THE SUB-FRAMES  $\Omega_3$  USING MRC FOR EXAMPLE 1.

Focal elements	$m_1^{\Omega_3}(\cdot)$	$m_2^{\Omega_3}(\cdot)$	$m_3^{\Omega_3}(\cdot)$
$\omega_{111} \triangleq \theta_2$	$\frac{2}{5}$	0.0	$\frac{1}{5}$
$\omega_{112} \triangleq \theta_4$	$\frac{3}{5}$	1.0	$\frac{4}{5}$

**Step 3:** The connection weights  $\lambda_i$  are computed from the assignments of coarsening elements. Hence, we obtain the following connecting weights in the bintree:

- At layer  $l = 1$ :  $\lambda_1 = 0.8333$ , and  $\lambda_2 = 0.1667$ .
- At layer  $l = 2$ :  $\lambda_3 = 0.5697$ ,  $\lambda_4 = 0.4303$ ,  $\lambda_5 = 0.5000$ , and  $\lambda_6 = 0.5000$ .
- At layer  $l = 3$ :  $\lambda_7 = 0.0669$ , and  $\lambda_8 = 0.9331$ .

**Step 4:** We finally get the following *combined and normalized* Bayesian BBA

$$m^\Theta(\cdot) = \{0.0833, 0.0318, 0.3586, 0.4430, 0.0834\}.$$

### C. Summary of the proposed method

The fusion method of BBAs to get a combined Bayesian BBA based on hierarchical decomposition of the FoD consists of several steps (**Algorithm 1**) illustrated in Fig 3.

### Algorithm 1: Modified Rigid Coarsening Method

---

**Input :** All original BBAs  
 $m_1^\Theta(\cdot), \dots, m_s^\Theta(\cdot), s = 1, 2, \dots, s$

**Output:** The final combined BBA  $m^\Theta(\cdot)$

- 1 **if** Compound focal elements in  
 $\Theta : \theta_i \cup \theta_j \neq \emptyset$  or  $\theta_i \cap \theta_j \neq \emptyset$  **then**
- 2     Probabilistic transformation:  
 $DSmP(m_1^\Theta(\cdot)), DSmP(m_2^\Theta(\cdot)), \dots, DSmP(m_s^\Theta(\cdot))$
- 3 **end**
- 4 **for**  $i \leq n$  **do**
- 5     **for**  $s \leq S$  **do**
- 6         Calculate  $D_{1-S}(\theta_i) \triangleq$   
 $\frac{1}{2} \sum_{i=1}^S \sum_{j=1}^S |d_{BI}(m_i^\Theta(\cdot), m_{\theta_i}(\cdot)) - d_{BI}(m_j^\Theta(\cdot), m_{\theta_j}(\cdot))|$
- 7     **end**
- 8 **end**
- 9 **for**  $i \leq n$  **do**
- 10     Sorting  $D_{1-S}(\theta_i)$  in an ascending order.
- 11 **end**
- 12 **while**  $|\Theta| \geq 2$  **do**
- 13     **if**  $n$  is an even number **then**
- 14          $m^{\Omega_i}(\omega_1) = \sum_{j=1, \dots, \frac{n}{2}} m^\Theta(\theta_j)$ ;
- 15          $m^{\Omega_i}(\omega_2) = \sum_{j=\frac{n}{2}+1, \dots, n} m^\Theta(\theta_j)$ ;
- 16     **else**
- 17          $m^{\Omega_i}(\omega_1) = \sum_{j=1, \dots, \lfloor \frac{n}{2} \rfloor + 1} m^\Theta(\theta_j)$ ;
- 18          $m^{\Omega_i}(\omega_2) = \sum_{j=\lfloor \frac{n}{2} \rfloor + 1 + 1, \dots, n} m^\Theta(\theta_j)$ ;
- 19     **end**
- 20     Then connection weights  $\lambda$  is calculated:  
 $PCR6(m^{\Omega_i}(\omega_1), m^{\Omega_i}(\omega_2))$
- 21 **end**
- 22 **foreach** focal element  $\theta_i, i \in 1, \dots, n$  **do**
- 23      $m^\Theta(\theta_i)$  equals to the product of path link weights  
 from root to terminal nodes.
- 24 **end**

---

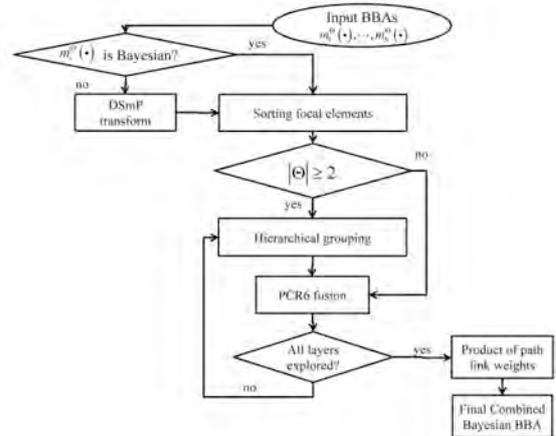


Figure 3. Modified rigid coarsening of FoD for fusion.

It is worth noting that when the given BBAs are not Bayesian, the first step is to use the existing Probabilistic Transformation (PT) to transform them to Bayesian BBAs. In order to use the proposed combination method in the RSs, modified rigid coarsening is mathematically denoted as  $\oplus$  in the following sections.

*D. Simulation considering accuracy and computational efficiency*

- Accuracy:

Assuming that the FoD is

$$\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}, \theta_{15}, \theta_{16}, \theta_{17}, \theta_{18}, \theta_{19}, \theta_{20}\}$$

then 1000 BBAs are randomly generated to be fused with three methods: modified rigid coarsening, rigid coarsening and also PCR6. And then distances of fusion results are computed using  $d_{BI}$  between two pairs: modified rigid coarsening and PCR6; rigid coarsening and PCR6. Comparisons are made in Fig 4, which show the superiority of our new approach proposed in this paper (The average value of the approximation of modified rigid coarsening is **97.5%** and original rigid coarsening is **94.5%**). Here, similarity represents the approximate degree between fusion results using hierarchical approximate method (both rigid and modified rigid coarsening) and PCR6.

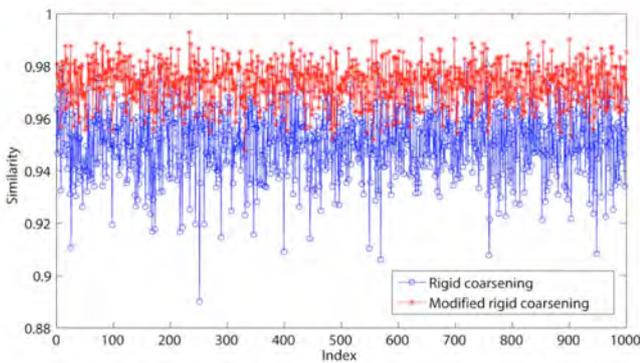


Figure 4. Accuracy comparisons between MRC and PCR6 (Only Singletons).

- Computational efficiency:

As we mentioned before, another advantage of the hierarchical combination method is the computational efficiency. Here, two experiments are conducted (*All experiments are implemented on a PC with I3 CPU, Integrated graphics chipsets and 4G DDR*): 1) the number of singletons is unchanged while the number of BBAs to be fused is increasing; 2) the number of BBAs is unchanged while the number of singletons in FoD is increasing. The results are illustrated in Fig 5 and 6. From experiment 1, all these three methods (classical PCR6, rigid coarsening and also modified rigid coarsening) calculate quickly (**less than 1.2s**) even the number of BBAs

increases from 100 to 1000. However, such situation deteriorates when the number of focal elements increases. In Fig 6, when the number of focal elements increases to 500, time consumption of three combinations is: PCR6: 20.6857s; modified rigid coarsening: 7.3320s; rigid coarsening: 5.9748s. This phenomenon also proves that it is reasonable to map original FoD to the coarsening FoD, with the aim of reducing the number of focal elements at the time of fusion. But in any case, computing efficiency of rigid coarsening or modified rigid coarsening is still better than PCR6. On the other hand, modified rigid coarsening makes a significant improvement (accuracy) at the expense of parts of the computational efficiency.

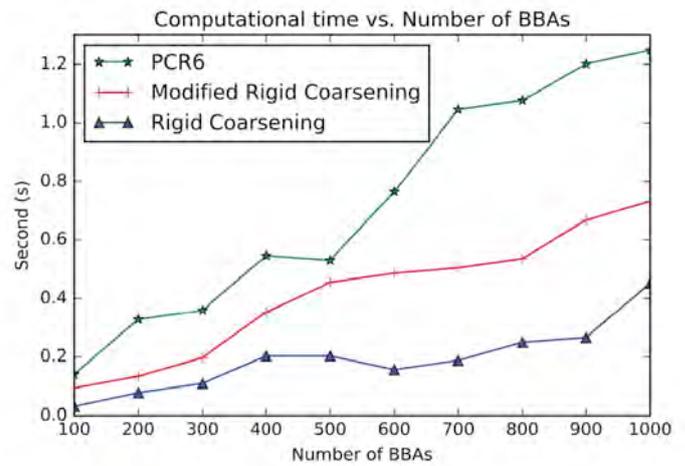


Figure 5. Efficiency comparisons between MRC, RC and PCR6 (With the number of BBAs increasing).

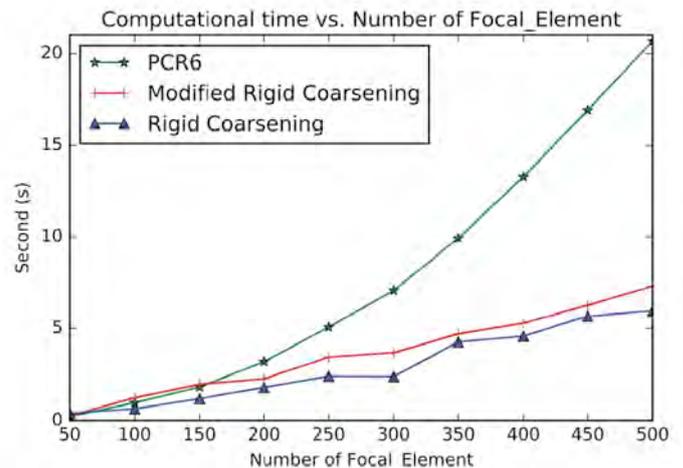


Figure 6. Efficiency comparisons between MRC, RC and PCR6 (With the number of focal elements increasing).

#### IV. A RECOMMENDER SYSTEM INTEGRATING WITH HIERARCHICAL COARSENING COMBINATION METHOD

In today's e-commerce, online providers often recommend proper goods or services to each consumer based on their personal opinions or preferences [21], [22]. However, it is a tough task to provide appropriate recommendation which may confront several difficulties. One difficulty is that users' preferences are usually characterized as uncertain, imprecise or incomplete [23], [24], which cannot be used directly in RSs. Besides, it is easy to understand that when the more information about user preferences are, the more accurate prediction of RSs will be [25], [26]. But, the problem is that which method we adopt to integrate multi-source uncertain information?

As a general framework for information fusion, DST can not only model uncertain information, but also provide an efficient way to combine multi-source information. These mentioned features make this theory a wide range of applications [27]–[29], especially in RSs [23], [25], [30]–[32]. According to DST, users' comments on products in RSs are described by using mass functions and rules of combination method are used frequently in order to provide appropriate recommendation.

As mentioned in previous sections, both the performances of combination rules in DST or in DSMT suffer from computational complex which is obviously ignored in [23], [25]. Thus, in this paper, modified rigid coarsening method is applicable to combine the imprecise users' preferences in RSs. First, we are required to introduce the relevant knowledge of RSs. Actually, almost all characteristics of RSs have been introduced in [23], [25], [30]–[32].

First, we give the corresponding representation of the mathematical notation in RSs based on DSMT. RSs usually contain two objects: *Users*, *Items*. A set of  $M$  users and a set containing  $N$  items is respectively denoted by  $\mathbf{U} = \{U_1, U_2, \dots, U_M\}$  and  $\mathbf{I} = \{I_1, I_2, \dots, I_N\}$ . Besides, we assume that users can give the corresponding ratings to the items, which include  $L$  rating levels ( $\Theta = \{\theta_1, \theta_2, \dots, \theta_L\}$ ). Here,  $L$  preference levels means multi-level evaluation results. For example, four-levels of user evaluation on the product are *Excellent*, *Good*, *Fair*, *Poor*.  $r_{i,k}$  means a rating of user  $U_i$  on item  $I_k$  and a rating matrix  $\mathbf{R} = r_{i,k}$  comprises all the ratings of users on items. It should be noted that  $r_{i,k}$  is originally modeled as a mass function  $m_{i,k} : D^\Theta \rightarrow [0, 1]$ . Additionally, let  $\mathbf{I}_i^R$  and  $\mathbf{U}_k^R$  denote the set of items rated by user  $U_i$  and the set of users having rated item  $I_k$ , respectively.

Contextual information can often be summarized into several genres that significantly affect user's rating of items. Normally, we represent contextual information by a set containing  $P$  genres, denoted by  $\mathbf{S} = \{S_1, S_2, \dots, S_P\}$ . And each genre  $S_p$ , with  $1 \leq p \leq P$  contains at most  $Q$  groups, denoted by  $S_p = \{g_{p,1}, g_{p,2}, \dots, g_{p,q}, \dots, g_{p,Q}\}$ ,  $1 \leq q \leq Q$ . For a genre  $S_p \in \mathbf{S}$ , a user  $U_i \in \mathbf{U}$  can be interested in several groups and also an item  $I_i \in \mathbf{I}$  can belong to one or some groups of this genre, which can be seen in Fig 7.

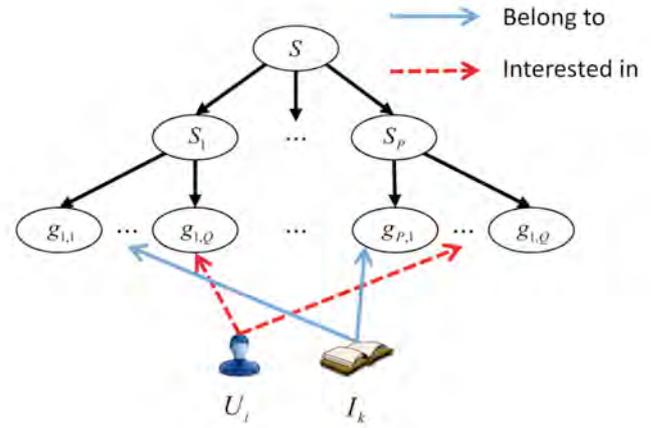


Figure 7. Contextual information.

**Definition 4:** In order to facilitate such expression, two functions  $\kappa(\cdot)$  and  $\varphi(\cdot)$  are defined to determine the groups in which user  $U_i$  is interested and the groups to which item  $I_k$  belongs, respectively:

$$\kappa_p : U_i \mapsto \kappa_p(U_i) \subseteq S_p \quad (12)$$

$$\varphi_p : I_k \mapsto \varphi_p(I_k) \subseteq S_p \quad (13)$$

Generally, the main steps of a recommendation system is illustrated in Fig 8.

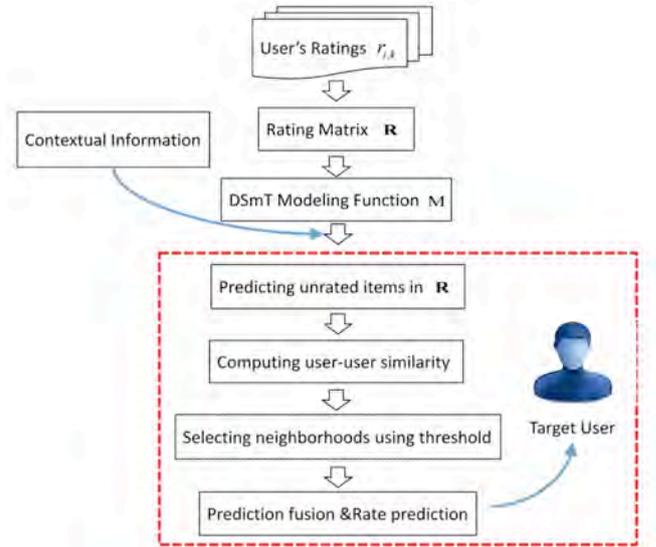


Figure 8. General process of recommendations.

The functional blocks of Fig. 8 are as follows:

1) DSMT-Modeling Function

Regarding the DS-partial probability models proposed in [23], the existing ratings  $r_{i,k}$ , of user  $U_i$  on item  $I_k$ , are

modeled by DSMT-modeling function  $M(\cdot)$  in order to transform such hard ratings into the corresponding soft ratings represented as  $m_{i,k}$  as below:

**Definition 5:**

$$m_{i,k} = \begin{cases} \alpha_{i,k}(1 - \sigma_{i,k}), & \text{for } A = \theta_l; \\ \frac{1}{2}\alpha_{i,k}\sigma_{i,k}, & \text{for } A = B; \\ \frac{1}{2}\alpha_{i,k}\sigma_{i,k}, & \text{for } A = C; \\ 1 - \alpha_{i,k}, & \text{for } A = \Theta; \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

with

$$B = \begin{cases} \theta_1 \cup \theta_2, & \text{if } l = 1; \\ \theta_{L-1} \cup \theta_L, & \text{if } l = L; \\ \theta_{l-1} \cup \theta_l \cup \theta_{l+1}, & \text{otherwise.} \end{cases}$$

$$C = \begin{cases} \theta_1 \cup \theta_2, & \text{if } l = 1; \\ \theta_{L-1} \cup \theta_L, & \text{if } l = L; \\ (\theta_{l-1} \cap \theta_l, \theta_l \cap \theta_{l+1}), & \text{otherwise.} \end{cases}$$

where  $\alpha_{i,l} \in [0, 1]$  and  $\sigma_{i,k}$  are a trust factor and a dispersion factor, respectively [23].

Referring to the partial probability model analysis in [23], we also give the corresponding user profiles which can be seen in Fig 9.



Figure 9. DSMT modeling function.

Compared to [23], the difference is that we not only consider the union  $n$  (black and gray rectangle), but also consider the intersection (red rectangle) of the hard ratings, which is also the distinction between DS theory and DSMT theory.

**Lemma 1:** Referring to Definition 5, we can also generate the relative refined BBA in the framework of DS theory:

$$m_{i,k}^{Refined} = \begin{cases} \alpha_{i,k}(1 - \sigma_{i,k}), & \text{for } A = \theta_l; \\ \alpha_{i,k}\sigma_{i,k}, & \text{for } A = B; \\ 1 - \alpha_{i,k}, & \text{for } A = \Theta; \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

with

$$B = \begin{cases} \theta_1 \cup \theta_2, & \text{if } l = 1; \\ \theta_{L-1} \cup \theta_L, & \text{if } l = L; \\ \theta_{l-1} \cup \theta_l \cup \theta_{l+1}, & \text{otherwise.} \end{cases}$$

where  $\alpha_{i,k} \in [0, 1]$  and  $\sigma_{i,k}$  are a trust factor and a dispersion factor, respectively [23].

After soft ratings are generated, DSMP [33] is applied to decouple non-Bayesian  $m_{i,k}$ , since the hierarchical fusion algorithm is currently just available for Bayesian BBAs.

**Definition 6:** DSMP is a new generalized pignistic transformation defined by  $DSMP_\varepsilon(\emptyset) = 0$  and for any singleton  $\theta_i \in \Theta$  by

$$DSMP_\varepsilon(\theta_i) \triangleq m(\theta_i) + (m(\theta_i) + \varepsilon) \times \sum_{A \in 2^\Theta, \theta_N \subset A, |A| \geq 2} \frac{m(A)}{\sum_{B \in 2^\Theta, B \subset A, |B|=1} m(B) + \varepsilon \cdot |A|} \quad (16)$$

As shown in [33], DSMP makes a remarkable improvement compared with BetP and CuzzP, since a more judicious redistribution of the ignorance masses to the singletons has been adopted by DSMP.  $\varepsilon$  is a small positive number, typically  $\varepsilon = 0.001$ .

2) Predicting unrated items:

Assuming that users who are keen on the similar groups tend to have common preferences. In this RS, it is necessary to predict the unrated items first. Considering a group  $g_{p,q} \in S_p$  with  $g_{p,q} \in \varphi(I_k)$ , every soft rating,  $m_{i,k}$ , of user  $U_i$ , who is keen on group  $g_{p,q}$ , on item  $I_k$  is regarded as a block of common preference for group  $g_{p,q}$ . Thus,  $G_{m_{p,q,k}} : D^\Theta \rightarrow [0, 1]$  which represents all users' group preferences on item  $I_k$  regarding group  $g_{p,q}$ , is computed as follows

$$G_{m_{p,q,k}} = \bigoplus_{\{j | I_k \in I_j^R, g_{p,q} \in \kappa_p(U_j), g_{p,q} \in \varphi_p(I_k)\}} m_{j,k} \quad (17)$$

Supposing that item  $I_k$  has not been rated by user  $U_i$ , it usually contains three steps to generate unprovided rating  $r_{i,k}$  of user  $U_i$  which are shown as below

- Step one: Considering a concept  $S_p$ , for each group  $g_{p,q} \in \kappa_p(U_i) \cap \varphi_p(I_k)$ , it is assumed that all users' group preferences on item  $I_k$  regarding group  $g_{p,q}$  imply common preference of  $U_i$  on  $I_k$  regarding group  $g_{p,q}$ . Furthermore, this group preference is regarded as a piece of user  $U_i$ 's concept preference on item  $I_k$  regarding concept  $S_p$ . Therefore, concept

preference of user  $U_i$  on item  $I_k$  regarding concept  $S_p$ , denoted by mass function  $S_{m_{p,q,k}} : D^\Theta \rightarrow [0, 1]$ , can be computed as below

$$S_{m_{p,q,k}} = \bigoplus_{\{q|g_{p,q} \in \kappa_p(U_j), g_{p,q} \in \varphi_p(I_k)\}} G_{m_{p,q,k}}. \quad (18)$$

- Step two: If there exists at least one common group in concept  $S_p$  which item  $I_k$  belongs to and also user  $U_i$  is interested in, then  $U_i$ 's concept preference on item  $I_k$  regarding concept  $S_p$  is regarded as a piece of context preference. Therefore, this user's contextual preference on item  $I_k$ , denoted by mass function  $S_{m_{i,k}} : D^\Theta \rightarrow [0, 1]$ , is achieved as follows

$$S_{m_{i,k}} = \bigoplus_{p=1, \dots, P} S_{m_{p,i,k}}. \quad (19)$$

- Step three: Context preference of  $U_i$  on item  $I_k$  is assigned to unprovided rating  $\bar{m}_{i,k}$  as below

$$\bar{m}_{i,k} = S_{m_{i,k}}. \quad (20)$$

So far, all unprovided ratings are predicted in this RS. Subsequently, user-user similarities are computed depending on both provided and predicted ratings in the following steps.

### 3) Computing user-user similarities:

Here, we use the distance measure proposed in [34] to calculate distances between two users  $U_i$  and  $U_j$  with  $i \neq j$ , which is defined as below

$$D(U_i, U_j) = \sum_{k=1}^N \left( \ln \max_{\theta \in \Theta} \frac{m_{j,k}(\theta)}{m_{i,k}(\theta)} - \ln \min_{\theta \in \Theta} \frac{m_{j,k}(\theta)}{m_{i,k}(\theta)} \right), \quad (21)$$

where  $m_{i,k}$  and  $m_{j,k}$  are the soft ratings of user  $U_i$  and user  $U_j$  on item  $I_k$  respectively. Afterwards, the degree of similarity between  $U_i$  and  $U_j$ , denoted by  $s_{i,j}$ , is calculated as follows

$$s_{i,j} = e^{-\gamma \times D(U_i, U_j)}, \text{ where } \gamma \in (0, \infty). \quad (22)$$

Obviously, if the value of  $s_{i,j}$  is high, it means the user  $U_i$  and user  $U_j$  are very close, and vice versa. Eventually, a mathematical matrix  $\mathbf{S} = \{s_{i,j} | U_i, U_j \in \mathbf{U}, i \neq j\}$  is employed to represent the similarities among all users.

### 4) Selecting neighbors based on user-user similarities:

Taking into account an active user  $U_i$ , for each unrated item  $I_k$  by user  $U_i$ , a set containing  $K$  nearest neighborhoods, denoted by  $\mathfrak{R}_{i,k}$ , is chosen by using the method proposed in [35]. Two simple steps of this method are shown below

- Step one: the process of such selection depends on two criteria: 1. Those users who rated  $I_k$  and 2. The corresponding user-user similarities with user  $U_i$  are equal or greater than the threshold  $\tau$ .  $\mathfrak{R}_{i,k}$  denotes the selected set, which is acquired as follows:

$$\mathfrak{R}_{i,k} = \{U_j \in \mathbf{U} | I_k \in \mathbf{I}_j^R, s_{i,j} \geq \tau\}. \quad (23)$$

- Step two: all of members in  $\mathfrak{R}_{i,k}$  is descending sorted by  $s_{i,j}$  and top  $K$  members are selected as the neighborhood set  $\mathfrak{R}_{i,k}$ .

### 5) Estimating ratings according to neighborhoods:

Supposing that item  $I_k$  has not been rated by user  $U_i$ . The predicted rating of  $U_i$  on item  $I_k$  is denoted as  $\hat{m}_{i,k}$ . Thus,  $\hat{m}_{i,k}$  is calculated according to the ratings of user  $U_i$ 's nearest users. Mathematically,  $\hat{m}_{i,k}$  is given as below

$$\hat{m}_{i,k} = m_{i,k} \oplus \tilde{m}_{i,k}, \quad (24)$$

where  $\tilde{m}_{i,k}$  is the mass regarding the neighborhoods' whole preference in the set Eq (23) on item  $I_k$ . Considering user  $U_i \in \mathfrak{R}_{i,k}$ , and supposing that  $s_{i,j}$  is the similarity between user  $U_i$  and user  $U_j$ . We use a discount rate  $1 - s_{i,j}$  to discount the rating of user  $U_j$  on item  $I_k$ .

Therefore,  $\tilde{m}_{i,k}$  is:

$$\tilde{m}_{i,k} = \bigoplus_{\{j | U_j \in \mathfrak{R}_{i,k}\}} \dot{m}_{j,k}^{s_{i,j}}, \quad (25)$$

where

$$\dot{m}_{j,k}^{s_{i,j}} = \begin{cases} s_{i,j} \times m_{j,k}(A), & \text{for } A \subset \Theta; \\ s_{i,j} \times m_{j,k}(\Theta) + (1 - s_{i,j}), & \text{for } A = \Theta. \end{cases}$$

### 6) Generating recommendations:

In order to generate appropriate recommendations for the candidate user  $U_i$ , predicted ratings of  $U_i$  on all unprovided items are sorted, and then based on the sorted list, the appropriate recommendations are generated.

## V. EXPERIMENTS

To evaluate the performance of modified rigid coarsening in precision of recommendation and computational time, original rigid coarsening method and also classical PCR6 combination method are selected to be regarded as baselines. Besides, we use DS-MAE [23] to measure the precision of recommendations.

**Definition 7:** DS-MAE is mathematically given as follows

$$DS - MAE(\theta_j) = \frac{1}{|D_j|} \sum_{(i,k) \in D_j, \theta_i \in \Theta} |\hat{m}_{i,k}(\theta_j) - M(\theta_j)|, \quad (26)$$

where  $D_j$  is the testing set identifying the user-item pairs whose true rating is  $\theta_j \in \Theta$ .

Those specific users' interested information about genres is unknown. Thus, we define a rule that if a user has rated an item then this user is interested in all genres to which the item belongs.

#### 1) Experiment One:

Movielens<sup>4</sup> is a movie recommendation dataset widely used for benchmarking process. There are nearly 100,000 hard ratings on 19 different types of movies

<sup>4</sup><http://grouplens.org/datasets/movielens>

(Action, Comedy, and so on). The domain of such rating given in Movielens includes 5 levels, denoted as  $\Theta = \{1, 2, 3, 4, 5\}$ . At the same time, each user is required to evaluate at least 20 movies, so as to ensure adequate rating information. The relevant parameters used in RSs are set:  $\gamma = 10^{-4}$  and  $\forall(i, k)\{\alpha_{i,k}, \sigma_{i,k}\} = \{0.9, 2/3\}$ . However, Setting parameter  $\tau$  to be a fixed value is obviously unreasonable because the similarity between two users is quite different when using different combination methods. Hence, in this paper, the value of parameter  $\tau$  will not be set in advance. Instead, it is determined based on the similarity in matrix  $S$ . Specifically, the highest value of top 30% in  $S$  is selected for  $\tau$ .

Additionally, we adopt the robust strategy of 10-fold cross validation to conduct experiments, which is widely applied in experimental verification. Specific steps are as follows: original ratings in Movielens are first randomly divided into 10-folds and the experiments are thus carried out 10 times: in each sub-experiment, nine tenths of the ratings are chosen as training data and the remaining ratings are regarded as testing data. It's worth noting that all results illustrated in the following experiments are the average values of 10 times.

The figure 10 demonstrates the values of overall DS-MAE varying with changing neighborhood size  $K$ . And the smaller values of DS-MAE indicate the better ones. As can be seen in Fig 10, with  $K \leq 70$  performances of the three methods increase sharply as well as being the same as each other. With  $K \geq 70$ , performances of both methods become stable. Especially, performance of modified rigid coarsening method is very close to classical PCR6 rules. However, original rigid coarsening is slightly worse than the other two algorithms.

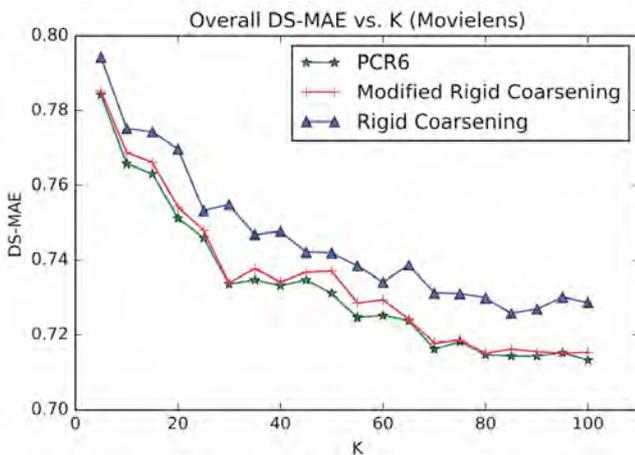


Figure 10. Overall DS-MAE between three combination methods. (Movielens).

The figure 11 depicts the computational time varying with changing neighborhood size  $K$ . In this figure, the time taken by hierarchical coarsening combination methods (both rigid coarsening and modified rigid coarsening method) is quite faster compared to classical PCR6. Besides, modified rigid coarsening is relatively slower than original rigid coarsening. All these results illustrate that modified rigid coarsening method sacrifices some of the computational efficiency, in exchange for upgrading the accuracy of approximation.

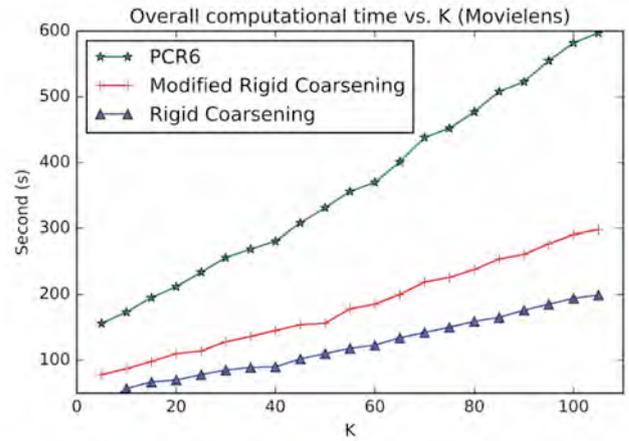


Figure 11. Overall computational time between three combination methods. (Movielens).

## 2) Experiment Two:

Flixster<sup>5</sup> is a classical recommendation dataset which nearly contains 535013 hard ratings on 19 different types of movies (Drama, Comedy, and so on). The domain of such rating given in Flixster includes 10 levels, denoted as  $\Theta = \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$ . At the same time, each user is required to evaluate at least 15 movies, so as to ensure adequate rating information. The relevant parameters used in RSs are set:  $\gamma = 10^{-4}$  and  $\forall(i, k)\{\alpha_{i,k}, \sigma_{i,k}\} = \{0.9, 2/3\}$ . However, Setting parameter  $\tau$  to be a fixed value is obviously unreasonable because the similarity between two users is quite different when using different combination methods. Hence, in this paper, the value of parameter  $\tau$  will not be set in advance. Instead, it is determined based on the similarity in matrix  $S$ . Specifically, the highest value of top 50% in  $S$  is selected for  $\tau$ .

The figure 12 demonstrates the values of overall DS-MAE varying with changing neighborhood size  $K$ . And the smaller values of DS-MAE indicate the better ones. As can be seen in Fig 12 we can get a similar result to the previous data set (Movielens). Especially, perfor-

<sup>5</sup><http://datasets.syr.edu/datasets/Flixster.html>

mance of modified rigid coarsening method is in the middle of the comparison methods. However, original rigid coarsening is worse than the other two algorithms.

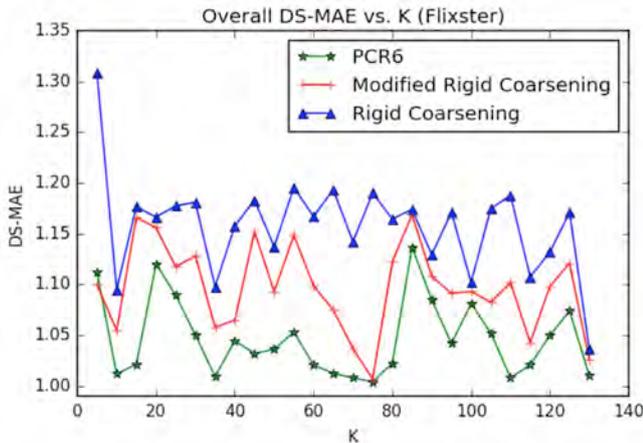


Figure 12. Overall DS-MAE between three combination method. (Flixster).

The figure 13 depicts the computational time varying with changing neighborhood size  $K$ . From this figure, we can also get the same conclusion that the time taken by hierarchical coarsening combination methods (both rigid coarsening and modified rigid coarsening method) is quite faster compared to classical PCR6.

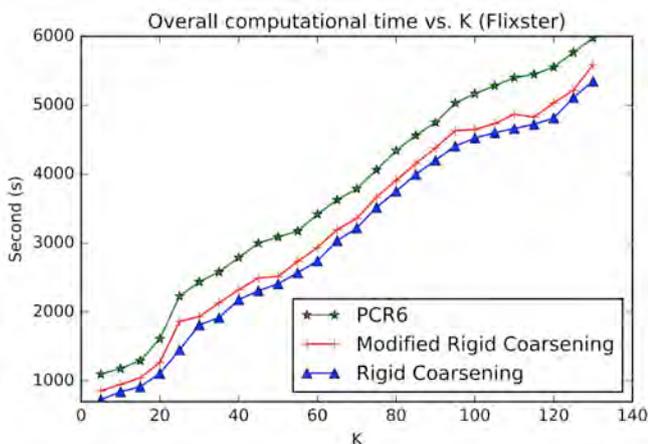


Figure 13. Overall computational time between three combination methods. (Flixster).

## VI. CONCLUSION

In this paper, we propose a new combination method, called modified rigid coarsening method. This new method can map the original refined FoD to the new coarsening FoD in the process of combination. Compared to traditional fusion method PCR6 in DS<sub>m</sub>T, this approach can not only reduce computational complexity, but also ensure high approximation accuracy. Besides, in order to verify the practicality of our

approach, we apply this approach to fuse soft ratings in RSs. To be specific, user preferences are first transformed by DS<sub>m</sub>T-partial probability model to accurately represent uncertain information. Then, information about user preferences from different sources can be easily combined. In the future work, more helpful information will be mined to discern focal element in FoD so as to improve the accuracy of approximation and more data sets will be applied.

## ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61573097, 91748106, in part by Key Laboratory of Integrated Automation of Process Industry (PAL-N201704), in part by the Qing Lan Project and Six Major Top-talent Plan, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions. The authors thank the reviewers and editors for giving valuable comments, which are very helpful for improving this manuscript.

## REFERENCES

- [1] G. Shafer, *A mathematical theory of evidence*, Princeton Univ. Press, 1976.
- [2] A. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, *Annals of Mathematical Statistics*, Vol. 38, pp. 325–339, 1967.
- [3] W. Jiang, S. Wang, X. Liu, H. Zheng, B. Wei, *Evidence conflict measure based on OWA operator in open world*, *PLoS ONE*, Vol.12(5):e0177828, 2017.
- [4] P. Smets, *Practical uses of belief functions*, in K.B. Laskey and H. Prade Editors, 15th Conf. on Uncertainty in Artificial Intelligence, pp. 612–621, Stockholm, Sweden, 1999.
- [5] J. Dezert, *Foundations for a new theory of plausible and paradoxical reasoning*, *Information & Security: An Int. Journal*, Vol. 9, 2002.
- [6] F. Smarandache, J. Dezert (Editors), *Advances and applications of DS<sub>m</sub>T for information fusion*, American Research Press, Rehoboth, NM, U.S.A., Vol. 1–4, 2004–2015. Available at webpage 2 of [7].
- [7] <http://www.onera.fr/staff/jean-dezert>
- [8] R. Kennes, *Computational aspects of the Möbius transform of graphs*, *IEEE Trans. on SMC*, Vol. 22, pp. 201–223, 1992.
- [9] G. Shafer, R. Logan, *Implementing Dempster's rule for hierarchical evidence*, *Artificial Intelligence*, Vol. 33, pp. 271–298, 1987.
- [10] Y. Yang, Y. Liu, *Iterative approximation of basic belief assignment based on distance of evidence*, *PLoS ONE*, Vol. 11(2):e0147799, 2016
- [11] T. Denœux, *Inner and outer approximation of belief structures using a hierarchical clustering approach*, *Int. J. of Uncertainty, Fuzziness and Knowledge-Based System*, Vol. 9(4), pp. 437–460, 2001.
- [12] Y. Yang, D. Han, C. Han, F. Cao, *A novel approximation of basic probability assignment based on rank-level fusion*, *Chinese Journal of Aeronautics*, Vol. 26(4), pp. 993–999, 2013.
- [13] D. Han, Y. Yang, J. Dezert, *Two novel methods of BBA approximation based on focal element redundancy*, in Proc. of Int. Conf on Information Fusion (Fusion 2015), Washington, D.C., USA, July 2015.
- [14] M. Li, Q. Zhang, Y. Deng, *A New Probability Transformation Based on the Ordered Visibility Graph*, *International Journal of Intelligent Systems*, Vol. 31(1), pp. 44–67, 2016.
- [15] Y. Dong, X. Li, J. Dezert, *A Hierarchical Flexible Coarsening Method to Combine BBAs in Probabilities*, in Proc. of the 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 1013, 2017.
- [16] P. Smets, *Analyzing the combination of conflicting belief functions*, *Information Fusion*, Vol.8(4), pp. 387–412, 2007.
- [17] X. Li, J. Dezert, X. Huang, Z. Meng, X. Wu, *A fast approximate reasoning method in hierarchical DS<sub>m</sub>T*, *Acta Electronica Sinica*, Vol. 38(11), pp. 2567–2572, 2010.
- [18] X. Li, W. Yang, X. Wu, J. Dezert, *A fast approximate reasoning method in hierarchical DS<sub>m</sub>T*, *Acta Electronica Sinica*.2011; 39(3A), pp. 32–36, 201.

- [19] X. Li, W. Yang, J. Dezert, *A fast approximate reasoning method in hierarchical DSMT*. J. Huazhong Univ. of Sci. and Tech., (Natural Science Edition), Vol. 39, pp. 151–156, 2011.
- [20] D. Han, J. Dezert, Y. Yang, *Belief Interval-Based Distance Measures in the Theory of Belief Functions*, IEEE Trans. on Systems, Man and Cybernetics: Systems, pp. 1–18, 2016.
- [21] J. Bobadilla, F. Orrega, A. Hernando, A. Gutierrez, *Recommender systems survey*, Knowledge-Based System, Vol. 46, pp. 109–132, 2013.
- [22] T. Chen, *Ubiquitous Hotel Recommendation Using a Fuzzy-Weighted-Average and Backpropagation Network Approach*, Int. J. of Intelligent Systems, Vol. 32(4), 2017.
- [23] T. Wichramaratne, K. Premaratne, M. Kubat, D. Jayaweera, *CoFiDS: a belief-theoretic approach for automated collaborative filtering*, IEEE Trans. on Knowledge and Data Engineering, Vol. 23(2), pp. 175–189, 2011.
- [24] P. Ladyzynski, P. Grzegorzewski, *Vague preferences in recommender systems*, Expert Systems With Applications, Vol. 42(24), pp. 9402–9411, 2015.
- [25] V. Nguyen, V. Huynh, *Two-probabilities focused combination in recommender systems*, Int. J. of Approximate Reasoning, Vol. 80, pp. 225–238, 2017.
- [26] R. Bagher, H. Hassanpour, H. Mashayekhi, *User trends modeling for a content-based recommender system*, Expert Systems With Applications, Vol. 87, pp. 209–219, 2017.
- [27] T. Denœux, *Maximum likelihood estimation from uncertain data in the belief function framework*, IEEE Trans. on Knowledge and Data Engineering, Vol. 25(1), pp. 119–130, 2013.
- [28] O. Kanjanarakul, S. Sriboonchitta, T. Denœux, *Forecasting using belief functions: an application to marketing econometrics*, Int. J. of Approximate Reasoning, Vol. 55(5), pp. 1113–1128, 2014.
- [29] M. Masson, T. Denœux, *Ensemble clustering in the belief functions framework*, Int. J. of Approximate Reasoning, Vol. 52(1), pp. 92–109, 2011.
- [30] L. Troiano, L. Rodriguez-Muniz, J. Diaz, *Discovering user preferences using Dempster-Shafer Theory*, Fuzzy Sets Systems, Vol. 278, pp. 98–117, 2015.
- [31] V. Nguyen, V. Huynh, *A reliably weighted collaborative filtering system*, in Proc. of ECSQARU 2015, pp. 429–439, 2015.
- [32] J. Iglesias, A. Bernardos, J. Casar, *An evidential and context-aware recommendation strategy to enhance interactions with smart spaces*, in Proc. of HAIS 2013, pp. 242–251, Lecture Notes in Computer Science, Vol. 8073. Springer, Berlin, Heidelberg, 2013.
- [33] J. Dezert, F. Smarandache, *A new probabilistic transformation of belief mass assignment*, in Proc. of 11th Int. Conf. on Information Fusion (Fusion 2008, Cologne, Germany, July 2008).
- [34] H. Chan, A. Darwiche, *A distance measure for bounding probabilistic belief change*, Int. J. of Approximate Reasoning, Vol. 38(2), pp. 149–174, 2005.
- [35] J. Herlocker, J. Konstan, A. Borchers, J. Riedl, *An algorithmic framework for performing collaborative filtering*, in Proc. of SIGIR'99, ACM, pp. 230–237, 1999.



# Geophysical and Geotechnical Data Fusion for Levee Assessment - Interface Detection with Biased Geophysical Data

Théo Dezert<sup>a,b</sup>, Sérgio Palma Lopes<sup>a</sup>, Yannick Fargier<sup>b,c</sup>, Philippe Côté<sup>a</sup>

<sup>a</sup>IFSTTAR, GERS, GeoEND, F-44344 Bouguenais, France.

<sup>b</sup>CEREMA Direction Territoriale Normandie-Centre, F-41000 Blois, France.

<sup>c</sup>Univ. Lyon, IFSTTAR, GERS, RRO, F-69675 Bron, France.

Emails: theo.dezert@ifsttar.fr, sergio.lobes@ifsttar.fr, yannick.fargier@ifsttar.fr, philippe.cote@ifsttar.fr

Originally published as: T. Dezert, S. Palma Lopes, Y. Fargier, P. Cote, *Geophysical and Geotechnical Data Fusion for Levee Assessment - Interface Detection with Biased Geophysical Data*, Near Surface Geoscience Conference & Exhibition 2018, 9–12 September 2018, Porto, Portugal, and reprinted with permission.

**Abstract**—To prevent disastrous consequences imputed to levee breakage, assessment methodologies have to be improved. Geophysical and geotechnical investigation methods are usually used to make such assessments. However, the effective combination of these two specific types of data remains a challenge. We propose the fusion of geophysical and geotechnical data by means of Belief Functions. Here we demonstrate our approach on a synthetic case study including geophysical (electrical resistivity) and geotechnical (cone-bearing) data and by implementing Smets and PCR5 normalization rules. This new data combination approach allows the characterization of horizontal interfaces and of a geological structure initially hidden by the effects of a highly conductive body.

**Keywords:** data fusion, belief functions, geophysical data, geotechnical data, experimental test bench, electrical resistivity tomography.

## I. INTRODUCTION

Fluvial levees are elevated partitions between channels and floodplains [1], built for flood protection. These structures are considered as hazardous and may fail, leading to disastrous consequences such as human and material loss and economic disasters. Levee assessment acknowledged methodologies usually include geotechnical and geophysical investigation methods [2]. While geotechnical investigation methods are intrusive, they provide quite accurate and punctual information. Conversely, geophysical methods are non-intrusive and provide physical information on large volumes of subsoil with high output (according to the chosen method and acquisition mode) and potentially significant uncertainties. These associated uncertainties can particularly be attributed to the indirect and integrating nature of the methods as well as to the non-uniqueness of inverse problems solution. One of the important issues when assessing levees is the combination of geotechnical and geophysical data [3]. Furthermore, one should take into consideration their respective imprecisions, uncertainties and contrasting spatial distributions. We suggest the use of Belief Functions (BFs) and combination rules to merge geotechnical (cone bearing) and geophysical (electrical resistivity) data. Our data fusion methodology is being optimized and tested on both

real and synthetic data. In this paper, we aim to demonstrate the potential of this methodology by means of a synthetic study that exemplifies a simplified levee case study. Indeed, we show the ability of combined geotechnical (cone bearing) and geophysical (electrical resistivity) data to discriminate three type of geological materials in the presence of a conductive anomaly, which has a significant bias effect on the geophysical data. The reader can refer to the theoretical basis of BFs, introduced by Shafer [4]. The use of BFs needs: (1) to select a common frame of discernment (FoD) of the considered problem, (2) to determine the masses of belief or Basic Belief Assignments (BBAs) from available data (geotechnical and geophysical), and (3) to choose a rule of combination.

## II. FOD AND BBAs CONSTRUCTION

For the addressed levee assessment issue, we consider three classes of distinct materials  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ . Since the FoD,  $\Theta$ , must consist of a set of exclusive and exhaustive hypotheses, we will be using a fourth class  $\theta_4$  to cover the physical characteristics of materials not included in the three first sets. Thus we use  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ . The construction of the BBAs for each data source consists in assigning each data type (geophysical and geotechnical) to  $\Theta$ .

## III. CONSTRUCTION FROM GEOPHYSICAL DATA - ELECTRICAL RESISTIVITY VALUES

Since electrical resistivity (ER) tomography is one of the most widely used methods for levee investigation, we propose the use of ER as geophysical data. As a framework to exemplify a fluvial levee problematic we consider two soil layers: an upper resistive layer ( $10^3 \Omega.m$ ) standing for sands [5] and a subjacent and more conductive one ( $10 \Omega.m$ ) standing for a clayey layer starting at 6 m depth. An inhomogeneity ( $10^2 \Omega.m$ ) standing for a silty lens of about 1.3 m high and 40.5 m wide is positioned at 7 m depth (Figure 1.a). We then associate ER classes of specific soils (split into ranges of ER in  $\Omega.m$ ) to  $\Theta$  so that:  $\theta_1 = [5, 20]$ ,  $\theta_2 = [50, 2 \cdot 10^2]$ ,  $\theta_3 = [5 \cdot 10^2, 2 \cdot 10^3]$  and  $\theta_4 = [0.2, 5[ \cup ]20, 50[ \cup ]2 \cdot 10^2, 5 \cdot 10^2[ \cup ]2 \cdot 10^3, 5 \cdot 10^3]$ .

We finally place a very small ( $0.5 \times 0.5\text{m}$ ) and highly conductive body ( $10^{-6} \Omega\cdot\text{m}$ ) centered at 1.35 m depth (represented in green in Figure 1.a), that could be considered as a metallic pipe. Even though this body is small in the  $XY$  plane, it might have an elongated shape in the  $Z$  direction. We use Res2Dmod free software [6] to simulate noiseless data acquisition from the chosen resistivity model (Figure 1.a) and then use the Res2Dinv software (ver. 3.71.118) [7] to get the inverted ER section as one would obtain from the processing of survey data (Figure 1.b).

The discrimination between sands and clays is obvious while the distinction of the silty lens is not. Indeed, the conductive local body generates resistive artefacts close to its position and a very large and conductive artifact deeper in the subsoil, hiding the presence of the silty lens. The image given by the inverted ER (Figure 1.b) strongly departs from the true model (Figure 1.a). We finally use the Res2dinv discretization grid for the BBA  $m_1(\cdot)$  corresponding to each event of  $2^\Theta$  (each event of  $\Theta$  plus their possible unions and the empty set). The values of the masses are set using the Wasserstein distances [8] between (a) an inverted ER value  $\pm$  its uncertainty, issued from the Res2dinv result, and (b) the interval corresponding to each event, so as each cell of the grid gets a normalized BBA.

#### IV. CONSTRUCTION FROM GEOTECHNICAL DATA - CONE-BEARING VALUES

We use artificial cone bearing values (expressed in MPa) as geotechnical data. These physical values could have been obtained from a cone penetrometer test (CPT) investigation campaign. Indeed, CPT campaigns are widely used to investigate embankment levees [2]. We simulate a data acquisition from four boreholes with an interspacing of 20 m, drilled to 17 m depth with a vertical acquisition every 50 cm (dashed lines in Figure 1.a). Two of the boreholes happen to go through the silty lens. We assign intervals of cone bearing values (in MPa) to  $\Theta$  so that:  $\theta_1 = [2, 8]$ ,  $\theta_2 = [20, 80]$ ,  $\theta_3 = [2 \cdot 10^2, 8 \cdot 10^2]$  and  $\theta_4 = [0.1, 2 \cup] 8, 20 \cup] 80, 2 \cdot 10^2 \cup] 8 \cdot 10^2, 10^3]$ . These intervals can be associated to specific soil types [9] such as clays for low values, silty soils for intermediate values and sands for higher ones. We assume a mass of belief equal to 1 in the borehole and impose an exponential lateral decrease of the trust in the data (following the mean horizontal scale of fluctuation of about 50 m proposed by Phoon and Kulhawy [10]). The geotechnical grid depends on the distance between the boreholes and the vertical sampling. Thus, for each cell, a second BBA  $m_2(\cdot)$  is proposed, entering in the fusion process

#### V. COMBINATION OF BBAS AND PRELIMINARY RESULTS

We suggest the use of a fusion mesh containing all the meshes from both the geophysical and geotechnical grids so that we avoid data interpolation that might lead to unnecessary data alteration. The data merging consists in combining  $m_1(\cdot)$  and  $m_2(\cdot)$  assigned to each cell of the grid. While many rules of BBA combination have been proposed, in this work we

focus on two of them: Smets' rule [11] and the Proportional Conflict Redistribution rule no. 5 (PCR5) [12]. Smets' rule (conjunctive rule under an open-world assumption) allows the quantification of the conflict level of our two information sources (geotechnical and geophysical sources) represented by (Eq. 1):

$$m_{12}(\emptyset) = \sum_{X_1, X_2 \subseteq \Theta | X_1 \cap X_2 = \emptyset} m_1(X_1)m_2(X_2) \quad (1)$$

Thanks to the latter rule, we are able to point out the conflictual zones in the vicinity of: the horizontal interfaces, the silty lens, the local very conductive body and the resistive and conductive artifacts (in red, Figure 2.a). The fusion, following Dempster-Shafer's rule (Eq. 2) (closed world assumption) [4] with the PCR5 normalization [12] (Figure 2.c) is fairly close to the true model we used (Figure 1.a). This normalization allows the spreading of the conflict masses  $m_{12}(\emptyset)$  to other events of  $2^\Theta$ .

$$m_{12}^{\text{DS}}(A) = \frac{1}{1 - m_{12}(\emptyset)} \sum_{X_1, X_2 \subseteq \Theta | X_1 \cap X_2 = A} m_1(X_1)m_2(X_2) \quad (2)$$

It exhibits a quite clear view of the interface between sands and clays and allows the visualization of the silty lens despite the blind zone generated by the conductive anomaly (Figure 1.b). As a decision-making support, we propose to display the events having the highest belief masses (Figure 2.a and 2.c) and their associated degrees of belief (Figure 2.b and 2.d).

Via our procedure, even though the highly conductive anomaly (that can be associated to a metallic pipe) is not clearly detected and characterized, Figure 2.a still points out a conflictual zone around the position of that anomaly. Unfortunately, we still have incorrect material type determination on the 19 first horizontal meters and 15 last horizontal meters about the sand/clay interface because of the wrong ER values proposed by the inverted geophysical model (Figure 1.b). In the future, this kind of under-determination may be minimized by reconsidering the way to decrease the lateral trust of the geotechnical data.

#### VI. CONCLUSION

The use of BFs for the fusion of geophysical and geotechnical investigation data is promising. Indeed, it enables to highlight the presence of an interface between two geological media much more precisely than the geophysical method alone, using Res2Dinv. Furthermore, it enables the reliable estimation of the complete extension of a lens with intermediate ER and cone bearing values, even though the effects of a local and highly conductive body (that can be associated to a transversal metallic pipe) hides the geological lens. Without normalization, Smets' combination rule easily spotlights the conflicting zones. Such information could also be precious during an investigation campaign, indicating zones where survey has to be strengthened.

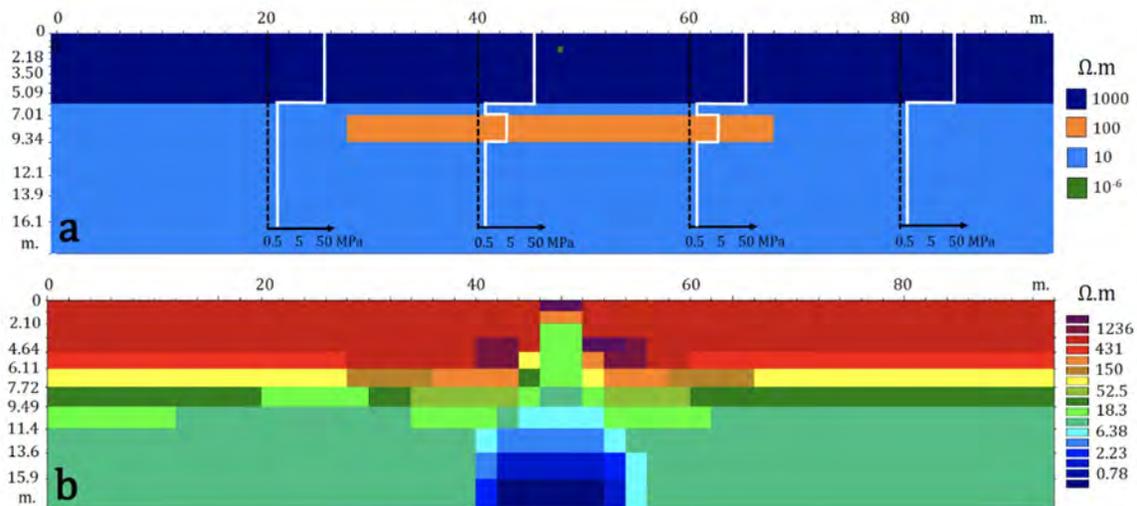


Fig. 1. 2D section of subsoil displaying a) true ER with borehole positions in dashed lines and associated cone bearing values in white and b) inverted ER model displayed in model data blocks with RMS error = 1.11%.

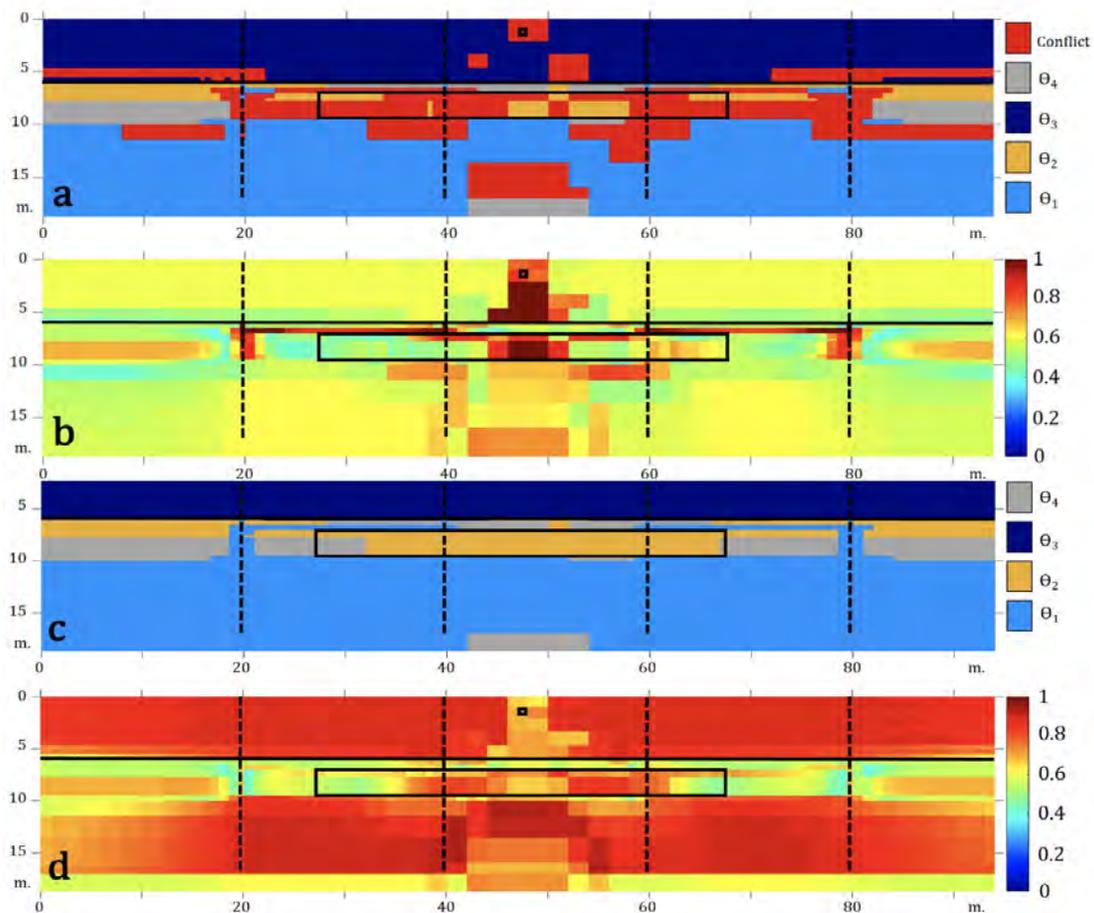


Fig. 2. Data merging with Smets (a, b) and PCR5 normalization rule (c, d). (b) and (d) represent the BBAs associated to the most plausible events respectively presented in (a) and (c). The black lines stand for the interfaces and the inhomogeneities fixed in the model (Figure 1.a) while the dashed lines stand for the borehole positions.

This will be in the heart of further studies in order to be able to propose the most pertinent positions for geotechnical boreholes thanks to belief functions and combination rules, therewith to improve fluvial levee assessment. This algorithm will also be employed using real data acquired on a scale model as well as on a levee in order to propose a 3D modelling.

#### ACKNOWLEDGMENTS

We would like to thank the “Région Pays de la Loire” for its financial support.

#### REFERENCES

- [1] G.J. Brierley, R.J. Ferguson, K.J. Woolfe, *What is a fluvial levee?*, Sedimentary Geology, Vol. 114(1-4), pp. 1-9, 1997.
- [2] C. Fauchard, P. Mériaux, *Geophysical and geotechnical methods for diagnosing flood protection dikes: Guide for implementation and interpretation*, Quae, 2007.
- [3] P. Royet, S. Palma Lopes, C. Fauchard, P. Mériaux, L. Auriou, *Rapid and cost-effective dike condition assessment methods: geophysics and remote sensing*, FloodProBE Project, 2012.
- [4] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [5] G. Palacky, G.F. West, *Electromagnetic methods in applied geophysics*, Resistivity Characteristics of Geologic Targets, 52-129, 1987.
- [6] M.H. Loke, *RES2DMOD ver. 3.01: Rapid 2D resistivity forward modelling using the finite difference and finite-element methods*, Software manual, 2002.
- [7] M.H. Loke, R.D. Barker, *Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method*, Geophysical prospecting, Vol. 44(1), pp. 131-152, 1996.
- [8] A. Irpino, R. Verde, *Dynamic clustering of interval data using a Wasserstein-based distance* Pattern Recognition Letters, Vol. 29(11), pp. 1648-1658, 2008.
- [9] P.K. Robertson, *Soil classification using the cone penetration test*, Canadian Geotechnical Journal, Vol. 27(1), pp. 151-158, 1990.
- [10] K.K. Phoon, F.H. Kulhawy, *Characterization of geotechnical variability*, Canadian Geotechnical Journal, Vol. 36(4), pp. 612-624, 1999.
- [11] P. Smets, *The combination of evidence in the transferable belief model*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12(5), pp. 447-458, 1990.
- [12] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion - Collected works*, Volume 3, American Research Press, 2009.

# Human Heading Perception Based on Form and Motion Combination

Albena Tchamova<sup>a</sup>, Jean Dezert<sup>b</sup>, Pavlina Konstantinova<sup>c</sup>,  
Nadejda Bocheva<sup>d</sup>, Biliana Genova<sup>d</sup>, Miroslava Stefanova<sup>d</sup>

<sup>a</sup>Institute of Information and Communication Techn., Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>c</sup>European Polytechnical University, Pernik, Bulgaria.

<sup>d</sup>Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia, Bulgaria.

Emails: tchamova@bas.bg, jean.dezert@onera.fr, pavlina.konstantinova@gmail.com,  
nbbocheva@hotmail.com, b.genova@abv.bg, b.genova@abv.bg

Originally published as: A. Tchamova, J. Dezert, P. Konstantinova, N. Bocheva, B. Genova, M. Stefanova, *Human Heading Perception Based on Form and Motion Combination*, in Proc. of IEEE (SMC) INSTA 2018 Conf., Thessaloniki, Greece, July 3–5, 2018, and reprinted with permission.

**Abstract**—This paper presents a study on human perception of the heading on the base of motion and form visual cues integration. The authors examine how human age influences this process. Because the visual stimuli are in general uncertain, or in some cases even conflicting, the process of combination is estimated on the base on the well known Normalized Conjunctive Consensus fusion rule, as well as on the base of the more efficient Dezert-Smarandache Theory (DSMT) of plausible and paradoxical reasoning, and more precisely on the probabilistic Proportional Conflict Redistribution rule no.5 defined within it. The main goal is focused on how these fusion rules succeed to model consistent and adequate predictions about both individuals' behavior, and age-contingent groups of individuals.

**Keywords:** vision, heading perception, form cue, motion cue, cues combination, DSMT, probabilistic proportional redistribution rule no.5, normalized conjunctive rule.

## I. INTRODUCTION

Form and motion information are closely linked and continuously interacting in the human visual system, which takes the advantage to utilize both of these visual characteristics (or so called cues) to make decisions about human heading perception [1] described via the respective rapid eye movement (so called saccades) towards the object of interest position. The cooperation between the form and motion cues becomes very useful and even necessary, when: (i) each cue (motion, form) alone does not supply sufficient information to estimate the proper and accurate heading, or/and, (ii) the uncertainty, associated with the utilized visual cues and the possible conflicts between them influence negatively the process of decision making. The last case relates closely to the effect of the age-related changes throughout the life cycle and to deterioration in the cognitive processes, and consequently in visual information processing due to a variety of factors like cell death, cognitive de-differentiation, increase of internal noise in the visual system. As a result, the contrast sensitivity, self-motion perception, as well as eye movement characteristics are deteriorated in the elderly [2], [3], [4]. To overcome all these difficulties one needs to combine and utilize in an effective way both

of cues in order to achieve inferences, more informative and potentially more accurate than if they were obtained by means of a single cue. Integration of information from multiple sources (cues) in a single modality increases the precision of perceptual performance. Such a claim recently has been supported by a list of neurobiological studies, like [5], [6], [7], and also neurophysiological findings exist about neurons responding to both form and motion in some cortical sites (including early visual areas and extrastriate areas) [8], [9].

Inspired and based on these important biological findings of the cue combination effectiveness, the aim of this paper is to investigate how humans integrate motion and form information in the process of decision making about heading direction. The authors will focus on how the human age influences this process, and also whether the human visual system is enable to adapt during the life cycle in order to exploit all available information, providing a sensible and meaningful decision about the problem under concern. In our study we simulate only the directional flow occurring during the forward motion of the observer and not the changes in speed or size of the moving objects that accompanied it. The researcher team will compare human cue combination performance with modelled combination performance, based on particular fusion rules. In the presented study the authors will apply and compare the performances of the following fusion rules: the Normalized Conjunctive Consensus (NCC), and the very recent probabilistic Proportional Conflict Redistribution rule no.5 (pPCR5) defined within DSMT. The novelty of our study consists in applying especially this novel pPCR5 fusion rule to model the human process of form and motion cues integration.

This paper is organized as follows. In section II we briefly present the form and motion combination process, and the principles of the used fusion rules, applied to model the human cue integration. Section III is devoted to the experimental strategy, methods, procedures, stimulus, apparatus, and also

subjects participating in the experiments. The results obtained are described and analysed in Section IV. Conclusions are made in Section V.

## II. FUSION RULES FOR MODELLING VISUAL CUE COMBINATION

Various fusion rules exist in the literature to deal with uncertain or even conflicting evidence based on different mathematical models and on different methods for transferring the conflicting mass onto the sensible hypotheses about the problem under consideration. The classical one is Bayesian inference [10], [11] which deals with probabilistic information. The main idea of Bayesian inference is to obtain the most reliable estimate of the state of the world on the base of independent cues combination, i.e. the estimate in which the variance of the resulting combined cue is minimized. But being very sensitive to the sources with the bigger means, it could neglect part of available information, which is not adequate and reliable behavior in cases of conflicting visual cues combination. Bayesian inference has some difficulties to apply, related to the requirements of measurements' statistics and knowledge about the a priori information. Dempster-Shafer Theory (DST) [12], [13] was the first theory for combining uncertain information expressed as basic belief assignments with Dempster's rule. Although appealing in modelling the epistemic uncertainty this theory shows very questionable and controversial results in cases of high (and even low) conflicting sources of evidence [14], [15], [16], [17].

To overcome all these limitations of DST, Dezert-Smarandache Theory of Plausible and Paradoxical Reasoning was developed [18].

DSMT works for any model, which fits adequately with the true nature of the fusion problem under consideration. It is a general mathematical framework for managing and solving problems of uncertain, highly conflicting, imprecise knowledge representation and fusion, and decision making procedures, based on vague, imprecise models for a wide class of static or dynamic fusion problems.

### A. Normalized Conjunctive Consensus rule

The Normalized Conjunctive Consensus (NCC) rule is used to combine simultaneously assumed independent visual cues. In the case considered in our paper, the information obtained by the available form and motion cues is characterized by Gaussian likelihood functions with given means  $\mu_i, i = 1, 2, \dots$  and standard deviations  $\sigma_i, i = 1, 2, \dots$ , defining the uncertainty encountered in data. In case of two independent cues with one-dimensional Gaussian distributions  $p_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2$  and  $p_2(x) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2$ , the combined distribution based on NCC rule becomes:

$$p_{NCC}(x) = \frac{1}{\sigma_{NCC}\sqrt{2\pi}} \exp -\frac{1}{2}\left(\frac{x-\mu_{NCC}}{\sigma_{NCC}}\right)^2, \quad (1)$$

where  $\sigma_{NCC}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$  and  $\mu_{NCC} = \sigma_{NCC}^2\left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2}\right)$ .

It is characterized with a mean, biased toward the function with the bigger of the two means, similarly to Bayesian estimator. It is optimal, i.e. minimizes the variance of the error estimation, when the original distributions have close mean values. When both cues are in conflict, however, (characterized with distant distributions), NCC rule leads to neglecting part of the available information, because the source with the bigger mean is weighted more heavily. In this case it is reasonable to keep the original distributions in the fused probability density function until it is possible to make reliable decision. This has been done by pPCR5 fusion rule defined in DSMT.

### B. Probabilistic Proportional Conflict Redistribution rule no.5

The general principle of all Proportional Conflict Redistribution rules [18], Vol.3 is to: 1) calculate the conjunctive consensus between sources of evidence (different visual cues) 2) calculate the total or partial conflicting masses; 3) redistribute the conflicting mass (total or partial) proportionally on non-empty sets involved in the model according to all integrity constraints. The recently proposed non-Bayesian probabilistic Proportional Conflict Redistribution rule no.5 (pPCR5) [18] is based on the discrete Proportional Conflict Redistribution rule no.5 [18], Vol.3, for combining discrete basic belief assignments. For completeness, we will discuss in brief the main idea behind the discrete PCR5. It comes from the necessity to deal with both uncertain and conflicting information, transferring partial or total conflicting masses proportionally only to non-empty sets involved in the particular conflict and proportionally to their individual masses. Basic belief assignment (bba) represents the knowledge, provided by particular source of information about its belief in the true state of the problem under consideration. Given a frame of hypotheses  $\Theta = \{\theta_1, \dots, \theta_n\}$ , and the so called power set  $2^\Theta = \{\emptyset, \theta_1, \dots, \theta_n, \theta_1 \cup \theta_2, \dots, \theta_1 \cup \theta_2 \cup \dots \cup \theta_n\}$ , on which the combination is defined, the general basic belief assignment is defined as a mapping  $m_s(\cdot) : 2^\Theta \rightarrow [0, 1]$ , associated with the given source of information s, such that:  $m_s(\emptyset) = 0$  and  $\sum_{X \in 2^\Theta} m_s(X) = 1$ . The quantity  $m_s(X)$  represents the mass of belief exactly committed to X. Under Shafer's model assumption of the frame  $\Theta$  (requiring all the hypotheses to be exclusive and exhaustive), the PCR5 combination rule for only two sources of information is defined as:  $m_{PCR5}(\emptyset) = 0$  and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{PCR5}(X) = m_{12}(X) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right]. \quad (2)$$

All sets involved in the formula are in canonical form. The quantity  $m_{12}(X)$  corresponds to the conjunctive consensus, i.e:  $m_{12}(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1) m_2(X_2)$ . All denominators are different from zero. If a denominator is zero, that fraction is discarded. No matter how big or small the conflicting mass is, PCR5 mathematically does a proper redistribution of the conflicting mass. It is because PCR5 goes backwards

on the tracks of the conjunctive rule and redistributes the partial conflicting masses only to the sets involved in the conflict and proportionally to their masses put in the conflict, considering the conjunctive normal form of the partial conflict. PCR5 is quasi-associative and preserves the neutral impact of the vacuous belief assignment. The probabilistic PCR5 (pPCR5) is an extension of discrete PCR5 version to its continuous probabilistic counterpart. Basic belief assignment, involved in discrete PCR5 rule is extended to densities of probabilities of random variables. For two independent sources of information with given Gaussian distributions  $p_1(x)$  and  $p_2(x)$ , the obtained combined result becomes [18]:

$$p_{pPCR5}(x) = p_1(x) \int \frac{p_1(x)p_2(y)}{p_1(x) + p_2(y)} dy + p_2(x) \int \frac{p_2(x)p_1(y)}{p_2(x) + p_1(y)} dy. \quad (3)$$

The behavior of pPCR5 fusion rule in comparison to NCC rule (1) could be characterized by two cases below:

**Case 1:** both densities  $p_1(x)$  and  $p_2(x)$  are close (Fig.1-case 1). The combined density acts as an amplifier of the information by reducing the variance. Here pPCR5 acts as NCC fusion rule.

**Case 2:** the densities  $p_1(x)$  and  $p_2(x)$  are distant (Fig.1-case 2). Then the combined density keeps both original densities (not merging both densities into only one unimodal Gaussian density as NCC rule does), avoiding to neglect a part of the available information.

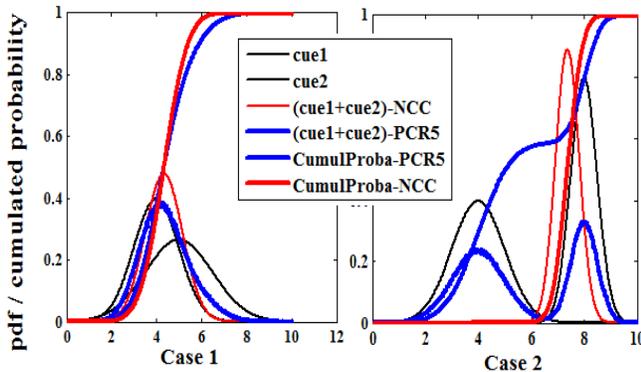


Figure 1. Performance of pPCR5 fusion rule vs. NCC rule.

This new (from a theoretical point of view) property is very interesting and it presents advantages for practical applications as it will be shown in our particular research. Application of pPCR5 fusion rule assures robustness to the potential errors and allows taking more reliable and adequate decisions in the process of integration of different cues in visual perception.

### III. EXPERIMENTS

#### A. Stimuli

The stimuli consisted of 50 dots. The dot patterns occupied an area of 15 angular degrees. The stimuli were

generated beforehand and contained 100 frames (except the static condition). Each frame lasted 33 msec. The lifetime of the dots was 3 frames, thus on every frame one-third of the dots were randomly re-positioned. For the motion and the combined condition the velocity of the dots was 4 degrees of arc/sec. The stimuli were radial patterns with a focus (center) positioned eccentrically to the middle of the screen. The center of the patterns defined by the orientation of the pairs or the trajectories of the dots could take 7 values to the left or to the right of the midpoint of the screen: 0.67 to 4.67 degrees of arc in steps of 0.67 degrees of arc. Ten different exemplars of patterns for each center and condition were generated. The dots subtended 0.2 degrees of arc.

#### B. Experimental conditions

Four different experimental conditions were performed:

- **Static (form) condition** The experimental stimuli (Fig.2) consist of dots pairs separated by 2 degrees of arc. The orientation of the virtual lines connecting the dots in 18 pairs intersected in a common point considered the center of the patterns, while the rest 7 pairs had random orientation.
- **Motion condition** In this experiment (Fig.3) 36 points had trajectories that intersected at a common point, while the rest 14 dots had random trajectories.
- **Flicker condition** In this condition (Fig.4) a sequence of random static patterns was presented. As in the static condition the orientation of 18 pairs of dots, separated by 2 degrees of arc pointed to a common center while the rest 7 pairs had different orientation. The sequential presentation of the static patterns created illusory motion, but the trajectories of the apparent motion were random.
- **Combined condition** In this experiment (Fig.5) 18 pairs of dots moved along trajectories towards a common center. The orientation of these pairs was along the motion trajectory. The rest 7 pairs had random trajectories, but again, the orientation defined by the pairs was along the trajectory of motion.

The figures 2-5 correspond to a single frame from the four experimental conditions. The four conditions of the experiment differ by the relative contribution and the order of temporal and spatial integration. In the static conditions the observers needed to find the correspondence of the dots to a pair and to globally integrate this information in order to find the focus of the radial pattern. In the flicker condition on every frame the observers had to integrate the spatial information from the pairs of dots but they could benefit from temporal integration of the sequential patterns that would be equivalent to the presence of a larger number of dot pairs. In the motion conditions the observers had to temporally integrate the displacement of dots in the sequential frames in order to determine their trajectory of motion and to integrate this information in space to determine the focus of the radial pattern. In the combined condition the observers had redundant information as both the trajectory of dot motion and the orientation of the dot pairs provided similar information.

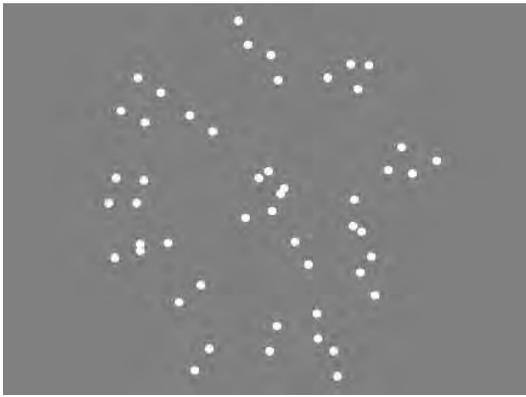


Figure 2. Static Condition.

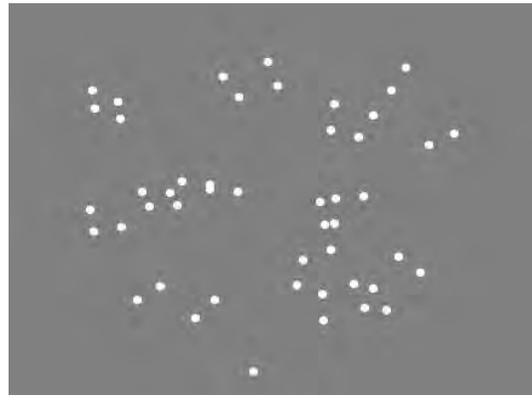


Figure 4. Flicker Condition.

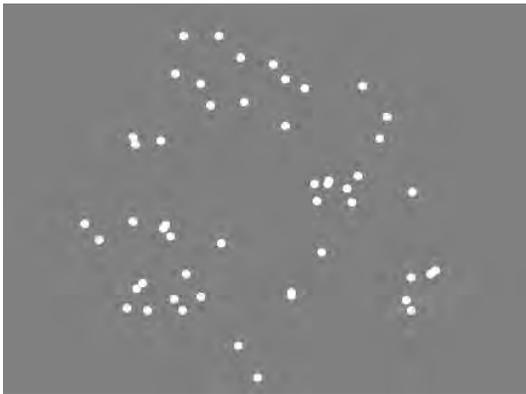


Figure 3. Motion Condition.

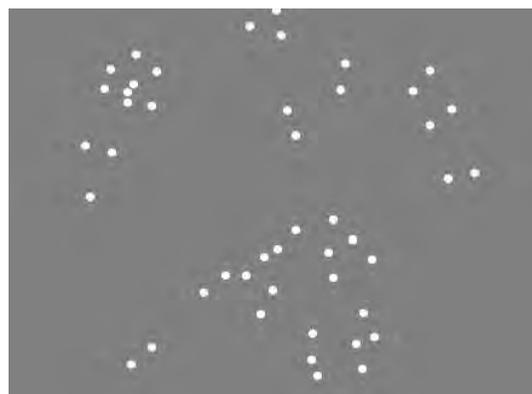


Figure 5. Combined Condition.

### C. Experimental Procedure

The subject sat at 57 cm from the monitor screen. The stimuli were presented on a gray screen with mean luminance  $50 \text{ cd/m}^2$ . Each stimulus presentation was preceded by a warning signal. A red fixation point with size of 0.8 degrees of arc appeared in the center of the screen for 500 msec. The stimuli were presented simultaneously with the disappearance of the fixation point. The Subjects performed a single-stimulus two-alternative force choice task. They had to continue looking at the position where the fixation point was presented until making a decision where the center of the pattern was (left or right relative to the fixation point). At this moment the subject had to move his/her eyes towards the position of the perceived center and to press the left or the right mouse button depending on whether the perceived center appeared to the left or to the right from the fixation point. If the subject could not make a decision during the 3.3 sec of the stimulus presentation (100 frames), the stimulus disappeared and the screen remained gray until the subject made a response.

### D. Method

The method of constant stimuli was used. Each condition was presented in a separate block consisting of 10 presentations for each position of the pattern center (a total of 140 presentations, 7 positions for a center shifted to the left and

7 positions for a center shifted to the right). The order of stimulus presentation was random. Each Subject took part in at least two experiments with 4 blocks for each of the 4 experimental conditions. All conditions were presented in a random order in a single day. The duration of each block depended on the subject performance, but the experiment did not exceed 1 hour. The eye movements of the subjects were registered with Jazz-novo multisensor measurement system (Ober Consulting Sp. z o.o) [20].

### E. Apparatus

The stimuli were presented on a 20.1 inch NEC MultiSync LCD monitor with NvidiaQuadro 900XGL graphic board at a refresh rate of 60 Hz and screen resolution 1280/1024 pixels. The experiments were controlled by a custom program developed under Visual C++ and OpenGL.

### F. Subjects

The subjects participating in the experiments are divided in three age groups: young (aged from 20 to 34 years), middle (aged 35 to 55 years) and elderly (aged 57 to 84 years). They did not have a whole training session, but they were given examples of stimuli to check whether they understood the task and to get an idea of the stimuli in a given condition.

#### IV. PERFORMANCE EVALUATION OF AGE-RELATED OBSERVERS GROUPS

The experimental goal of our study is directed to characterize the human heading perception influenced by: (i) form information only (ii) motion information only (iii) flicker information, i.e temporal integration of form information (iv) combined form and motion information. The question is if people rely and base their responses on a single source of information, or on combined one, and also which type of information utilized is more informative in the decision process. The participants belong to three age groups: Young, Middle aged, and Old. Hence, also the influence of human age on the assessment of heading perception will be evaluated. The evaluation is made on the base of experimental psychometric functions, obtained for all different experimental conditions and for each subject in all age-contingent groups. The psychometric function reflects the dependence between a given physical quantity (in this case, the pattern shift from the middle of the screen) and the proportion of subjects' responses of a given type, in our case "the proportion of responses "Center to the right".

- **Evaluation of heading perception in Young observers group**

The comparison of the performance in the static, motion and flicker conditions show that in Young group only 2 out of 10 observers have best performance for the static condition, 4 observers effectively utilized the motion information showing best performance in this case, and 4 out of 10 observers show best performance in the flicker condition. For 4 out of 10 observers the null hypothesis of equal psychometric functions for both motion and flicker information could not be rejected, i.e they could be considered as equivalent. These results suggest that the young observers effectively integrate the available information in time. The contribution of the information available in each of these three conditions to the performance of the combined condition differs. Only 1 out of 10 subjects relies mainly on motion, 1 - on the information available in the flicker condition, while 7 out of 10 combined effectively the independent sources of information available in the static and motion condition. The performance of averaged (on the base of 10 subjects in the group) young subject is shown on Fig.6.

For the averaged young subject the psychometric curves associated with static, motion, and flicker information are not distant and the null hypothesis that they do not differ could not be rejected.

- **Evaluation of heading perception in Middle aged observers group**

In this age-related group only 1 out of 6 subjects shows

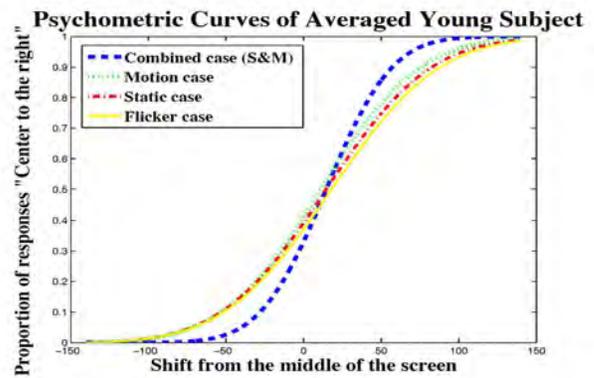


Figure 6. Psychometric Curves of Averaged Young Subject.

better performance in the static condition and 1 out of 6 observers - in the flicker condition. For 1 out of 6 observers the null hypotheses of equal psychometric functions for both motion and static information could not be rejected. For 4 out of 6 observers the null hypothesis for equal psychometric functions for motion and flicker conditions could not be rejected too. As general, the results suggest a small effect of the static information. The results for 4 out of 6 observers show that the results in the combined condition could be successfully predicted based on the performance of the static and motion conditions. The performance of averaged middle aged subject is shown on Fig.7.

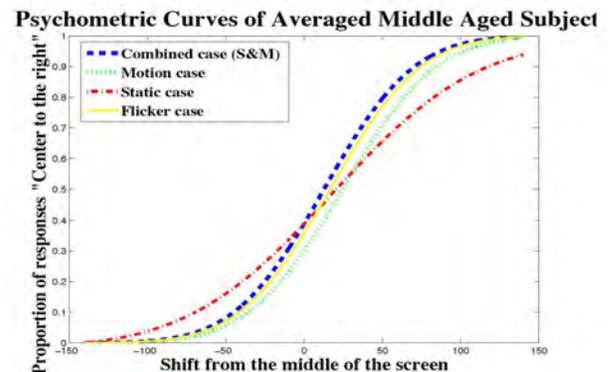


Figure 7. Psychometric Curves of Averaged Middle aged Subject.

The averaged middle aged observer does not rely mainly on the static information. For him the combined and flicker condition do not differ significantly.

- **Evaluation of heading perception in Old observers group**

The obtained results in Old-age group show that 3 out of 10 observers show best performance in the static, 3 out of 10 - in motion, and 4 out of 10 in the flicker condition. The null hypothesis for equal psychometric

functions is valid for: 1 out of 10 for motion and static condition, and for 2 out of 10 - for motion and flicker condition. Six out of 10 subjects utilize combined static and motion information to make their final decision in the combined condition. The performance of averaged old subject is shown on Fig.8. For averaged old subject the null hypotheses that the static and flicker cases do not differ is valid. The averaged old observer relies more on motion information.

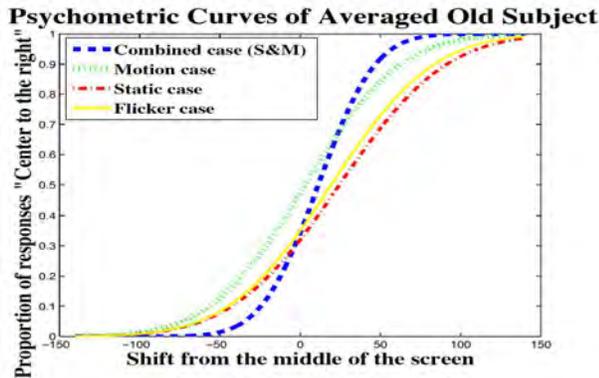


Figure 8. Psychometric Curves of Averaged Old Subject.

V. pPCR5 AND NCC RULES PERFORMANCE FOR PREDICTING HUMAN'S WAY OF FORM AND MOTION COMBINATION

The main question here is which fusion rule - pPCR5 or NCC used to combine available static and motion information predicts more adequately human cue integration? In order to answer this question we need to make a comparison between experimentally obtained and predicted (via pPCR5 and NCC rules) psychometric functions for combined condition (static and motion), for the three age contingent groups. This comparison is provided on the base of *goodness-of-fit test* [19], one important application of chi-squared criteria:  $\chi^2 = \sum_{j=1}^J \frac{(O_j - E_j)^2}{E_j}$  where  $\chi^2$  is an index of the agreement between an observed(*O*)/experimental and expected(*E*)/predicted via particular fusion rule sample values of psychometric function. For our case  $J = 14$  represents the number of pattern's shifts from the middle of the screen. The critical value of the test for  $\nu = J - 1 = 13$  degrees of freedom at assumed  $p = 0.1$  is  $\chi^2 = 19.81$  [19]. The respective results are given in Table I - for young group, in Table II - for middle aged group, and in Table III - for old persons' group.

In general, the results show that the pPCR5 fusion rule predicts more adequately than NCC rule human performance for the three age groups.

For young and for middle aged persons (Tables I and II) both fusion rules predict psychometric functions that do not differ significantly from the experimental ones, but the differences in the fits are smaller in case of pPCR5 rule than in case of NCC rule application. The same findings are valid for old people (Table III), but in this group NCC rule show

Table I  
CHI-SQUARED VALUES FOR YOUNG SUBJECTS.

Subject	(Form and Motion) pPCR5	(Form and Motion) NCC
1	0.8587	1.8482
2	0.4801	0.8456
3	0.3045	1.2690
4	0.1509	0.9716
5	0.1655	0.1458
6	0.3342	0.7013
7	0.0912	0.1810
8	0.5103	0.8381
9	0.1943	0.2090
10	0.0913	0.1494

Table II  
CHI-SQUARED VALUES FOR MIDDLE AGED SUBJECTS.

Subject	(Form and Motion) pPCR5	(Form and Motion) NCC
1	0.3698	0.9854
2	0.1856	0.4934
3	0.4192	0.9341
4	0.9872	1.4716
5	0.2380	1.0143
6	0.2425	0.8456

worse performance for subject no.4 (put in bold in Table III) showing the exceeded critical value of  $\chi^2 = 19.81$ . The reason for this result reflects the situations, when the experimentally obtained psychometric functions, associated with single static and single motion conditions are characterized with distant underlying Gaussian distributions. In this case pPCR5 makes prediction, which models more correctly and adequately human combination behavior. Using NCC rule however, part of available information has been neglected, because the cue with bigger mean was weighted more heavily than the cue with a smaller one (as it was described in Section II).

VI. COMMON TRENDS OF AGE RELATED OBSERVER GROUPS

The goal here is to find the common trend, concerning the performance of the three groups. In order to achieve it, we consider each group as a set of different sources of evidence, associated with each person in the group. That way young group consists of 10 (middle aged of 6, old aged of 10) sources (subjects) of evidence, which should be combined all together via pPCR5 and NCC fusion rules.

The combined individual behaviors in particular group are estimated, revealing its intrinsic behavior as a whole, reducing uncertainties associated with individual performances. All the tested subjects in age groups are considered as independent and equally reliable sources of information, because each subject provides his/her own psychometric function, associated with the static and motion condition and should be taken into account with equal weights to derive these trends.

Our goal is to find out which combinational rule (pPCR5 or NCC) is able to model correctly and adequately such human age-contingent group trends in the process of decision making. The results obtained for experimental and estimated (via the

Table III  
CHI-SQUARED VALUES FOR OLD SUBJECTS.

Subject	(Form and Motion) pPCR5	(Form and Motion) NCC
1	0.3751	0.7693
2	0.2762	0.5721
3	0.3691	0.4078
4	<b>2.9287</b>	<b>21.0845</b>
5	0.5418	0.8592
6	0.1652	0.3021
7	0.2013	0.3103
8	0.3984	0.5932
9	0.6712	1.6964
10	0.7152	1.8598

fusion rules) trends, concerning the cues combination groups' performance are presented in Figures 9, 10, and 11.

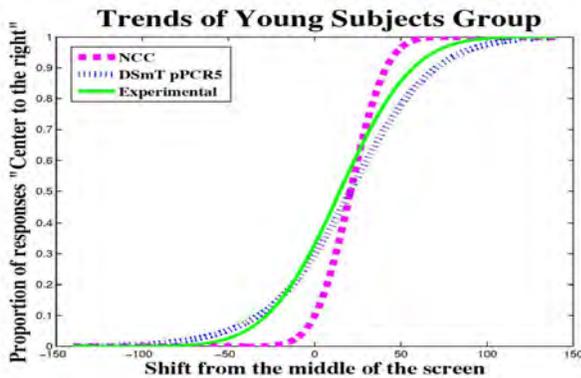


Figure 9. Trends of Young Subjects Group.

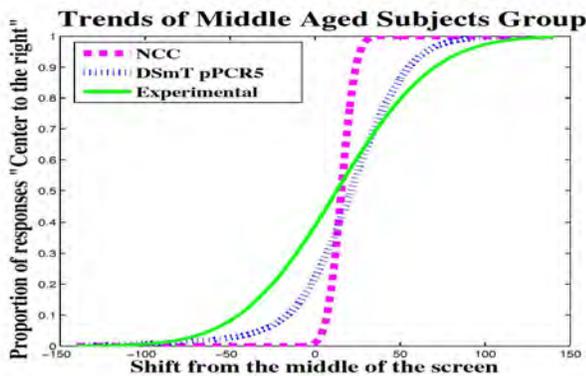


Figure 10. Trends of Middle aged Subjects Group.

In order to compare the performance of both fusion rules in estimating common trends' prediction the city-block errors between the corresponding triples (*young/middle/old* group experimental form and motion combination) - *young/middle/old* group estimated (via pPCR5 and NCC) form and motion combination are given in Table IV. Results show ultimately that experimentally obtained and those, based on pPCR5 fusion rule are closer and for the three age-contingent groups are more than two times less than those, obtained via NCC

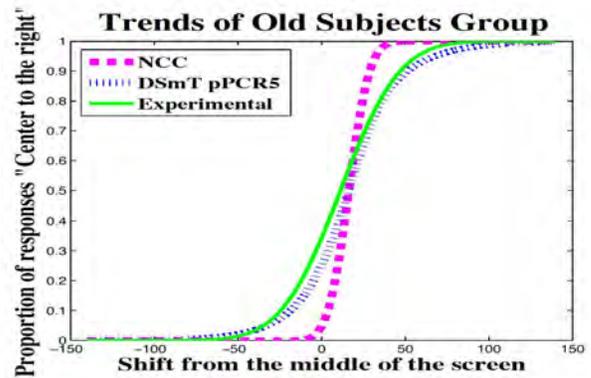


Figure 11. Trends of Old Subjects Group.

fusion rule. pPCR5 fusion rule predicts more correctly the human model of decision making, than NCC rule, utilizing all the available information (**Form and Motion**), even in case of conflict. NCC based trends are very sensitive to the sources (different subjects' psychometric functions) with the bigger means, neglecting that way part of the available information and acting as an amplifier of the information by reducing the variances.

Table IV  
CITY BLOCK ERRORS BETWEEN EXPERIMENTAL AND PREDICTED TRENDS.

	PCR5	NCC
<i>FM Young</i>	0.03	0.10
<i>FM Middle</i>	0.06	0.13
<i>FM Old</i>	0.04	0.12

## VII. CONCLUSIONS

This paper presented a study on human heading perception obtained on the base of motion and form visual cues integration. The influence of human age on this process was evaluated. The results obtained show age-related difference in the performance of the subjects in estimating the heading direction based on the combined static (form) and motion information.

Our experimentally obtained data for young observers suggest smaller effect of the static information case and provides indirect evidence that their performance is based more on the temporal integration of information in the motion and flicker conditions. The experimental results for Middle-aged group suggest less effect of the static information and an effect of the order of temporal and spatial integration. The old subjects used to rely more on the motion information. All age-related groups rely on combined (motion and form) information to take their final decisions for heading perception.

A comparison between experimentally obtained and predicted (via pPCR5 and NCC rules) psychometric functions for combined condition (static and motion), for the three age contingent groups was made and estimated on the base of *goodness-of-fit test*, one important application of chi-squared criteria. Results proved that pPCR5 makes prediction, which

models more correctly and adequately human combination behavior than NCC, especially in cases of conflicts between the different visual cues.

The combined individual behaviors (the trends) in particular age groups were estimated, revealing its intrinsic behavior as a whole, reducing uncertainties associated with individual observers' performance. Results show ultimately that pPCR5 fusion rule, utilizing all the available information - static (form) and motion, even in case of conflict, predicts more correctly the human model of decision making, than NCC rule. That way pPCR5 fusion rule assures preserving the richness of cues data in the process of visual stimuli combination and assures improvement of decision accuracy. pPCR5 describes better the characteristics of the different age groups in decision making based on the motion and form information in heading perception.

### VIII. ACKNOWLEDGEMENT

The reported work is partially supported by the project DN02/3/2016 'Modelling of voluntary saccadic eye movements during decision making' funded by the Bulgarian Science Fund.

### REFERENCES

- [1] Z. Kourtzi, B. Krekelberg, R.J.A. Van Wezel, *Linking form and motion in the primate brain*, Trends in Cognitive Sciences, Vol. 12, pp. 230–236, 2008.
- [2] S. Pardhan, *Contrast sensitivity loss with aging: sampling efficiency and equivalent noise at different spatial frequencies*, JOSA A, Vol. 21(2), pp. 169–175, 2004.
- [3] S. Dowiasch, S. Marx, W. Einhauser, F. Bremmer, *Effects of aging on eye movements in the real world*, Frontiers in Human Neuroscience, Vol 9, pp. 12, 2015.
- [4] M. Lich, F. Bremmer, *Self-Motion perception in the elderly*, Frontiers in Human Neuroscience, Vol. 8, pp. 681, 2014.
- [5] A. Pavan, G. Mather, *Distinct position assignment mechanisms revealed by cross-order motion*, Vision Research, Vol. 48, pp. 2260–2268, 2008.
- [6] H.B. Barlow, B.A. Olshausen, *Convergent evidence for the visual analysis of optic flow through anisotropic attenuation of high spatial frequencies*, Journal of Vision, Vol. 4(6), pp.1, 2004.
- [7] M. Edwards, M.F. Crane, *Motion streaks improve motion detection*, Vision Research, Vol. 47, pp. 828–833, 2007.
- [8] D. Jancke, *Orientation formed by a spots trajectory: A two-dimensional population approach in primary visual cortex*, Journal of Neuroscience, Vol. 20, pp. 1–6, 2000.
- [9] T.D. Albright, *Direction and orientation selectivity of neurons in visual area MT of the macaque*, Journal of Neurophysiology, Vol. 52, pp. 1106–1130, 1984.
- [10] T. Bayes, *An Essay towards solving a Problem in the Doctrine of Chances*, Philosophical Transactions of the Royal Society of London, pp. 330–418, 1763.
- [11] D. Sivia, *Data Analysis, a Bayesian Tutorial*, Clarendon (Oxford), 1996.
- [12] A.P. Dempster, *A generalization of Bayesian inference*, J. Roy. Stat. Soc., B30, 1968.
- [13] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [14] L.A. Zadeh, *Mathematical Theory of Evidence (book review)*, AI Magazine, Vol. 5(3), pp. 81–83, 1984.
- [15] F. Voorbraak, *On the justification of Dempster's rule of combination*, Utrecht Univ., Netherlands, Logic Group Preprint Series, No. 42, 1988.
- [16] P. Wang, *A defect in Dempster-Shafer theory*, in Proc. of 10th Conf. on Uncertainty in AI, pp. 560–566, 1994.
- [17] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., Vol. 29(3), pp. 223–252, March 2014.
- [18] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion - Collected works*, Volumes 1–4, American Research Press, 2004–2015.  
<https://www.onera.fr/staff/jean-dezert/references>
- [19] J. Matre, G. Gilbreath, *Statistics for Business and Economics*, 3rd Edition, 1987.
- [20] <http://www.ober-consulting.com/9/lang/1/>

# Fusion of Multispectral Imagery and DSMs for Building Change Detection Using Belief Functions and Reliabilities

Jiaojiao Tian<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Remote Sensing Technology Institute, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

Emails: jiaojiao.tian@dlr.de, jean.dezert@onera.fr

Originally published as: J. Tian, J. Dezert, *Fusion of Multispectral Imagery and DSMs for Building Change Detection Using Belief Functions and Reliabilities*, Int. Journal of Image and Data Fusion (IJIDF), Vol. 10, No.1, pp. 1–27, 2019, and reprinted with permission.

**Abstract**—The extraction of building changes from very high resolution satellite images is an important but challenging task in remote sensing. Digital surface models (DSMs) generated from stereo imagery have proved to be valuable additional data sources for this task. In order to efficiently use the change information from the DSMs and spectral images, belief functions have been introduced. In this article, two-step building change detection fusion models based on both Dempster-shafer theory (DST) and Dezert-Smarandache Theory (DSMT) frameworks are proposed. In the first step, basic belief assignments (BBAs) of the change indicators from images and DSMs are calculated by using a refined sigmoidal BBA model. Then these BBAs are employed for the new proposed building change detection decision fusion approach. In order to cover the miss-detections introduced by the wrong height values of the DSMs and incomplete information from images, disparity maps from the DSM generation procedure and shadow maps from the multispectral channels are adopted to generate reliability maps, which are further integrated to the fusion models. In the last step, building change masks are generated based on four decision-making criteria. In the experimental part of this work, we evaluate the performance of this new building change detection method on real satellite images thanks to a building change reference mask representing the ground truth. Substantial accuracy improvements are achieved when comparing the new results with those obtained from classical 3D change detection approaches.

**Keywords:** change detection, belief functions, DSMT, DST, DSM.

## I. INTRODUCTION

Efficient and accurate detection of building changes using remote sensing data is of great importance for urban monitoring and disaster monitoring. It is one of the fundamental tasks in remote sensing and is attracting more interests due to the high and accelerated rate of urban growing and more frequent natural disasters with climate changes.

In the last decades, 2D change detection methods on large scale land cover monitoring have been extensively studied and applied on satellite images [1], [2], [3]. There are many excellent approaches available which can extract landcover changes from multi-temporal images [4], [5]. However, highlighting only building changes in urban area remains difficult due to the mixture of other background changes, for instance the changes introduced by different illumination conditions or

human activities. The influence of these changes is growing as higher resolution images show more details of the landcover objects. In addition, even with very high resolution data it is sometimes impossible to distinguish buildings and roads using simple spectral change.

Therefore, height information derived from Digital Surface Model (DSM) is posing new possibilities for building change detection. Benefiting from improved data quality and advanced computer vision techniques, the accuracy of the DSMs from satellite stereo imagery has been largely improved and enables building change detection in a larger region and with high frequency. However, the DSMs may exhibit some inaccurate height values resulting from failed matching and occlusions within the stereo and multiple views. Thus the fusing of changes from multispectral image and DSMs would be an effective solution for building change detection. The comparison of DSMs can locate the changes of high-level objects efficiently and robustly and the spectral images have rich spectral and texture features which can highlight more changes among the multi-temporal datasets. On the other hand, as the DSMs have been generated from the multispectral data, there is no time difference between them. The 2D and 3D information can be combined through post-refinement, region-based approaches or decision fusion [6]. In more recent researches, DSMs from multi-sensors and time-series data were involved [7], [8].

Regarding to feature fusion due to the diverse building characteristics and background information, the urban building monitoring approaches may perform variedly for different test regions. Thus recently some researches are trying to combine different change features and change classification methods, and fuse the results with a decision model. For instance, [9] have proposed a probabilistic framework to fuse the results from four local feature vectors for building detection. Based on an adaptive network-based fuzzy inference system, [10] have fused the change detection results from different feature combinations. Besides fusing the detection result, decision fusion can be also directly used for classification and change detection [11]–[15].

Thus until now there is no decision fusion model that

directly takes the change indices from images and height maps for building change detection. In our previous research works [16], belief functions have performed very well for 3D building change detection. As aforementioned, the accuracy of 2D change detection of specific objects is limited due to the misdetections caused by irrelevant changes. These irrelevant changes have a larger effect on very high resolution (VHR) images than on low and moderately high resolution images, since in VHR images their higher details are more sensitive to viewing and solar angle differences. DSMTs generated from satellite stereo imagery can largely help to solve this problem. Unfortunately, the fusion model proposed in [16] is rather basic, and it is not robust in dealing with high conflict situations. Therefore, the belief functions have been further investigated and improved in this article. Besides Dempster-Shafer Theory (DST) [17], [18], an extended Dezert-Smarandache Theory (DSMT) [19] will be adopted in this article to generate the building change detection models. One of the difficulties of using Dempster-Shafer theory is the definition of uncertainty and the calculation of the basic belief assignments (BBAs). [16] used one sigmoid function to distribute the values of one change feature to the BBAs ranging from 0 to 1. The symmetry point which indicates a certainty of 50% was automatically calculated with a thresholding method. However, the accuracy and robustness of the thresholding approach will directly influence the correctness of the obtained BBAs. Thus, as well as the fusion models, the BBAs construction approach should be updated to further improve the change detection result. These problems have been well addressed in our modified approach [20]. In addition, the uncertainty of change indicators was measured in order to improve the accuracy of BBAs. Due to space limitation constraint of conference paper format, the methodology part has been only shortly described in [20] and only small patches have been tested in the experimental part. A better description of this methodology with more experiments of our approach is presented in this article with the improvement of the reliability discounting approach.

Focusing on building change detection by fusing spectral and height information extracted from satellite stereo imagery, this article is organised as follows. First, the belief functions of DST and DSMT are briefly reviewed. Then, the building change models are proposed for these theoretical frameworks. The belief functions are used in both BBAs preparation and change detection procedure. Two sigmoid functions are simulated for each change feature to obtain the BBAs. In order to further improve the BBAs values reliability discounting techniques are presented. We use the unfilled disparity map and shadow maps to generate the reliability map of the changes from the height and 2D images, respectively. The reliability maps are then used in the fusion process to refine the initial BBAs. We generate four sets of global BBAs. With four decision criteria the final change detection masks can be generated. In the end, these refined fusion models are tested on four sets of real satellite images, and a comprehensive comparison is included to validate the new approaches.

## II. BELIEF FUNCTIONS, DST AND DSMT

### A. Basics of belief functions

The details of DST and DSMT have been presented by [18], [19] and [21]. Let  $\Theta$  be a frame of discernment of a problem under consideration.  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  consists of a list of  $N$  exhaustive and mutually exclusive elements  $\theta_i$ ,  $i = 1, 2, \dots, N$ . Each  $\theta_i$  represents a possible state related to the problem we want to solve. The assumption of exhaustivity and mutual exclusivity of elements of  $\Theta$  is classically referred as *Shafer's model* of the frame  $\Theta$ . A BBA also called a belief mass function (or just a mass for short), is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  from the power set<sup>1</sup> of  $\Theta$  (denoted  $2^\Theta$ ) to  $[0, 1]$ , that verifies [18]:

$$m(\emptyset) = 0, \quad \text{and} \quad \sum_{X \in 2^\Theta} m(X) = 1. \quad (1)$$

$m(X)$  represents the mass of belief exactly committed to  $X$ . An element  $X \in 2^\Theta$  is called a focal element if and only if  $m(X) > 0$ . The belief and plausibility functions based on DST theory are defined respectively as:

$$Bel(A) = \sum_{B \in 2^\Theta, B \subseteq A} m(B), \quad (2)$$

$$Pl(A) = \sum_{B \in 2^\Theta, B \cap A \neq \emptyset} m(B). \quad (3)$$

In DST, the combination (fusion) of several independent sources of evidences is done with Dempster-Shafer<sup>2</sup> (DS) rule, assuming that the sources are not in total conflict<sup>3</sup>. DS combination of two independent BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$ , denoted symbolically by  $DS(m_1, m_2)$ , is defined by  $m^{DS}(\emptyset) = 0$ , and for all  $X \in 2^\Theta \setminus \{\emptyset\}$  by:

$$m^{DS}(X) = \frac{1}{1 - K^{DS}} \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2), \quad (4)$$

where the total degree of conflict  $K^{DS}$  is given by

$$K^{DS} \triangleq \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2). \quad (5)$$

A discussion on the validity of DS rule and its incompatibility with Bayes fusion rule for combining Bayesian BBAs can be found in the literature [21], [22], [23]. To circumvent the problems of DS rule, Smarandache and Dezert ([19], Vol. 2, Chap. 1), then Martin and Osswald ([19], Vol. 2, Chap. 2) have developed in DSMT [19] two fusion rules called PCR5 and PCR6 based on the proportional conflict redistribution (PCR) principle which consists

<sup>1</sup>The power set is the set of all subsets of  $\Theta$ , empty set included.

<sup>2</sup>Although the rule has been proposed originally by Dempster, we call it Dempster-Shafer rule as it has been widely promoted by Shafer in DST [18].

<sup>3</sup>otherwise DS rule is mathematically not defined because of 0/0 indeterminacy.

- 1) apply the conjunctive rule
- 2) calculate the total or partial conflicting masses
- 3) then redistribute the (total or partial) conflicting mass proportionally on non-empty sets involved in the conflict according to the integrity constraints one has for the frame  $\Theta$ .

This PCR principle transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is not degraded. Because the proportional transfer can be done in different ways, this has yielded to several different fusion rules. It has been proved by [24] that only PCR6 rule is compatible with frequentist probability estimation, and that is why we recommend its use in the applications. PCR5 and PCR6 rules simplify greatly and their formulas coincide for the combination of two sources. In this case, the PCR6 combination is obtained by taking  $m^{PCR6}(\emptyset) = 0$ , and for all  $X \neq \emptyset$  in  $2^\Theta$  by

$$m^{PCR6}(X) = \sum_{\substack{x_1, x_2 \in 2^\Theta \\ x_1 \cap x_2 = X}} m_1(x_1)m_2(x_2) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right], \quad (6)$$

where all denominators in Eq. (6) are different from zero. If a denominator is zero, that fraction is discarded.

If a denominator, e.g.,  $m_1(X) + m_2(Y)$  tends towards 0, then also the conflicting mass  $m_1(X)m_2(Y)$  that is transferable tends to zero because  $m_1(X)$  and  $m_2(Y)$  tend to zero (since they are positive); therefore, the redistribution of masses also tends to zero. That reflects the continuity of PCR6.

### B. Reliability discounting

The reliability discounting has been described and discussed in the references [25], [26]. Briefly, if an additional knowledge about the reliability ( $\alpha$ ) of certain source of evidence is available, it can be adopted to refine the initial BBAs. For instance the height change and image change indicators may not perform well under some situations. This situation can be measured, and used as reliability factors. Each factor  $\alpha$  would be a value ranging from 0 to 1. And  $\alpha = 1$  means fully reliable, while  $\alpha = 0$  means the indicator is totally unreliable. And all the remaining discounted mass are transferred to the full ignorance  $\Theta$ . Based on Shafer's discounting model [18], the reliability discounting factor  $\alpha$  is introduced to discount any BBA  $m(\cdot)$  defined on the power set  $2^\Theta$  as follows  $\forall X \in 2^\Theta$ :

$$\begin{cases} m_\alpha(X) = \alpha \cdot m(X), \text{ for } X \neq \Theta, \\ m_\alpha(\Theta) = \alpha \cdot m(\Theta) + (1 - \alpha). \end{cases} \quad (7)$$

## III. BUILDING CHANGE DETECTION FUSION MODEL

### A. Choice of the frame of discernment

Focusing on change detection, as a data preparation step, DSMs are calculated from satellite stereo imagery based on

semi-global matching approach [27], [28]. It follows two main steps. First, the epipolar image pair is generated through a pyramidal local least squares matching. Then the matching is cast into dynamic programming to minimise the cost function. We use census feature to measure the similarity between two pixels [27]. The challenges and opportunities of the DSMs assisted building change detection have been well described in [16]. The geo-information is employed to co-register these data, which enables a sub-pixel accuracy. Focusing on building change detection, two change indicators, one from images and one from DSMs are extracted. Changes from spectral images are highlighted by using the Iteratively Reweighted Multivariate Alteration Detection (IRMAD) [5]. Consequently, height changes from DSMs are shown after robust height difference [29], [16]. We suppose that new, demolished or rebuilt buildings may exhibit both height and spectral changes. But the spectral changes can also be introduced by seasonal changes and other irrelevant changes. After excluding building changes, changed pixels exclude building regions are named here as *OtherChange*. Therefore, three classes are considered to define the frame of discernment satisfying Shafer's model (i.e. the elements of the frame of discernment are disjoint):

$$\begin{aligned} \Theta = \{ & \theta_1 \triangleq \text{Pixel} \in \text{BuildingChange}, \\ & \theta_2 \triangleq \text{Pixel} \in \text{OtherChange}, \\ & \theta_3 \triangleq \text{Pixel} \in \text{NoChange} \}, \end{aligned} \quad (8)$$

and

$$\theta_1 \cap \theta_2 \cap \theta_3 = \emptyset. \quad (9)$$

In image domain, each pixel represents a single sample, thus in Eq. (8), we have directly used the word 'Pixel'. Based on the three exclusive classes, the set of potential focal elements *FE* that enter in our application is:

$$FE = \{\theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3\}. \quad (10)$$

It is worth noting that even if we work with Shafer's model of the frame of discernment for this application (which is the basis of DST), we can also use PCR6 rule developed in DSMT because PCR6 works also with Shafer's model as shown in [19].

The whole procedure of the proposed building change detection model is shown in Fig. 1. After the changes from DSMs and images are extracted, they will be reprojected using the sigmoid function to calculate the concordance index  $a$  and discordance index  $b$ . Then the decision fusion rules will be performed to generate the BBAs for height change and image change, respectively. After that, global BBAs can be calculated by using both DST and DSMT fusion rules. Finally, change mask can be obtained with various decision-making criteria.

### B. BBAs construction for building change detection

In [30] a sigmoidal model for both concordance and discordance indexes has been briefly presented. The details and advantages of this approach are described in [31]. The concordance index measures the concordance of change indicator and BBA in the assertion, while the discordance measures the

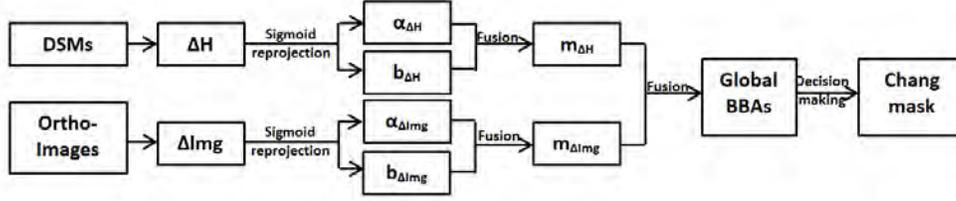


Fig. 1. Workflow of the proposed method.

opposition of change indicator to the BBAs in the assertion. In our previous works [16], the BBAs were built based on sigmoid curves related with the concordance index only. As explained in [16], the original sigmoid curve is defined as

$$f_{(\tau, T)}(x) = 0.99 / (1 + e^{-\frac{x-T}{\tau}}), \quad (11)$$

where  $x$  is the original value of each indicator ( $\Delta H, \Delta Img$ ), where  $\Delta H$  means the change in the height and  $\Delta Img$  means the change between two spectral images at a given pixel location. Two parameters  $T$  and  $\tau$  are used to control the symmetry point and the slope of the sigmoid function. The symmetry point indicates a certainty of 50%. In this article, we improve our model to construct the BBAs thanks to sigmoidal models for both concordance and discordance indexes following the idea proposed by [31]. The concordance index is similar as the indicator of our previous research. The green line in Fig. 2 shows an example of the concordance index from height changes. A higher height change indicator leads to a higher probability to be building change. The discordance index is defined as an indication for the opposite argument. The discordance index in Fig. 2 is shown in red color, which means that a higher height change reflects a lower probability to be *not building change*. The blue curve shows the conflict between the concordance and the discordance index. Both concordance index and discordance index are projected to the sigmoid curve distribution characterised by parameters  $T$  and  $\tau$ .

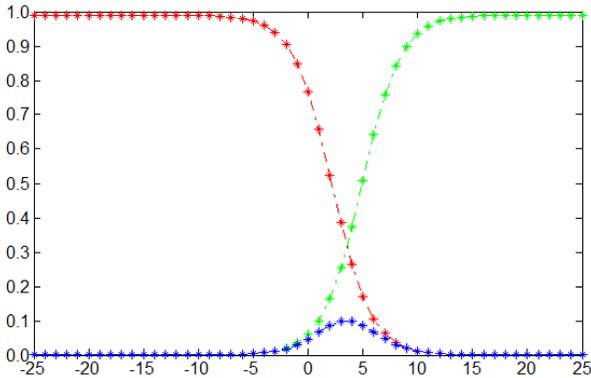


Fig. 2. Concordance and discordance index.

In [31] these two parameters  $T$  and  $\tau$  were manually selected. Here, as an improvement the multi-level Otsu's thresh-

olding method [32], [33] is used for automatically getting the symmetry points for both concordance index and discordance index. Otsu's algorithm assumes that an image is composed of objects and background. A discriminant analysis is performed by minimising the intra-class variance. When three classes are of interest, two thresholds  $T_1$  and  $T_2$  are expected, and Otsu's method can be extended to

$$\sigma_{\omega}^2(T_1, T_2) = \omega_1 \sigma_1^2(T_1, T_2) + \omega_2 \sigma_2^2(T_1, T_2) + \omega_3 \sigma_3^2(T_1, T_2). \quad (12)$$

The weights  $\omega_i$  are the probabilities obtained from the image histogram that are separated by the thresholds  $T_1$  and  $T_2$ .  $\sigma_i$  is the standard deviation of the  $i$ -th class, for  $i = 1, 2, 3$ .  $T_1$  and  $T_2$  can be used as the symmetry points of discordance and concordance index, respectively. Thus, using the height change index as in the example, the BBAs for concordance and discordance height change index are functions of values  $a_{\Delta H}$  and  $b_{\Delta H}$  defined by

$$a_{\Delta H} = f_{\tau, T_1}(\Delta H), \quad \text{and} \quad b_{\Delta H} = f_{-\tau, T_2}(\Delta H). \quad (13)$$

The discordance index can be considered as a reflection of the concordance index along the mirror line. Therefore, they are sharing the same  $\tau$ . Here, the factor  $\tau$  is calculated with a sample value ( $\Delta H = 1, a_{\Delta H} = 0.1$ ), which means 1 m height change indicates 10% probability to be building changes. The BBAs for discordance and concordance image change index are built similarly. Differences appearing in 2D images give a concordance indication for all changes, which include the building changes or other changes ( $\theta_1 \cup \theta_2$ ). In this article, the changes from images are named  $\Delta Img$ .

In the Tables I and II, we present the two ways of construction of the BBAs from the sources of evidence based either on DS or on PCR6 rules of combination for the height change indicator (i.e. the first source of evidence) and the image change indicator (i.e. the second source of evidence). It has to be noted that  $\theta_1 \cup \theta_3$  is not mentioned in the fusion model, as they do not share similar characters within the used feature space. In Table I,  $m_1(\cdot)$  and  $m'_1(\cdot)$  represent the concordance and discordance BBAs from  $\Delta H$ , whereas in Table II  $m_2(\cdot)$  and  $m'_2(\cdot)$  represent the concordance and discordance BBAs from images.  $K_{\Delta H}$  is the total conflicting mass value between  $m_1(\cdot)$  and  $m'_1(\cdot)$ , and  $K_{\Delta Img}$  in Table II is the total conflicting mass value between  $m_2(\cdot)$  and  $m'_2(\cdot)$ ,

TABLE I  
BBA CONSTRUCTION FOR HEIGHT CHANGE INDICATOR  $\Delta H$ . [CONFLICT:  $K_{\Delta H} = a_{\Delta H}b_{\Delta H}$ ]

Focal Elem.	$m_1(\cdot)$	$m'_1(\cdot)$	$m_1^{DS}(\cdot)$	$m_1^{PCR6}(\cdot)$
$\theta_1$	$a_{\Delta H}$	0	$\frac{a_{\Delta H}(1-b_{\Delta H})}{1-K_{\Delta H}}$	$a_{\Delta H}(1-b_{\Delta H}) + \frac{a_{\Delta H}K_{\Delta H}}{a_{\Delta H}+b_{\Delta H}}$
$\theta_2$	0	0	0	0
$\theta_3$	0	0	0	0
$\theta_1 \cup \theta_2$	0	0	0	0
$\theta_2 \cup \theta_3$	0	$b_{\Delta H}$	$\frac{(1-a_{\Delta H})b_{\Delta H}}{1-K_{\Delta H}}$	$(1-a_{\Delta H})b_{\Delta H} + \frac{b_{\Delta H}K_{\Delta H}}{a_{\Delta H}+b_{\Delta H}}$
$\theta_1 \cup \theta_2 \cup \theta_3$	$1-a_{\Delta H}$	$1-b_{\Delta H}$	$\frac{(1-a_{\Delta H})(1-b_{\Delta H})}{1-K_{\Delta H}}$	$(1-a_{\Delta H})(1-b_{\Delta H})$

TABLE II  
BBA CONSTRUCTION FOR IMAGE CHANGE INDICATOR  $\Delta Img$ . [CONFLICT:  $K_{\Delta Img} = a_{\Delta Img}b_{\Delta Img}$ ]

Focal Elem.	$m_2(\cdot)$	$m'_2(\cdot)$	$m_2^{DS}(\cdot)$	$m_2^{PCR6}(\cdot)$
$\theta_1$	0	0	0	0
$\theta_2$	0	0	0	0
$\theta_3$	0	$b_{\Delta Img}$	$\frac{(1-a_{\Delta Img})b_{\Delta Img}}{1-K_{\Delta Img}}$	$(1-a_{\Delta Img})b_{\Delta Img} + \frac{b_{\Delta Img}K_{\Delta Img}}{a_{\Delta Img}+b_{\Delta Img}}$
$\theta_1 \cup \theta_2$	$a_{\Delta Img}$	0	$\frac{a_{\Delta Img}(1-b_{\Delta Img})}{1-K_{\Delta Img}}$	$a_{\Delta Img}(1-b_{\Delta Img}) + \frac{a_{\Delta Img}K_{\Delta Img}}{a_{\Delta Img}+b_{\Delta Img}}$
$\theta_2 \cup \theta_3$	0	0	0	0
$\theta_1 \cup \theta_2 \cup \theta_3$	$1-a_{\Delta Img}$	$1-b_{\Delta Img}$	$\frac{(1-a_{\Delta Img})(1-b_{\Delta Img})}{1-K_{\Delta Img}}$	$(1-a_{\Delta Img})(1-b_{\Delta Img})$

### C. Reliability discounting

In the DSM assisted building change detection, false alarms arise if wrong heights are presenting in the DSM for large regions [16]. And these wrong heights are mostly introduced not in the stereoscopic images matching procedure, but in the gaps filling step. In the last step of the DSM generation procedure, the height of un-matched pixels is interpolated using the height values of neighbourhood pixels. Normally a reliable height value can be achieved for small gaps. But when large gaps appear in the disparity map, for example, for a whole building roof, the height of that building can not be correctly interpolated. Thus, the percentage of available successfully matched pixels inside a predefined neighbourhood region can be used to generate the height reliability. Fig. 3 shows an example of the generated reliability map. Fig. 3a is the gap mask. The gaps region of the disparity map is represented with black colour. Pixels with proper elevation values are displayed with white colour. It can be observed, based on our approach that pixels in the centre of a gap get lower reliability factor values than pixels next to the gap boundary (see Fig. 3b).

In the building change detection procedure, the reliability map of two DSMs ( $\alpha_{DSM1}$  and  $\alpha_{DSM2}$ ) are calculated, respectively. They are then fused together to generate a final reliability map  $\alpha_{\Delta H}$  for the height change mass.

$$\alpha_{\Delta H} = \alpha_{DSM1} \cdot \alpha_{DSM2}. \quad (14)$$

Shadow has played an important role when analysing very

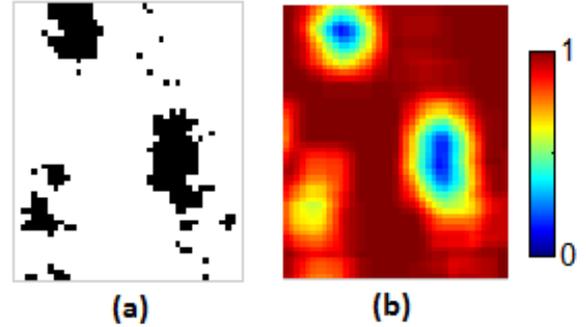


Fig. 3. Reliability map (b) generated from the gaps mask (a).

high resolution images in urban region. Both of the changes of shadow and coverage of shadow will bring false alarms for change detection. Therefore, the 2D changes that are detected in shadow regions are less reliable than in non-shadow regions. Benefit to this character, we can adopt the shadow map as the reliability map of BBA from the image change indicator. For this purpose, first a shadow map is generated by calculating the average brightness of the multi-spectral image, as normally a dark colour indicates the existence of shadows. We take an easy and fast shadow detection approach as shown in Eq. (15) to highlight the shadow class. It is a pixel-based approach, therefore,  $B_k$  in Eq. (15) represents the intensity values at one pixel location in different multi-spectral band images.

And  $n$  is the number of the multispectral bands. The detected shadow map from brightness is enough for our purpose. In this shadow map, a smaller value indicates higher probability to be shadows; thus, the 2D changes detected in these regions are less reliable.

$$\text{Brightness} = \frac{1}{n} \sum_{k=1}^n B_k. \quad (15)$$

A further process is proposed to obtain a valid reliability map from the shadow map. First, it has been projected to a sigmoid curve. The lower threshold value from the two-level Ostu threshold is used as the symmetry point of the sigmoid curve. The obtained probability map is denoted as *ShadowMap*. In order to control the influence of the *ShadowMap*, we have only kept the values less than 0.5.

$$\alpha_{img} = \begin{cases} 0.5 + I_{ShadowMap}, & \text{if } I_{ShadowMap} < 0.5, \\ 1, & \text{otherwise.} \end{cases} \quad (16)$$

where  $I_{ShadowMap}$  is the pixel intensity of the shadow map in  $[0, 1]$ . The reliability map generated from the shadow map is then recorded as  $\alpha_{\Delta Img}$ , and it is the combination of the shadow maps of two dates.

$$\alpha_{\Delta Img} = \alpha_{img1} \cdot \alpha_{img2}. \quad (17)$$

#### D. Global BBAs

The BBAs related with the concordance and discordance indexes are combined to get the global BBA regarding each source of evidence. These global BBAs will then be used as input for solving the change detection problem thanks to their combination. From the previous step of BBAs modelling, each pixel will get two sets of BBAs to combine results from Table I and II. More precisely, we will have to combine either  $\{m_1^{DS}(\cdot), m_2^{DS}(\cdot)\}$  if DS rule is preferred for the BBA modeling, or  $\{m_1^{PCR6}(\cdot), m_2^{PCR6}(\cdot)\}$  if the PCR6 rule is adopted. These BBAs from Table I and II are represented by  $a_1, a_2, a_3$  and  $b_1, b_2, b_3$ . In this article, the mass values  $a_1, a_2$  and  $a_3$  are further discounted by the generated reliability map  $\alpha_{\Delta H}$  and denoted respectively as  $A_1, A_2$  and  $A_3$ . The mass values from the image change indicator  $b_1, b_2$  and  $b_3$  are discounted by the vegetation and shadow indicators  $\alpha_{\Delta Img}$  obtained in formula Eq. (17) to  $B_1, B_2$  and  $B_3$ .

More precisely, one computes

$$\begin{cases} A_1 = \alpha_{\Delta H} \cdot a_1, \\ A_2 = \alpha_{\Delta H} \cdot a_2, \\ A_3 = \alpha_{\Delta H} \cdot a_3 + (1 - \alpha_{\Delta H}). \end{cases} \quad (18)$$

$$\begin{cases} B_1 = \alpha_{\Delta Img} \cdot b_1, \\ B_2 = \alpha_{\Delta Img} \cdot b_2, \\ B_3 = \alpha_{\Delta Img} \cdot b_3 + (1 - \alpha_{\Delta Img}). \end{cases} \quad (19)$$

Table III and Table IV describe the final building change detection models based either on DS or on PCR6 rules. Here, the discounted height change indicator is denoted as  $m_{1\alpha_{\Delta H}}(\cdot)$ ,

and the discounted image change indicator is denoted as  $m_{2\alpha_{\Delta Img}}(\cdot)$ .

TABLE III  
DS FUSION MODEL FOR BUILDING CHANGE DETECTION.

Focal Elem.	$m_{1\alpha_{\Delta H}}(\cdot)$	$m_{2\alpha_{\Delta Img}}(\cdot)$	$m_{12}^{DS}(\cdot)$
$\theta_1$	$A_1$	0	$\frac{A_1(B_1+B_3)}{1-A_1B_2}$
$\theta_2$	0	0	$\frac{A_2B_1}{1-A_1B_2}$
$\theta_3$	0	$B_2$	$\frac{(A_2+A_3)B_2}{1-A_1B_2}$
$\theta_1 \cup \theta_2$	0	$B_1$	$\frac{A_3B_1}{1-A_1B_2}$
$\theta_2 \cup \theta_3$	$A_2$	0	$\frac{A_2B_3}{1-A_1B_2}$
$\Theta$	$A_3$	$B_3$	$\frac{A_3B_3}{1-A_1B_2}$

TABLE IV  
PCR6 FUSION MODEL FOR BUILDING CHANGE DETECTION.

Foc. Elem.	$m_{1\alpha_{\Delta H}}(\cdot)$	$m_{2\alpha_{\Delta Img}}(\cdot)$	$m_{12}^{PCR6}(\cdot)$
$\theta_1$	$A_1$	0	$A_1(B_1 + B_3) + \frac{A_1A_1B_2}{A_1+B_2}$
$\theta_2$	0	0	$A_2B_1$
$\theta_3$	0	$B_2$	$(A_2 + A_3)B_2 + \frac{B_2A_1B_2}{A_1+B_2}$
$\theta_1 \cup \theta_2$	0	$B_1$	$A_3B_1$
$\theta_2 \cup \theta_3$	$A_2$	0	$A_2B_3$
$\Theta$	$A_3$	$B_3$	$A_3B_3$

$m_{1\alpha_{\Delta H}}(\cdot)$  can be obtained from the discounting of the fusion results presented in Table I. Thus they have been denoted respectively as  $m_{1\alpha_{\Delta H}}^{DS}(\cdot)$  and  $m_{1\alpha_{\Delta H}}^{PCR6}(\cdot)$ . These discounted height change indicators are fused in the second step with the image change indicator  $m_{2\alpha_{\Delta Img}}(\cdot)$  to generate the final global BBAs. From the Tables III and IV, four sets of global BBAs can be computed based on different BBAs and fusion models. The flow diagram in Fig. 4 summarises the different fusion schemes tested in our application.

As one sees, if both the BBA modelling procedure and global BBAs are constructed based on DS fusion rule, the generated global BBA is recorded as  $G_1$ . If the global BBAs are constructed based on PCR6 fusion rule, they are recorded as  $G_2$ . The basic BBAs can also be calculated with PCR6 fusion rule, as shown in Table II. Based on these BBAs, the global BBAs can be also constructed using DS theory  $G_3$  and PCR6 rule  $G_4$ . It has to be mentioned that these four fusion schemes have different computational cost and  $G_1$  is the simplest one and  $G_4$  is the most expensive one in terms of computational burden.

#### E. Change mask generation

The final building change mask is our decision-making procedure. After the second step of fusion, each pixel in the images will get a certain degree of belief for all focal elements. The value of global BBAs in  $\theta_1$  gives a direct building change probability map. A decision criterion is required in generating

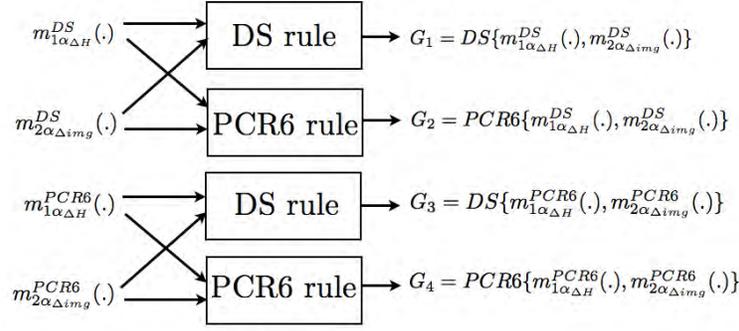


Fig. 4. Four fusion schemes based DS and PCR6 rules.

the final building change detection masks. A change mask can be generated after giving a threshold value [16]. However, BBAs on the partial ignorance and full ignorance set should also be considered in the decision-making procedure. The building change probability map is only a part of the global BBAs. DST and DSmt propose different approaches to make the final decision. Several decision criteria are available. In this article, four decision criteria are tested. They are: 1) maximum of belief (Max\_Bel), 2) maximum of plausibility (Max\_Pl), 3) maximum of betting probabilities (Max\_BetP) and 4) the maximum of DSmtP (Max\_DSmtP) [18], [19](Vol. 3, Chap. 3).

1) *Maximum of Belief (Max\_Bel)*: Valid for different strategies of BBA modelings and fusions according to Fig. 4. More precisely, for a strategy  $G$  generating a combined mass  $\in G_1, G_2, G_3, G_4$ , the label (decision) is obtained by comparing the final global mass values obtained from Table. III and IV.

$$Label = \operatorname{argmax}\{G(\theta_1), G(\theta_2), G(\theta_3)\}. \quad (20)$$

2) *Maximum of plausibility (Max\_Pl)*: Plausibility is defined in Eq. (3). Max\_Pl compares the plausibility of each class.

$$Label = \operatorname{argmax}\{Pl(\theta_1), Pl(\theta_2), Pl(\theta_3)\}. \quad (21)$$

3) *Maximum of betting probabilities (Max\_BetP)*: The pignistic probabilities, denoted as  $BetP$ , is making decisions on the pignistic level. In the betting probabilities, global masses of joint focal elements are averagely redistributed to each class.

$$BetP(A) = \sum_{B \in \Theta} \frac{|A \cap B|}{|B|} m(B), A \in \Theta. \quad (22)$$

$$Label = \operatorname{argmax}\{BetP(\theta_1), BetP(\theta_2), BetP(\theta_3)\}. \quad (23)$$

4) *Maximum of DSmtP (Max\_DSmtP)*: DSmtP probabilistic transformation is an important alternative to the pignistic transformation [34]. The basic idea of DSmtP is to redistribute the mass of (partial and total) ignorances proportionally to the masses of singletons involved in the ignorances.

$$DSmtP_\varepsilon(A) = \sum_{B \in \Theta} \frac{\sum_{\substack{Z \subset A \cap B \\ |Z|=1}} m(Z) + \varepsilon |A \cap B|}{\sum_{\substack{Z \subset B \\ |Z|=1}} m(Z) + \varepsilon |B|} m(B). \quad (24)$$

where  $\varepsilon \geq 0$  is a small positive number (typically 0.001) that avoids numerical indeterminacies in very degenerated cases occurring if the mass in the denominator of Eq. (24) is zero. More detailed information about DSmtP is given in [34]–[35].

$$Label = \operatorname{argmax}\{DSmtP(\theta_1), DSmtP(\theta_2), DSmtP(\theta_3)\}. \quad (25)$$

Among the four decision-making rules, max of belief or max of plausibility have the advantage to be very simple to calculate but they represent respectively two extreme pessimistic or optimistic decisional attitudes. The choice of one of these extreme attitudes depends on the consequence of decision error we are ready to take which is conditioned by the type of application under concern. Moreover, it has been shown by [34] that the more sophisticated transformation DSmtP outperform BetP transformation at a price of much higher computational complexity, which can be a bottleneck in some real-time image processing applications.

## IV. EXPERIMENTS

### A. Datasets

The belief function-based building change detection models have been tested on four pairs of satellite images. Each of the first three experimental datasets consist of two pairs of IKONOS stereo imagery captured in February 2006 and May 2011 over an industrial region in Dong-an, North Korea. These three sub-test regions are shown in Fig. 5 and 6 and 7, respectively. The original IKONOS stereo imagery has 1 m pixel size in the panchromatic band and 4 m pixel size in the multispectral bands. The fourth experimental dataset (shown in Fig. 8) was captured over the centre of Munich, Germany, which is a typical European urban region. The two pairs of stereo data of this dataset were captured by IKONOS on July 15, 2005 and WorldView-2 on July 12, 2010, respectively. In Fig. 5 to Fig. 8, the first two images are the panchromatic images of before- and after-change. (c) and (d) are the generated DSMs. They have been generated based on the method explained by [27]. The elevation values from low to high are represented with the colours from dark blue to dark red as described in the colour bar. These images are co-registered through camera model parameter corrections before the DSM

generation procedure with block adjustment among all datasets [27], [36]. A sub-pixel accuracy in planimetry and 1 to 2 m in height can be achieved. The Gram-Schmidt pan-sharpening method which has been widely used and implemented in ENVI software is applied to the multispectral channels of all three test regions [37]. In the first three subsets the generated DSMs have been re-sampled to 1 m resolution. As the IKONOS and WorldView-2 data for the Munich test region have different resolutions, the IKONOS images are up sampled to 0.5 m resolution, to be equal to WorldView-2 data. Instead of down-scale [38], an up-scale re-sampling is selected here to keep the sharp boundaries in the WorldView-2 data. The resulting DSMs also have a resolution of 0.5 m.

Fig. 5 and Fig. 6 show normal building change examples with DSMs in high accuracy. The size of these two test regions are  $450 \times 700 \text{ m}^2$ , and  $1000 \times 400 \text{ m}^2$ , respectively. In Fig. 5 some seasonal changes are visible. The generated DSMs are displayed in Fig. 5c and 5d. The second test region (Fig. 6) shows much larger sized buildings, and these buildings are well separated from each other.

The third test region consists of two images with the size of  $160 \times 340$  pixels. This region is characterised by small sized buildings (Fig. 7). It has to be mentioned, the largest building with a dark colour roof does not have the correct height in the first DSM, as is shown in Fig. 7c. This test region is especially selected to prove the robustness of our fusion models. The image size of the fourth test region is  $1600 \times 1600$  pixels, which is  $640,000 \text{ m}^2$ . It has mainly large size buildings with complex roof shapes. From 2005 to 2010, besides newly constructed buildings, there are also rebuilt/demolished buildings. Especially, many roofs have been renovated with another material. Without height information, it is very difficult to separate the newly constructed buildings from other kinds of changes.

## B. Results

The proposed DS fusion model and PCR6 fusion model have been applied to all datasets. In the first step, the four sets of global BBAs for all three focal elements and joint elements are generated based on various fusion rules and fusion rule combinations. In the second step, building change masks are generated by using four decision criteria. All three classes including *BuildingChange*, *OtherChange* and *NoChange* are generated. But this article focuses on the newly constructed buildings, thus only the *BuildingChange* results are analysed and evaluated. The proposed models have two novel properties. The first one is the improved fusion model, and the second one is the reliability discounting. In the experimental part, the minimal value of the reliability map generated from DSM gaps is manually modified to 0.1 to remove too small values. In the height change reliability map generation procedure, a window size of  $9 \times 9$  is selected.

To prove the advantages of the proposed method, firstly the best building change detection results are displayed together with the original height change map. The results of all four test regions are displayed in Fig. 9, Fig. 10, Fig. 11 and Fig.

12, respectively. In each figure, different colours represent different height changes in Figs. 9-12(a). Figs. 9-12(b) are the generated building change masks. To show the quality of these building change masks, these masks have been overlaid with the change reference data, which have been manually extracted for all four test regions. In Figs. 9-12(b) the green colour represents the correctly detected building changes. The false alarms which indicate pixels that are wrongly detected as building changes are presented with red colours. The blue colour objects are the misdetracted changed buildings, which are named as *false negatives* in this article.

Generally speaking, the proposed models are able to extract the newly constructed buildings in high accuracy. Noise effects from the height change map have been largely reduced in the final change results. The four selected test regions present four different situations. In the first test region most of the buildings are relatively low in the height and well separated from each other. The second test region has much higher and larger buildings, which produce large regions of shadow. The third test region is a special case. As we observe, in the first DSM of test region 3 the height of one big building is not correctly extracted. Actually the same building has been detected as false alarm and been discussed in the reference [16]. It has been explained in [16], due to the large region size and height change values, the false alarm can not be avoided. The fourth test region is much more complicated than the others, exhibiting very high building density, complex roof shapes and various building change types.

Benefiting from the improved fusion models and the reliability discounting procedure, some false alarms can be successfully avoided. Especially for the building in the left-bottom corner in the test region 1 and that big building in test region 3. In both situations, the first DSM is not able to get the correct height values. Based on the traditional feature fusion approach or our initial fusion model [16], this kind of buildings will very possibly be detected as *BuildingChange*. However, as we observe in the presented change detection results, these buildings are correctly detected as *NoChange*. It has to be noted that vegetation change is not considered in this model. Thus, in the centre of the first region, these two large regions of false alarms, which are newly planted trees from visual interpretation, are not able to be avoided. Another difficult to detect region is one building in construction. Half of the building has been finished in the after-change data; thus, this region has both height and spectral changes. As it can not be called a finished building yet, we did not include it as *BuildingChange* in our reference data.

Many false negatives (blue regions/pixels) in Fig. 10(b) are visible. Most of these false negatives can be explained by the quality of the DSMs. A subset of the gaps mask of test region 2 in date2 is displayed in Fig. 13. As it shows all of the four missed buildings (shown in blue colour) are actually gaps in the unfilled DSM. After gaps filling, they are not interpreted with correct height values, as shown in Fig. 6 (d). Thus, these four buildings only feature spectral changes, therefore are falsely identified as *OtherChange*.

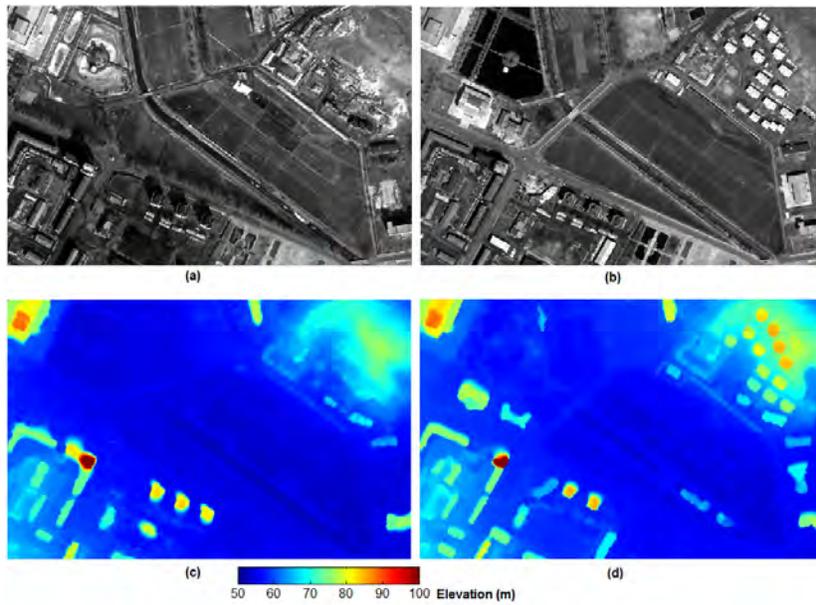


Fig. 5. Datasets of the test region 1: a) panchromatic image from date1; b) panchromatic image from date2; c) DSM from date1; (d) DSM from date2.

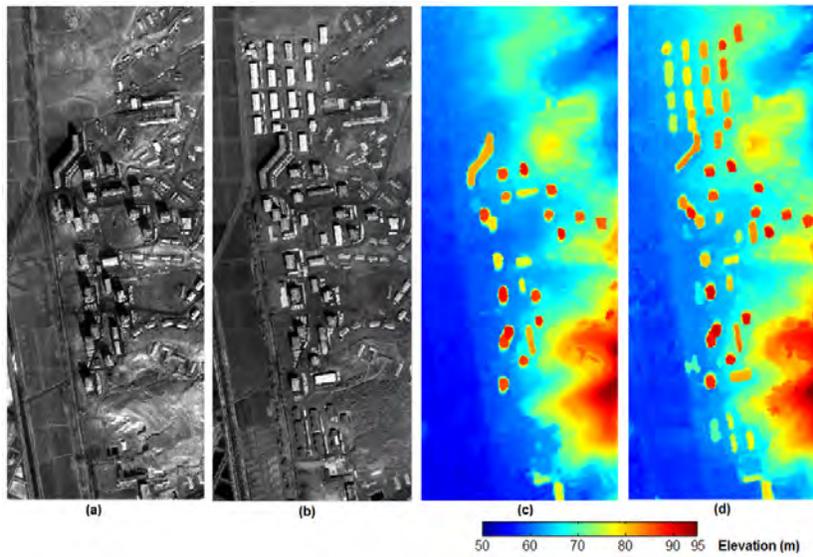


Fig. 6. Datasets of the test region 2: a) panchromatic image from date1; b) panchromatic image from date2; c) DSM from date1; (d) DSM from date2.

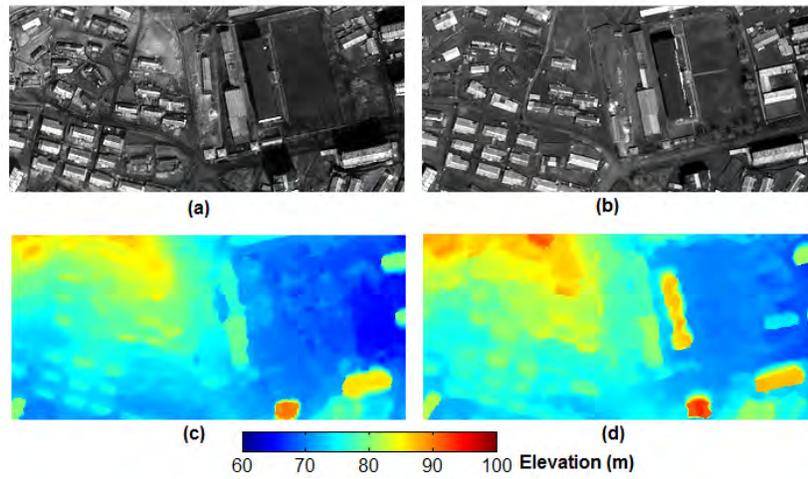


Fig. 7. Datasets of the test region 3: a) panchromatic image from date1; b) panchromatic image from date2; c) DSM from date1; (d) DSM from date2.

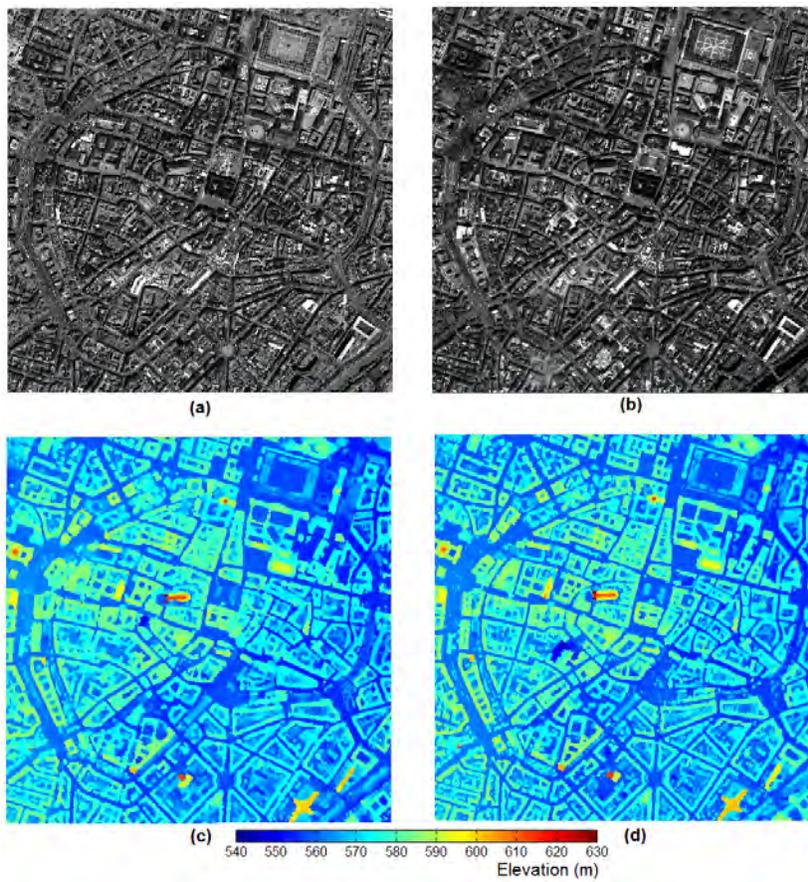


Fig. 8. Datasets of the test region 4: a) panchromatic image from date1; b) panchromatic image from date2; c) DSM from date1; (d) DSM from date2.

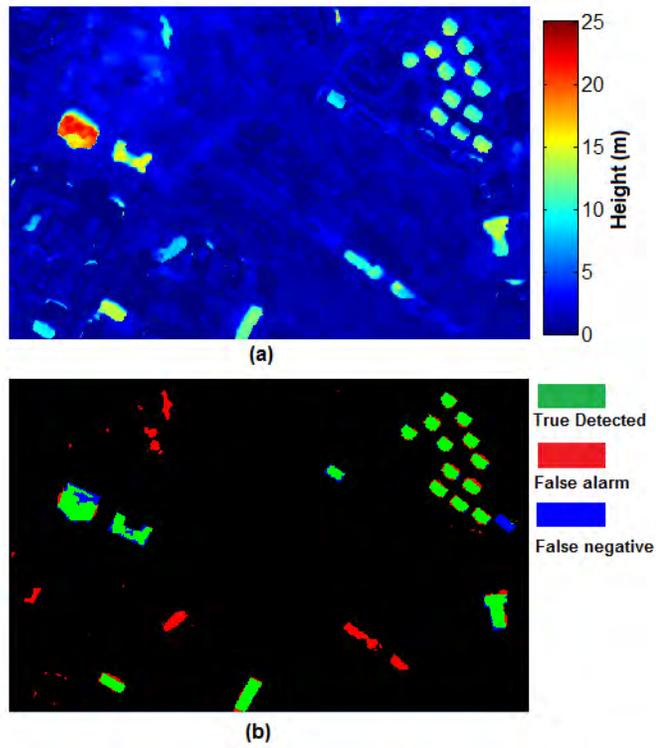


Fig. 9. Change detection results of test region 1 (a) original height change map (b) building change result  $Max\_PI(G_4)$  overlaid with change reference data.

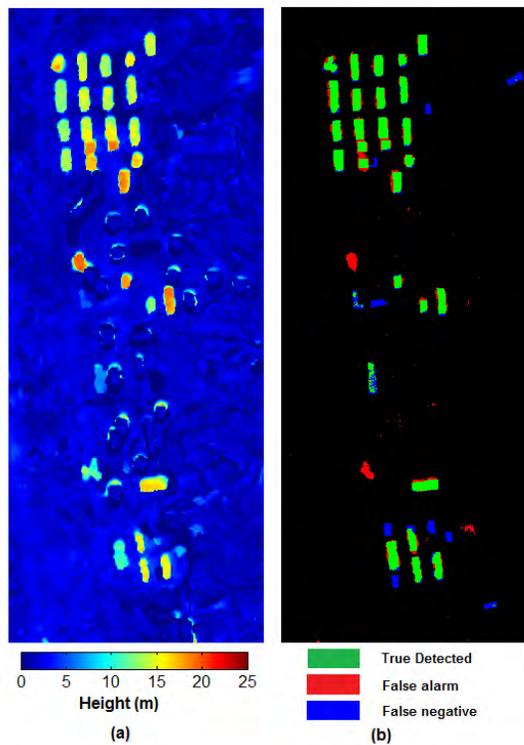


Fig. 10. Change detection results of test region 2 (a) original height change map (b) building change result  $Max\_PI(G_4)$  overlaid with change reference data.

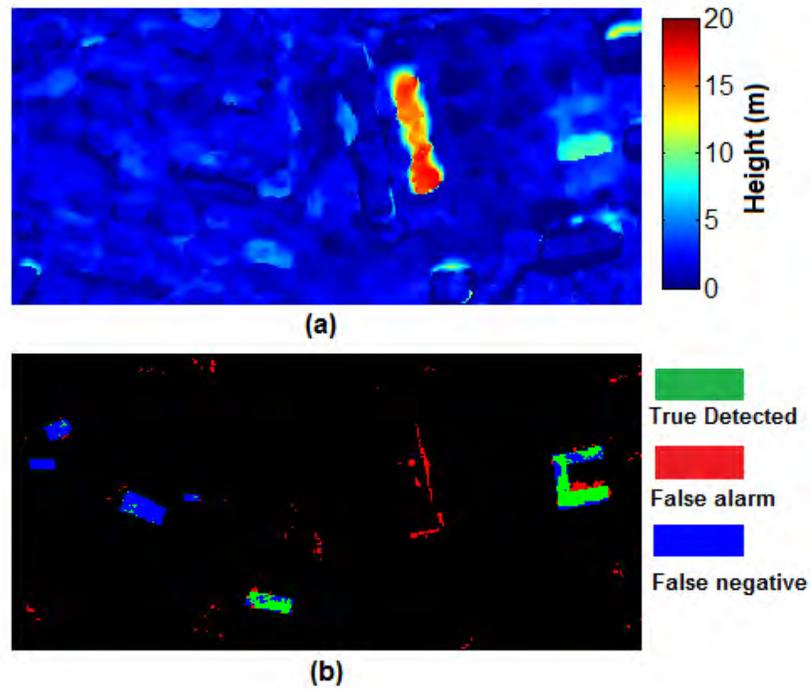


Fig. 11. Change detection results of test region 3 (a) original height change map (b) building change result  $Max\_DSmT(G_2)$  overlaid with change reference data.

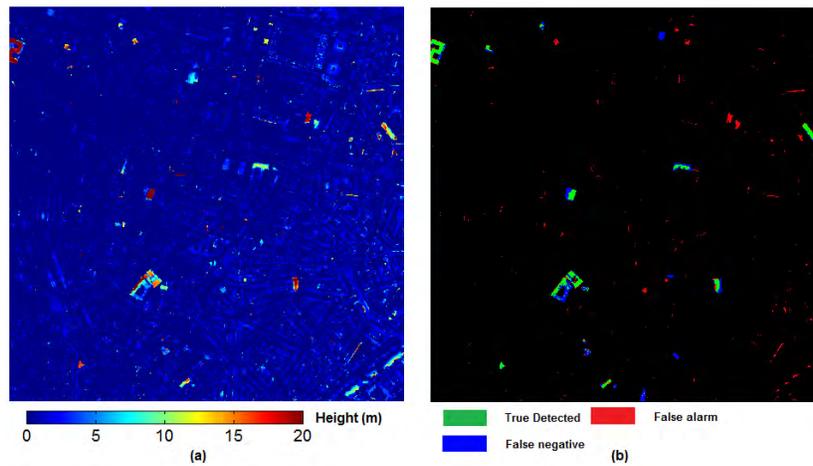


Fig. 12. Change detection results of test region 4 (a) original height change map (b) building change result  $Max\_DSmT(G_2)$  overlaid with change reference data.



Fig. 13. DSMs gaps of part of test region 2 (black holes).

### C. Results evaluation

To further understand the quality of these results and the advantages of the method, more evaluation and analysis are proposed. First, the building change masks extracted from these four global BBA sets are compared and evaluated. Each global BBA set results four building change masks based on the four decision criteria. The building change masks are compared with the masks from [30]. The accuracy of these results have been evaluated by comparing them with ground truth images, which have been manually prepared by visually comparing the pre- and post-event images and referring additional Google Earth history data [39]. The similarity between the obtained result and the ground truth is measured in terms of Kappa Accuracy (KA) [40]. The evaluation results of test region 1, 2, 3 and 4 are shown in Table. V, VI, VII and VIII, respectively. Limited to the available reference data, only the *BuildingChange* class is evaluated.

1) *Comparison of the fusion and decision rules:* Table V to VIII mainly aim to describe and compare the performance of the DS fusion and DSMT fusion rules and the four decision criteria. Unfortunately, the differences among these four global BBA sets of all four test regions are indistinguishable. Our quantitative evaluations results allow comparing the different fusion and decision-making strategies for building change mask construction in different types of region under analysis. As we have observed, there is no unique best fusion and decision strategy working for all types of regions which is an interesting result to be aware of and the different fusion methods (with a chosen decision strategy) perform always better with our refined approach than the previous (original) works which is the main contribution of this work for all type of regions tested.

2) *Validation of the reliability discounting:* The global BBAs obtained with and without reliability discounting are listed under the name of *Refined* and *Original* in Table V to VIII. *Original* refers to the approach presented by [30], in which the reliability discounting is not involved.

In the first test region, the advantage of the reliability discounting is not obvious. By using the Max\_Pl and Max\_BetP decision rules, the refined models perform better than the original models. However, the original models get higher KA values when using the Max\_Pl and Max\_BetP decision rules. This can be partly explained by the shadow detection results, as one dark colour building roof (middle left in the test region) get higher probability to be shadow; thus, a lower probability to be *BuildingChange*.

The second test region is characterised mainly by large and high buildings; thus, the influences of shadows are stronger than in the first test region. The refined models with reliability discounting get generally better accuracy than the original fusion models. Here, we will compare the Max\_DSMT of  $G_4$  of this test region, as it shows the highest difference among these four decision criteria in Table VI. Fig. 14 shows building change masks of the top left part of the test region 2. Fig.

14 (a) and (b) display the change masks obtained from the original model and the refined model overlaid with the change reference mask respectively. The same as Fig. 10 (b), the green colour represents the true detected, the red colour shows the false alarms, while the blue colour pixels are the false negatives. As it shows, based on the refined model, building boundary regions of the change mask obtain less false alarms than the results from the original fusion model.

The advantage of the improved decision fusion models has been well proved by Table VII. The first DSM of this test region contains a large region of pixels with incorrect height introduced by stereo image matching failures. The improved models can solve this problem by adopting the reliability map of height change. Therefore, the increase of KA value of this region is much higher than for the other two test regions. More precisely, under all fusion rules the KAs have improved from around 0.30 to 0.50. For better understanding of this improvement, the global BBAs of *BuildingChange* without and with reliability discounting are displayed in Fig. 15 (a) and (b), respectively. We display here only the  $Prob(\theta_1)$  of  $G_1$ . Both probability maps are less noisy than the original height change map, which are displayed in Fig. 11. By observing the original panchromatic images in Fig. 7, it is not difficult to find out that this building exists in the panchromatic images of both dates. This is the same building that has been mentioned in [16], for which only the DSM of pre-change contains the correct height values. In Fig. 7 (c), this building can not be recognised as a high-level object. A higher value in  $m_1(.)$  leads to a larger global BBA in the class of *BuildingChange*. Thus, this building would be incorrectly detected as *BuildingChange* if no reliability discounting is applied (Fig. 15(a)). Fig. 16 shows the generated height change reliability map. As can be seen, that building region get very low reliability values, that means the height changes of this region cannot be trusted. Therefore, the proposed model is able to remove this kind of errors and correctly recognise this region as *NoChange* (Fig. 11).

The Munich test region has a much larger size and includes several kinds of building changes. The proposed method is able to fuse the spectral and height information efficiently; thus, to identify the newly constructed buildings. The main false negatives are produced in the rebuilt buildings and construction sites. As shown in Fig. 17, the labelled four buildings represent four types of changes. Building A is labelled as a newly constructed building in our reference data. However, half of that building has similar shape and height as the original one, which brings false negatives to our result. Building B, C and D are buildings in different construction phases. By referring [39], in the reference data only D is identified as *OtherChange* as it is almost completed in Fig. 17(a). In the result we are able to correctly identify B as a newly constructed building and D as *OtherChange*. But building C is falsely labeled as *OtherChange* due to low height change values.

TABLE V  
CHANGE MASKS EVALUATION FROM FOUR GLOBAL BBAS OF TEST REGION 1 (KA).

	$G_1$		$G_2$		$G_3$		$G_4$	
	Original	Refined	Original	Refined	Original	Refined	Original	Refined
Max_Bel	0.7392	0.7150	0.7369	0.7138	0.7419	0.7144	0.7391	0.7130
Max_Pl	0.7619	<b>0.7648</b>	0.7607	<b>0.7642</b>	0.7623	<b>0.7652</b>	0.7609	<b>0.7641</b>
Max_BetP	0.7533	0.7442	0.7515	0.7423	0.7541	0.7428	0.7522	0.7412
Max_DSmP	0.7468	0.7200	0.7450	0.7189	0.7490	0.7190	0.7465	0.7181

TABLE VI  
CHANGE MASKS EVALUATION FROM FOUR GLOBAL BBAS OF TEST REGION 2 (KA).

	$G_1$		$G_2$		$G_3$		$G_4$	
	Original	Refined	Original	Refined	Original	Refined	Original	Refined
Max_Bel	0.7401	0.7821	0.7399	0.7816	0.7401	0.7826	0.7401	0.7821
Max_Pl	0.7380	0.7800	0.7391	0.7818	0.7380	0.7812	0.7393	0.7831
Max_BetP	0.7413	<b>0.7853</b>	0.7409	<b>0.7853</b>	0.7412	<b>0.7868</b>	0.7409	<b>0.7867</b>
Max_DSmP	0.7402	0.7842	0.7403	0.7841	0.7405	0.7857	0.7403	0.7855

TABLE VII  
CHANGE MASKS EVALUATION FROM FOUR GLOBAL BBAS OF TEST REGION 3 (KA).

	$G_1$		$G_2$		$G_3$		$G_4$	
	Original	Refined	Original	Refined	Original	Refined	Original	Refined
Max_Bel	0.3356	<b>0.5432</b>	0.3356	<b>0.5418</b>	0.3351	<b>0.5415</b>	0.3345	<b>0.5419</b>
Max_Pl	0.2396	0.3689	0.2416	0.3703	0.2391	0.3694	0.2409	0.3713
Max_BetP	0.2860	0.4726	0.2885	0.4756	0.2869	0.4761	0.2882	0.4786
Max_DSmP	0.3043	0.5082	0.3057	0.5094	0.3008	0.5072	0.3030	0.5066

TABLE VIII  
CHANGE MASKS EVALUATION FROM FOUR GLOBAL BBAS OF TEST REGION 4 (KA).

	$G_1$		$G_2$		$G_3$		$G_4$	
	Original	Refined	Original	Refined	Original	Refined	Original	Refined
Max_Bel	0.5158	0.5217	0.5159	0.5219	0.5154	0.5193	0.5158	0.5195
Max_Pl	0.5122	0.5229	0.5125	0.5232	0.5120	0.5224	0.5128	0.5232
Max_BetP	0.5137	0.5268	0.5140	0.5267	0.5135	0.5258	0.5137	0.5258
Max_DSmP	0.5161	<b>0.5285</b>	0.5163	<b>0.5284</b>	0.5157	<b>0.5274</b>	0.5162	<b>0.5275</b>

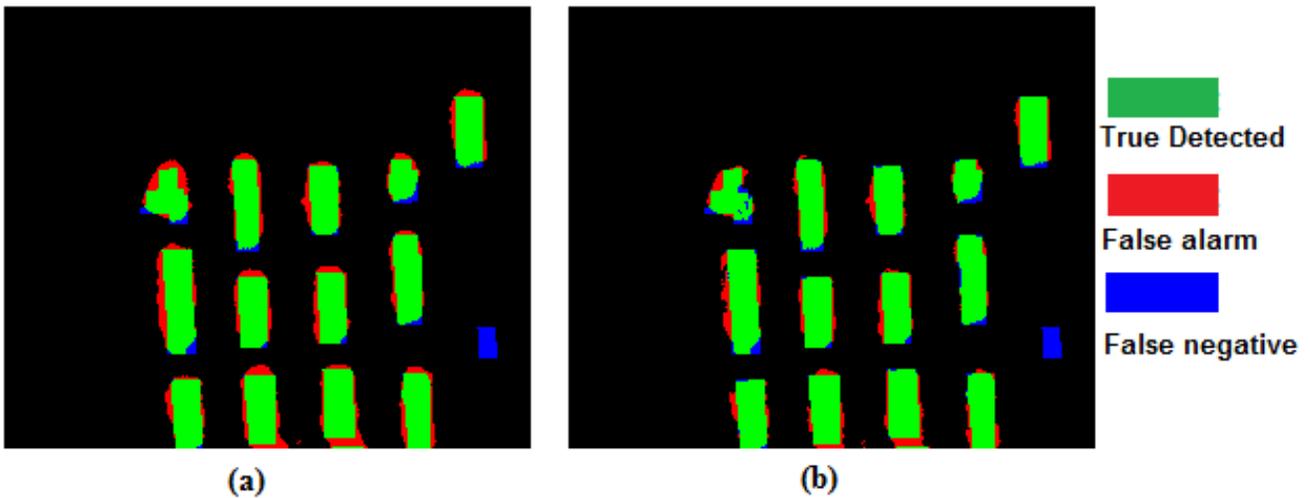


Fig. 14. Building change masks from the original model (a) and refined model (b) of a subset of test region 2.

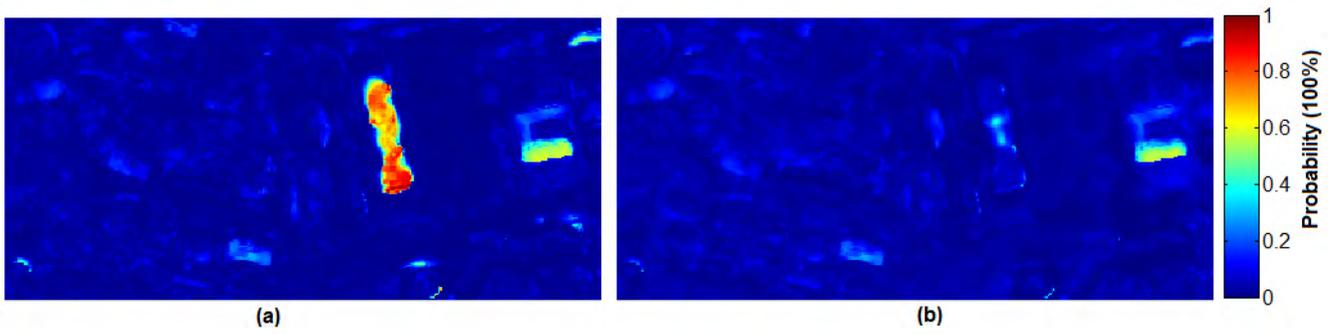


Fig. 15. Comparison of building change global BBAs  $Prob_{\theta_1}$  of  $G_1$  based on the fusion models without reliability discounting (a) and with reliability discounting (b).

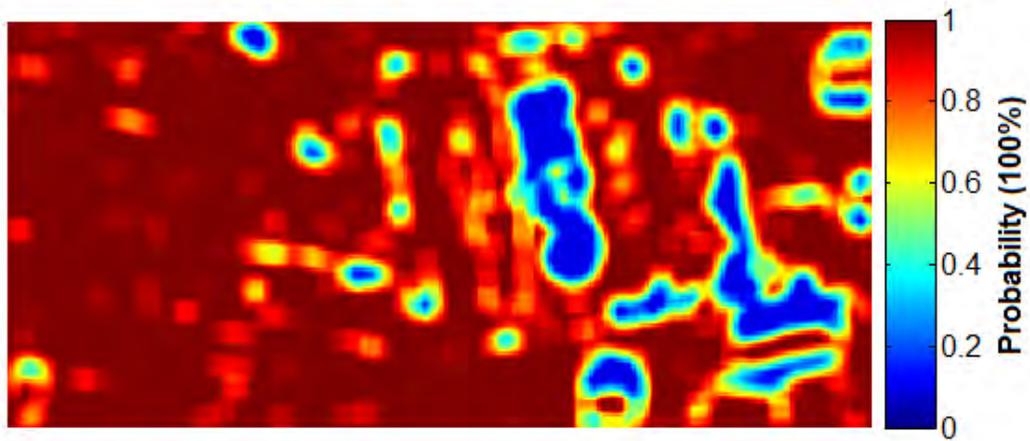


Fig. 16. Reliability discounting map of the height changes of test region 3.



Fig. 17. Example of the various building change types in test region 4.

A window size of  $9 \times 9$  has been used to generate the  $\alpha_{\Delta H}$ . In order to test the sensitivity of our fusion model to the window width used, we have changed the width parameter from 3 to 13 by steps of 2. For each size, we generate the global BBA  $G_1$ . Thus four final building change masks based on the four decision criteria can be regenerated. We provide the KA for each mask as show in Fig. 18. As a comparison, we have also provided the KAs without using  $\alpha_{\Delta H}$ . This test shows that the final results benefit largely from the reliability discounting procedure, but the KA rate did not change significantly with various window sizes for all four test regions.

#### D. Comparison with existing methods

In this section, the improved belief fusion models are compared with the directly feature fusion method [41] and the initial fusion model that described in [16].

As a typical feature fusion approach, [41] adopted the kernel Minimum Noise Fraction (kMNF) approach to fuse change features from the DSMs and panchromatic images. Based on the resulting kMNF components, a change mask was extracted with iterated canonical discriminant analysis (ICDA). [30] randomly selected the training data from the ground truth, as the experiments were devoted to algorithm comparison. However, it was not a practical procedure, because in real situations the ground truth is unknown. Therefore, in this article as well as using the set of random pixels from the ground truth, another set of training data for each test region is prepared by manually selecting changed regions. All pixels in these regions are then used as training samples.

The results generated based on these two sets of training data are described as  $kMNF_{random}$  and  $kMNF_{manual}$ , respectively, in Table IX in the term of KA and Overall Accuracy (OA). All training data in the first three test regions contain around 200 pixels/samples. In the fourth test region, 500 pixels are used to fit with the large image size. If the training data are selected from the ground truth, the newly proposed approach can deliver a slightly better result than the approach in [41]. When using the manually selected training data, the advantages of the newly developed approach are obvious. As in real applications the ground truth is normally unknown, we conclude that the proposed fusion method is more robust for larger test regions with diverse characterised objects.

In addition, the approach proposed by [16] is tested on the same test data, and the results are shown in the third and fourth columns. In that approach, after the fusion approach a shape-based refinement was proposed to reach the final building change mask. Thus, the resulting masks before and after the refinement procedure are both calculated and evaluated. In the North Korea test region, we have used  $T_{height} = 3m$ ,  $T_{area} = 50m^2$  and  $T_{convexity} = 0.55$  as thresholds. And in the Munich test region, as the buildings have a larger size and complicated roof shapes than North Korea, we manually modified these threshold values to  $T_{area} = 100m^2$  and  $T_{convexity} = 0.50$  to improve the results. The accuracies are recorded in the columns  $[16]_{before}$  and  $[16]_{after}$  in Table IX. The refinement

is not included in this article to avoid unnecessary threshold parameters; thus, to achieve an automatic and robust workflow. By comparing the KAs with Tables V, VI and VII, one can see that the shape-based refinement can further improve the result accuracy. But the fusion model in [16] performs rather weakly. All obtained KAs are lower than values from the proposed refined decision fusion approaches, especially for test regions 2 and 3.

TABLE IX  
COMPARISON WITH EXISTING METHODS.

	$kMNF_{random}$		$kMNF_{manual}$		[16] <sub>before</sub>		[16] <sub>after</sub>	
	KA	OA	KA	OA	KA	OA	KA	OA
Region1	<b>0.7178</b>	0.9799	0.5477	<b>0.9803</b>	0.5929	0.9628	0.6312	0.9683
Region2	<b>0.6791</b>	<b>0.9822</b>	0.2458	0.9688	0.6433	0.9681	0.6718	0.9718
Region3	0.2195	0.9794	0.2272	<b>0.9799</b>	0.3060	0.9375	<b>0.3287</b>	0.9447
Region4	0.2057	0.9878	0.1937	0.9876	0.4909	0.9912	<b>0.5641</b>	<b>0.9941</b>

It has to be mentioned that vegetation change is not noted as false alarm in the improved decision fusion model. As the vegetation change and building change can be easily separated by using a vegetation index. [16] has adopted vegetation index as *no-building change* indicators to highlight building changes. This step is not considered in this article as not many forest changes are available in the test regions. Moreover, if forest changes are of interest, we can easily modify this model using vegetation index to separate forest changes from building changes.

#### V. CONCLUSIONS AND PERSPECTIVES

Building change detection is a difficult topic, to solve uncertain change information from images and DSMs, decision fusion methods have been introduced as a new concept and proved to be efficient and appropriate. The innovative contribution of this article is the improvement of the decision fusion models. DS as well as DSMT decision fusion models are further developed to solve the building change detection problem in this article. Another contribution lies in the BBA calculation procedure, and the sigmoid distribution is further improved by taking both concordance and discordance situations. As a third contribution, the reliability of each indicator is introduced according to the change objects of interest.

The proposed building change detection models enable an improved result by comparing to the original fusion model and other change detection methods. A comparative analysis of the results shows that there is not a so big difference of performances between DS and DSMT fusion methods based on the best decisional strategy and so we can in practise use the simplest fusion method to reduce to computational burden without degrading too much the performance. Of course the most critical question is to select beforehand the decisional strategy based on type of region under analysis, for this we need to define efficient indicators for characterising each type of region which then will help us to automatically select the best criterion to use. Our future research works will address, and hopefully help, to solve this important question.

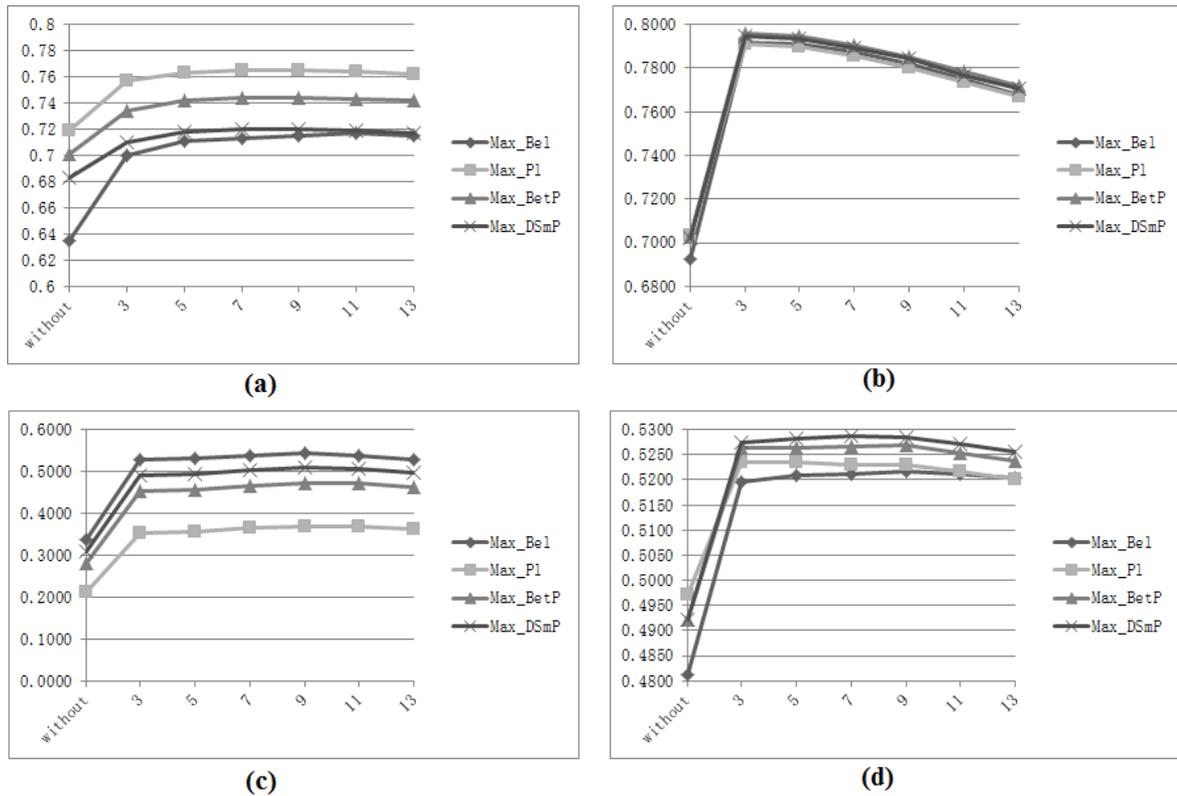


Fig. 18. Effect of window size on KA for test region 1 (a), test region 2 (b), test region 3 (c) and test region 4 (d).

## REFERENCES

- [1] D. Lu, et al., Change detection techniques, *International Journal Remote Sens.*, Vol. 25(12), pp.2365–2407, 2004.
- [2] A.P. Tewkesbury, et al., *A critical synthesis of remotely sensed optical image change detection techniques*, *Remote Sensing of Environment*, Vol. 160, pp. 1–14, 2015.
- [3] Z.-G. Liu, et al., *Change detection in heterogenous remote sensing images via homogeneous pixel transformation*, *IEEE Trans. on Image Processing*, Vol. 27(4), pp. 1822–1834, 2018.
- [4] L. Bruzzone, D.F. Prieto, *Automatic analysis of the difference image for unsupervised change detection*, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 38(3), pp. 1171–1182, 2000.
- [5] A.A. Nielsen, *The regularized iteratively reweighted MAD method for change detection in multi-and hyperspectral data*, *IEEE Trans. on Image Processing*, Vol. 16(2), pp. 463–478, 2007.
- [6] R. Qin, J. Tian, P. Reinartz, *3D change detection—approaches and applications*, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 122, pp. 41–56, 2016.
- [7] W. Li, et al., *A new approach to performing bundle adjustment for time series UAV images 3D building change detection*, *Remote Sensing*, Vol. 9(6), 625, 2017.
- [8] S. Pang, *Building change detection from bi-temporal dense-matching point clouds and aerial images*, *Sensors*, Vol. 18(4), 966, 2018.
- [9] B. Sirmacek, C. Unsalan, *A probabilistic framework to detect buildings in aerial and satellite images*, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 49(1), pp. 211–221, 2011.
- [10] M. Janalipour, M. Taleai, *Building change detection after earthquake using multi-criteria decision analysis based on extracted information from high spatial resolution satellite images*, *Int. Journal of Remote Sensing*, Vol. 38(1), pp. 82–99, 2017.
- [11] S. Le Hégarat-Masclé, I. Bloch, D. Vidal-Madjar, *Application of Dempster-Shafer evidence theory to unsupervised classification in multi-source remote sensing*, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 35(4), pp. 1018–1031, 1997.
- [12] F. Rottensteiner, et al., *Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection*, *Information Fusion*, Vol. 6(4), pp. 283–300, 2005.
- [13] F. Rottensteiner, *Building change detection from digital surface models and multi-spectral images*, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (IAPRS)*, Vol. 36(3), pp. 145–150, 2007.
- [14] Z.-G. Liu, et al., *Change detection in heterogeneous remote sensing images based on multidimensional evidential reasoning*, *IEEE Geoscience and Remote Sensing Letters*, Vol. 11(1), pp. 168–172, 2014.
- [15] Z.-G. Liu, *Credal classification of uncertain data based on belief function theory*, Ph.D. Thesis, Télécom Bretagne; Université de Bretagne Occidentale, Brest, France, 2014.
- [16] J. Tian, S. Cui, P. Reinartz, *Building change detection based on satellite stereo imagery and digital surface models*, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 52(1), pp. 406–417, 2014.
- [17] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, *Annals of Mathematical Stat.*, Vol. 38(2), pp. 325–339, 1967.
- [18] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [19] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion - Collected Works*, Volumes 1–4, American Research Press, Rehoboth, NM, USA, 2004–2015.
- [20] J. Tian, J. Dezert, P. Reinartz, *Refined building change detection in satellite stereo imagery based on belief functions and reliabilities*, in *Proc. of 2015 IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pp. 160–165, San Diego, CA, USA, 14–16 September, 2015.
- [21] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, *Int. Journal Intelligent Systems*, Vol. 29(3), pp. 223–252, 2014.
- [22] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in *Proc. of the 15th Int. Conf. on Information Fusion (Fusion 2012)*, pp. 655–660, Singapore, July 2012. Available from: [www.onera.fr/staff/jean-dezert/references](http://www.onera.fr/staff/jean-dezert/references).

- [23] A. Tchamova, J. Dezert, On the behavior of Dempster's rule of combination and the foundations of Dempster-Shafer theory, in Proceedings of IEEE IS'2012, pp. 108–113, Sofia, Bulgaria, Sept. 6-8, 2012.
- [24] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of the 16th Int. Conf. on Information Fusion (Fusion 2013), pp. 1119–1126, Istanbul, Turkey, July 9–12, 2013
- [25] D. Mercier, B. Quost, B., T. Denœux, *Contextual discounting of belief functions*, In: Lluís Godo, ed. Symbolic and quantitative approaches to reasoning with uncertainty, Berlin, Heidelberg: Springer, pp. 552–562, 2005.
- [26] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of the 13th Int. Conf. on Information Fusion (Fusion 2010), pp. 1–8, Edinburgh, Scotland, UK, 26-29 July 2010.
- [27] P. d'Angelo, P. Reinartz, P., *DSM based orientation of large stereo satellite image blocks*, International Archives Photogramm. Remote Sens. Spatial Information Sciences, Vol. 39 (B1), pp. 209–214, 2012.
- [28] J. Tian, P. Reinartz, P. d'Angelo, M. Ehlers, *Region-based automatic building and forest change detection on Cartosat-1 stereo imagery*, ISPRS Journal Photogramm. and Remote Sensing, Vol. 79, pp. 226–239, 2013.
- [29] J. Tian, et al., *Automatic 3D change detection based on optical satellite stereo imagery*, In: ISPRS TC VII Symposium – 100 Years ISPRS, Vol. XXXVIII - Part 7B, pp. 586–591, Vienna, Austria, 2010.
- [30] J. Tian, P. Reinartz, J. Dezert, *Building change detection in satellite stereo imagery based on belief functions*, in Proc. of 2015 Joint Urban Remote Sensing Event (JURSE), pp. 1–4, Lausanne, Switzerland, 30 March - 01 April, 2015.
- [31] J. Dezert, J.-M. Tacnet, *Sigmoidal model for belief function-based ELECTRE TRI method*, in Proc. of Belief 2012 Int. Conf., Springer, pp. 401–408, Compiègne, France, May 2012.
- [32] N. Otsu, *A threshold selection method from gray-level histograms*, IEEE Trans. on Systems, Man and Cybernetics, Vol. 9(1), pp. 62–66, 1975.
- [33] P.S. Liao, T.S. Chen, P.C. Chung, *A fast algorithm for multilevel thresholding*, Journal Information Sciences Engineering, Vol. 17(5), pp. 713–727, 2001.
- [34] J. Dezert, F. Smarandache, *A new probabilistic transformation of belief mass assignment*, in Proc. of the 11th Int. Conf. on Information Fusion (Fusion 2008), pp. 1–8, Cologne, Germany, July 2008.
- [35] J. Dezert, Z.-G. Liu, G. Mercier, *Edge detection in color images based on DSMT*, in Proc. of the 14th Int. Conf. on Information Fusion (Fusion 2011), pp. 1–8, Chicago, IL, USA, July 5–8, 2011.
- [36] R. Qin, J. Tian, P. Reinartz, *Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images*, Int. Journal of Remote Sensing, Vol. 37(15), pp. 3455–3476, 2016.
- [37] C.A. Laben, B.V. Brower, *Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening*, U.S. patent # 6,011,875, Jan. 4, 2000.
- [38] M.N. Klaric, et al., *GeoCDX: an automated change detection and exploitation system for high resolution satellite imagery*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 51(4), pp. 2067–2086, 2013.
- [39] GoogleEarth, 2018. Available from: <https://support.google.com/earth/answer/148094?hl=en>.
- [40] R.G. Congalton, *A review of assessing the accuracy of classifications of remotely sensed data*, Remote Sensing of Environment, Vol. 37(1), pp. 35–46, 1991.
- [41] J. Tian, A.A. Nielsen, P. Reinartz, *Improving change detection in forest areas based on stereo panchromatic imagery using kernel MNF*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 52(11), pp. 7130–7139, 2014.

# Tracking Uncertainty Propagation from Model to Formalization: Illustration on Trust Assessment

Valentina Dragos<sup>a</sup>, Jean Dezert<sup>a</sup>, Kellyn Rein<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>Fraunhofer FKIE, Wachtberg, Germany.

Emails: valentina.dragos@onera.fr, jean.dezert@onera.fr, kellyn.rein@fraunhofer.fkie.de

Originally published as: V. Dragos, J. Dezert, K. Rein, *Tracking Uncertainty Propagation from Model to Formalization: Illustration on Trust Assessment*, Journal of Advances in Information Fusion, Vol. 13, No. 2, pp. 216–234, December 2018, and reprinted with permission.

**Abstract**—This paper investigates the use of the URREF ontology to characterize and track uncertainties arising within the modeling and formalization phases. Estimation of trust in reported information, a real-world problem of interest to practitioners in the field of security, was adopted for illustration purposes. A functional model of trust was developed to describe the analysis of reported information, and it was implemented with belief functions. When assessing trust in reported information, the uncertainty arises not only from the quality of sources or information content, but also due to the inability of models to capture the complex chain of interactions leading to the final outcome and to constraints imposed by the representation formalism. A primary goal of this work is to separate known approximations, imperfections and inaccuracies from potential errors, while explicitly tracking the uncertainty from the modeling to the formalization phases. A secondary goal is to illustrate how criteria of the URREF ontology can offer a basis for analyzing performances of fusion systems at early stages, ahead of implementation. Ideally, since uncertainty analysis runs dynamically, it can use the existence or absence of observed states and processes inducing uncertainty to adjust the tradeoff between precision and performance of systems on-the-fly.

**Keywords:** uncertainty, reported information, trust, belief functions, information fusion, DSMT, URREF ontology.

## I. INTRODUCTION

A key element when designing information fusion systems is the way the system designer isolates and analyzes real world phenomena. A model is abstracted into a simpler representation, in which components, modules, interactions, relationships and data flows are easier to express. Uncertainty tracking highlights approximations induced by model construction and its formalization, as well as providing a checklist to ensure that all uncertainty factors have been identified and considered ahead of system implementation.

This paper illustrates the use of the uncertainty representation and reasoning framework (URREF) ontology [1] to identify and assess uncertainties arising during the modeling and formalization phases of an information fusion system intended to estimate trust in reported information.

Trust assessment is a real-world problem grounded in many applications relying on reported items, with different persons observing and then reporting on objects, individuals, actions or events. For such contexts, using inaccurate, incomplete or distorted items can result in unfortunate consequences and

analysts need to ensure the consistency of reported information by collecting multiple items from several sources.

From the perspective of an information analyst, trust can be analyzed along two dimensions: the subjective evaluation of items reported by the source itself, called self-confidence, and the evaluation of source by the analyst, called reliability. While self-confidence encompasses features of subjectivity, the reliability of a source is related to the quality of previously reported items, the competence of the source for specific topics, and the source's capacity for misleading intentions. Trust estimation aims at capturing, in an aggregated value, the combined effects of self-confidence and reliability on the perceived quality of information. The model is represented with belief functions, a formalism which offers a sound mathematical basis to implement fusion operators which estimate trust by combining self-confidence and reliability.

The model developed for trust assessment focuses on the global characterization of information and provides a better understanding of how trust is to be estimated from various dimensions. The overall process has humans as a central element in both the production and the analysis of information.

Trust in reported information offers a good illustration for tracking uncertainty: the phenomenon is complex, so any model adopted is generally a simplification of the real world interactions. Uncertainties can be made explicit not only for static elements of the model, such as sources or items, but also for the dynamic processes of combining items with one another. Moreover, adopting belief functions as representation formalism will have an impact on the way an information system could be implemented and on the accuracy of its results.

The contribution of this paper is twofold: first, it presents a trust estimation model which combines the reliability of sources and self-confidence of reported items, and, second, the paper analyzes types of uncertainty occurring during modeling and formalization by relating elements of the model to uncertainty criteria defined by the URREF ontology.

The remainder of this paper is divided into 8 sections: section II discusses related approaches for trust modeling and uncertainty assessment. The problem tackled in this paper is presented in section III. Section IV describes the model developed for trust estimation, while its implementation with

belief functions is presented in section V. The analysis of uncertainty is discussed in VI, while examples and scenarios for trust assessment are presented in section VII. Strengths and limitations of belief-based formalization are discussed in section VIII and section IX concludes this paper.

## II. RELATED APPROACHES

The work presented in this paper is related to approaches for trust modeling and assessment as well as solutions for uncertainty analysis for information fusion systems. Trust modeling is not a new research topic; it spans diverse areas such as agent systems [2] and logical modeling and argumentation [3]. The Internet and social media offer new application contexts for trust assessment; this topic is addressed in relation to service provision on the Internet [4], social networks analysis [5], and crowdsourcing applications [6]. Trust analysis is also of interest in the military field where techniques have been developed in order to identify clues of veracity in interview statements [7].

The concept of trust in these communities varies in how it is represented, computed and used. Although having an obvious social dimension, trust is not only understood with regard to other humans, but also towards information pieces [6], information sources [8], Internet sites [9], algorithms for data and knowledge fusion [10], intelligent agents [2], and services for the Internet of things [11].

While definitions of trust vary from one domain to another, there are some common elements. The first commonality for all research areas cited above is to consider trust as a user-centric notion that needs to be addressed in integrated human-machine environments which rely heavily on information collected by humans, even if further processing can be executed automatically. Moreover, all definitions associate some degree of uncertainty with trust, which is then captured by concepts such as subjective certainty [12] and subjective probability [13].

Trust goes hand in hand with the concepts veracity [14] and deception. [15] addresses veracity along the dimensions of truthfulness / deception, objectivity / subjectivity and credibility / implausibility. The authors developed a veracity index ranging from true/objective/credible to untrustworthy/subjective/implausible to characterize texts in the context of big data analysis. Deception is defined as a message knowingly transmitted with the intent to foster false beliefs or conclusions. The topic is addressed in studies from areas such as interpersonal psychology and communication [16], [17] and it is also considered in the field of natural language processing, as part of a larger research direction tackling subjectivity analysis and the identification of private states (emotions, speculations, sentiments, beliefs). These solutions stem from the idea that humans express various degrees of subjectivity [18] that are marked linguistically and can be identified with automatic procedures [19].

Contributions on trust estimation keep the distinction between analyzing the source of information, the item reported and reasoning about trust. Approaches developed for trust

in information sources consider that trust is not a general attribute of the source but rather related to certain properties: competence [20], sincerity and willingness to cooperate [3]. On this basis, it becomes possible to consider the competence of a source not in general but with respect to specific topics [21]. Trust can be also analyzed in relation to roles, categories or classes [22].

Research efforts on reasoning about trust analyze information sources from past behaviors rather than directly from their properties [23], or they infer trust from estimations already computed for a set of properties [24]. These approaches generally focus on building trust by using argumentation [25] or beliefs functions [26], or investigating the joint integration of those techniques [27]. Taking this work a step further, [28] identified several patterns for reasoning about trust and its provenance while the notion of conflict in handling trust is discussed in [29].

As shown by approaches above, trust is a multifaceted concept and, in practice, this complex notion can be decomposed into two components: communication or interaction trust, and data trust [30]. The model developed deals with data trust and keeps the distinction between sources and items provided by those sources, although several approaches consider these elements as a whole [26], estimating the trust of information sources [24], [29] rather than information items. The model does not require statistical data to infer the behavior of the source [23] and introduces reliability to characterize the source. More specifically, reliability encompasses not only competence [22], [20] and reputation [21] - two attributes already considered by previous approaches - but also intentions which constitute an original aspect of the model. Intention is of important significance in the context of human-centered systems, including open-sources, and supports the analysis of emerging phenomena such as on-line propaganda or disinformation. Another original aspect of the model is consideration of the characterization of items by the source itself, thus overcoming a main limitation of the solution presented in [31]. Our approach can be considered as partially overlapping solutions investigating trust propagation in direct and indirect reporting [28], [25], and the model enables a particular kind of trust estimation, based both on more or less complete characterizations of the source by the analyst, and more or less accurate characterizations of the items by the source. The model also addresses disagreement and the fusion of diverging opinions, not in a panel of experts as described in [27], but rather between items showing high levels of confidence according to the source and sources having low reliability according to the analyst. By ascribing characterizations to both information sources and reported items, the model allows analysts to make use of both prior experience and their own beliefs in order to assess various degrees of trust.

From a different perspective, the evaluation of uncertainty regarding the inputs, reasoning and outputs of the information fusion is the goal of Evaluation Techniques for Un-

certainty Representation Working Group<sup>1</sup> (ETURWG). The group developed an ontology for this purpose [1]. The URREF ontology defines the main subjects under evaluation [32], such as uncertainty representation and reasoning components of fusion systems. Furthermore, the frame also introduces criteria for secondary evaluation subjects: sources and pieces of information, fusion methods and mathematical formalisms. URREF criteria have generic definitions and therefore can be instantiated for applications with coarser or finer granularity levels. This means evaluation metrics can be defined for data analysis [33], increased particularity for data specific types [34] or attributes, reliability and credibility [35], self-confidence [36] or veracity [37].

In addition to allowing a continuous analysis of uncertainty representation, quantification and evaluation, as described in [38], URREF criteria are detailed enough to capture model-embedded uncertainties [39], imperfection of knowledge representations [40], and their propagation in the context of the decision loop [41]. The frame also offers a basis to compare different fusion approaches [42]. URREF criteria were used for uncertainty tracking and investigation in several applications: vessel identification for maritime surveillance [43], activity detection for rhino poaching [44] and imagery analysis for large area protection [45].

Beyond developing a model for trust estimation, this paper also fills a gap within the ETURWG community by illustrating how uncertainty analysis tracks imperfections occurring from problem definition to model abstraction and formalization.

### III. HUMAN SOURCES AND REPORTED INFORMATION

Many applications rely on human sources which are used to continuously supply observations, hypotheses, subjective beliefs and opinions about what they sense or learn. In such applications reports are often wrong, due to environment dynamics, simple error, malicious act or intentions, [46]. From the analyst standpoint, decisions have to be made based on indirect reporting and trust relies upon the in-depth investigation of items and sources, thus the analysis of reported items is a critical step. This analysis is a multilevel process, relying on the ability of analysts to understand the content of messages and assess their quality from additional clues. The use cases described below highlight levels of indirection occurring when collecting information and their with impact on trust estimation.

#### A. Assertions, opinions and reported information

For illustration, let's consider  $X$ , the analyst receiving information provided by a human source  $Y$ .

**Case 1: direct reporting**  $X$  is an analyst collecting evidence in order to decide whether or not an individual is involved in terrorist activities. In particular, he takes into account reports submitted by  $Y$ , a human source. Those reports usually consist on a mixed set of assertions (e.g., descriptions of events or states observed by  $Y$ ) and opinions

(i.e., judgments, assessments, or beliefs) expressed by  $Y$  about assertion which give the analyst an insight into how strongly the source commits to the assertion, see fig. 1.

It is very likely that John is a terrorist.

Opinion                      Primary information

Fig. 1. Assertions and opinions in human messages.

In the statement contained in fig. 1, the source  $Y$  lets us know that she does not commit her full belief to the assertion that *John is a terrorist*, otherwise the reporter would have used phrasing such as *I am completely convinced* or *it is without doubt* or simply reported *John is a terrorist* as an unadorned statement.

The information item is the sentence, which contains the assertion *John is a terrorist* and the uncertainty degree to be assigned because the analyst knows that  $Y$  is not completely certain about her own statements. The analyst must make a judgment about the veracity of John being a terrorist based upon factors such as previous experience with  $Y$ 's assessments in the past, or, perhaps, on the fact that other sources are relating the same information.

**Case 2: indirect reporting** Again, let  $X$  be an analyst collecting evidence in order to decide whether or not an individual is involved in terrorist activities. In this case, he takes into account reports submitted by  $Y$ , a human source who is herself relating information obtained from a secondary source named Mary, see fig. 2.

Mary told me it is very likely that John is a terrorist.

Secondary source      Reporting                      Opinion                      Primary information

Fig. 2. Hearsay, assertions and opinions in human messages.

The source  $Y$  does not report on her direct observations or her deductions or beliefs, but conveys information received from a second source, in this case Mary, in the statement in fig. 2.

In this report the information item is again the sentence containing the assertive part *John is a terrorist* but this use case introduces more levels of complexity in uncertainty to deal with. The information that the assertion comes from Mary, who has added her own opinion, is a distancing mechanism on the part of the source  $Y$  as (unlike in fig. 1), she is neither claiming the opinion nor the assertion.

This case introduces yet more layers of uncertainty. How sure can we be that the reporter  $Y$  has accurately repeated what Mary said? For example, did Mary really say *it is likely*

<sup>1</sup><http://eturwg.c4i.gmu.edu/>

or did the reporter insert this (intentionally or unintentionally) based upon the reporter's assessment of the reliability of Mary as a source of information? Or perhaps, subtly, *Y* is expressing her own uncertainty by putting words in Mary's mouth. Furthermore, it is possible Mary made this statement under circumstances which would strengthen or weaken this statement, but those conditions have not been passed on by the reporter.

The goal of the analyst is to take this assertion into account, but also to encode his own belief about the quality of the source further in the analysis. All these different attitudes have to be evaluated by the analyst, who may have additional background information or prior evaluation of the source that have to be considered.

In both cases discussed above, the outcome of the analyst is the assertive part of the information item, augmented with a coefficient that helps to measure and track the different levels of trust for their future exploitation. For the purpose of this work, this quality is called *trust in reported information*.

### B. Concepts and notions for trust assessment

This section introduces several notions that are relevant for trust analysis.

*Trustworthiness of information sources* is considered, for the purpose of this work, as confidence in the ability and intention of an information source to deliver correct information, see [47]. Trustworthiness is an attribute of information sources who have the competences to report information, and who can be relied upon to share sincerely and clearly their beliefs on the uncertainty level of reported information. An item provided by such a source is then trusted by analysts.

*Self-confidence* [36] captures the explicit uncertainty assigned to reported assertions by the source. Statements may include the source's judgments when lacking complete certainty; these judgments are generally identified through the use of various lexical clues such as *possibly*, *probably*, *might be*, *it is unlikely*, *undoubtedly*, etc., all of which signal the source's confidence (or lack thereof) in the veracity of the information being conveyed. It should be noted that self-confidence, in our usage understood as the linguistic dimension of the certainty degree that the source assigns to reported items, is an aspect exhibited by the source, but it will be considered from the analyst's standpoint during trust analysis.

*Reliability of sources* indicates how strongly the analyst is willing to accept items from a given source at their face-value. As an overall characterization, reliability is used in this work to rate how much a source can be trusted with respect to their reputation, competence and supposed intentions.

*Reputation of sources* [48] captures a commonly accepted opinion about how the source performs when reporting information, and is generally understood as the degree to which prior historical reports have been consistent with fact. For human sources, reputation is considered by the analyst for each source based on previous interactions with the source and on the source's history of success and failure in delivering accurate information. Reputation relies, to a large extent, upon

negative and positive experiences provided to the analyst by the source in the past.

*Competence of sources* [20] is related to a source's possession of the skills and knowledge in reporting on various topics: This aspect defines to what extent a human source can understand the events they report on, whether the source has the ability to accurately describe those events, and how capable the source is of following the logic of processes producing the information.

*Intentions* correspond to specific attitudes toward the effect of one's actions or conduct. Reporting information can become *more a means to manipulate others than a means to inform them* [49] and thus can be carried out with the express purpose of inducing changes in another person's beliefs and understanding. Intentions are specific to human sources as only humans have the capacity to deliberately provide false or misleading information. Sensors may provide erroneous data due to a number of factors such as device failure or environmental conditions, but never due to intention.

In addition to the above facets, *credibility of information* and *reliability of sources* are two notions introduced by the STANAG 2511 [50], which standardizes the terminology used in analysis of intelligence reports used by NATO Forces with distinct focus on sources and information provided. STANAG reliability is understood with respect to the quality of information that has been delivered by sources in the past. STANAG credibility relies on the intuition that a joint analysis of items in combination with each other will likely reveal inconsistencies, contradictions or redundancies. Reliability and credibility are independent criteria for evaluation. Definitions for both reliability and credibility are in natural language.

Attributes of sources and information items adopted for the model of trust are related to the notions introduced by the STANAG 2511 but are addressed differently: reliability of sources is understood here in terms of source competence, reputation and intentions, while credibility is restricted to features of self-confidence as described above.

## IV. A FUNCTIONAL MODEL OF TRUST

This section introduces the model developed to estimate trust in reported information by taking into account the reliability of the source and the source's own characterization of reported items. The advantage of this distinction is to better dissociate the impact of both beliefs of sources and opinions of analysts on the source on the information provided.

Even if the primary function of a source is to provide information, we keep the distinction between the source and the information by considering separate dimensions for each element. The rationale behind this is the observation that even reliable sources can sometimes provide inaccurate or imprecise information from one report to another, which is even more plausible in the case of human sources.

The model, illustrated in fig. 3., is composed of a source which provides an information item augmented with a degree of uncertainty captured by self-confidence to an analyst. Based upon his direct assessment of the reliability of the source,

the analyst constructs his own estimation of trust in the item reported.



Fig. 3. Model for trust analysis.

In the following section, the model is discussed using a granularity that is detailed enough to describe its elements, but still rough enough to avoid the adoption of a representation formalism.

#### A. Elements of the trust model

The model is composed of two elements: an information source and reported items from that source. The analyst is considered to be outside the model, although she has multiple interactions with its elements.

**Definition of information source:** An information source is an agent who provides an information item along with a characterization of its level of uncertainty. "Source" is a relative notion, depending on the perspective of analysis. In general, information is propagated within a chain relating real world information to some decision maker, and agents along the path can be both trained observers, whose job is to provide such reports, as well as witnesses or lay observers who may add items, in spite of not being primarily considered as information sources, but rather as opportunistic ones.

The notion of source is central in many information fusion applications and numerous research efforts aimed at modeling the properties of those applications. A general analysis of sources is undertaken by [51], who identify three main classes: S-Space, composed of physical sensors, H-Space for human observers and I-Space for open and archived data on the Internet. In [52], a unified characterization of hard and soft sources is described, along with a detailed description of their qualities and processing capabilities.

Processing hard sensor information is widely covered [53] in the research community, and can be considered quite mature, while the integration of human sources brings many new challenges. Our model addresses human sources, and reported items can refer to actions, events, persons or locations of interest.

Information reported by humans is unstructured, vague, ambiguous and subjective, and thus is often contrasted with

information coming from physical sensors, described as structured, quantitative and objective. While humans can deliberately change the information or even lie, sensors are also prone to errors and therefore hard information items are not always accurate.

For human agents, the source is part of the real world, (a community, a scene, an event) and can be either directly involved in the events reported, or just serving as a witness.

**Definition of reported information:** Reported information is a couple  $(I, \chi(I))$ , where  $I$  is an item of information and  $\chi(I)$  the confidence level as assigned by the source. Items are information pieces that can be extracted from natural language sentences, although the extraction and separation from subjective content are out of the scope for the model developed. Each item  $I$  has assertive  $i_a$  and subjective  $i_s$  components conveying factual and subjective contents respectively.

The analysis of reported information continues to be an open topic as the fusion of information from soft sources receives increasing attention in recent years. Although some authors have developed logic-based approaches for modelling distortions of items exchanged between agents who have both the intention and the ability to deceive [31], there are still more challenges arising when the information is analyzed in its textual form.

Features of uncertainty, as expressed in natural language statements, are analyzed in [54] while [55] provides a broader discussion of pitfalls and challenges related to soft data integration for information fusion.

#### B. Functions of the trust model

The model introduces several functions estimating features of reliability, self-confidence and trust, as described hereafter.

**Definition of a reliability function:** A reliability function is a mapping which assigns a real value to an information source.

This real value is a quantitative characterization of the source, inferred with respect to the source's previous failures, its reputation and the relevance of its skills for specific domains. For this model, the reliability of human sources combines three features: competence, reputation and intention. Competence captures the intuition that the quality of information reported by a source depends on the level of training and expertise, which may be designated as satisfactory or not, depending upon the task. Reputation is the overall quality of a source, estimated by examination of the history of its previous failures. Intentions refer to attitudes or purposes, often defined with respect to a hidden purpose or plan to achieve.

Reliability is a complex concept and, from a practical standpoint, it is difficult to have complete information about the global reliability of a source. Thus, this model describes reliability along the three attributes (competence of a source, its reputation and its intentions) described above. In practical applications, this solution allows for compensation for insufficient information on one or several aspects of reliability and

to conduct, if necessary, the analysis of reliability based on just one attribute.

**Evaluation of reliability:** Assessing reliability is of real interest when opportunistic sources are considered because the analyst has neither an indication of how the source might behave nor the ability to monitor or control either the human providing the information or the environment in which the source operates. Various methods can be developed to estimate competence, reputation and intentions of the source. For example, competence is closely related to the level of training of an observer or can be defined by domain knowledge. Values can be expressed either in a linguistic form (*bad, good, fair, unknown*) or by a number. Reputation is an attribute which can be constructed not just by examining previous failures of the source but also by considering its level of conflict with other sources; this too can be expressed by numeric or symbolic values.

While reputation and competence can be, at least in some cases, estimated from prior knowledge, characterizing the intentions of a source is subject to human perception and analysis. Judgment of human experts is needed not just because there usually is no *a priori* characterization of the source with respect to its intentions but also because it is important to assess those aspects from the subjective point of view of an expert in the form of binary values only.

From a practical standpoint, it is suitable to provide an expert with a description of source competence, reputation and intentions as assessed independently. This way, experts can have the opportunity to develop different strategies of using reliability: they can decide to assign different importance to those attributes under different contexts or can use their own hierarchy of attributes. For instance, an expert may consider as irrelevant the information provided by a source whose competences is lower than a specific threshold or if he suspects the source of having malicious intentions.

**Definition of a self-confidence function:** A self-confidence function is a mapping linking a real value and an information item. The real value is a measure of the information credibility as evaluated by the sensor itself and is of particular interest for human sources, as often such sources provide their own assessments of the information conveyed. Identifying features of self-confidence requires methods related to a research task of natural language processing: the identification of assertions and opinions in texts. In this field, the commonly adopted separation of those notions considers assertions as statements that can be proven true or false, while opinions are hypotheses, assumptions and theories based on someone's thoughts and feelings and cannot be proven.

**Evaluation of self-confidence:** The estimation of self-confidence aims at assigning a numerical value which captures how strongly the author stands behind assertions in the statement, on the basis of lexical clues he has included in the utterance. More generally, markers of an author's commitment are in the form of hedges, modal verbs and forms of passive/

active language. A hedge is a mitigating word that modifies the commitment to the truth of propositions, i.e., certainly, possibly. Its impact can be magnified by a booster (highly likely) or weakened by a downtoner (rather certain).

Modal verbs indicate if something is plausible, possible, or certain (*John could be a terrorist, you might be wrong*). Moreover, in some domains sentences making use of the passive voice are considered as an indicator of uncertainty, in the sense that author seeks to distance himself from the assertions in the items reported through use of passive voice. Quantifying self-confidence is a topic of particular interest for intelligence analysis, and it was early addressed by Kent in 1962, [56] who created a standardized list of words of estimative probability which were widely used by intelligence analysts. This list has continued to be a common basis to be used by analysts to produce uncertainty assessments. Kesselman describes in [57] a study conducted to analyze the way the list was used by analysts over the past, and identifies new trends to convey estimations and proposes a new list having the verb as a central element. Given the variety of linguistic markers for uncertainty, the estimation of a numerical value based on every possible combination seems unrealistic, as the same sentence often contains not just one but multiple expressions of uncertainty. Additionally, assigning numerical values to lexical expressions is not an intuitive task, and Rein shows that there are no universal values to be associated in a unique manner to hedges or other uncertainty markers, see [58]. As the author argues further, it is, however, possible to order those expressions and use this relative ordering as a more robust way to compare combinations of uncertainty expressions, and thus highlight different levels of uncertainty in natural language statements.

**Using the model for trust analysis:** The model proposed in this work proposed in this work combines various attributes of the source (discussed previously under "reliability") with "self-confidence" in order to capture trust of information as conveyed by the human. The model is source-centric predominantly focused on the source's ability to correct, alter or qualify the information report. Although the rules for ranking, prioritizing and combining the attributes introduced by the model can be drafted empirically, the estimation of a trust value requires a formal representation of the model.

A possible solution for estimating a unified value for trust is to consider reliability and self-confidence within the framework of an uncertainty theory and to rely on the set of combination rules the theory defines - for example, those developed in probability theory, in possibility theory, or in belief functions theory. All these theories provide various operators to combine reliability and self-confidence in order to estimate trust.

In the following the model is represented by using belief functions and several scenarios are used to illustrate trust estimation.

## V. TRUST FORMALIZATION WITH BELIEF FUNCTIONS

The aim of trust formalization is to provide a formal representation of the model, combining the capability to exploit the structure and relationship of elements of the model with the ability to express degrees of uncertainty about those elements. Of particular interest to this paper is the observation that the developed model introduces a cognitive view of trust as a complex structure of beliefs that are influenced by the individual's opinions about certain features and elements, including their own stances. Such a structure of beliefs determines various degrees of trust, which are based on personal choices made by analyst, on the one hand, and the source, on the other hand. Therefore, the formalization requires a formalism that is more general than probability measures or fuzzy category representation, which are more suitable for applications considering trust in the context of interactions between agents. Moreover, the limitations of using subjective probabilities to formalize trust from this cognitive standpoint are clearly stated in [13]. As a result, the model was represented with belief functions, a formalism that is consistent with the cognitive perspective of trust adopted by the model. This belief-based representation provides the most direct correspondence with elements of the model and their underlying uncertainty, while being able to quantify subjective judgments.

After introducing main concepts of belief functions, this section shows how the formalism is used to represent the trust model.

## A. Basic Belief Assignment

Belief Functions (BF) have been introduced by Shafer in his his mathematical theory of evidence [59], also referred to Dempster-Shafer Theory (DST), to model epistemic uncertainty. The frame of discernment (FoD) of the decision problem under consideration, denoted  $\Theta$ , is a finite set of exhaustive and mutually exclusive elements. The powerset of  $\Theta$  denoted  $2^\Theta$  is the set of all subsets of  $\Theta$ , empty set included. A body of evidence is a source of information characterized by a Basic Belief Assignment (BBA), or a mass function, which is the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  that satisfies  $m(\emptyset) = 0$ , and the normalization condition  $\sum_{A \in 2^\Theta} m(A) = 1$ . The belief (a.k.a credibility)  $Bel(\cdot)$  and plausibility  $Pl(\cdot)$  function, usually interpreted as lower and upper bounds of unknown (subjective) probability measure  $P(\cdot)$ , are defined from  $m(\cdot)$  respectively by

$$Bel(A) = \sum_{B \subseteq A | B \in 2^\Theta} m(B), \quad (1)$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset | B \in 2^\Theta} m(B). \quad (2)$$

An element  $A \in 2^\Theta$  is called a focal element of the BBA  $m(\cdot)$ , if and only if  $m(A) > 0$ . The set of all focal elements of  $m(\cdot)$  is called the core of  $m(\cdot)$  and is denoted  $\mathcal{K}(m)$ . This formalism allows for modeling a completely ignorant source

by taking  $m(\Theta) = 1$ . The Belief Interval (BI) of any element  $A$  of  $2^\Theta$  is defined by

$$BI(A) \triangleq [Bel(A), Pl(A)]. \quad (3)$$

The width of belief interval of  $A$ , denoted  $U(A) = Pl(A) - Bel(A)$  characterizes the degree of imprecision of the unknown probability  $P(A)$ , often called the uncertainty of  $A$ . We define the uncertainty (or imprecision) index by

$$U(m) \triangleq \sum_{A \in \Theta} U(A), \quad (4)$$

to characterize the overall imprecision of the subjective (unknown) probabilities committed to elements of the FoD bounded by the belief intervals computed with the BBA  $m(\cdot)$ .

Shafer proposed using Dempster's rule of combination for combining multiple independent sources of evidence [59] which is the normalized conjunctive fusion rule. This rule has been strongly disputed in the BF community after Zadeh's first criticism in 1979, and since the 1990s many rules have been proposed to combine (more or less efficiently) BBAs; the reader is advised to see discussions in [60], in particular the proportional conflict redistribution rule number 6 (PCR6). To combine the BBAs we use the proportional conflict redistribution (PCR) rule number 6 (denoted PCR6) proposed by Martin and Osswald in [60] because it provides better fusion results than Dempster's rule in situations characterized by both high and low conflict as explained in detail in [61], [62].

The PCR6 rule is based on the PCR principle which transfers the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses, so that the specificity of the information is entirely preserved. The steps in applying the PCR6 rule are:

- 1) apply the conjunctive rule;
- 2) calculate the total or partial conflicting masses; and
- 3) redistribute the (total or partial) conflicting mass proportionally on non-empty sets.

The general PCR6 formula for the combination of  $n > 2$  BBAs is very complicated (see [60] Vol. 2, Chap. 2). For convenience's sake, we give here just the PCR6 formula for the combination of only two BBAs. When we consider two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined on the same FoD  $\Theta$ , the PCR6 fusion of these two BBAs is expressed as  $m_{PCR6}(\emptyset) = 0$  and for all  $X \neq \emptyset$  in  $2^\Theta$

$$m_{PCR6}(X) = \sum_{\substack{x_1, x_2 \in 2^\Theta \\ x_1 \cap x_2 = X}} m_1(x_1)m_2(x_2) + \sum_{\substack{Y \in 2^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right], \quad (5)$$

where all denominators in (5) are different from zero. If a denominator is zero, that fraction is discarded. A very basic (not optimized) Matlab code implementing the PCR6 rule can

be found in [60] and [63], and also in the toolboxes repository on the web<sup>2</sup>.

Instead of working with quantitative (numerical) BBA, it is also possible to work with qualitative BBA expressed by labels using the linear algebra of refined labels proposed in Dezert-Smarandache Theory (DSMT), [60] (Vol. 2 & 3).

### B. Trust formalization model

Because beliefs are well defined mathematical concepts in the theory of belief functions, we prefer to use self-confidence terminology to represent the confidence declared by a source  $Y$  on its own assertion  $A$ . Let's denote by  $A$  the assertion given by the source, for instance  $A = \text{John is a terrorist}$ . With respect to elements of the model,  $A$  (the assertion) corresponds to  $i_a$ , the assertive part of the item  $I$  and  $v(A)$  is a numeric estimation of the subjective  $i_s$  component of  $I$ .

The valuation  $v(A)$  made by the source  $Y$  about the assertion  $A$  can be done either quantitatively (by a probability or a BBA) or qualitatively (by a label associated to a linguistic form). This paper considers quantitative representation of  $v(A)$  for simplicity<sup>3</sup>.

The basic information items provided by a source  $Y$  consists of  $A$  (the assertion), and  $v(A)$  (its valuation). To be as general as possible, we suppose that  $v(A)$  is a basic belief mass assignment defined with respect to the very basic frame of discernment  $\Theta_A \triangleq \{A, \bar{A}\}$  where  $\bar{A}$  denotes the complement of  $A$  in  $\Theta_A$ , that is  $v(A) = (m(A), m(\bar{A}), m(A \cup \bar{A}))$ . Note that only two values of the triplet are really necessary to define  $v(A)$  because the third one is automatically derived from the normalization condition  $m(A) + m(\bar{A}) + m(A \cup \bar{A}) = 1$ . So one could also have chosen equivalently  $v(A) = [Bel(A), Pl(A)]$  instead of the BBA. In a probabilistic context, one will take  $m(A \cup \bar{A}) = 0$  and so  $v(A) = P(A)$  because  $Bel(A) = Pl(A) = P(A)$  in such a case.

The self-confidence of the source  $Y$  is an extra factor  $\alpha_Y \in [0, 1]$  which characterizes the self-estimation of the quality of the piece of information  $(A, v(A))$  provided by the source itself.  $\alpha_Y = 1$  means that the source  $Y$  is 100% confident in his valuation  $v(A)$  about assertion  $A$ , and  $\alpha_Y = 0$  means that the source  $Y$  is not at all confident in his valuation  $v(A)$ . In the theory of belief functions, this factor is often referred as the discounting factor of the source because this factor is usually used to discount the original piece of information  $(A, v(A))$  into a discounted one  $(A, v'(A))$  as follows [59]:

$$m'(A) = \alpha_Y \cdot m(A), \quad (6)$$

$$m'(\bar{A}) = \alpha_Y \cdot m(\bar{A}), \quad (7)$$

$$m'(A \cup \bar{A}) = \alpha_Y \cdot m(A \cup \bar{A}) + (1 - \alpha_Y). \quad (8)$$

The idea of Shafer's discounting technique is to diminish the belief mass of all focal elements with the factor  $\alpha_Y$  and redistribute the missing discounted mass  $(1 - \alpha_Y)$  to the whole

ignorance  $A \cup \bar{A}$ . Note that the valuation of the discounted piece of information is always degraded because its uncertainty index is always greater than the original one, that is,  $U(m') > U(m)$ , which is normal.

The reliability factor  $r$  estimated by the analyst  $X$  on the piece of information  $(A, v(A))$  provided by the source  $Y$  must take into account both the competence  $C_Y$ , the reputation  $R_Y$  and the intention  $I_Y$  of the source  $Y$ . A simple model to establish the reliability factor  $r$  is to consider that  $C_Y$ ,  $R_Y$  and  $I_Y$  factors are represented by numbers  $[0, 1]$  associated to select subjective probabilities, that is  $C_Y = P(Y \text{ is competent})$ ,  $R_Y = P(Y \text{ has a good reputation})$  and  $I_Y = P(Y \text{ has a good intention (i.e. is fair)})$ . If each of these factors has equal weight, then one could use  $r = C_Y \times R_Y \times I_Y$  as a simple product of probabilities. However, in practice, such simple modeling does not fit well with what the analyst really needs to take into account epistemic uncertainties in Competence, Reputation and Intention. In fact, each of these factors can be viewed as a specific criterion influencing the level of the global reliability factor  $r$ . This is a multi-criteria valuation problem. Here we propose a method to solve the problem.

We consider the three criteria  $C_Y$ ,  $R_Y$  and  $I_Y$  with their associated importance weights  $w_C$ ,  $w_R$ ,  $w_I$  in  $[0, 1]$  with  $w_C + w_R + w_I = 1$ . We consider the frame of discernment  $\Theta_r = \{r, \bar{r}\}$  about the reliability of the source  $Y$ , where  $r$  means that the source  $Y$  is reliable, and  $\bar{r}$  means that the source  $Y$  is definitely not reliable. Each criteria provides a valuation on  $r$  expressed by a corresponding BBA. Hence, for the competence criteria  $C_Y$ , one has  $(m_C(r), m_C(\bar{r}), m_C(r \cup \bar{r}))$ , while for the reputation criteria  $R_Y$ , one has  $(m_R(r), m_R(\bar{r}), m_R(r \cup \bar{r}))$  and for the intention criteria  $I_Y$ , one has  $(m_I(r), m_I(\bar{r}), m_I(r \cup \bar{r}))$ .

To get the final valuation of the reliability  $r$  of the source  $Y$ , one needs to efficiently fuse the three BBAs  $m_C(\cdot)$ ,  $m_R(\cdot)$  and  $m_I(\cdot)$ , taking into account their importance weights  $w_C$ ,  $w_R$ , and  $w_I$ . This fusion problem can be solved by applying the importance discounting approach combined with PCR6 fusion rule of DSMT [63] to get the resultant valuation  $v(r) = (m_{PCR6}(r), m_{PCR6}(\bar{r}), m_{PCR6}(r \cup \bar{r}))$  from which the decision ( $r$ , or  $\bar{r}$ ) can be drawn (using BI distance, for instance). If a firm decision is not required, an approximate probability  $P(r)$  can also be inferred with some lossy transformations of BBA to probability measure [60]. Note that Dempster's rule of combination cannot be used here because it does not respond to the importance discounting, as explained in [63].

The trust model consists of the piece of information  $(A, v(A))$  and the self-confidence factor  $\alpha_Y$  provided by the source  $Y$ , as well as the reliability valuation  $v(r)$  expressed by the BBA  $(m(r), m(\bar{r}), m(r \cup \bar{r}))$  to infer the trust valuation about the assertion  $A$ . For this, we propose using the mass  $m(r)$  of reliability hypothesis  $r$  of the source  $Y$  as a new discounting factor for the BBA  $m'(\cdot)$  reported by the source  $Y$ , taking into account its self-confidence  $\alpha_Y$ . Hence, the trust valuation  $v_t(A) = (m_t(A), m_t(\bar{A}), m_t(A \cup \bar{A}))$  of assertion

<sup>2</sup>[http://bfaswiki.iut-lannion.fr/wiki/index.php/Main\\_Page](http://bfaswiki.iut-lannion.fr/wiki/index.php/Main_Page)

<sup>3</sup>Without loss of generality one can always map a qualitative representation to a quantitative one by a proper choice of scaling and normalization (if necessary).

$A$  for the analyst  $X$  is defined by

$$m_t(A) = m(r) \cdot m'(A), \quad (9)$$

$$m_t(\bar{A}) = m(r) \cdot m'(\bar{A}), \quad (10)$$

$$m_t(A \cup \bar{A}) = m(r) \cdot m'(A \cup \bar{A}) + (1 - m(r)), \quad (11)$$

or equivalently by

$$m_t(A) = m(r)\alpha_Y \cdot m(A), \quad (12)$$

$$m_t(\bar{A}) = m(r)\alpha_Y \cdot m(\bar{A}), \quad (13)$$

$$m_t(A \cup \bar{A}) = m(r)\alpha_Y \cdot m(A \cup \bar{A}) + (1 - m(r)\alpha_Y). \quad (14)$$

The DSMT framework using the PCR6 fusion rule and the importance discounting technique provides an interesting solution for the fusion of attributes having different degrees of importance while making a clear distinction between those attributes.

The discounting method proposed in this work is directly inspired by Shafer's classical discounting approach [59]. In our application, the classical discounting factor that we propose integrates both the mass of reliability hypothesis  $m(r)$  and the self-confidence factor  $\alpha_Y$ . It is worth noting that more sophisticated (contextual) belief discounting techniques [64] exist and they could also have been used, in theory, to refine the discounting but these techniques are much more complicated and they require additional computations. The evaluation of contextual belief discounting techniques for such types of application is left for further investigations and research works.

## VI. UNCERTAINTY ANALYSIS UNDER URREF CRITERIA

Tracking uncertainties from problem description to model construction and formalization is done under criteria of the uncertainty representation and reasoning evaluation framework.

The goal of URREF is to place the focus on the evaluation of uncertainty representation and reasoning procedures. The URREF ontology defines four main classes of evaluation criteria: Data Handling, Representation, Reasoning and Data Quality. These criteria make distinctions between the evaluation of the fusion system, the evaluation of its inputs and outputs, and the evaluation of the uncertainty representation and reasoning aspects.

Listing all criteria is an extensive task and in this paper the authors will provide one piece of the puzzle by considering criteria that relate to the evaluation of uncertainty induced by the proposed model. In the model developed in this paper, uncertainty is due to imperfections of information gathering and reporting as well as constraints of the representation formalism.

Uncertainty analysis is carried out by assigning uncertainty criteria to elements and functions of the trust model in order to make explicit the uncertainty arising when the problem is abstracted by the model and the model is then simplified in order to fulfill constraints of specific formalism, fig. 6.

The URREF criteria selected are subclasses of two main concepts: *Credibility*, a subclass under *DataCriteria*, and *EvidenceHandling*, a subclass of *RepresentationCriteria*.



Fig. 4. Trust estimation from source to analyst.

To summarize, uncertainties of the model will be captured by the following URREF criteria :

- **Objectivity**, subconcept of **Credibility**: indicates a source providing unbiased information;
- **ObservationalSensitivity**, subconcept of **Credibility**: characterizes the skills and competences of sources;
- **SelfConfidence**, subconcept of **Credibility**: measures the certainty degree about the piece of information reported, according to the source;
- **Ambiguity**, subconcept of **EvidenceHandling**: captures if the sources provide data supporting different conclusions;
- **Dissonance**, subconcept of **EvidenceHandling**: captures the ability of formalism to represent inconsistent evidence;
- **Completeness**, subconcept of **EvidenceHandling**: is a measure of how much is known given the amount of evidence; and
- **Conclusiveness**, subconcept of **EvidenceHandling**: indicates how strong the evidence supports a conclusion;

Besides selecting uncertainty criteria relevant for trust estimation, the analysis also discusses the mapping of URREF criteria to attributes of the model and sheds a light on imperfect matchings. This mapping offers a basis for identifying the limitations of the URREF ontology, by emphasizing those elements whose characterizations in terms of uncertainty are out of the ontology's reach or beyond the ontology's intended scope.

### A. Uncertainties from problem definition to model abstraction

Let  $M$  be the model for trust estimation, with elements introduced in paragraph IV: the source  $Y$ , the reported item  $I$  with its assertive  $i_a$  and subjective  $i_s$  parts, and  $\chi(I)$  the confidence level assigned by the source  $Y$  to  $I$ .

From an information fusion standpoint, inputs of the model are the source and the information items, along with their uncertainty, captured with the following URREF criteria: *Objectivity*, *ObservationalSensitivity* and *SelfConfidence*. These criteria are subclasses of the concept *InputCriteria*.

*Objectivity* is an attribute of the source, related to its ability to provide factual, unbiased items, without adding their own points of view or opinions. For a source  $Y$  providing

information item  $i$ , having  $i_s$  and  $i_a$  as the subjective and factual parts respectively, objectivity can be expressed as:

$$\text{Objectivity}(Y, I) = \psi_o(i_s, i_a), \quad (15)$$

where  $\psi_o(i_s, i_a)$  represents the mathematically quantified expression of the subjective over the factual content of  $i$ .

*ObservationalSensitivity* is an attribute of the source which represents the source's ability to provide accurate reports. In the proposed model, this criterion is an aggregation of competence  $C$  and reputation  $R$ , two attributes of the model.

$$\text{ObservationalSensitivity}(Y, i) = \psi_{os}(C, R), \quad (16)$$

where  $\psi_{os}(C, R)$  is a function aggregating values of competence and reputation.

Information items entering the system are described by *SelfConfidence*. Again, considering  $i_s$  and  $i_a$  as the subjective and factual items conveyed by  $I$ , *SelfConfidence* can be expressed as:

$$\text{SelfConfidence}(I) = \psi_{sc}(i_s), \quad (17)$$

with  $\psi_{sc}(i_s)$  a function quantifying the subjective content of item  $I$ .

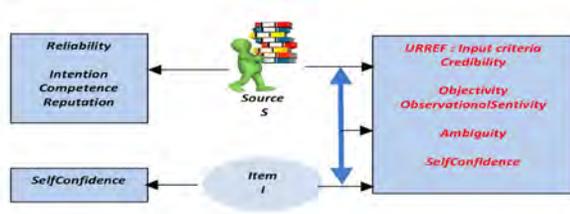


Fig. 5. Mapping of model attributes to URREF criteria.

Fig. 5 shows the mapping between the elements of the model and the set of relevant URREF uncertainty criteria. The mapping shows a perfect match between *SelfConfidence* as introduced by the model and the eponymous URREF criterion as well as several imperfect matches described later in this paper.

At source level, URREF criteria are not able to capture in a distinct manner the features of competence, reputation and intentions, the main attributes of the sources added by the model under Reliability. To some extent, competence and reputation can be related to *ObservationalSensitivity*, but intentions clearly remains out of reach for URREF criteria.

### B. Uncertainties from model to formal representation

Let  $F$  be the DST formalization of the trust estimation model, with parameters introduced in paragraph V. The formalism induces two types of uncertainty related to its capacity to handle incomplete, ambiguous or contradictory evidence. The uncertainty of evidence handling is captured by *Ambiguity*, *Dissonance*, *Conclusiveness* and *Completeness*. Those criteria are subclasses of the concept *EvidenceHandling*.

*Ambiguity* measures the extent to which the formalism can handle data sets which support different conclusions.

$$\text{Ambiguity}(F) = \phi_a(\alpha_Y, R_Y), \quad (18)$$

where the function  $\phi_a(\alpha_Y, R_Y)$  considers the self-confidence factor  $\alpha_Y$  provided by the source  $Y$  and the reliability of  $Y$  provided by the analyst  $R_Y$  to estimate the degree of ambiguity. The measure is of particular interest in the case where items having high values of self-confidence are provided by unreliable sources.

*Dissonance* captures the ability of the formalism to represent inconsistent evidence. For BBA representations, dissonance can be related to the capacity of the formalism to assign belief mass to an element and its negation, and can therefore be assessed for every BBA representation build for the model. For example, the dissonance for a source's competence can be in the form:

$$\text{Dissonance}(F) = \phi_d(m_C(r), m_C(\bar{r})), \quad (19)$$

where  $\phi_d(m_C(r), m_C(\bar{r}))$  is a function combining the belief mass assigned to whether the source is considered to be competent or incompetent, respectively.

Dissonance is useful for highlighting situations in which there are significant differences in belief masses assigned at the attribute level, such as when a source is considered to be incompetent (low  $m_C(r)$ , high  $m_C(\bar{r})$ ) but has a good reputation (high  $m_R(r)$ , low  $m_R(\bar{r})$ ).

*Conclusiveness* is a measure expressing how strongly the evidence supports a specific conclusion or unique hypothesis:

$$\text{Conc.}(F) = \phi_{cc}(m_t(A), m_t(\bar{A}), m_t(A \cup \bar{A})), \quad (20)$$

where  $\phi_{cc}(m_t(A), m_t(\bar{A}), m_t(A \cup \bar{A}))$  is a function combining the belief masses estimated for truthful, untruthful and unknown qualifications of assertion  $A$  respectively. This measure indicates to which extent the result of inferences can support a conclusion, in this case whether the hypothesis that the assertion under analysis is trustworthy or not. It can be used during the inference process to show how taking into account additional elements such as the competence of the source, its reputation or intentions impact the partial estimations of trust.

*Completeness* is a measures of the range of the available evidence, and captures the ability of formalism to take into account how much is unknown. The measures is somewhat similar to *Dissonance*, as is can be assessed for every BBA representation build for the model. Thus, completeness of source's reliability is described as:

$$\text{Completeness}(F) = \phi_{cp}(m(r \cup \bar{r})), \quad (21)$$

where  $\phi_{cp}(m(r \cup \bar{r}))$  is a function depending on the belief mass assigned to unknown.

The measure is used for estimation and analysis before entering the fusion process, in order to have a picture of how complete the evidence describing the various elements of the model is, and to avoid performing fusion on highly incomplete data sets. Both *EvidenceHandling* and *KnowledgeHandling* are subclasses of *RepresentationCriteria*.

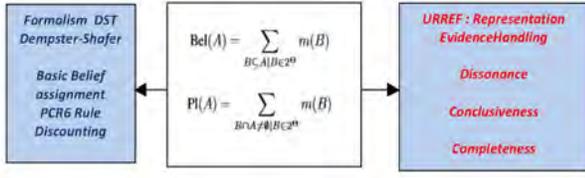


Fig. 6. Mapping of formalism uncertainties to URREF criteria.

This section has analyzed the nature of uncertainties arising when going from problem to model definition and then on to formalization with belief functions. The next section shows how uncertainties can be highlighted for particular scenarios of trust estimation.

## VII. UNCERTAINTY ANALYSIS FOR TRUST ESTIMATION

### A. Running example and method for uncertainty tracking

As a running example, let's consider an assertion  $A$  and its valuation  $v(A)$  provided by the source  $Y$  as follows:  $m(A) = 0.7$ ,  $m(\bar{A}) = 0.1$  and  $m(A \cup \bar{A}) = 0.2$ . Its self-confidence factor is  $\alpha_Y = 0.75$ . Hence, the discounted BBA  $m'(\cdot)$  is given by

$$\begin{aligned} m'(A) &= 0.75 \cdot 0.7 = 0.525 \\ m'(\bar{A}) &= 0.75 \cdot 0.1 = 0.075 \\ m'(A \cup \bar{A}) &= 1 - m'(A) - m'(\bar{A}) = 0.4 \end{aligned}$$

Let's assume that the BBAs about the reliability of the source based on Competence, Reputation and Intention criteria are given as follows:

$$\begin{aligned} m_C(r) &= 0.8, m_C(\bar{r}) = 0.1, m_C(r \cup \bar{r}) = 0.1, \\ m_R(r) &= 0.7, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0.2, \\ m_I(r) &= 0.6, m_I(\bar{r}) = 0.3, m_I(r \cup \bar{r}) = 0.1, \end{aligned}$$

with importance weights  $w_I = 0.6$ ,  $w_R = 0.2$  and  $w_C = 0.2$ .

After applying the importance discounting technique presented in [63] which consists of discounting the BBAs with the importance factor and redistributing the missing mass onto the empty set, then combining the discounted BBAs with PCR6 fusion rule, we finally get, after normalization, the following BBA

$$\begin{aligned} m(r) &= 0.9335, \\ m(\bar{r}) &= 0.0415, \\ m(r \cup \bar{r}) &= 1 - m(r) - m(\bar{r}) = 0.025. \end{aligned}$$

The final trust valuation of assertion  $A$  reported by the source  $Y$  taking into account its self-confidence  $\alpha_Y = 0.75$  and the reliability factor  $m(r)$  is therefore given by Eqs. (12)–(14) and obtaining

$$\begin{aligned} m_t(A) &= 0.4901, \\ m_t(\bar{A}) &= 0.0700, \\ m_t(A \cup \bar{A}) &= 1 - m_t(A) - m_t(\bar{A}) = 0.4399. \end{aligned}$$

Note that if  $m_C(r) = m_R(r) = m_I(r) = 1$ , then we will always get  $m(r) = 1$  regardless of the choice of weightings factors, which is normal. If there is a total conflict between valuations of reliability based on Competence, Reputation and Intention criteria, then Dempster's rule cannot be applied to get the global reliability factor  $m(r)$  because of  $0/0$  indeterminacy in the formula of Dempster's rule. For instance, if one has  $m_C(r) = m_R(r) = 1$  and  $m_I(\bar{r}) = 1$ , then  $m(r)$  is indeterminate with Dempster's rule of combination, whereas it corresponds to the average value  $m(r) = 2/3$  using PCR6 fusion rule (assuming equal importance weights  $w_C = w_R = w_I = 1/3$ ), which makes more sense.

The following subsections explore several scenarios for trust assessment, corresponding to different situations of BBAs distributions, and track the uncertainty according to URREF criteria. Each scenario illustrates specific instances of the model developed for trust estimation.

The method adopted to track uncertainty defines the following measures to estimate URREF criteria:

$$\begin{aligned} SelfConfidence &= \alpha_Y, \\ Ambiguity &= |\alpha_Y - m(r)|, \\ Objectivity &= m_I(r), \\ ObservationalSensitivity &= \min(m_C(r), m_R(r)). \end{aligned}$$

As shown in previous formulas, URREF criteria are estimated based on features of the BBA formalization and are assigned to the static elements of the model, i.e., the source and the information item. While *Objectivity* and *ObservationalSensitivity* captures imperfections of observations, *SelfConfidence* and *Ambiguity* reflect inaccuracies in reporting information to analysts. These criteria are assessed before entering the fusion phase, and describe the initial uncertainty present in the system before inferences.

In addition, *Dissonance*, *Conclusiveness* and *Completeness* will be estimated at the scenario level by adopting the following formulas:

$$\begin{aligned} Dissonance &= 1 - |m_t(A) - m_t(\bar{A})|, \\ Conclusiveness &= |m_t(A) - m_t(\bar{A})|, \\ Completeness &= 1 - m(A \cup \bar{A}). \end{aligned}$$

Criteria above will be assessed for elements impacted by the fusion process: the reliability of the source, the updated BBAs of the initial assertion and estimated trust. In the following subsection we illustrate several scenarios for trust estimation and the uncertainty analysis underlying each scenario.

### B. Scenarios for trust assessment and uncertainty analysis

Scenarios introduced below provide examples of trust construction using various operators and highlight the uncertainty assigned to elements of the model and its propagation during the fusion process.

**Scenario 1 - Consensus:** Suppose that  $Y$  provides the assertion  $A$ , while stating that  $A$  certainly holds and that  $X$  considers  $Y$  to be a reliable source.

In this case, the trust will be constructed on the basis of two consensual opinions: the analyst  $X$  that considers  $Y$  as a

reliable source, and the source’s conviction that the information provided is certain. In this case,  $m(A) = 1$ ,  $\alpha_Y = 1$  and  $m(r) = 1$ , so that  $m'(A) = 1$  and  $m_t(A) = m(r) \cdot m'(A) = 1$ . The result will be in the form  $(A, v(A))$  initially provided by the source.

Uncertainty of inputs		
2*Observation	Objectivity	1
	ObservationalSensitivity	1
2*Reporting	SelfConfidence	1
	Ambiguity	0

TABLE I  
CONSENSUS: INPUT UNCERTAINTY.

Fusion uncertainty	Dissonance	Conclusiv.	Comple.
Updated BBAs	0	1	1
Reliability	0	1	1
Trust	0	1	1

TABLE II  
CONSENSUS: FUSION UNCERTAINTY.

This scenario illustrates an ideal situation for trust assessment, where the source is trustworthy and well known to the analyst, and observations are reported in perfect conditions. As shown in table I, there is no uncertainty induced by the source, and once fusion is performed the items impacted show high values for conclusiveness and completeness, while dissonance is 0 for the updates BBAs for values, source’s reliability and estimated trust, as shown in table II.

**Scenario 2 - Uncertain utterances:**  $Y$  is considered by  $X$  to be a reliable source and reports the assertion  $A$ , while showing a low level of certainty  $v(A)$  about the veracity of  $A$ . This example is relevant for situations where a reliable source provides (possibly) inaccurate descriptions of events due to, say, bad conditions for observation. This scenario corresponds by example to the following case for inputs:  $\alpha_Y = 0.6$

$$m(A) = 0.8, m(\bar{A}) = 0.1, m(A \cup \bar{A}) = 0.1,$$

$$m_C(r) = 0.9, m_C(\bar{r}) = 0, m_C(r \cup \bar{r}) = 0.1,$$

$$m_R(r) = 0.9, m_R(\bar{r}) = 0, m_R(r \cup \bar{r}) = 0.1,$$

$$m_I(r) = 0.3, m_I(\bar{r}) = 0.3, m_I(r \cup \bar{r}) = 0.6,$$

and  $w_C = 0.5$ ,  $w_R = 0.5$  and  $w_I = 0$ .  
This results in

$$m'(A) = 0.48, m'(\bar{A}) = 0.06, m'(A \cup \bar{A}) = 0.46,$$

and

$$m(r) = 0.9846, m(\bar{r}) = 0, m(r \cup \bar{r}) = 0.0154.$$

Therefore, one finally obtains the trust valuation

$$m_t(A) = 0.47, m_t(\bar{A}) = 0.05, m_t(A \cup \bar{A}) = 0.46.$$

This case shows that self-confidence has an important impact on the values of discounted BBA, as  $m'(A)$  is decreased

from 0.8 to 0.48, and thus the remaining mass is redistributed on  $m'(A \cup \bar{A})$ .

The combination of competence, reliability and intention are in line with the assumption of the scenario, which states that  $Y$  is a reliable source. After normalization, values for trust assessment clearly highlight the impact of uncertain utterances, as the BBA shows a mass transfer from  $m_t(A)$  to  $m_t(A \cup \bar{A})$ . Still, values of trust are close to BBA integrating the self-confidence, which confirms the intuition that when the analyst  $X$  considers  $Y$  to be a reliable source, the assertion  $A$  is accepted with an overall trust level almost equal to the certainty level stated by the source.

Uncertainty of inputs		
2*Observation	Objectivity	0.3
	ObservationalSensitivity	0.9
2*Reporting	SelfConfidence	0.6
	Ambiguity	0.38

TABLE III  
UNCERTAIN UTTERING: INPUT UNCERTAINTY.

Fusion uncertainty	Dissonance	Conclusiv.	Comple.
Updates BBAs	0.3	0.7	0.9
Reliability	0.02	0.98	0.98
Trust	0.59	0.41	0.54

TABLE IV  
UNCERTAIN UTTERANCE: FUSION UNCERTAINTY.

This scenario illustrates uncertainty induced by observations failures, as *Objectivity*, and *SelfConfidence* are low, see table III.

While the quality of the source is highlighted by high values of *Conclusiveness* and *Completeness*, showing the analyst’s confidence in the reports analyzed, the impact of imperfect observation is shown in the overall estimation of trust, through a combination of *Dissonance*, *Conclusiveness* and *Completeness* which have values close to 0.5, see table IV.

**Scenario 3 - Reputation:** Suppose that  $Y$  provides  $A$  and  $v(A)$  and  $X$  has no global description of  $Y$  in terms of reliability. As the reliability of  $Y$  is not available,  $Y$ ’s reputation will be used instead, as derived from historical data and previous failures. This scenario corresponds by example to the following case for inputs:  $\alpha_Y = 1$

$$m(A) = 0.8, m(\bar{A}) = 0.1, m(A \cup \bar{A}) = 0.1,$$

$$m_C(r) = 0.1, m_C(\bar{r}) = 0.1, m_C(r \cup \bar{r}) = 0.8,$$

$$m_R(r) = 0.9, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0,$$

$$m_I(r) = 0.1, m_I(\bar{r}) = 0.1, m_I(r \cup \bar{r}) = 0.8,$$

and  $w_C = 0.1$ ,  $w_R = 0.8$  and  $w_I = 0.1$ .

Hence, one gets

$$m'(A) = 0.8, m'(\bar{A}) = 0.1, m'(A \cup \bar{A}) = 0.1,$$

and

$$m(r) = 0.94, m(\bar{r}) = 0.01, m(r \cup \bar{r}) = 0.03.$$

Therefore, one finally obtains the trust valuation

$$m_t(A) = 0.75, m_t(\bar{A}) = 0.09, m_t(A \cup \bar{A}) = 0.14.$$

For this scenario, the source is confident about their own assertions, and therefore

$$m(A) = 0.8, m(\bar{A}) = 0.1, m(A \cup \bar{A}) = 0.1$$

and

$$m'(A) = 0.8, m'(\bar{A}) = 0.1, m'(A \cup \bar{A}) = 0.1$$

have identical BBA distributions. The reliability of the source is built namely on its reputation, as there are clues about the competence and intentions of the source. Hence, the overall BBA

$$m(r) = 0.9449, m(\bar{r}) = 0.0196, m(r \cup \bar{r}) = 0.0355$$

is close to the initial reputation distribution

$$m_R(r) = 0.9, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0$$

Values of trust show the impact of using not completely reliable sources, which decreased the certainty level of the initial BBA

$$m'(A) = 0.8, m'(\bar{A}) = 0.1, m'(A \cup \bar{A}) = 0.1$$

to

$$m_t(A) = 0.75, m_t(\bar{A}) = 0.09, m_t(A \cup \bar{A}) = 0.14$$

They also support the intuition that the trust assigned by the analyst to  $A$  will have an upper limit equal to the reputation of the source.

Uncertainty of inputs		
2*Observation	Objectivity	0.10
	ObservationalSensitivity	0.10
2*Reporting	SelfConfidence	1
	Ambiguity	0.60

TABLE V  
REPUTATION: INPUT UNCERTAINTY.

Fusion uncertainty	Dissonance	Conclusiv.	Compleat.
Updated BBAs	0.30	0.70	0.90
Reliability	0.07	0.93	0.95
Trust	0.34	0.66	0.84

TABLE VI  
REPUTATION: FUSION UNCERTAINTY.

This scenario is similar the previous one as, in both cases, there are incomplete descriptions of the source. For this particular case, a historical recording of source's failures offers a basis to overcome the missing pieces and, in spite of low values for *Objectivity* and *ObservationalSensitivity* (see table V), the final trust evaluation is improved with respect to the previous scenario and shows a better combination of *Dissonance*, *Conclusiveness* and *Completeness*, as shown in table VI.

**Scenario 4 - Misleading report:** In this case,  $Y$  provides the assertion  $A$ , while stating that it certainly holds and  $X$  considers  $Y$  to be a completely unreliable source. For this case, the analyst knows that the report is somehow inaccurate, for example, it cannot be corroborated or it contradicts, at least in part. information from other (more reliable) sources. The analyst suspects the source of having misleading intentions, and can therefore assign a maximal uncertainty level to the information reported. This scenario corresponds by example to the following case for inputs:  $\alpha_Y = 1$

$$\begin{aligned} m(A) &= 1, m(\bar{A}) = 0, m(A \cup \bar{A}) = 0, \\ m_C(r) &= 0.1, m_C(\bar{r}) = 0.1, m_C(r \cup \bar{r}) = 0.8, \\ m_R(r) &= 0.1, m_R(\bar{r}) = 0.1, m_R(r \cup \bar{r}) = 0.8, \\ m_I(r) &= 0.1, m_I(\bar{r}) = 0.8, m_I(r \cup \bar{r}) = 0.1, \end{aligned}$$

and  $w_C = 0.1, w_R = 0.1$  and  $w_I = 0.8$ . Hence, one gets

$$m'(A) = 1, m'(\bar{A}) = 0, m'(A \cup \bar{A}) = 0,$$

and

$$m(r) = 0.02, m(\bar{r}) = 0.91, m(r \cup \bar{r}) = 0.06.$$

Therefore, one finally obtains as trust valuation

$$m_t(A) = 0.023, m_t(\bar{A}) = 0, m_t(A \cup \bar{A}) = 0.976.$$

The values for this scenario reflect the high self-confidence of the source and high accuracy of the assertion provided; therefore, the initial BBA is unchanged after fusion with self-confidence. Nevertheless, the impact of having misleading intention is visible first on the mass distribution assigned to reliability and then on the overall values of trust. With respect to the initial values

$$m(A) = 1, m(\bar{A}) = 0, m(A \cup \bar{A}) = 0,$$

and the partially fused ones

$$m'(A) = 1, m'(\bar{A}) = 0, m'(A \cup \bar{A}) = 0,$$

the integration of a misleading source transfers the mass assignment almost exclusively to  $m_t(A \cup \bar{A})$ . Intuitively, the assertion  $A$  will be ignored, as the reliability of the source is dramatically decreased by a high mass assignment on misleading intentions.

Uncertainty of inputs		
2*Observation	Objectivity	0.10
	ObservationalSensitivity	0.10
2*Reporting	SelfConfidence	1.00
	Ambiguity	0.97

TABLE VII  
MISLEADING REPORT: INPUT UNCERTAINTY.

This scenario illustrates the impact of misleading sources on trust estimation. Hence, the use case has very good values for reporting induced uncertainty, with high *SelfConfidence* and low *Ambiguity* (see table VII), but the overall trust

Fusion uncertainty	Dissonance	Conclusiv.	Compleat.
Assertion	0	1	1
Source	0.11	0.89	0.93
Trust	0.76	0.23	0.03

TABLE VIII  
MISLEADING: FUSION UNCERTAINTY.

characterization shows strong *Dissonance*, corroborated with low *Conclusiveness* and near zero *Completeness*, as shown in table VIII.

**Scenario 5 - Ambiguous report:** The source  $Y$  provides  $A$  and  $v(A)$ , the uncertainty level. Suppose that  $v(A)$  has a low value, as the source is not very sure about the events reported, and that  $X$  considers  $Y$  to be unreliable. This scenario corresponds by example to the following case for inputs:  $\alpha_Y = 0.3$

$$\begin{aligned} m(A) &= 0.6, m(\bar{A}) = 0.2, m(A \cup \bar{A}) = 0.2, \\ m_C(r) &= 0.1, m_C(\bar{r}) = 0.8, m_C(r \cup \bar{r}) = 0.1, \\ m_R(r) &= 0.1, m_R(\bar{r}) = 0.8, m_R(r \cup \bar{r}) = 0.1, \\ m_I(r) &= 0.1, m_I(\bar{r}) = 0.1, m_I(r \cup \bar{r}) = 0.8, \end{aligned}$$

and  $w_C = 0.2$ ,  $w_R = 0.4$  and  $w_I = 0.4$ .  
Hence, one gets

$$m'(A) = 0.18, m'(\bar{A}) = 0.06, m'(A \cup \bar{A}) = 0.76,$$

and

$$m(r) = 0.02, m(\bar{r}) = 0.43, m(r \cup \bar{r}) = 0.53.$$

Therefore, one finally obtains the trust valuation

$$m_t(A) = 0.0040, m_t(\bar{A}) = 0.0013,$$

and

$$m_t(A \cup \bar{A}) = 0.9946.$$

This scenario is an illustration for the worst practical case and is relevant when the analyst receives a report provided by a source that lacks the skills or competence to provide accurate descriptions of events. In this case, the reports are incomplete, ambiguous, or even irrelevant. In addition to low competence and reliability, the source himself is also unsure about the statement.

The first modification of BBA shows the strong impact of self-confidence, which changes drastically the BBA of the initial assertions, from

$$m(A) = 0.6, m(\bar{A}) = 0.2, m(A \cup \bar{A}) = 0.2,$$

to

$$m'(A) = 0.18, m'(\bar{A}) = 0.06, m'(A \cup \bar{A}) = 0.76.$$

Unsurprisingly, the overall reliability is low:

$$m(r) = 0.0223, m(\bar{r}) = 0.4398, m(r \cup \bar{r}) = 0.5379,$$

Uncertainty of inputs		
2*Observation	Objectivity	0.10
	ObservationalSensitivity	0.10
2*Reporting	SelfConfidence	0.30
	Ambiguity	0.27

TABLE IX  
AMBIGUOUS REPORT: INPUT UNCERTAINTY.

Fusion uncertainty	Dissonance	Conclusiv.	Compleat.
Assertion	0.6	0.4	0.8
Source	0.583	0.417	0.47
Trust	0.973	0.027	0.006

TABLE X  
AMBIGUOUS REPORT: FUSION UNCERTAINTY.

and the results of the final combination show an important mass assigned to  $m_t(A \cup \bar{A}) = 0.9946$ . Intuitively, the information provided is useless, and considered as highly uncertain.

This scenario shows the combined effects of uncertain reporting and incomplete source description for trust estimation. First, the outcome is affected by high values of uncertainty induced during observation and reporting passes, table IX. Then, fusion leads to a trust estimation having high values of *Dissonance*, and very low values of *Conclusiveness* and *Completeness*.

The same criteria estimated for reliability show the main difference with respect to the previous case, which was also based on unreliable sources. While in scenario 4 the source still has important *Completeness*, this measure is drastically decreased for this scenario, as shown in table X.

#### VIII. STRENGTHS AND LIMITATIONS OF BELIEF-BASED FORMALIZATION FOR TRUST ASSESSMENT

This section discusses the strengths and limitations of the belief-based perspective in trust modeling in the light of results shown by previous scenarios. The main advantage of using belief functions is that the formalism is consistent with the cognitive perspective of trust adopted by the model, thanks to the notion of belief. It also captures uncertainties both of the analyst with respect to the source and of the source with respect to their own statements with different mechanisms. First, self-confidence is implemented thanks to a discounting coefficient, as, in practice, the values of self-confidence may rely upon linguistic clues of certainty/uncertainty that can be translated into numerical values. Second, the formalization introduces weighting factors in order to offer a flexible solution, which allow for situations in which the analyst has more or less complete knowledge about distinct attributes of the source, or wishes to emphasize one particular attribute. Moreover, the formalization is able to handle ignorance on various aspects, including missing data. The overall fusion mechanism performs trust estimation in several steps, which allows for a better traceability of the outcome and the mapping at different processing stages using URREF criteria. The results of these scenarios are in line with their specific hypotheses, reflecting the intuition that the fusion technique is appropriate for estimating trust.

As with any user-centric approach, the main limitation of the solution discussed in this paper is the lack of guidance for choosing the set of numerical values with which to instantiate the model. For example, two different analysts may choose differing mass distribution and weight coefficients with respect to the same source, and they may also use slightly different approaches to infer a numerical value from linguistic clues when handling self-confidence. Thus, the outcome depends crucially on the interventions of users and their ability to build a model able to capture the situation under analysis. Also, the solution requires preexisting knowledge about the source's reputation, competence, and intention, indeed, in practice, it is difficult to have access to information on those aspects. Provided that there is no other meta-data or domain knowledge available for use, the model is likely to fail to produce an accurate trust evaluation in some contexts due to the shortage of knowledge on critical aspects.

As such, the belief-based formalization has limited capabilities to explain the outcome. To overcome this limitation, a mapping to URREF uncertainty criteria is used. The mapping highlights when uncertainties are added into the system and which partial results are affected. It facilitates the interpretation of results by adding additional information as to why the item is to be trusted or no; for example, whereas the fusion process outputs low values of trust for a given item, the mapping to URREF criteria allows to underline problems related to evidence collection or reporting, dissonance or incompleteness during the fusion stages.

As shown in previous scenarios, using a belief-oriented formalism and URREF criteria mapping offers a pragmatic approach to develop a more comprehensive and easy to interpret solution for trust estimation.

## IX. CONCLUSION

This paper presents a computational model by which an analyst is able to assess trust in reported information based on several possible unknown attributes of the source as well as additional characterization of the informational content by the source itself. The paper also illustrates the use of URREF criteria to track uncertainty affecting the results, from model construction to its formalization with belief functions. First, a model for trust estimation has been developed that combines several attributes of sources and their own assessment of the items reported. The model is implemented using belief functions, and takes advantage of its mathematical background to define fusion operators for trust assessment. Several scenarios are presented to illustrate uncertainty analysis, illustrating when uncertainty occurs and how it affects partial results for different applications.

Tracking uncertainty is suitable for fusion systems in which various human sources send observations of questionable quality and there is a need to continuously update the trust associated with reports to be analyzed. The set of URREF criteria offers a unified basis to analyze inaccuracies affecting trust estimation during different phases: observation, reporting, and fusion. Select use cases clearly illustrated the benefits

of managing uncertainties arising during the modeling and formalization phases, with the twofold analysis offering additional details on results and improving their interpretation.

The general approach taken in this paper could be adapted to investigate the general mechanisms by which fusion processes integrate information from multiple sources. The solution is especially useful for comparing different fusion approaches with respect to their implications for uncertainty management.

## REFERENCES

- [1] P.C. Costa, K.B. Laskey, E. Blasch, A.-L. Jousset, *Towards unbiased evaluation of uncertainty reasoning: The URREF ontology*, in Proc. of 15th Int. Conf on Information Fusion (Fusion 2012), pp. 2301–2308, Singapore, July 2012
- [2] J. Granatyr, V. Botelho, O.R. Lessing, E.E. Scalabrin, J.-P. Barthès, F. Enembreck, *Trust and reputation models for multiagent systems*, ACM Computing Surveys (CSUR), Vol. 48(2), pp. 1–42, 2015.
- [3] F. Paglieri, C. Castelfranchi, C. da Costa Pereira, R. Falcone, A. Tettamanzi, S. Villata, *Trusting the messenger because of the message: feedback dynamics from information quality to source evaluation*, Computational and Mathematical Organization Theory, Vol. 20(2), pp. 176–194, 2014.
- [4] A. Jøsang, R. Ismail, C. Boyd, *A survey of trust and reputation systems for online service provision*, Decision Support Systems, Vol. 43(2), pp. 618–644, 2007.
- [5] W. Sherchan, S. Nepal, C. Paris, *A survey of trust in social networks*, ACM Computing Surveys (CSUR), Vol. 45(4), pp. 1–33, 2013.
- [6] M. Venanzi, A. Rogers, N.R. Jennings, *Trust-based fusion of untrustworthy information in crowdsourcing applications*, in Proc. of the 12th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2013), pp. 829–836, Ito, Jonker, Gini, and Shehory (eds.), Saint Paul, Minnesota, USA, May 6–10, 2013.
- [7] D.P. Twitchell, D.P. Biros, M. Adkins, N. Forsgren, J.K. Burgoon, J.F. Nunamaker, *Automated determination of the veracity of interview statements from people of interest to an operational security force*, in Proc. of the 39th Annual Hawaii Int. Conf. on System Sciences (HICSS'06), Vol. 1, Kauai, HI, USA, 04–07 January 2006.
- [8] A. Koster, A.L. Bazzan, M. de Souza, *Liar liar, pants on fire; or how to use subjective logic and argumentation to evaluate information from untrustworthy sources*, Artificial Intelligence Review, Vol. 48(2), pp. 219–235, 2017.
- [9] X.L. Dong, E. Gabrilovich, K. Murphy, V. Dang, W. Horn, C. Lugaresi, S. Sun, W. Zhang, *Knowledge-based trust: Estimating the trustworthiness of web sources*, Proceedings of the VLDB Endowment, Vol. 8(9), pp. 938–949, 2015.
- [10] X.L. Dong, E. Gabrilovich, G. Heitz, W. Horn, K. Murphy, S. Sun, W. Zhang, *From data fusion to knowledge fusion*, Proceedings of the VLDB Endowment, Vol. 7(10), pp. 881–892, 2014.
- [11] J. Guo, I.-R. Chen, *A classification of trust computation models for service oriented internet of things systems*, in Proc. of 2015 IEEE Int. Conf. on Services Computing (SCC), pp.324–331, New York, NY, USA, 2015.
- [12] R. Falcone, C. Castelfranchi, *Social trust: A cognitive approach*, Chap. in Trust and deception in virtual societies, Springer, 2001.
- [13] C. Castelfranchi, R. Falcone, *Trust is much more than subjective probability: Mental components and sources of trust*, in Proc. of the 33rd Annual Hawaii Int. Conf. on System Sciences, Maui, HI, USA, January 2000.
- [14] L. Berti-Équille, J. Borge-Holthoef, *Veracity of data: From truth discovery computation algorithms to models of misinformation dynamics*, Synthesis Lectures on Data Management, Vol. 7(3), pp. 1–155, 2015.
- [15] T. Lukoianova, V.L. Rubin, *Veracity roadmap: Is big data objective, truthful and credible?*, Advances in Classification Research Online, Vol. 24(1), pp. 4–15, 2014.
- [16] D.B. Buller, J.K. Burgoon, *Interpersonal deception theory*, Communication theory, Vol. 6(3), pp. 203–242, 1996.
- [17] J.T. Hancock, L.E. Curry, S. Goorha, M. Woodworth, *On lying and being lied to: A linguistic analysis of deception in computer-mediated communication*, Discourse Processes, Vol. 45(1), pp. 1–23, 2007.

- [18] R. Saurí, J. Pustejovsky, *Are you sure that this happened? assessing the factuality degree of events in text*, Computational Linguistics, Vol. 38(2), pp. 261–299, 2012.
- [19] V.L. Rubin, *Stating with certainty or stating with doubt: Intercoder reliability results for manual annotation of epistemically modalized statements*, in Proc of the North American Chapter of the Association for Computational Linguistics (NAACL HLT 2007), Companion Volume, Short Papers, pp. 141–144, Rochester, NY, April 2007.
- [20] R. Falcone, A. Sapienza, C. Castelfranchi, *The relevance of categories for trusting information sources*, ACM Trans. on Internet Technology (TOIT), Vol. 15(4), pp. 1–21, 2015.
- [21] R. Falcone, M. Piunti, M. Venanzi, C. Castelfranchi, *From manifesta to krypta: The relevance of categories for trusting others*, ACM Transactions on Intelligent Systems and Technology (TIST), Vol. 4(2), pp. 1–24, 2013.
- [22] R. Hermoso, H. Billhardt, S. Ossowski, *Trust-based role coordination in task-oriented multiagent systems*, Knowledge-Based Systems, Vol. 52, pp. 78–90, 2013.
- [23] P.-A. Matt, M. Morge, F. Toni, *Combining statistics and arguments to compute trust*, in Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010), pp. 209–216, Van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), Toronto, Canada, May 10–14, 2010.
- [24] L. Amgoud, R. Demolombe, *An argumentation-based approach for reasoning about trust in information sources*, Argument & Computation, Vol. 5(2–3), pp. 191–215, 2014.
- [25] Y. Tang, K. Cai, P. McBurney, E. Sklar, S. Parsons, *Using argumentation to reason about trust and belief*, Journal of Logic and Computation, Vol. 22(5), pp. 979–1018, 2011.
- [26] P. Everaere, S. Konieczny, P. Marquis, *Belief merging versus judgment aggregation*, in Proc. of the 14th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2015), pp. 999–1007, Bordini, Elkind, Weiss, Yolum (eds.), Istanbul, Turkey, May, 4–8, 2015.
- [27] G. Pigozzi, *Belief merging and judgment aggregation*, Stanford Encyclopedia of Philosophy, 2015.
- [28] S. Parsons, K. Atkinson, Z. Li, P. McBurney, E. Sklar, M. Singh, K. Haigh, K. Levitt, J. Rowe, *Argument schemes for reasoning about trust*, Argument & Computation, Vol. 5(2–3), pp. 160–190, 2014.
- [29] S. Villata, G. Boella, D.M. Gabbay, L. Van Der Torre, *Arguing about the trustworthiness of the information sources*, in Proc. of European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU 2011), pp. 74–85, Belfast, UK, June 29–July 1, 2011.
- [30] M. Momani, S. Challa, R. Alhmouz, *BNWSN: Bayesian network trust model for wireless sensor networks*, in Proc. of 2008 Mosharaka Int. Conf. on Communications, Computers and Applications (MIC-CCA 2008), pp. 110–115, Amman, Jordan, 08–10 August 2008.
- [31] L. Cholvy, *How strong can an agent believe reported information?*, in Proc. of European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU 2011), pp. 386–397, W. Liu (Ed.), Springer, 2011.
- [32] J.P. de Villiers, G. Pavlin, P. Costa, A.-L. Joussetme, K. Laskey, V. Dragos, E. Blasch, *Subjects under evaluation with the URREF ontology*, in Proc. of 20th Int. Conf on Information Fusion (Fusion 2017), pp. 1–8, Xi’an, China, July 10–13, 2017.
- [33] J.P. de Villiers, R. Focke, G. Pavlin, A.-L. Joussetme, V. Dragos, K. Laskey, P. Costa, E. Blasch, *Evaluation metrics for the practical application of URREF ontology: An illustration on data criteria*, in Proc. of 20th Int. Conf on Information Fusion (Fusion 2017), pp. 1–8, Xi’an, China, July 10–13, 2017.
- [34] V. Dragos, *An ontological analysis of uncertainty in soft data*, in Proc. of 16th Int. Conf on Information Fusion (Fusion 2013), pp. 1566–1573, Istanbul, Turkey, July 9–12, 2013.
- [35] E. Blasch, K.B. Laskey, A.-L. Joussetme, V. Dragos, P. C. Costa, J. Dezert, *URREF reliability versus credibility in information fusion (STANAG 2511)*, in Proc. of 16th Int. Conf on Information Fusion (Fusion 2013), pp. 1600–1607, Istanbul, Turkey, July 9–12, 2013.
- [36] E. Blasch, A. Jøsang, J. Dezert, P.C. Costa, A.-L. Joussetme, *URREF self-confidence in information fusion trust*, in Proc. of 17th Int. Conf on Information Fusion (Fusion 2014), pp. 1–8, Salamanca, Spain, July 7–10, 2014.
- [37] E. Blasch, A. Aved, *URREF for veracity assessment in query-based information fusion systems*, in Proc. of 18th Int. Conf on Information Fusion (Fusion 2015), pages 58–65, Washington D.C, USA, July 6–9, 2015.
- [38] J.P. de Villiers, K. Laskey, A.-L. Joussetme, E. Blasch, A. de Waal, G. Pavlin, P. Costa, *Uncertainty representation, quantification and evaluation for data and information fusion*, in Proc. of 18th Int. Conf on Information Fusion (Fusion 2015), pp. 50–57, Washington D.C, USA, July 6–9, 2015.
- [39] A.-L. Joussetme, *Semantic criteria for the assessment of uncertainty handling fusion models*, in Proc. of 19th Int. Conf on Information Fusion (Fusion 2016), pp. 488–495, Heidelberg, Germany, July 5–8, 2016.
- [40] V. Dragos, Z. Juergen, J.P. de Villiers, *Applicaton of URREF criteria to assess knowledge representation in cyber threats models*, in Proc. of 21th Int. Conf on Information Fusion (Fusion 2018), Cambridge, UK, July 10–13, 2018.
- [41] J.P. de Villiers, A.-L. Joussetme, A. de Waal, G. Pavlin, K. Laskey, E. Blasch, P. Costa, *Uncertainty evaluation of data and information fusion within the context of the decision loop*, in Proc. of 19th Int. Conf on Information Fusion (Fusion 2016), pp. 766–773, Heidelberg, Germany, July 5–8, 2016.
- [42] V. Dragos, X. Lerouvreur, S. Gatepaille, *A critical assessment of two methods for heterogeneous information fusion*, in Proc. of 18th Int. Conf on Information Fusion (Fusion 2015), pp. 42–49, Washington D.C, USA, July 6–9, 2015.
- [43] A.-L. Joussetme, G. Pallotta, *Dissecting uncertainty-based fusion techniques for maritime anomaly detection*, in Proc. of 18th Int. Conf on Information Fusion (Fusion 2015), pp. 34–41, Washington D.C, USA, July 6–9, 2015.
- [44] H. Koen, J.P. de Villiers, G. Pavlin, A. de Waal, P. de Oude, F. Mignet, *A framework for inferring predictive distributions of rhino poaching events through causal modelling* in Proc. of 17th Int. Conf on Information Fusion (Fusion 2014), pp. 1–7, Salamanca, Spain, July 7–10, 2014.
- [45] E. Blasch, P.C. Costa, K.B. Laskey, H. Ling, G. Chen, *The URREF ontology for semantic wide area motion imagery exploitation*, in Proc. of 2012 IEEE Aerospace and Electronics Conference (NAECON 2012), pages 228–235, Dayton, OH, USA, 25–27 July 2012.
- [46] M.B. Sinai, N. Partush, S. Yadid, E. Yahav, *Exploiting social navigation*. arXiv preprint arXiv:1410.0151, 2014.
- [47] K.S. Barber, J. Kim, *Belief revision process based on trust: Simulation experiments*, in Proc. of Autonomous Agents ’01 Workshop on Deception, Fraud, and Trust in Agent Societies, Citeseer, 2001.
- [48] C. Castelfranchi, R. Falcone, G. Pezzulo, *Trust in information sources as a source for trust: a fuzzy approach*, in Proc. of the 2nd Int. Joint Conf. on Autonomous Agents and Multiagent Systems (AAMAS’03), pp. 89–96, Melbourne, Australia, July 14–18, 2003.
- [49] L. Cronk, *Communication as manipulation: Implications for biosociological research*, In Proc. of the American Sociological Association Annual Meetings, 1991.
- [50] NATO, *STANAG 2011*, Technical report, Intelligence reports, 2003.
- [51] D.L. Hall, J.M. Jordan, *Human-centered information fusion*, Artech House, 2010.
- [52] A.-L. Joussetme, A.-C. Boury-Brisset, B. Debaque, D. Prevost, *Characterization of hard and soft sources of information: A practical illustration*, in Proc. of 17th Int. Conf on Information Fusion (Fusion 2014), pp. 1–8, Salamanca, Spain, July 7–10, 2014.
- [53] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, *Multisensor data fusion: A review of the state-of-the-art*, Information Fusion, Vol. 14(1), pp. 28–44, 2013.
- [54] A. Auger, J. Roy, *Expression of uncertainty in linguistic data*, in Proc. of 11th Int. Conf on Information Fusion (Fusion 2008), pp. 1–8, Cologne, Germany, July 2008.
- [55] V. Dragos, K. Rein, *Integration of soft data for information fusion: Pitfalls, challenges and trends*, in Proc. of 17th Int. Conf on Information Fusion (Fusion 2014), pp. 1–8, Salamanca, Spain, July 7–10, 2014.
- [56] S. Kent, *Words of estimative probability*, Studies in Intelligence, Vol. 8(4), pp. 49–65, 1964.
- [57] R.F. Kesselman, *Verbal probability expressions in national intelligence estimates: a comprehensive analysis of trends from the fifties through post 9/11*, Ph.D. thesis, Mercyhurst College, Erie, PA, USA, May 2008.
- [58] K. Rein, *I believe it’s possible it might be so . . .*, Dissertation, University of Bonn, 2016
- [59] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, 1976.

- [60] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, Volumes 1–4, American Research Press (ARP), 2004–2015.
- [61] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of 15th Int. Conf on Information Fusion (Fusion 2012), Singapore, July 2012.
- [62] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of bayesian fusion rule*, International Journal of Intelligent Systems, Special Issue: Advances in Intelligent Systems, Vol. 29(3), pp. 223–252, 2014.
- [63] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of 13th Int. Conf on Information Fusion (Fusion 2010), pp. 1-8, Edinburgh, Scotland, UK, 26–29 July 2010.
- [64] D. Mercier, B. Quost, T. Denœux, *Refined modeling of sensor reliability in the belief function framework using contextual discounting*, Information Fusion, Vol. 9(2), pp. 246–258, 2008.



# DSMT Decision-Making Algorithms for Finding Grasping Configurations of Robot Dexterous Hands

Ionel-Alexandru Gal<sup>a</sup>, Dănuț Bucur<sup>a</sup>, Luige Vlădăreanu<sup>a</sup>

<sup>a</sup>Institute of Solid Mechanics of the Romanian Academy, 010141 Bucharest S1, Romania.

Emails: alexandru.gal@imsar.ro, xdan.bucur@yahoo.com, luigiv2007@yahoo.com.sg

Originally published as: I.-A. Gal, D. Bucur, L. Vladareanu, *DSMT Decision-Making Algorithms for Finding Grasping Configurations of Robot Dexterous Hands*, *Symmetry*, Vol.10(6), 198, 2018, and reprinted with permission.

**Abstract**—In this paper, we present a deciding technique for robotic dexterous hand configurations. This algorithm can be used to decide on how to configure a robotic hand so it can grasp objects in different scenarios. Receiving as input, several sensor signals that provide information on the object's shape, the DSMT decision-making algorithm passes the information through several steps before deciding what hand configuration should be used for a certain object and task. The proposed decision-making method for real time control will decrease the feedback time between the command and grasped object, and can be successfully applied on the robot dexterous hands. For this we have used the Dezert-Smarandache theory which can provide information even on contradictory or uncertain systems.

**Keywords:** neutrosophy, DSMT, decision-making algorithms, robotic dexterous hands, grasping configurations, grasp type.

## I. INTRODUCTION

The purpose of autonomous robotics is to build systems that can fulfill all kind of tasks without human intervention, in different environments which were not specially build for robot interaction. A major challenge for this autonomous robotics field comes from high uncertainty within real environments. This is because the robot designer can't know all the details regarding the environment. Most of the environment parameters are unknown, the position of humans and objects can't be previously anticipated and the motion path might be blocked. Beside these, the accumulated sensor information can be uncertain and error prone. The quality of this information is influenced by noise, visual field limitations, observation conditions and the complexity of interpretation technique.

The artificial intelligence and the heuristic techniques were used by many scientists in the field of robot control [1] and motion planning. Regarding the grasping and object manipulations, the main research activities were to design a mechanism for hand [2-4] and dexterous finger motion [5], which are a high complexity research tasks in controlling robotic hands.

Currently in the research area of robotics, it's desired to develop robotic systems with applications in dynamic and unknown environments, in which human lives would be at risk, like natural or nuclear disaster areas, and also in different fields of work, ranging from house chores or agriculture to military applications. In any of these research areas, the robotic system must fulfill a series of tasks which implies object manipulation and transportation, or using equipment and tools. From here arises the necessity of development grasping systems [6] to reproduce as well as possible human hand motion [7-9].

To achieve an accurate grasping system, grasp taxonomy of human hand was analyzed by Feix (et. al) [10] who found 33 different grasp types, sorted by opposition type, virtual finger assignments, type in terms of power, precision or intermediate grasp, and the position of the thumb. While Alvarez [11] (et. al) researched human grasp strategies within grasp types, Fermuller (et. al) [12] focused on manipulation action for human hand on different object types including hand pre-configuration. Tsai (et. al) [13] found that classifying objects into primitive shapes can provide a way to select the best grasping posture, but a general approach can also be used for hand-object geometry fitting [14]. This classification works well for grasping problems in constrained work space using visual data combined with force sensors [15] and also for under-actuated grasping which uses rotational stiffness [16]. But for unknown objects, scientists found different approaches to solve the hand grasping problem. Choi (et. al) [17] used two different neural networks and data fusion to classify objects, Seredynski (et. al) [18] achieved fast grasp learning with probabilistic models, while Song (et. al) [19] used a tactile-based blind grasping along with a discrete-time controller. The same approach is used by Gu (et. al) [20] which proposed a blind haptic exploration of unknown objects for grasp planning of dexterous robotic hand. Using grasping methods, Yamakawa (et. al) [21] developed a robotic hand for knot manipulation, while Nancy (et. al) [22] used artificial neural network algorithms for slip prevention and Zaidi (et. al) [23] used a multi-fingered robot hand to grasp 3D deformable objects, applying the method on spheres and cubes.

While other scientists developed grasping strategies for different robotic hands [21-23], an anthropomorphic robotic hand has the potential to grasp regular objects of different shapes and sizes [24, 25], but selecting the grasping method for a certain object is a difficult problem. A series of papers have approached this problem by developing algorithms for classifying the grasping by the contact points [26, 27]. These algorithms are focused on finding a fix number of contact areas without taking into consideration the hand geometry. Other methods developed grasping systems for a certain robotic hand architecture, scaling down the problem to finding a grasping method with the tip of the fingers [27]. These methods are useful in certain object manipulation, but can't be applied for a wide range of objects because it doesn't provide a stable grasping due to the face that it's not used, the finger's interior surface or the palm of the hand. A method for filtering the high

number of hand configurations is to use predefined grasping hand configurations. Before grasping an object, humans, unconsciously simplify the grasping action, choosing one of the few hand positions which match the object's shape and the task to accomplish. In the scientific literature there are papers which have tried to log in the positioning for grasping and taxonomy, and one of the most known papers is [28]. Cutkosky and Weight [29] have extended Napier's [28] classification by adding the required taxonomy in the production environment, by studying the way in which the weight and geometry of the object affects choosing the grasping positioning. Iberall [30] has analyzed different grasping taxonomies and generalized them by using the virtual finger concept. Stransfield [31] has chosen a simpler classification and built a system based on rules which provided a grasping positioning set, starting from a simplified description of the object gained from a video system.

The developed algorithm presented in this paper, has the purpose to determine the grasping position according to the object's shape. To prove the algorithm's efficiency we have chosen 3 types of grasping: cylindrical, spherical and prismatic. For this, we start from the hypothesis that the environment data are captured through a stereovision system [32] and a Kinect sensor [33]. On this data, which the two system observers provide, we apply a template matching algorithm [34]. This algorithm will provide a matching percentage of the object that needs to be grasped with a template object. Thus, each of the two sources will provide 3 matching values, for each of the three grasping types. These values represent the input for our detection algorithm, based of Dezert-Smarandache Theory (DSMT) [35] for data fusion. This algorithm has as input data from two or multiple observers and in the first phase they are processed through a process of neutrosification which is similar with the fuzzification process. Then, the neutrosophic observers' data are passed through an algorithm which applies the classic DSMT theory [35] in order to obtain a single data set on the system's states, by combining the observers' neutrosophic values. On this obtained data set, we apply the developed DSMT decision-making algorithm that decides on the category from which the target object is part of. This decision facilitates the detection-recognition-grasping process which a robotic hand must follow, obtaining in the end a real-time decision that doesn't stop or delay the robot's task.

In recent years, using more sensors for a certain applications and then using data fusion is becoming more common, in the military and nonmilitary research fields. The data fusion techniques combine the information received from different sensors with the purpose of eliminating disturbances and to improve precision compared to the situations when a single sensor is used [36, 37]. This technique works on the same principle used by humans to feel the environment. For example, a human being can't see over the corner or through vegetation, but with his hearing he can detect certain surrounding dangers. Beside the statistical advantage build from combining the details for a certain object (through redundant observations),

using more types of sensors increases the precision with which an object can be observed and characterized. For example, an ultrasonic sensor can detect the distance to an object, but a video sensor can estimate its shape, and combining these two information sources will provide two distinct data on the same object.

The evolution on the new sensors, the hardware's processing techniques and capacity improvements facilitate more and more the real time data fusion. The latest progress were made in the area of computational and detection systems, and provide the ability to reproduce, in hardware and software, the data fusion capacity of humans and animals. The data fusion systems are used for targets tracking [38], automatic targets identification [39] and automated reasoning applications [40]. The data fusion applications are widespread, ranging from the military [41] applications (target recognition, autonomous moving vehicles, distance detection, battlefield surveillance, automatic danger detection) to civilian application (monitoring the production processes, complex tools maintenance based on certain conditions, robotics [42], and medical applications [43]). The data fusion techniques undertake classic elements like digital signal processing, statistical estimation, control theory, artificial intelligence and numeric methods [44].

Combined data interpretation requires automated reasoning techniques taken from the area of artificial intelligence. The purpose of developing the recognition based systems, was to analyze issues like the data gathering context, the relationship between observed entities, hierarchical grouping of targets or objects and to predict future actions of these targets or entities. This kind of reasoning is encountered in humans, but the automated reasoning techniques can only closely reproduce it. Regardless of the used technique, for a knowledge based system, 3 elements are required: one or more reasoning diagrams, an automated evaluation process and a control diagram. The reasoning diagrams are techniques of facts representation, logical relations, procedural knowledge and uncertainty. For these techniques, uncertainty from the observed data and from the logical relations can be represented using probabilities, fuzzy theory [45, 46], Demspter-Shafer [47] evidence intervals or other methods. Dezert-Smarandache theory [35] comes to extend these methods, providing advanced techniques of uncertainty manipulation. The automated reasoning system's developing purpose is to reproduce the human capability of reasoning and decision making, by specifying rules and frames that define the studied situation. Having at hand an information database, it's required an evaluation process so this information can be used. For this there are formal diagrams developed on the formal logic, fuzzy logic, probabilistic reasoning, template based methods, case based reasoning and many others. Each of these reasoning diagrams has a consistent internal formalism which describes how to use the knowledge database for obtaining the final conclusion. An automated reasoning system needs a control diagram to fulfill the thinking process. The used techniques include searching methods, systems for maintaining the truth based on assumptions and justifications, hierarchical decomposition, control theory, etc. Each of these

methods has the purpose of controlling the reasoning evolution process.

The results presented in this paper, were obtained using the classic Dezert Smarandache Teory (DSMT) to combine inputs from two different observers that want to classify objects into three categories: Sphere, Parallelepiped and Cylinder. These categories were chosen to include most of the objects that a manipulator can grasp. The algorithm's inputs were transformed into belief values of certainty, falsity, uncertainty and contradiction values. Using these four values and their combinations according to DSMT we applied Petri net diagram logic for taking decisions on the shape type of the analysed objects. This type of algorithm has never been used before for real time decision on hand grasping taxonomy. Comparing to other algorithms [13-15] and methods [16-18],

ours has the advantage to detect high uncertainties and contradictions which in practice has a very low encounter rate but can have drastic effects on the decision type or robot, because if the object's shape is not detected properly, then the robot might not be able to grasp it, which can lead to serious consequences. In deciding how to grasp objects, researchers have used different methods to choose the grasping taxonomy using a blind haptic exploration [20] or in different applications for tying knots [21] or grasp deformable objects [23]. Because the proposed algorithm can detect anomalies of contradicting and uncertain input values, we can say that the proposed method transforms the deciding process into a less difficult problem of grasping method [24, 25].

## II. OBJECTS GRASPING AND ITS CLASSIFICATION

Mechanical hands have been developed to provide the robots with the ability of grasping objects with different geometrical and physical properties [48]. To make an anthropomorphic hand seem natural, its movement and the grasping type must match the human hand.

On this regard, grasping position taxonomy for human hands has been long studied and applied for robotic hands. Seventeen different categories of human hands grasping positions were studied. But first we must consider two important things. The first one is that these categories are derived from human hand studies, which proved that they are more flexible and able to perform a multitude of movements than any other robotic hand, so that the grasping taxonomy for robot hands can be only a simple subset of the human hand. Secondly, the human behavior studies of real object grasping, have shown some differences between the real observations and the classified properties [49].

In conclusion, any proposed taxonomy is only a reference point which the robot hand must attain. Below are described the most used grasping positions (extracted from [50]), which should be considered when developing an able robotic hand:

- 1) Power grasping. The contact with the objects is made on large surfaces of the hand, including hand phalanges and the palm of the hand. For this kind of grasping high forces can be exerted on the object.
  - Spherical grasping: used to grasp spherical objects;

- Cylindrical grasping: used to grasp long objects which can't be completely surrounded by the hand;
- Lateral grasping: the thumb exert a force towards the lateral side of the index finger.

- 2) Precision grasping: the contact is made only with the tip of the fingers.
  - Prismatic grasping (pinch): used to grasp long objects (with small diameter) or very small. Can be achieved with two to five fingers.
  - Circular grasping (tripod): used in grasping circular or round objects. Can be achieved with three, four or five fingers.

- 3) No grasping:
  - Hook: the hand forms a hook on the object and the hand force is exerted against an external force, usually gravity.
  - Button pressing or pointing
  - Pushing with open hand.

In the table I, are shown manipulation activities that the robotic hand can achieve, correlated with the required activity grasping positions [51].

## III. OBJECT DETECTION USING STEREO-VISION AND KINECT SENSOR

Object recognition in artificial sight represents the task of searching a certain object in a picture or a video sequence. This problem can be approached as a learning problem. At first, the system is trained with sample images which belong to the target group, the system being taught to spot these among other pictures. Thus, when the system receives new images, it can 'feel' the presence of the searched object/sample/template.

Template matching is a techniques used to sort objects in an image. A model is an image region, and the goal is to find instances of this model in a larger picture. The template matching techniques represent a classic approach for localization problems and object recognition in a picture. These methods are used in applications like object tracking, image compression, stereograms, image segmentation [52], and other specific problems of artificial vision [53]. Object recognition is very important for a robot that must fulfill a certain task. To complete its task, the robot must avoid obstacles, to obtain the size of the object, to manipulate it, etc. For the case of detected object manipulation, the robot must detect the object's shape, size and position in the environment. The main methods for achieving the depth information use stereoscopic cameras, laser scanners and depth cameras. To achieve the proposed decision-making algorithm, we assumed that the environment information is captured with a stereoscopic system and a Kinect sensor.

Stereovision systems [32] represents a passive technique of achieving a virtual 3D image of the environment in which the robot moves, by matching the common features of an image set of the same scene. Because this method works with images, it needs a high computational power. The depth information can be noisy in certain cases, because the method depends

TABLE I: Grasping position for certain tasks.

Object	Activity	Grasping position
Bottles, cups and mugs	Transport Pouring/ filling	Force: Cylindrical grasping (from the side or the top)
Cups (using handles)	Pouring/filling	Force: Lateral grasping Precision: Prismatic grasping
Plates/trays	Transport Receiving from humans	Power: Lateral grasping Precision: Prismatic grasping No grasp: pushing (open hand)
Pens, cutlery	Transport	Precision: Prismatic grasping
Door handle	Open/Close	Force: Cylindrical grasping No grasp: Hook
Small objects	Transport	Power: Spherical grasping Precision: Circular grasping (tripod)
Switches, buttons	Pushing	No grasp: Button pressing
Round switches, bottle caps	Rotation	Force: Lateral grasping Precision: Circular grasping (tripod)

on the texture of the environment objects and on the ambient light.

Kinect [33] is a fairly easy to obtain platform, which makes it widespread. It uses a depth sensor based on structured light. By using an ASIC board, the Kinect sensor generates a depth map on 11 bits with a resolution of  $640 \times 480$  pixels, at 30Hz. Given the price of the device, the information quality is pretty good, but it has both advantages and disadvantages, meaning that the depth images contain areas where the depth reading couldn't be achieved. This problem appears from the fact that some materials don't reflex infrared light. When the device is moved really fast, like any other camera, it records blurry pictures, which also lead to missing information from the acquired picture.

#### IV. NEUTROSOPHIC LOGIC AND DSM THEORY

##### A. Neutrosophic Logic

The neutrosophic triplet (truth, falsity and uncertainty) idea appeared in 1764 when J.H. Lambert investigates a witness credibility which was affected by the testimony of another person. He generalized Hooper's rule of sample combination (1680), which was a Non-Bayesian approach for finding a probabilistic model. Koopman in 1940 introduces the low and high probability, followed by Good and Dempster (1967) who gave a combination rule of two arguments. Shafer (1976) extended this rule to Dempster-Shafer Theory for Trust Functions by defining the Trust and Plausibility functions and using the inference rules of Dempster for combining two samples from two different sources. The trust function is a connection between the fuzzy reasoning and probability. Dempster-Shafer theory for Trust functions is a generalization of Bayesian Probability (Bayes 1760, Laplace 1780). It uses the mathematical probability in a more general way and it is based on the probabilistic combination of artificial intelligence samples. Lambert one said that "there is a chance  $p$  that the witness can be trustworthy and fair, a chance  $q$  that he will be deceiving and a chance  $1 - p - q$  that he will be indifferent". This idea was taken by Shafer in 1986 and later, used by Smarandache to further develop the neutrosophic logic [54, 55].

1) *Neutrosophic Logic Definition:* A logic in which each proposition has its percentage of truth in a subset  $T$ , its percentage of uncertainty in a subset  $I$  and its percentage of falsity in a subset  $F$  is called neutrosophic logic [54,55]. This paper extends the general structure of the Neutrosophic Robot Control (RNC), known as the Vlădăreanu–Smarandache method [55]–[57] for the robot hybrid force-position control in a virtual platform [58,59], which applies neutrosophic science to robotics using the neutrosophic logic and set operators. Thus, using two observers, a stereovision system and a Kinect sensor, will provide 3 matching values for DSMT decision-making algorithms. A subset of truth, uncertainty and falsity is used instead of a single number because in many cases one cannot know with precision the percentage of truth or falsity. But these can be approximated. For example, a supposition can be 30% to 40% true and 60% to 70% false [60].

2) *Neutrosophic components definition:* Let  $T, I, F$  be three standard or non-standard subsets of  $]^{-0}, 1^{+}[$  with

$$\begin{aligned} \sup T &= t_{sup} & \inf T &= t_{inf}, \\ \sup I &= i_{sup} & \inf I &= i_{inf}, \\ \sup F &= f_{sup} & \inf F &= f_{inf}, \end{aligned}$$

and

$$\begin{aligned} n_{sup} &= t_{sup} + i_{sup} + f_{sup} \\ n_{inf} &= t_{inf} + i_{inf} + f_{inf} \end{aligned}$$

The  $T, I, F$  sets are not always intervals, but can be subsets: discrete or continuum; with a single element; finite or infinite (the elements are countable or uncountable); subsets union or intersection. Also, these subsets can overlap, and the real subsets represent the relative errors in finding the  $t, i, f$  values (when the  $T, I, F$  subsets are reduced to single points).

$T, I, F$  are called the neutrosophic components and represents the truth, uncertainty and falsity values, when referring to neutrosophy, neutrosophic logic, neutrosophic sets, neutrosophic probability or neutrosophic statistics. This representation is closer to the human reasoning and defines knowledge imprecision or linguistic inaccuracy received from different

observers (this is why  $T, I, F$  are subsets and can be more than a set of points), the uncertainty given by incomplete knowledge or data acquisition errors (for this we have the set  $I$ ) and the vagueness caused by missing edges or limits.

After defining the sets, we need to specify their superior ( $x_{sup}$ ) and inferior ( $x_{inf}$ ) limits because in most of the cases they will be needed [61,62].

3) *Dezert-Smarandache Theory (DSMT)*: To develop artificial cognitive systems a good management of sensor information is required. When the input data are gathered by different sensors, according to the environment certain situations may appear when one of the sensors cannot give correct information or the information is contradictory between sensors. To resolve this issue a strong mathematical model is required, especially when the information is inaccurate or uncertain.

The Dezert-Smarandache Theory (DSMT) [53,54,60] can be considered an extension of Dempster-Shafer theory (DST) [46]. DSMT allows information combining, gathered from different and independent sources as trust functions. DSMT can be used for solving information fusion on static or dynamic complex problems, especially when the information differences between the observers are very high.

DSMT starts by defining the notion of *DSM free model*, denoted by  $\mathcal{M}^f(\Theta)$  and says that  $\Theta$  is a set of exhaustive elements  $\theta_i, i = 1, \dots, n$  which cannot overlap. This model is free because there are no other suppositions over the hypothesis. As long as the DSM free model is fulfilled, we can apply the associative and commutative DSM rule of combination.

DSM theory [62] is based on defining the Dedekind lattice, known as the *hyper power set* of frame  $\Theta$ . In DSMT,  $\Theta$  is considered a set  $\{\theta_1, \dots, \theta_n\}$  of  $n$  exhaustive elements, without adding other constraints. DSMT can tackle information samples, gathered from different information sources which don't allow the same interpretation of the set  $\Theta$  elements. Let  $\Theta = \{\theta_1, \theta_2\}$  be the simple case, made of two assumptions, then [54]:

- the probability theory works (assuming exclusivity and completeness assumptions) with basic probability assignments (bpa)  $m(\cdot) \in [0, 1]$  such that

$$m(\theta_1) + m(\theta_2) = 1$$

- the Dempster-Shafer theory works, (assuming exclusivity and completeness assumptions) with basic belief assignments (bba)  $m(\cdot) \in [0, 1]$  such that

$$m(\theta_1) + m(\theta_2) + m(\theta_1 \cup \theta_2) = 1$$

- the DSM theory works (assuming exclusivity and completeness assumptions) with basic belief assignment (bba)  $m(\cdot) \in [0, 1]$  such that

$$m(\theta_1) + m(\theta_2) + m(\theta_1 \cup \theta_2) + m(\theta_1 \cap \theta_2) = 1$$

**The  $D^\Theta$  hyperpower set notion**

One of the base elements of DSM theory is the notion of hyper power set. Let  $\Theta = \{\theta_1, \dots, \theta_n\}$  be a finite set (called

frame) with  $n$  exhaustive elements. The Dedekind lattice, called hyper power set  $D$  within DSMT frame, is defined as the set of all built statements from the elements of set  $\Theta$  with the  $\cup$  and  $\cap$  operators such that:

- 1)  $\emptyset, \theta_1, \dots, \theta_n \in D^\Theta$ ;
- 2) If  $A, B \in D^\Theta$ , then  $A \cap B \in D^\Theta$  and  $A \cup B \in D^\Theta$ ;
- 3) No other element is included in  $D^\Theta$  with the exception of those mentioned at 1 and 2.

$D^\Theta$  dual's (obtained by changing within expressions the operator  $\cap$  with the operator  $\cup$ ) is  $D^\Theta$ . In  $D^\Theta$  there are elements that are dual with themselves. The cardinality of  $D^\Theta$  increases with  $2^n$  when the cardinality of  $\Theta$  is  $n$ . Generating the  $D^\Theta$  hyper power set is close connected with the Dedekind [54,55] known problem of isotone Boolean function set. Because for any finite set  $\theta, |D^\Theta| \geq |2^\Theta|$ ,  $D^\Theta$  is called the hyper power set of  $\Theta$ .

The  $\theta_i, i = 1, \dots, n$  elements from  $\Theta$  form the finite set of suppositions/concepts that characterize the fusion problem.  $D^\Theta$  represents the free model of DSM  $\mathcal{M}^f(\Theta)$  and allows working with fuzzy concepts that describe the intrinsic and relative character. This kind of concepts cannot be accurately distinguished within an absolute interpretation because of the unapproachable universal truth.

With all of this, there are certain particular fusion problems that imply discrete concepts, where the  $\theta_i$  elements are exclusively true. In this case, all the exclusivity constraints of  $\theta_i, i = 1, \dots, n$  must be included in the previous model, to properly characterize the truthiness character of the fusion problem and to match reality. For this, the hyper power set  $D^\Theta$  is decreased to the classic power set  $2^\Theta$  forming the smallest hybrid DSM model, noted with  $\mathcal{M}^0(\Theta)$ , and coincide with Shafer's model.

Besides the problem types that correspond with the Shafer's model  $\mathcal{M}^0(\Theta)$  and those that correspond with the DSM free model  $\mathcal{M}^f(\Theta)$ , there is an extensive class of fusion problems that include in  $\Theta$  states, continuous fuzzy concepts and discrete hypothesis. In this case we must take into consideration certain exclusivity constraints and some non-existential constraints. Each fusion hybrid problem is described by a DSM hybrid model  $\mathcal{M}(\Theta)$  with  $\mathcal{M}(\Theta) \neq \mathcal{M}^f(\Theta)$  and  $\mathcal{M}(\Theta) \neq \mathcal{M}^0(\Theta)$ .

**The generalized belief functions**

Starting from a general frame  $\Theta$ , we define a  $D^\Theta \rightarrow [0, 1]$  transformation associated with an information source  $\mathcal{B}$  like [54]:

$$m(\emptyset) = 0, \text{ and } \sum_{A \in D^\Theta} m(A) = 1. \tag{1}$$

The  $m(A)$  value is called *generalized basic belief assignment* of  $A$ .

The generalized trust and plausibility are defined in the same way as in Dempster-Shafer theory [47]:

$$Bel(A) = \sum_{\substack{BA \\ B \in D^\Theta}} m(B), \tag{2}$$

$$Pl(A) = \sum_{\substack{B \cap A \neq \emptyset \\ B \in D^\Theta}} m(B). \quad (3)$$

These definitions are compatible with the classic trust function definition from the Dempster-Shafer theory when  $D^\Theta$  is reduced to  $2^\Theta$  for fusion problems where the Shafer model  $\mathcal{M}^0(\Theta)$  can be applied. We're still having for all  $A \in D^\Theta$ ,  $Bel(A) \leq Pl(A)$ . To notice that when we work with the free DSm  $\mathcal{M}^f(\Theta)$  model, we will always have  $Pl(A) = 1$ ,  $\forall A \neq \emptyset \in D^\Theta$ , which is normal [54].

### The DSm classic rule of combination

When the DSm free model  $\mathcal{M}^f(\Theta)$  can be applied, the combination rule  $m_{M^f(\Theta)} \equiv m(\cdot) \triangleq [m_1 \otimes m_2](\cdot)$  of two independent sources  $\mathcal{B}_1$  and  $\mathcal{B}_2$  that provide information on the same frame  $\Theta$  with the belief functions  $Bel_1(\cdot)$  and  $Bel_2(\cdot)$  associated to gbbas  $m_1(\cdot)$  and  $m_2(\cdot)$  correspond to the conjunctive consensus of sources. Data combinations are done by using the formula [54]:

$$\forall C \in D^\Theta, m_{M^f(\Theta)}(C) = m(C) = \sum_{\substack{A, B \in D^\Theta \\ A \cap B = C}} m_1(A)m_2(B). \quad (4)$$

Because  $D^\Theta$  is closed under  $\cap$  and  $\cup$  operators, this new combination rule guarantees that  $m(\cdot)$  is a generalized trust value, meaning that  $m(\cdot) : D^\Theta \rightarrow [0, 1]$ . This rule of combination is commutative and associative and can be used all the time for sources fusion which implies fuzzy concepts. This rule can be extended with ease for combining  $k > 2$  independent information sources [55,56].

Because of the high number of elements in  $D^\Theta$ , when the cardinality of  $\Theta$  increases, the need of computational resources also increases for processing the DSm combination rule. This observation is true only if the core (the set of generalized basic belief assignment for the needed elements)  $\mathcal{K}_1(m_1)$  and  $\mathcal{K}_2(m_2)$  coincide with  $D^\Theta$ , meaning that when  $m_1(A) > 0$  and  $m_2(A) > 0$  for any  $A \neq \emptyset \in D^\Theta$ . For the most practical applications, the  $\mathcal{K}_1(m_1)$  and  $\mathcal{K}_2(m_2)$  dimensions are much smaller than  $|D^\Theta|$  because the information sources provide most of the time the generalized basic belief assignment for only one subset of hyper power set. This facilitates the DSm classic rule implementation.

Figure 1 presents the DSm combination rule architecture. The first layer is formed by all the generalized basic belief assignment values of the needed elements  $A_i, i = 1, \dots, n$  of  $m_1(\cdot)$ . The second layer is made out of all the generalized basic belief assignment values  $B_j, j = 1, \dots, k$  of  $m_2(\cdot)$ . Each node from the first layer is connected with each node of the second layer. The output layer is created by combining the generalized basic belief assignment values of all the possible intersections  $A_i \cap B_j, i = 1, \dots, n$  and  $j = 1, \dots, k$ . If we would have a third source to provide generalized basic belief assignment values  $m_3(\cdot)$ , this would have been combined by placing it between the output layer and the second one that provides the generalized basic belief assignment values  $m_2(\cdot)$ . Due to the commutative and associative properties of DSm

classic rule of combination, in developing the DSm network, a particular order of layers is not required [54].

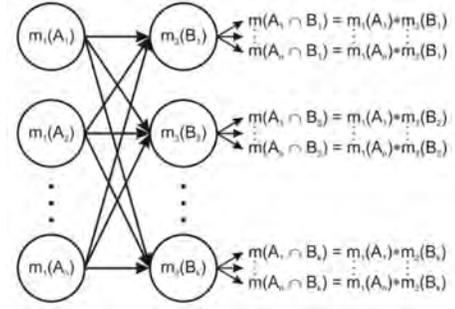


Fig. 1: Graphical representation of DSm classic rule of combination for  $M^f(\Theta)$  [35].

### V. DECISION-MAKING ALGORITHM

As observed in this paper, according to the object shape and assigned task, grasping is divided into 8 categories [63]: spherical grasping, cylindrical grasping, lateral grasp, prismatic grasp, circular grasping, hook grasping, button pressing and pushing. From these grasping types, the most used ones are cylindrical and prismatic grasping (see table I). These can be used in almost any situation and we can say that spherical grasping is a particular grasping of these two. The spherical grasp is used for power grasping, when the contact with the object is achieved with all the fingers' phalanges and the hand's palm. This is why a requirement for classification by the shape of the object is needed. Due to the fact that these types are more often encountered, they were taken into consideration for the studied fusion problems.

The fusion problem aims to achieve a classification, by shape of objects to grasp, so that these can match with the other three types of grasping studied. The target objects are classified into three categories: sphere, parallelepiped and cylinder. For each category, a grasping type is assigned [56].

Following the presented theory in section IV, the information is provided by two independent sources (observers): a stereovision system and a Kinect sensor. The observers are presented in section III, and are used to scan the robot's work environment. By using the information provided by the two observers, a 3D virtual image of the environment is achieved, from which the human operator chooses the object to be grasped, thus defining the grasping task that must be achieved by the robot. The 3D image of the object, isolated from the scene, is compared with three templates, formed by similar methods, which represents a sphere, a parallelepiped and a cylinder. Afterwards, a template matching algorithm is applied to place the object in one of the three categories, with a certain matching percentage. This percentage can vary according to the conditions in which the images are obtained (weak light, object from which the light is reflected, etc.). The data taken from each sensor are then individual processed

with a neutrosophication algorithm, with the purpose of obtaining the generalized basic belief assignment values for each hypothesis that can characterize the system. In the next step, having the basic belief assignment values, we combine the data provided by the two observers by using the classic DSMT rule of combination. The next step is to apply a de-neutrosophication algorithm on the obtained values, to achieve the decision on the shape of the object by placing it into the three categories mentioned above. The entire process is graphical represented in figure 2.

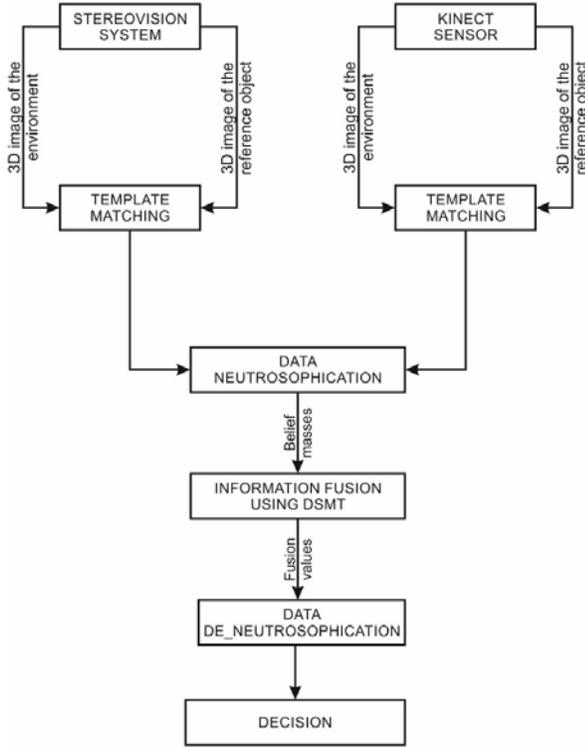


Fig. 2: Diagram of the proposed algorithm.

#### A. Data neutrosophication

Each observer provides a truth percentage for each system's state. The state set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  that characterizes the fusion problem is:

$$\Theta = \{Sp, Pa, Cy\}, \quad (5)$$

where  $Sp$  = Sphere,  $Pa$  = Parallelepiped and  $Cy$  = Cylinder.

To compute the belief values for the hyper power set  $D^\Theta$  elements we developed an algorithm based on the neutrosophic logic. The hyper power set  $D^\Theta$  is formed by using the method presented in the paragraph devoted to the  $D^\Theta$  hyperpower set notion, and has the form:

$$\begin{aligned} D^\Theta = \{ & \emptyset, Sp, Pa, Cy, Sp \cup Pa, Sp \cup Cy, Cy \cup Pa, \\ & Sp \cap Pa, Sp \cap Cy, Cy \cap Pa, Sp \cap (Cy \cup Pa), \\ & Cy \cap (Sp \cup Pa), Pa \cap (Cy \cup Sp), \\ & Sp \cup Cy \cup Pa, Sp \cap Cy \cap Pa \}. \end{aligned} \quad (6)$$

The statements of each observer are handled in ways of truth ( $T$ ), uncertainty ( $I$ ) and falsity ( $F$ ), specific to the neutrosophic logic. Due to the fact that  $F = 1 - T - I$ , the statements of falsity are not taken into consideration.

The neutrosophic algorithm has as input the certainty probabilities (truth) provided by the observers on the system's states. These probabilities are then processed using the described rules in figure 3. If the difference between the certainties probabilities used at a certain point by the processing algorithm is larger than a certain threshold found by trial and error, then we'll consider that the uncertainty percentage between the compared states is null, and the probability that one of the states is true increases. In the case where this difference is not a set threshold, we compute the uncertainty probability by using the formula

$$m(A \cup B) = 1 - \frac{m(A) - m(B)}{const}, \quad (7)$$

where  $A, B \in \Theta$ , and "const" depends of the chosen threshold. While the point determined by the two probabilities approaches the main diagonal, the uncertainty approaches the maximum probability value.

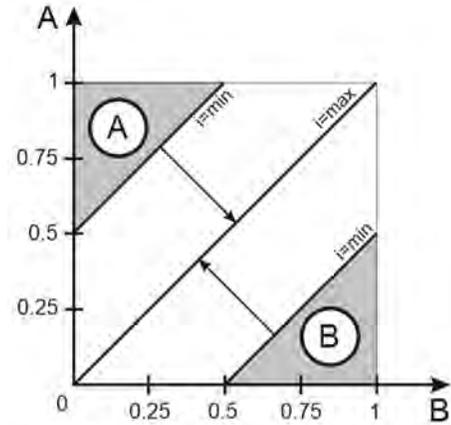


Fig. 3: Data neutrosophication rule for the observer's data.

From the hyper power set  $D^\Theta$ , we can determine the belief masses only for the values  $Obs_i(D^\Theta)$  (information obtained after observer's data interpretation) presented below, because the intersection operation  $\cap$  represents contradiction in DSM theory and we cannot compute the contradiction values for a single observer using:

$$\begin{aligned} Obs_i(D^\Theta) = \{ & Sp, Pa, Cy, Sp \cup Pa, Sp \cup Cy, \\ & Cy \cup Pa, Sp \cap Cy \cup Pa \}. \end{aligned} \quad (8)$$

The neutrosophic probabilities are detailed in Table II.

#### B. Information fusion

Having known the trust values of the hyper power set elements  $Obs_i(D^\Theta)$ , presented in table II, we apply the fusion algorithm, using the classic DSM combination rule described previously.

TABLE II: Grasping position for certain tasks.

Mathematical representation	Description
$Sp$	Certainty that the target object is a 'sphere'
$Pa$	Certainty that the target object is a 'parallelepiped'
$Cy$	Certainty that the target object is a 'cylinder'
$Sp \cup Pa$	Uncertainty that the target object is a 'sphere' or 'parallelepiped'
$Sp \cup Cy$	Uncertainty that the target object is a 'sphere' or 'cylinder'
$Cy \cup Pa$	Uncertainty that the target object is a 'cylinder' or 'parallelepiped'
$Sp \cup Cy \cup Pa$	Uncertainty that the target object is a 'sphere', 'cylinder' or 'parallelepiped'

TABLE III: Contradictions that may appear between the neutrosophic probabilities.

Mathematical representation	Description
$Sp \cap Pa$	Contradiction between the certainties that the target object is a 'sphere' and 'parallelepiped'
$Sp \cap Cy$	Contradiction between the certainties that the target object is a 'sphere' and 'cylinder'
$Cy \cap Pa$	Contradiction between the certainties that the target object is a 'cylinder' and 'parallelepiped'
$Sp \cap (Cy \cup Pa)$	Contradiction between the certainty that the target object is a 'sphere' and the uncertainty that the target object is a 'cylinder' or 'parallelepiped'
$Pa \cap (Sp \cup Cy)$	Contradiction between the certainty that the target object is a 'parallelepiped' and the uncertainty that the target object is a 'sphere' or 'cylinder'
$Cy \cap (Pa \cup Sp)$	Contradiction between the certainty that the target object is a 'cylinder' and the uncertainty that the target object is a 'parallelepiped' or 'sphere'
$Sp \cap Cy \cap Pa$	Contradiction between the certainties that the target object is a 'sphere' and 'cylinder', and 'parallelepiped'

Applying equation (4), we get the following formulas for the combination values:

$$\begin{aligned}
 m(Sp) = & m_1(Sp)m_2(Sp) + m_1(Sp)m_2(Sp \cup Pa) \\
 & + m_1(Sp \cup Pa)m_2(Sp) + m_1(Sp)m_2(Sp \cup Cy) \\
 & + m_1(Sp \cup Cy)m_2(Sp) \\
 & + m_1(Sp)m_2(Sp \cup Cy \cup Pa) \quad (9) \\
 & + m_1(Sp \cup Cy \cup Pa)m_2(Sp) \\
 & + m_1(Sp \cup Pa)m_2(Sp \cup Cy) \\
 & + m_1(Sp \cup Cy)m_2(Sp \cup Pa),
 \end{aligned}$$

$$\begin{aligned}
 m(Pa) = & m_1(Pa)m_2(Pa) + m_1(Pa)m_2(Sp \cup Pa) \\
 & + m_1(Sp \cup Pa)m_2(Pa) + m_1(Pa)m_2(Cy \cup Pa) \\
 & + m_1(Cy \cup Pa)m_2(Sp) \\
 & + m_1(Pa)m_2(Sp \cup Cy \cup Pa) \quad (10) \\
 & + m_1(Sp \cup Cy \cup Pa)m_2(Pa) \\
 & + m_1(Sp \cup Pa)m_2(Cy \cup Pa) \\
 & + m_1(Cy \cup Pa)m_2(Sp \cup Pa),
 \end{aligned}$$

$$\begin{aligned}
 m(Cy) = & m_1(Cy)m_2(Cy) + m_1(Cy)m_2(Cy \cup Pa) \\
 & + m_1(Cy \cup Pa)m_2(Cy) + m_1(Cy)m_2(Sp \cup Cy) \\
 & + m_1(Sp \cup Cy)m_2(Cy) \\
 & + m_1(Cy)m_2(Sp \cup Cy \cup Pa) \quad (11) \\
 & + m_1(Sp \cup Cy \cup Pa)m_2(Cy) \\
 & + m_1(Sp \cup Cy)m_2(Cy \cup Pa) \\
 & + m_1(Cy \cup Pa)m_2(Sp \cup Cy)
 \end{aligned}$$

$$\begin{aligned}
 m(Sp \cup Pa) = & m_1(Sp \cup Pa)m_2(Sp \cup Pa) \\
 & + m_1(Sp \cup Pa)m_2(Sp \cup Cy \cup Pa) \quad (12) \\
 & + m_1(Sp \cup Cy \cup Pa)m_2(Sp \cup Pa)
 \end{aligned}$$

$$\begin{aligned}
 m(Sp \cup Cy) = & m_1(Sp \cup Cy)m_2(Sp \cup Cy) \\
 & + m_1(Sp \cup Cy)m_2(Sp \cup Cy \cup Pa) \quad (13) \\
 & + m_1(Sp \cup Cy \cup Pa)m_2(Sp \cup Cy)
 \end{aligned}$$

$$\begin{aligned}
 m(Cy \cup Pa) = & m_1(Cy \cup Pa)m_2(Cy \cup Pa) \\
 & + m_1(Cy \cup Pa)m_2(Sp \cup Cy \cup Pa) \quad (14) \\
 & + m_1(Sp \cup Cy \cup Pa)m_2(Cy \cup Pa)
 \end{aligned}$$

$$m(Sp \cup Cy \cup Pa) = m_1(Sp \cup Cy \cup Pa)m_2(Sp \cup Cy \cup Pa) \quad (15)$$

During the fusion process, between the information provided by the two observers contradiction situations may appear. These are included in the hyper power set  $D^\ominus$  and are described in table III.

Fusion values for contradiction are determined as following:

$$m(Sp \cap Pa) = m_1(Sp)m_2(Pa) + m_1(Pa)m_2(Sp) \quad (16)$$

$$m(Sp \cap Cy) = m_1(Sp)m_2(Cy) + m_1(Cy)m_2(Sp) \quad (17)$$

$$m(Cy \cap Pa) = m_1(Cy)m_2(Pa) + m_1(Pa)m_2(Cy) \quad (18)$$

$$\begin{aligned}
 m(Sp \cap (Cy \cup Pa)) = & m_1(Sp)m_2(Cy \cup Pa) \\
 & + m_1(Cy \cup Pa)m_2(Sp) \quad (19)
 \end{aligned}$$

$$\begin{aligned}
 m(Pa \cap (Sp \cup Cy)) = & m_1(Pa)m_2(Sp \cup Cy) \\
 & + m_1(Sp \cup Cy)m_2(Pa) \quad (20)
 \end{aligned}$$

$$\begin{aligned}
 m(Cy \cap (Sp \cup Pa)) = & m_1(Cy)m_2(Sp \cup Pa) \\
 & + m_1(Sp \cup Pa)m_2(Cy) \quad (21)
 \end{aligned}$$

C. Data de-neutrosophication and decision-making

The combination values found in the previous section are de-neutrosophicated using the logic diagram presented in figure 4. For the decision-making algorithm we opted to use Petri nets [64], for it's easier to notice the system's states transitions.

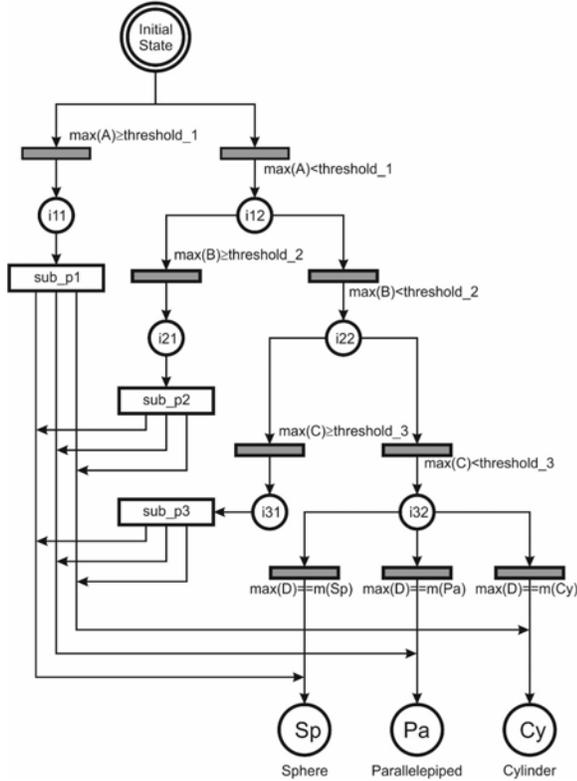


Fig. 4: Petri diagram for decision-making algorithm.

The decision-making diagram proved to have a certain difficulty level, which required adding three sub diagrams:

1) **sub\_p1** (figure 5) – this sub diagram deals with the contradiction between:

- the certainty probability that the target object is a 'sphere' and the uncertainty probability that the target object is either a 'parallelepiped' or a 'cylinder';
- the certainty probability that the target object is a 'cu parallelepiped be' and the uncertainty probability that the target object is either a 'sphere' or a 'cylinder';
- the certainty probability that the target object is a 'cylinder' and the uncertainty probability that the target object is either a 'parallelepiped' or a 'sphere'.

2) **sub\_p2** (figure 6) – this sub diagram deals with the contradiction between:

- The certainty probability that the target object is a 'sphere' and a 'parallelepiped';
- The certainty probability that the target object is a 'sphere' and a 'cylinder';

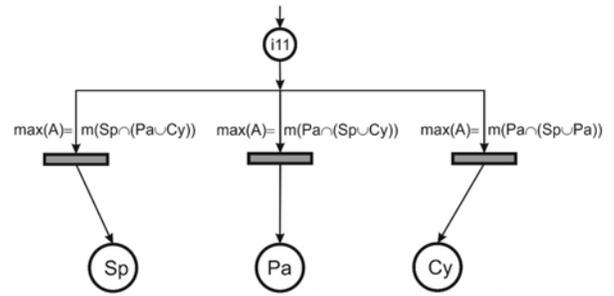


Fig. 5: Petri net for sub\_p1.

- The certainty probability that the target object is a 'cylinder' and a 'parallelepiped'.

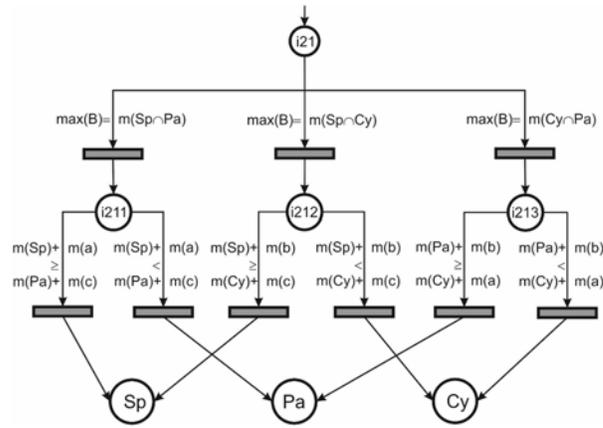


Fig. 6: Petri net for sub\_p2.

3) **sub\_p3** (figure 7) – this sub diagram deals with the uncertainty that the target object is:

- a 'sphere' or a 'parallelepiped';
- a 'sphere' or a 'cylinder';
- a 'cylinder' or a 'parallelepiped'.

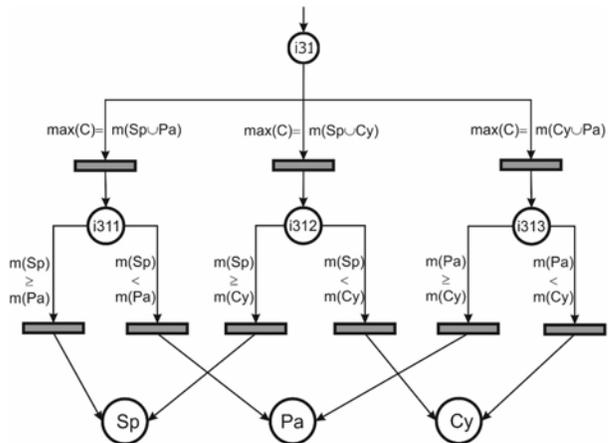


Fig. 7: Petri net for sub\_p3.

To not overload figures 4–7 we have used the following notations:

$$A = \{m(Sp \cap (Cy \cup Pa)), m(Pa(SpCy)), m(Cy \cap (Pa \cup Sp))\},$$

$$B = \{m(Sp \cap Pa), m(Sp \cap Cy), m(Cy \cap Pa)\},$$

$$C = \{m(Sp \cup Pa), m(Sp \cup Cy), m(Cy \cup Pa)\},$$

$$D = \{m(Sp), m(Pa), m(Cy)\},$$

$$a = m(Sp \cup Cy),$$

$$b = m(Pa \cup Sp),$$

$$c = m(Cy \cup Pa).$$

With the help of the Petri diagram (figure 4), we take the decision of sorting the target object in one of the three categories, as follows:

1) Determine

$$m^{\max} \triangleq \max(m(Sp \cap (CyPa)), m(Pa \cap (SpCy)), m(Cy \cap (PaSp))).$$

- If  $m^{\max} = m(Sp \cap (Cy \cup Pa))$ , the contradiction between the certainty value that the target object is ‘sphere’ and the uncertainty value that the target object is ‘cylinder’ or ‘parallelepiped’ is compared with a threshold determined through an experimental trial-error process. If this is higher or equal with the chosen threshold, the target object is a ‘sphere’.
- If  $m^{\max} = m(Pa \cap (Sp \cup Cy))$ , the contradiction between the certainty value that the target object is ‘parallelepiped’ and the uncertainty value that the target object is ‘sphere’ or ‘cylinder’ is compared with the threshold mentioned above. If this is higher or equal with the chosen threshold, the target object is a ‘parallelepiped’.
- If  $m^{\max} = m(Cy \cap (Pa \cup Sp))$ , the contradiction between the certainty value that the target object is ‘cylinder’ and the uncertainty value that the target object is ‘parallelepiped’ or ‘sphere’ is compared with the threshold mentioned above. If this is higher or equal with the chosen threshold, the target object is a ‘cylinder’.

If none of the three conditions are met, we proceed to the next step:

2) Determine

$$m^{\max} \triangleq \max(m(Sp \cap Pa), m(Sp \cap Cy), m(Cy \cap Pa)).$$

- If  $m^{\max} = m(Sp \cap Pa)$ , the contradiction between the certainty values that the target object is ‘sphere’ and ‘parallelepiped’ is compared with a threshold determined through an experimental trial-error process. If this is higher or equal with the chosen threshold, we check if  $m(Sp) + m(Sp \cup Cy) > m(Pa) + m(Cy \cup Pa)$ . If this condition is fulfilled, then the target objects is ‘sphere’. Otherwise, the target object is ‘parallelepiped’.

- If  $m^{\max} = m(Sp \cap Cy)$ , the contradiction between the certainty values that the target object is ‘sphere’ and ‘cylinder’ is compared with the threshold mentioned above. If this is higher or equal with the chosen threshold, we check if  $(Sp) + m(Sp \cup Pa) > m(Cy) + m(Cy \cup Pa)$ . If this condition is fulfilled, then the target objects is ‘sphere’. Otherwise, the target object is ‘cylinder’.
- If  $m^{\max} = m(Cy \cap Pa)$ , the contradiction between the certainty values that the target object is ‘cylinder’ and ‘parallelepiped’ is compared with the threshold mentioned above. If this is higher or equal with the chosen threshold, we check if  $m(Cy) + m(Sp \cup Cy) > m(Pa) + m(Sp \cup Pa)$ . If this condition is fulfilled, then the target objects is ‘cylinder’. Otherwise, the target object is ‘parallelepiped’.

If in none of the situations, the contradiction is not larger than the chosen threshold, we go to the next step:

3) Determine

$$m^{\max} \triangleq \max(m(Sp \cup Pa), m(Sp \cup Cy), m(Cy \cup Pa)).$$

- If  $m^{\max} = m(Sp \cup Pa)$ , the uncertainty probability that the target object is ‘sphere’ or ‘parallelepiped’ is larger than a threshold determined through an experimental trial-error process, we check if  $m(Sp) > m(Pa)$ . If the condition is fulfilled, the target object is ‘sphere’. Otherwise, the target object is ‘parallelepiped’.
- If  $m^{\max} = m(Sp \cup Cy)$ , the uncertainty probability that the target object is ‘sphere’ or ‘cylinder’ is larger than the threshold mentioned above, we check if  $m(Sp) > m(Cy)$ . If the condition is fulfilled, the target object is ‘sphere’. Otherwise, the target object is ‘cylinder’.
- If  $m^{\max} = m(Cy \cup Pa)$ , the uncertainty probability that the target object is ‘cylinder’ or ‘parallelepiped’ is larger than the threshold mentioned above, we check if  $m(Cy) > m(Pa)$ . If the condition is fulfilled, the target object is ‘cylinder’. Otherwise, the target object is ‘parallelepiped’.

If none of the hypotheses mentioned above are not fulfilled, we go to the next step:

4) Determine

$$m^{\max} \triangleq \max(m(Sp), m(Pa), m(Cy)).$$

- If  $m^{\max} = m(Sp)$ , the target object is a ‘sphere’.
- If  $m^{\max} = m(Pa)$ , the target object is a ‘parallelepiped’.
- If  $m^{\max} = m(Cy)$ , the target object is a ‘cylinder’.

## VI. DISCUSSION

As mentioned in the introduction chapter, the main goal of this paper is to find a way to grasp objects according to their shape. This is done by classifying the target objects into three main classes: sphere, parallelepiped and cylinder. To determine the shape of the target objects, the robot work environment was scanned with a stereovision system and a Kinect sensor, with the purpose of creating a 3D image of the surrounding space in which the robot must fulfill its task. From the two created images, the target object is selected and then it is compared with 3 templates, which represents a sphere, a cube and a cylinder. With a template matching algorithm the matching percentage is determined for each of the templates. These percentages (figure 8), represents the data gathered from the observers, for the fusion problem. Because we wanted to test and verify the decision-making algorithm for as many cases as possible, the observers' values were simulated using sine signals with different frequency and amplitude of 1 (figure 8). This amplitude represents the maximum probability percentage that a certain type of object is found by the template matching algorithm.

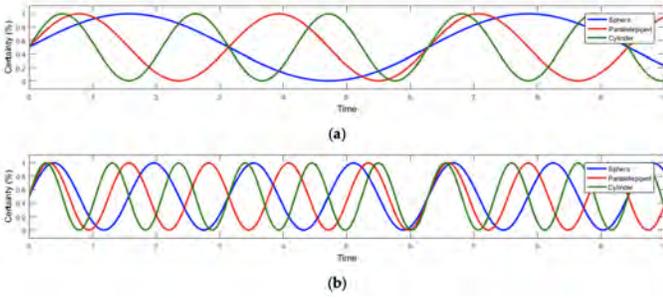


Fig. 8: Simulation of the information provided by the two sensors/observers: (a) first observer detection; (b) second observer detection.

On these input data we then apply a neutrosophication algorithm with the purpose of obtaining the generalized belief assignment values for each of the statements an observer is doing:

- The certainty probability that the object is a 'sphere' (figure 9.a, h)
- The certainty probability that the object is a 'parallelepiped' (figure 9.b, i)
- The certainty probability that the object is a 'cylinder' (figure 9.c, j)
- The uncertainty probability that the object is a 'sphere' or a 'parallelepiped' (figure 9.d, k)
- The uncertainty probability that the object is a 'sphere' or a 'cylinder' (figure 9.e, l)
- The uncertainty probability that the object is a 'cylinder' or a 'parallelepiped' (figure 9.f, m)
- The uncertainty probability that the object is a 'sphere' or a 'cylinder' or a 'parallelepiped' (figure 9.g, n)

After the belief values were computed for each statements of the observers, we go to the data fusion step (figure 10).

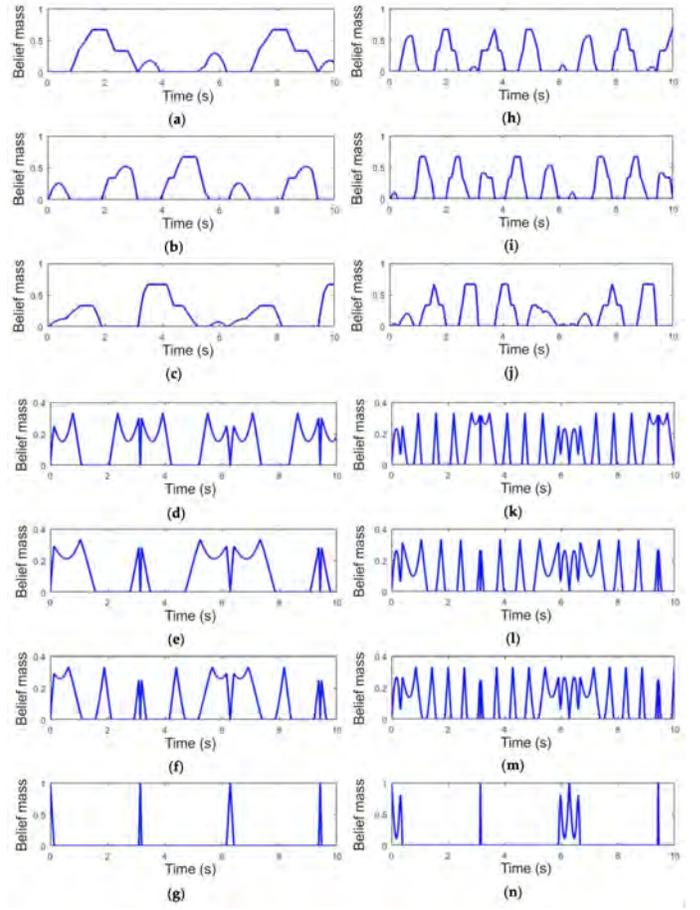


Fig. 9: Generalized trust values. From a to g correspond to observer 1 and from h to n for observer 2 as follows: (a)  $m_1(Sp)$ ; (b)  $m_1(Pa)$ ; (c)  $m_1(Cy)$ ; (d)  $m_1(Sp \cup Pa)$ ; (e)  $m_1(Sp \cup Cy)$ ; (f)  $m_1(Pa \cup Cy)$ ; (g)  $m_1(Sp \cup Pa \cup Cy)$ ; (h)  $m_2(Sp)$ ; (i)  $m_2(Pa)$ ; (j)  $m_2(Cy)$ ; (k)  $m_2(Sp \cup Pa)$ ; (l)  $m_2(Sp \cup Cy)$ ; (m)  $m_2(Pa \cup Cy)$ ; (n)  $m_2(Sp \cup Pa \cup Cy)$ .

With the help of belief values presented in figure 10 and computed using the neutrosophication algorithm described in section V-A, we find the fusion values, presented in figure 11. As one can see in figure 11, the fusion values for certainty, uncertainty and contradiction are minimum. The only exception is the fusion value for the uncertainty that the target object is 'sphere' or 'parallelepiped' or 'cylinder',  $m(Sp \cup Cy \cup Pa)$ , when the data received from the observers are identical and not contradicting, the uncertainty is maximum. This means:

$Obs_1$ : 33.33% sphere, 33.33% parallelepiped, 33.33% cylinder,

and

$Obs_2$ : 33.33% sphere, 33.33% parallelepiped, 33.33% cylinder.

Therefore, the system cannot decide on a single state. This is why the robotic hand will maintain its starting position until the system will decide the target object's category. This indecision period of time takes about 0.07 seconds. When the sensors values about the target object are changed from the

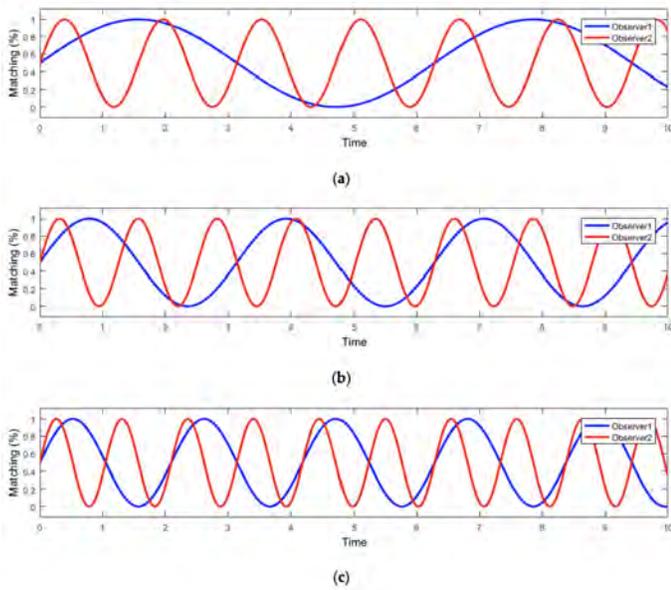


Fig. 10: Data fusion: (a) Observer 1 vs. Observer 2 for Sphere objects; (b) Observer 1 vs. Observer 2 for Parallelepiped objects; (c) Observer 1 vs. Observer 2 for Cylinder objects.

equal values presented above, the algorithm is able to provide a solution.

The indecision also reaches high values at the time 3.14s, 6.28s and 9.42s of the simulation, in the conditions that the observer's statements are close in value with the already presented case from above,

$Obs_1$ : 33.35% sphere, 33.46% parallelepiped, 33.19% cylinder

$Obs_2$ : 33.21% sphere, 33.10% parallelepiped, 33.69% cylinder

for the moment 3.14s,

$Obs_1$ : 33.44% sphere, 33.23% parallelepiped, 33.33% cylinder

$Obs_2$ : 33.44% sphere, 33.23% parallelepiped, 33.33% cylinder

for the moment 6.28s and

$Obs_1$ : 33.39% sphere, 33.70% parallelepiped, 32.91% cylinder

$Obs_2$ : 32.96% sphere, 32.64% parallelepiped, 34.40% cylinder

for the moment 9.42s.

In tables IV, V, and VI we present the percentage of the states' occurrence, the general belief assignment values, the fusion values, and the decision made by the algorithm for the situations previously mentioned.

TABLE IV: Percentages of states' occurrence for each source at different time steps.

state\time	$Obs_1$ 3.14s	$Obs_2$ 3.14s	$Obs_1$ 6.28s	$Obs_2$ 6.28s	$Obs_1$ 9.42s	$Obs_2$ 9.42s
$Sp$	33.35%	33.21%	33.44%	33.44%	33.39%	32.96%
$Pa$	33.46%	33.10%	33.23%	33.23%	33.70%	32.64%
$Cy$	33.19%	33.69%	33.33%	33.33%	32.91%	34.40%

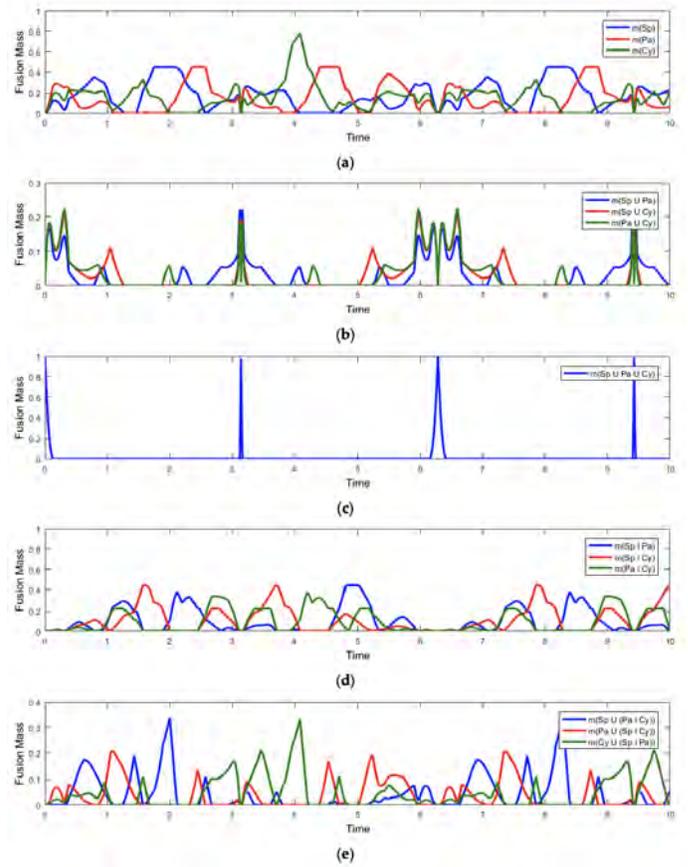


Fig. 11: Fusion values: (a) Fusion values of  $m(Sp)$ ,  $m(Pa)$  and  $m(Cy)$ ; (b) Fusion values of  $m(Sp \cup Pa)$ ,  $m(Sp \cup Cy)$ ,  $m(Pa \cup Cy)$ ; (c) Fusion value of  $m(Sp \cup Pa \cup Cy)$ ; (d) Fusion values of  $m(Sp \cap Pa)$ ,  $m(Sp \cap Cy)$ ,  $m(Pa \cap Cy)$ ; (e) Fusion values of  $m(Sp \cup (Pa \cap Cy))$ ,  $m(Pa \cup (Sp \cap Cy))$ ,  $m(Cy \cup (Sp \cap Pa))$ .

In all three cases, the uncertainty is quite large, and the algorithm ask for restarting the decision process and keeps the decision taken in previous decision process. In our case the decision was that the object is a 'cylinder', 'sphere' and 'cylinder' for the three analyzed points.

Analyzing figure 11(a), at the time of 4.08 seconds the object is decided to be a 'cylinder' because the probability that the target object is a cylinder is very high, but  $m(Cy) = 0.7777$ .

For the time interval of 4.3–4.9 seconds, where in figure 11(d) the contradiction between the target object being a 'sphere' or a 'parallelepiped' is larger than that the target object is a 'sphere' or a 'cylinder' respectively a 'cylinder' or a 'parallelepiped', the object is decided to be a 'parallelepiped' at first because the probability for it being a 'parallelepiped' is larger than the probability of it being a 'sphere' or a 'cylinder'. This situation is changed starting with second 5 of the simulation, when the probability that the target object is a 'sphere' increase, the probability that the same object is a 'cylinder' remains low and the probability that the target

TABLE V: Generalized belief assignment values.

Observations at time	$Obs_1$ 3.14s	$Obs_2$ 3.14s	$Obs_1$ 6.28s	$Obs_2$ 6.28s	$Obs_1$ 9.42s	$Obs_2$ 9.42s
$m_{Obs_i}(Sp)$	0.0001	0.0001	0.0001	0.0001	0.0005	0.0007
$m_{Obs_i}(Pa)$	0.0001	0	0	0	0.0011	0.0067
$m_{Obs_i}(Cy)$	0	0.0008	0	0	0	0
$m_{Obs_i}(Sp \cup Pa)$	0.0106	0.0234	0.0085	0.0085	0.0317	0.0692
$m_{Obs_i}(Sp \cup Cy)$	0.0106	0.0231	0.0085	0.0085	0.0315	0.0669
$m_{Obs_i}(Cy \cup Pa)$	0.0106	0.0231	0.0085	0.0085	0.0312	0.0662
$m_{Obs_i}(Sp \cup Cy \cup Pa)$	0.9680	0.9295	0.9744	0.9744	0.9040	0.7903

TABLE VI: Fusion values and decision at different time steps.

Fusion	$m_{Obs_1} \otimes m_{Obs_2}$ at time 3.14s	$m_{Obs_1} \otimes m_{Obs_2}$ at time 6.28s	$m_{Obs_1} \otimes m_{Obs_2}$ at time 9.42s
$m(Sp)$	0.0006	0.0001	0.0054
$m(Pa)$	0.0006	0.0003	0.0053
$m(Cy)$	0.0012	0.0002	0.0106
$m(Sp \cup Pa)$	0.0328	0.0166	0.0898
$m(Sp \cup Cy)$	0.0325	0.0166	0.0875
$m(Cy \cup Pa)$	0.0324	0.0166	0.0866
$m(Sp \cup Cy \cup Pa)$	0.8999	0.9495	0.7145
$m(Sp \cap Pa)$	0	0	0
$m(Sp \cap Cy)$	0	0	0
$m(Cy \cap Pa)$	0	0	0
$m(Sp \cap (Cy \cup Pa))$	0	0	0.0001
$m(Pa \cap (Sp \cup Cy))$	0	0	0.0001
$m(Cy \cap (Sp \cup Pa))$	0	0	0.0002
<b>Decision</b>	Cylinder	Sphere	Cylinder

object is a ‘parallelepiped’ decrease below the value of the sphere probability.

Using the fusion values and the decision-making diagram (figure 4), from section V-B, we can sort the desired object into the three categories: sphere, parallelepiped and cylinder. The obtained results are presented in figure 12.

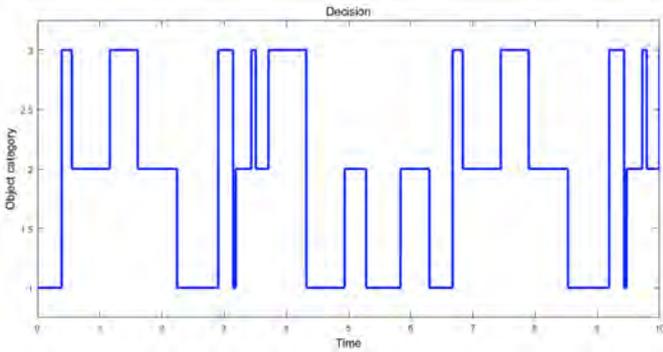


Fig. 12: Object category decision, obtained from the proposed algorithm. Value 1 represents decision for Sphere, value 2 represents decision for Parallelepiped and value 3 represents decision for Cylinder.

## VII. CONCLUSIONS

Any robot, no matter of its purpose, has a task to fulfill. That task can be either of grasping and manipulation or just a transport task. To successfully complete its task, the robot must be equipped with a number of sensors that will provide

enough information about the work environment in which the work is being done.

In this paper we studied the situation in which the robot is equipped with a stereovision system and a Kinect sensor, to detect the environment. The robot’s job was to grab and manipulate certain objects. With the help of two different systems, two 3D images of the environment can be created, each one for the two sensor type. In these images, we isolated the target object and it’s compared with three template images, obtained through similar methods as the environment images. The three template images represent the 3D virtual model of a sphere, a parallelepiped and a cylinder. The comparison is achieved with a template matching method, and following that we obtain a matching percentage for each template tested against the desired image. Because we wanted to develop the decision-making algorithm based on information received from certain template matching methods, we considered as known the information that these algorithms can provide. Moreover, to test different cases, we selected as input for our decision-making algorithm and output for the template matching methods, several sine signals that can provide all the different cases that can occur in practice.

The goal of this paper is in part a data fusion problem with the purpose of classifying the objects in visual range of a humanoid robot, so it can fulfill his grasping and manipulation task. We also wanted to label the target object in one of three categories mentioned above, so that during the approach phase on the target object, the robotic hand can prepare for grasping the object, lowering the time needed to complete the task.

The stereovision system and the Kinect sensor presented in

section III, represent the information sources, called in this paper, the observers, name taken from the neutrosophic logic. These observers specify the state in which the system is. One observer can specify 7 states for the searched object.

With the help of neutrosophic logic, we determine the generalized belief values for each of the 7 states. The neutrosophic algorithm is applied to information gathered from both of the sensors. We have chosen the neutrosophic logic, because it extends fuzzy logic, providing instruments for approaching also the uncertainty situations besides the truth and falsity ones.

Using these belief values, we compute the fusion values on which we apply the classic DSMT combination rule, and build the decision-making algorithm presented in section V. To help develop this decision-making algorithm we used a Petri net which provided us a clear method of switching through system states under certain conditions.

The decision-making algorithm analyzes the probability of completing all the possible tasks that may appear in sensor data fusion and tackles these possibilities so that for every input the system will have an output.

The presented method can be used successfully in real time applications, because it provides a decision in all the cases in a very short time (table VII). The algorithm can be extended so that it can use information received from multiple sources or provide a decision starting from a high number of system states. The number of observer/data sources is not limited nor is the system's states. But while increasing the number of observers and system's states, the data to be processed is increased and the decision-making algorithm design is becoming a highly difficult task to achieve.

TABLE VII: Average execution time of the presented algorithm.

Method	Execution time (s)
Data neutrosophication for Obs. 1	0.0026
Data neutrosophication for Obs. 2	0.0026
Data fusion using DSMT	0.0002
Data de-neutrosophication/decision-making	0.0092
<b>Total time</b>	<b>0.0146</b>

In the case of autonomous robots, these must be taught what to do and how to complete their tasks. From this the necessity of developing new intelligent and reasoning system arises. The developed algorithm in this paper can be used successfully for target identification applications, object sorting, image labeling, motion tracking, obstacle avoidance, edge detection, etc.

#### ACKNOWLEDGMENTS

This work was developed with the support of the Institute of Solid Mechanics of the Romanian Academy, the European Commission Marie Skłodowska-Curie SMOOTH project, Smart Robots for Fire-Fighting, H2020-MSCA-RISE-2016-734875, Yanshan University: 'Joint Laboratory of Intelligent Rehabilitation Robot' project, KY201501009, Collaborative research agreement between Yanshan University, China

and Romanian Academy by IMSAR, RO, and the UEFIS-CDI Multi-MonD2 Project, Multi-Agent Intelligent Systems Platform for Water Quality Monitoring on the Romanian Danube and Danube Delta, PCCDI 33/2018, PN-III-P1-1.2-PCCDI2017-2017-0637.

#### REFERENCES

- [1] V. Vlădăreanu, O. I.Sandru, L. Vlădăreanu, H. Yu, *Extension dynamical stability control strategy for the walking robots*, SKIMA 2012, ISSN online: 1741-5276, ISSN print: 0267-5730.
- [2] Z. Xu, V. Kumar, E. Todorov, *A low-cost and modular, 20-DOF anthropomorphic robotic hand: design, actuation and modeling*, Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids), Atlanta, GA, USA, 15–17 Oct. 2013.
- [3] A. Jaffar, M.S. Bahari, C.Y. Low, R. Jaafar, *Design and control of a multifingered anthropomorphic robotic hand international*, IJMME-IJENS 2011, Vol. 11(4), 2011. <https://doi.org/10.1.1.418.1415>
- [4] M.A. Roa, M. J. Argus, D. Leidner, C. Borst, G. Hirzinger, *Power grasp planning for anthropomorphic robot hands*, IEEE ICRA 2012, pp.663-569, 2012. <https://doi.org/10.1109/ICRA.2012.6225068>
- [5] V. Lippiello, B. Siciliano, L. Villani, *Multi-fingered grasp synthesis based on the object dynamic properties*, RAS 2013. Vol. 61, pp. 626–636, 2013. <https://doi.org/10.1016/j.robot.2013.02.003>
- [6] I.M. Bullock, R.R. Ma, A.M. Dollar, *A hand-centric classification of human and robot dexterous manipulation*, IEEE Trans. on Haptics 2013. Vol. 6(2), pp. 129–144, 2013. <https://doi.org/10.1109/TOH.2012.53>
- [7] R.C. Ormaechea, *Robotic Hands*, AMRS 2011. Vol. 1, pp. 19–39, 2011. [https://doi.org/10.1007/978-0-85729-588-0\\_2](https://doi.org/10.1007/978-0-85729-588-0_2)
- [8] *Enabling the Future*, Available online: <http://enablingthefuture.org> (accessed on 26 March 2018).
- [9] D. Prattichizzo, J.C. Trinkle, *Grasping*, In Springer Handbook of Robotics; Siciliano, B; Khatib, O, Eds; Publisher: Springer-Verlag Berlin, Germany, 2016; pp. 955–988, ISBN: 978-3-319-32552-1; 978-3-319-32550-7. <https://doi.org/10.1007/978-3-319-32552-1>
- [10] T. Feix, J. Romero, H.B. Schmedmayer, A.M. Dollar, D. Kragic, *The GRASP taxonomy of human grasp types*, THMS 2016. Vol. 46, pp. 66–77, 2016. <https://doi.org/10.1109/THMS.2015.2470657>
- [11] A.G. Alvarez, A. Roby-Brami, J. Robertson, N. Roche, *Functional classification of grasp strategies used by hemiplegic patients*, PLOS ONE 2017, Vol. 12, 2017. <https://doi.org/10.1371/journal.pone.0187608>
- [12] C. Fermuller, F. Wang, Y.Z. Yang, K. Zampogiannis, Y. Zhang, F. Baranco, M. Pfeiffer, *Prediction of manipulation actions*, IJCV 2018, Vol. 126, pp. 358–374, 2018. <https://doi.org/10.1007/s11263-017-0992-z>
- [13] J.R. Tsai, P.C. Lin, *A low-computation object grasping method by using primitive shapes and in-hand proximity sensing*, In IEEE ASME Int. Conf. on Advanced Intelligent Mechatronics, pp. 497–502, Munich, Germany, 03–07 Jul 2017. (ISBN: 978-1-5090-6000-9)
- [14] P. Song, Z.Q. Fu, L.G. Liu, *Grasp planning via hand-object geometric fitting*, VS 2018, Vol. 34, pp. 257–270, 2018. <https://doi.org/10.1007/s00371-016-1333-x>
- [15] C. Ma, H. Qiao, R. Li, X.Q. Li, *Flexible robotic grasping strategy with constrained region in environment*, IJAC 2017, Vol. 14, pp. 552–563, 2017. <https://doi.org/10.1007/s11633-017-1096-5>
- [16] R.A.J. Stavenhuter, L. Birglen, J.L. Herder, *A planar underactuated grasper with adjustable compliance*, MMT 2017, Vol. 112, pp. 295–306, 2017. <https://doi.org/10.1016/j.mechmachtheory.2016.08.001>
- [17] C. Choi, S.H. Yoon, C.N. Chen, K. Ramani, *Robust hand pose estimation during the interaction with an unknown object*, in IEEE Int. Conf. on Computer Vision, pp. 3142–3151, Venice, Italy, 22–29 Oct, 2017. (ISBN: 978-1-5386-1032-9)

- [18] D. Seredynski, W. Szykiewicz, *Fast grasp learning for novel objects*, Advances in intelligent systems and computing, Proceedings of Int. Conf. Challenges in Automation, Robotics and Measurement Techniques, Vol. 440, pp. 681–692, Warsaw, Poland, 02–04 Mar, 2016. (ISBN: 978-3-319-29357-8; 978-3-319-29356-1), [https://doi.org/10.1007/978-3-319-29357-8\\_59](https://doi.org/10.1007/978-3-319-29357-8_59)
- [19] W. Shaw-Cortez, D. Oetomo, C. Manzie, P. Choong, *Tactile-based blind grasping: a discrete-time object manipulation controller for robotic hands*, IEEE RAL 2018, Vol. 3, pp. 1064–1071, 2018. <https://doi.org/10.1109/LRA.2018.2794612>
- [20] H.W. Gu, Y.F. Zhang, S.W. Fan, M.H. Jin, H. Liu, *Grasp configurations optimization of dexterous robotic hand based on haptic exploration information*, IJHR 2017, Vol. 14, 2017. <https://doi.org/10.1142/S021984361750013X>
- [21] Y. Yamakawa, A. Namiki, M. Shikawa, M. Shimojo, *Planning of knotting based on manipulation skills with consideration of robot mechanism/motion and its realization by a robot hand system*, SB 2017, Vol. 9, 2017. <https://doi.org/10.3390/sym9090194>
- [22] S.M. Nacy, M.A. Tawfik, I.A. Baqer, *A novel approach to control the robotic hand grasping process by using an artificial neural network algorithm*, JIS 2017, Vol. 26, pp. 215–231, 2017. <https://doi.org/10.1515/jisys-2015-0115>
- [23] L. Zaidi, J.A. Corrales, B.C. Bouzgarrou, Y. Mezouar, L. Sabourin, *Model-based strategy for grasping 3D deformable objects using a multi-fingered robotic hand*, RAS 2017. Vol. 95, pp. 196–206, 2017. <https://doi.org/10.1016/j.robot.2017.06.011>
- [24] H. Liu, P. Meusel, G. Hirzinger, M. Jin, Y. Liu, Z. Xie, *The modular multisensory DLR-HIT-Hand: hardware and software architecture*, IEEE/ASME Trans. on Mechatronics 2008, Vol. 13(4), pp. 461–469, 2008. <https://doi.org/10.1109/TMECH.2008.2000826>
- [25] L. Zollo, S. Roccella, E. Guglielmelli, M.C. Carrozza, P. Dario, *Biomechatronic design and control of an anthropomorphic artificial hand for prosthetic and robotic applications*, IEEE/ASME Transactions On Mechatronics, Vol.12(4), pp. 418–429, 2007. <https://doi.org/10.1109/TMECH.2007.901936>
- [26] E. Lopez-Damian, D. Sidobre, R. Alami, *Grasp planning for non-convex objects*, ISR 2005, Vol. 36, 2005. <https://doi.org/10.1.1.107.2382>
- [27] A.T. Miller, S. Knoop, H.I. Christensen, P.K. Allen, *Automatic grasp planning using shape primitives*, IEEE Rob. & Aut. 2003, Vol. 2, 2003. <https://doi.org/10.1109/ROBOT.2003.1241860>
- [28] J. Napier, *The prehensile movements of the human hand*, JBJS 1956, Vol. 38B(4), pp. 902–913, 1956. [https://doi.org/10.1007/978-1-4471-5451-8\\_85](https://doi.org/10.1007/978-1-4471-5451-8_85)
- [29] M.R. Cutkosky, P.K. Wright, *Modeling manufacturing grips and correlation with the design of robotic hands*, Proc. of the 1986 IEEE Int. Conf. on Robotics and Automation, pp. 1533–1539, San Francisco, CA, USA, 1986.
- [30] T. Iberall, *Human prehension and dexterous robot hands*, IJRR1997, Vol. 16(3), pp. 285–299, 1997. <https://doi.org/10.1177/027836499701600302>
- [31] S.A. Stansfield, *emphRobotic grasping of unknown objects: A knowledge-based approach*, IJRR 1991, Vol. 10(4), pp. 314–326, 1991. <https://doi.org/10.1177/027836499101000402>
- [32] X. Lai, H. Wang, Y. Xu, *A real-time range finding system with binocular stereo vision*, Int. J. of Advanced. Robotic Systems, Vol. 9(1), March–May 2012. <https://doi.org/10.5772/50921>
- [33] A. Oliver, B.C. Wünsche, S. Kang, B. MacDonald, *Using the Kinect as a navigation sensor for mobile robotics*, in Proc. of the 27th Conf. on Image and Vision Computing, pp. 509–514, New Zealand, 2012. (ISBN: 978-1-4503-1473-2)
- [34] I.A. Aljarrah, A.S. Ghorab, I.M. Khater, *Object recognition system using template matching based on signature and principal component analysis*, IJDIWC 2012, Vol. 2(2), pp. 156–163, 2012. (ISSN 2225-658X)
- [35] F. Smarandache, J. Dezert, *Applications and advances of DSMT for information fusion*, Vol. 3, American Research Press, Rehoboth, 2009. (ISBN: 1-59973-073-1, ISSN: 1938-5544)
- [36] B. Khaleghi, A. Khamis, F. Karray, S.N. Razavi, *Multisensor data fusion: A review of the state-of-the-art*, Information Fusion, INFFUS 2013, Vol. 14(1), pp. 28–44, 2013. <https://doi.org/10.1016/j.inffus.2011.08.001>
- [37] Z. Jian, C. Hongbing, S. Jie, L. Haitao, *Data fusion for magnetic sensor based on fuzzy logic theory*, IEEE Int. Conf. on Intelligent Computation Technology and Automation, Vol. 1, Shenzhen, Guangdong, China, 28–29 March 2011. <https://doi.org/10.1109/ICICTA.2011.29>
- [38] M. Munz, K. Dietmayer, M. Mahlich, *Generalized fusion of heterogeneous sensor measurements for multi target tracking*, in Proc. of 13th Int. Conf. on Information Fusion (Fusion 2010), pp. 1–8, 2010.
- [39] Y. Jiang, H.G. Wang, N. Xi, *Target object identification and location based on multi-sensor fusion*, IJAST 2013, Vol. 3(1), pp. 57–67, 2013. <https://doi.org/10.5875/ausmt.v3i1.171>
- [40] D. L.Hall, B. Hellar, M.D. McNeese, *Rethinking the data overload problem: closing the gap between situation assessment and decision-making*, in Proc. of the National Symposium on Sensor Data Fusion, June, 2007.
- [41] J. Esteban, A. Starr, R. Willetts, P. Hannah, P. Bryanston-Cross, *A review of data fusion models and architectures: towards engineering guidelines*, NC&A 2005, Vol. 4(4), pp. 273–281, 2005. <https://doi.org/10.1007/s00521-004-0463-7>
- [42] A. Chilian, H. Hirschmuller, M. Gerner, *Multisensor data fusion for robust pose estimation of a six-legged walking robot*, IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), Sept. 2011. <https://doi.org/10.1109/IROS.2011.6048125>
- [43] B.V. Dasarathy, *Editorial: Information fusion in the realm of medical applications - a bibliographic glimpse at its growing appeal*, Information Fusion 2012, Vol. 13 (1), pp. 1–9, 2012. <https://doi.org/10.1016/j.inffus.2011.06.003>
- [44] D.L. Hall, R.J. Linn, *Survey of commercial software for multisensor data fusion*, in Proc. of SPIE Conf. Sensor Fusion and Aerospace Applications, Orlando, FL, USA, April 1991.
- [45] L. Vlădăreanu, V. Vlădăreanu, P. Schiopu, *Hybrid force-position dynamic control of the robots using fuzzy applications*, AMM 2013, Vol. 245, pp. 15–23, 2013. <https://doi.org/10.4028/www.scientific.net/AMM.245.15>
- [46] B.R. Gaines, *Fuzzy and probability uncertainty logics*, Inf. & Ctrl. 1978, Vol. 38(2), pp. 154–169, 1978. [https://doi.org/10.1016/S0019-9958\(78\)90165-1](https://doi.org/10.1016/S0019-9958(78)90165-1)
- [47] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Vol. 73, no. 363, 1976.
- [48] A. Sahbani, S. El-Khoury, P. Bidaud, *An overview of 3D object grasp synthesis algorithms*, RAS 2012, Vol. 60(3), pp. 326–336, 2012. <https://doi.org/10.1016/j.robot.2011.07.016>
- [49] R. De Souza, S.El. Khoury, A. Billard, *Towards comprehensive capture of human grasping and manipulation skills*, Thirteenth International Symposium on the 3-D Analysis of Human Movement, No. EPFL-CONF-198301, 2014.
- [50] I.M. Bullock, A.M. Dollar, *Classifying human manipulation behavior*, IEEE Int. Conf. on Rehabilitation Robotics Rehab Week, Zurich, ETH Zurich Science City, Switzerland, pp. 1–6, 2011.
- [51] A. Morales, P. Azad, T. Asfour, D. Kraft, S. Knoop, R. Dillmann, A. Kargov, Ch. Pylatiuk, S. Schulz, *An anthropomorphic grasping approach for an assistant humanoid robot*, Int. Symposium of Robotics, January 2006.
- [52] R. Etienne-Cwmings, P. Pouliquen, M.A. Lewis, *Single chip for imaging, color segmentation, histogramming and template matching*, Electron Letters, Vol. 2(4), pp. 172–174, 2002. <https://doi.org/10.1109/ISSCC.2002.992090>
- [53] I. Guskov, *Kernel-based template alignment*, in Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 610–617, 2006. <https://doi.org/10.1109/CVPR.2006.162>
- [54] F. Smarandache, J. Dezert, *Applications and advances of DSMT for information fusion*, Vol. 1, American Research Press, Rehoboth, USA, 2004.
- [55] F. Smarandache, L. Vlădăreanu, *Applications of neutrosophic logic to robotics: An introduction*, In Proc. GrC, pp. 607–612, Kaohsiung, Taiwan, 8–10 Nov., 2011. <https://doi.org/10.1109/GRC.2011.6122666>
- [56] I.-A. Gal, L. Vlădăreanu, H. Yu, *Applications of neutrosophic logic approaches in 'RABOT' real time control*, SISOM 2012 and Session of the Commission of Acoustics, Bucharest, 2013.
- [57] I.-A. Gal, L. Vlădăreanu, R.I. Munteanu, *Sliding motion control with bond graph modeling applied on a robot leg*, RRST/SEE 2010, Vol. 60(2), pp. 215–224, 2010. (ISSN: 0035-4066)

- [58] V. Vlădăreanu, I. Dumitrache, L. Vlădăreanu, I.S. Sacala, G. Tont, M.A. Moiescu, *Versatile intelligent portable robot control platform based on cyber physical systems principles*, SIC 2015, Vol. 24(4), pp. 409–418, 2015. (WOS:000366543700005, ISSN: 1220-1766)
- [59] D.O. Melinte, L. Vlădăreanu, R.A. Munteanu, H. Wang, F. Smarandache, M. Ali, *Nao robot in virtual environment applied on VIPRO platform*, Acta Electrotehnica, Vol.57, 2016. (WOS:000399457100072, ISSN 2344-5637)
- [60] F. Smarandache, *A unifying field in logics: neutrosophic field*, Multiple-valued logic, International Journal 2002. Vol. 8(3), pp. 385–438, 2002.
- [61] H. Wang, F. Smarandache, Y.Q. Zhang, R. Sunderraman, *Interval Neutrosophic sets and logic: Theory and application in computing*, HEXIS Neutrosophic Book Series, No. 5, 2005. (ISBN: 1-931233-94-2) <https://doi.org/10.6084/m9.figshare.1014223.v1>
- [62] F. Smarandache, *Neutrosophy : neutrosophic probability, set, and logic; analytic synthesis & synthetic analysis*, 105 pages, American Research Press, Gallup, NM, USA, 1998. (ISBN 1-87958-563-4)
- [63] J. Rosell, X. Sierra, F.L. Palomo, *Finding grasping configurations of a dexterous hand and an industrial robot*, R&A 2005, pp. 1178–1183, 2005. (ISSN: 1050-4729) <https://doi.org/10.1109/ROBOT.2005.1570275>
- [64] S. Emadi, F. Shams, *Modeling of component diagrams using Petri Nets*, IJST 2010, Vol. 3(12), pp. 1151–1161, 2010. (ISSN: 0974- 6846) <https://doi.org/10.17485/ijst/2010/v3i12/29851>

# Combination of Geophysical and Geotechnical Data Using Belief Functions: Assessment with Numerical and Laboratory Data

Théo Dezert<sup>a,b</sup>, Sérgio Palma Lopes<sup>a</sup>, Yannick Fargier<sup>b,c</sup>, Philippe Côte<sup>a</sup>

<sup>a</sup>IFSTTAR, GERS, GeoEND, F-44344 Bouguenais, France.

<sup>b</sup>CEREMA Direction Territoriale Normandie-Centre, F-41000 Blois, France.

<sup>c</sup>Univ. Lyon, IFSTTAR, GERS, RRO, F-69675 Bron, France.

Emails: theo.dezert@ifsttar.fr, sergio.lobes@ifsttar.fr, yannick.fargier@ifsttar.fr, philippe.cote@ifsttar.fr

Originally published as: T. Dezert, S. Palma Lopes, Y. Fargier, P. Cote, *Combination of Geophysical and Geotechnical Data Using Belief Functions: Assessment with Numerical and Laboratory Data*, Journal of Applied Geophysics, Vol. 170, November 2019, and reprinted with permission.

**Abstract**—The identification of the subsoil constitutive materials, as well as the detection of possible interfaces and anomalies, are crucial for many site characterization applications. During investigation campaigns, complementary geophysical and geotechnical methods are usually used. These two sets of methods yield data with very different spatial scales and different levels of incompleteness, uncertainty and inaccuracy. In this work, a mathematical combination of geophysical and geotechnical information is proposed in order to produce a better subsoil characterization. It is shown that belief functions can be used for such a fusion process. A specific methodology is developed in order to manage conflictual information and different levels of uncertainties and inaccuracies from different investigation methods. In order to test and validate this methodology, we focus on the use of two selected methods, Electrical Resistivity Tomography (ERT) and Cone Penetration Test. First, a synthetic model with artificial data is considered, taking advantage of the results obtained to conduct a comparative study (effect of parameters and noise level). Then, an experimental test bench is considered, in which a two-layered model is placed (plaster and saturated sands) and geophysical and geotechnical data are generated, using a mini-ERT device and insertion depth values. This work also aims at providing a better graphical representation of a subsoil section with associated degrees of belief. The results highlight the ability of this fusion methodology to correctly characterize the considered materials as well as to specify the positions of the interfaces (both vertical and horizontal) and the associated levels of confidence.

**Keywords:** data fusion, belief functions, geophysical data, geotechnical data, experimental test bench, electrical resistivity tomography.

## I. INTRODUCTION

For subsoils characterization, investigation campaigns are set up, usually consisting of geophysical and geotechnical methods. These two families of methods are complementary and are used for various issues such as the characterization of slope stability [1-4] the characterization of potentially dangerous sites [5], the characterization of sites at construction [6] or the characterization of river embankments [7].

On the one hand, geophysical methods are non-intrusive and provide physical information on large volumes of soils but with significant potential uncertainties. These uncertainties are

due in particular to the integrative and indirect aspects of the methods as well as to the resolution of the inverse problems. On the other hand, the geotechnical investigation methods are intrusive and provide more punctual information but also more accurate. An important issue for the assessment of subsoils is to be able to combine acquired geophysical and geotechnical data, while taking into account their respective uncertainties, inaccuracies and spatial distributions [8]. The complementarity of these two sets of methods is often underused since the uncertainty and inaccuracy associated with each method are rarely considered. Furthermore, the results are usually only graphically superimposed [9] instead of being mathematically merged.

To characterize a section of subsoil and its potentially risky areas, it is essential to distinguish the different materials in place. The horizontal and vertical interfaces, as well as possible anomalies, have to be located. For levee embankment, as an example, it is in these locations that internal erosion is likely to develop, which may lead to the complete rupture of the levee [10]. Such a section characterization, with associated confidence indexes, could be included in failure hazard models.

The use of belief functions [11-12] and different information combination rules to combine geotechnical and geophysical data is proposed. This makes it possible to take into account at the same time the uncertainties, inaccuracies and incompleteness of data related to each method. In the field of geosciences, belief functions have already been used and provide interesting results for slope instability mapping [13-14], detection of precious metal [15], groundwater [16] or flood susceptibility mapping [17]. To our best knowledge, no work has been proposed, considering the combination of two sources of information with different spatial distribution (spatialized and punctual) and for an investigation campaign in the vertical section.

Here, an innovative method of information fusion to combine electrical resistivity tomography results and cone penetrometer test data is proposed. First, work on data obtained from synthetic models is displayed. The obtained results allow

to conduct a comparative study, evaluating the effect of different parameters (like the data noise level) on the fusion result. The fusion methodology is then tested from data acquired on a test bench. In this work, the potential of such a methodology is shown by using insertion depth data, acquired by a laboratory penetration cone, and electrical resistivity data acquired by a mini Electrical Resistivity Tomography (ERT) device. The depth of penetration data corresponds to geotechnical information while the electrical resistivity data correspond to geophysical information. The main concern is to highlight the ability of this information fusion algorithm to characterize the interfaces between materials and to discriminate three different types of materials with variation in thickness of one of them, and to present the variation of the results according to the number and position of the simulated boreholes.

The main contributions of this work are as follows. First, this new methodology makes it possible to take into account the uncertainties, inaccuracies and incompleteness associated with the different methods of investigation used, proposing a modeling of the Basic Belief Assignments (BBAs) specifically adapted to the problematic. Then, the proposed graphic representation is innovative since it allows both to present the different geological sets that would be present in the subsoil and their layout, while presenting the confidence associated with these results. This methodology is particularly suitable for the characterization of interfaces and anomalous zones, which may correspond to areas where the risk of instability is potentially the greatest. This work also allows the implementation of a small physical model to validate the fusion approach with real data.

This article is organized as follows. In section II a presentation of the approach of fusion used in the methodology is given, which introduces the use of the evidence theory and the combination methods used here. In section III, a synthetic study will then present the fusion approach from artificial data. It will also present the comparative results associated to two parametric studies. Then, in section IV, a presentation of the investigation methods used in the introduced experiment (laboratory penetration cone and mini ERT device) is given. Finally, the test bench fusion results are presented in section V and discussed in section VI, in order to understand the interests, limitations and perspectives of such a methodology.

## II. FUSION METHODOLOGY

### A. Belief functions and combination rules

The belief functions have been introduced by Shafer [11] in 1976 in the development of his mathematical theory of evidence inspired by previous works of Dempster [12]. Shafer's theory is also referred as Dempster-Shafer theory (DST) in the literature. This theory (proposes a method to) calculate(s) the belief and the plausibility of an event (here a soil material class) from distinct source of evidence (measured data). The practical advantage of using such a theory lies in its ability to manage information from different sources, associated with variable uncertainties and inaccuracies. In this work, only two sources of information will be considered: geotechnical and

geophysical. Another advantage of this theory is its ability to assess the degree of conflict between sources (ex: contradictory information between data obtained from large scale geophysical campaign and from punctual geotechnical investigation). Uncertainties correspond to degrees of confidence that are given to a value, whereas inaccuracies correspond to intervals of values that can be directly associated with measurement errors related to the method. For example, the uncertainty of measuring a geotechnical parameter identical to the one measured in a borehole increases with the distance to that point. The inaccuracy can for its part, be associated with the error bar of the result. The belief functions allow to take into account the ignorance and incompleteness of the information. It is indeed possible to grant credit on all the possible results in order to quantify the ignorance. For the reader eager to learn more, the theory is detailed in [18].

A Bayesian approach as part of a subjective probability approach [19] could have been considered for geophysical and geotechnical data combination. However, the main limitation of such an approach is that probabilities essentially represent uncertainty and only very poorly the level of inaccuracy. Moreover, in the probabilistic modeling stage, the different decisions (events) are only represented on singletons (i.e. single events) and are necessarily considered exhaustive and exclusive. The exclusivity is implied by the assumption of the additivity of probabilities. However, this hypothesis may be too strong and limit the representation of the knowledge. Furthermore, with a Bayesian approach, it is difficult to model the lack of knowledge or the knowledge that is not expressed in probability distributions.

In order to define and to use the belief functions, it is necessary (i) to set a frame of discernment, (ii) to assign belief mass values to the events of this framework (Basic Belief Assignments - BBAs), (iii) to choose a fusion rule for combining information; and (iv) to represent the combined information.

The Frame of Discernment (FoD)  $\Theta$  is made of all the possible events about the problem under concern, the elements of the FoD are exclusive and exhaustive, so that for  $n$  events:

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}. \quad (1)$$

In the considered problematic, the possible events of the FoD correspond to intervals of values of geophysical and geotechnical parameters that can be associated with classes of geological materials (for example,  $\theta_1$ =clays,  $\theta_2$ =sands, ...). The space of belief mass functions, the set of all subsets of  $\Theta$ , written  $2^\Theta$ , is fixed by all the disjunctions and by the possible conflict between the sources of information (written  $\emptyset$ ) such that:

$$2^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_1 \cup \theta_2, \theta_3, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3, \dots, \theta_1 \cup \theta_2 \cup \theta_3 \cup \dots \cup \theta_n\}. \quad (2)$$

As in the probability theory, the belief mass function  $m_j$  is defined, for a source of evidence  $S_j$  (for  $j = 1, 2$ ), attributed

to  $A$  (defined on  $2^\Theta$ ) in  $[0, 1]$  such that the more  $m(A)$  tends to 1 and the more the confidence in  $A$  is important :

$$\sum_{A \in 2^\Theta} m(A) = 1. \quad (3)$$

The difference with the probability theory lies in the fact that  $A$  can represent the union of several events (for example, either  $\theta_1$  OR  $\theta_2$ ). It is therefore possible to model uncertainty and lack of knowledge. For instance, when no information is available about the achievement of an event member of  $\Theta$ , one can set  $m_j(\Theta) = 1$ , avoiding the uniform distribution that would have been considered in a probabilistic scheme. Combination rules, as part of the belief functions theory, can thus take different levels of uncertainty and imprecisions into account according to the source of information. If only defined on singletons, the belief mass function is similar to a probability distribution.

Smets fusion approach developed in his Transferable Belief Model (TBM) [20] (i.e. conjunctive fusion) allows the attribution of a mass of belief to the conflict, outside the FoD, so that (open-world assumption):

$$m_{12}(\emptyset) > 0. \quad (4)$$

Where  $m_{12}(\cdot)$  denotes the combined BBA resulting from the combination of information of sources 1 and 2. The belief mass resulting from the fusion of information from source 1 and 2 is written:

$$m_{12}(A) = \sum_{\substack{X, Y \subseteq \Theta \\ X \cap Y = A}} m_1(X)m_2(Y). \quad (5)$$

And the level of conflict between the two considered sources of information can therefore be quantified by:

$$m_{12}(\emptyset) = \sum_{\substack{X, Y \subseteq \Theta \\ X \cap Y = \emptyset}} m_1(X)m_2(Y), \quad (6)$$

with  $m_1(X)$  and  $m_2(Y)$  the belief masses respectively attributed to events  $X$  and  $Y$  by sources 1 and 2.

According to Shafer's approach and unlike Smets' rule, Dempster-Shafer's rule (DS) does not allow the attribution of a mass of belief to the conflict (closed-world assumption):

$$m_{12}^{DS}(\emptyset) = 0. \quad (7)$$

The conflict is there reallocated through a classical normalization factor. The mass of belief in  $A$ ,  $m_{12}^{DS}(A)$ , resulting from the fusion of information from sources 1 and 2 is written:

$$m_{12}^{DS}(A) = \frac{1}{1 - m_{12}(\emptyset)} \sum_{\substack{X, Y \subseteq \Theta \\ X \cap Y = A}} m_1(X)m_2(Y). \quad (8)$$

The disadvantage of this method is that the conflict between the sources is no longer represented and it is possible to obtain counterintuitive results if the conflict is important because of this normalization. Even more problematic, even if the distinct sources are both informative whatever the level of conflict is, Dempster-Shafer's fusion process can even not take into account the second source of information [21].

The disadvantage of this method is that the conflict between the sources is no longer represented and it is possible to obtain counterintuitive results if the conflict is important because of this normalization. Even more problematic, even if the distinct sources are both informative whatever the level of conflict is, Dempster-Shafer's fusion process can even not take into account the second source of information [21].

$$m_{12}^{PCR6}(A) = m_{12}(A) + \sum_{\substack{Y \in 2^\Theta \\ A \cap Y = \emptyset}} \left[ \frac{m_1(A)^2 m_2(Y)}{m_1(A) + m_2(Y)} + \frac{m_2(A)^2 m_1(Y)}{m_2(A) + m_1(Y)} \right]. \quad (9)$$

### B. Construction of BBAs from geophysical and geotechnical data

Belief masses have to be assigned to each considered event of the FoD, for both sources of information. The combination of the belief masses can only be initiated after this stage. In the following, the geophysical source of information will be identified as source 1 and the geotechnical source of information as source 2. A 2D model assumption will be made, corresponding to the  $x$  and  $z$  spatial axes, since vertical sections of subsoil are considered.

#### Geophysical data

The discretization of the considered subsoil section, as well as the depth of investigation and the resolution, depend on the acquisition method used [24]. It is the user who sets, using the inversion tool used, the shape and dimensions of the discretization grid used. It is about starting from this discretization and being able to associate for each cell, masses of beliefs for each event of the FoD.

The constitutive classes of the FoD are also fixed at the end of the inversion process by the geophysicist, with the help of a representation of the distribution of the set of inverted geophysical values, in the form of modal classes (Figure 1.a). The representation in this form makes it possible to highlight the centers, minima and maxima of the events considered in order to be able to fix the bounds of the intervals associated with the events of the FoD. The number of cells of the subsoil section are represented according to the geophysical parameter values. The infima and suprema must be fixed so that the intervals are of the same width in order to avoid the appearance of a bias when calculating Wasserstein distances (detailed under). To associate the belief masses with the FoD events, the intervals of inverted values of the physical parameter (in red, Figure 1.b) are considered. For some geophysical methods, these intervals can correspond to the value obtained at the end of the inversion with its associated inaccuracy.

It is then necessary to associate belief mass values  $m_1(\cdot)$  corresponding to each element of  $2^\Theta$ , for each cell of the inverted section. The masses are obtained from the calculation of Wasserstein distances [25], considering two geophysical intervals  $A = [a_1, a_2]$  and  $B = [b_1, b_2]$  with  $A$  and  $B$  belonging to  $\mathbf{R}$ ,  $A$  being the interval corresponding to an event

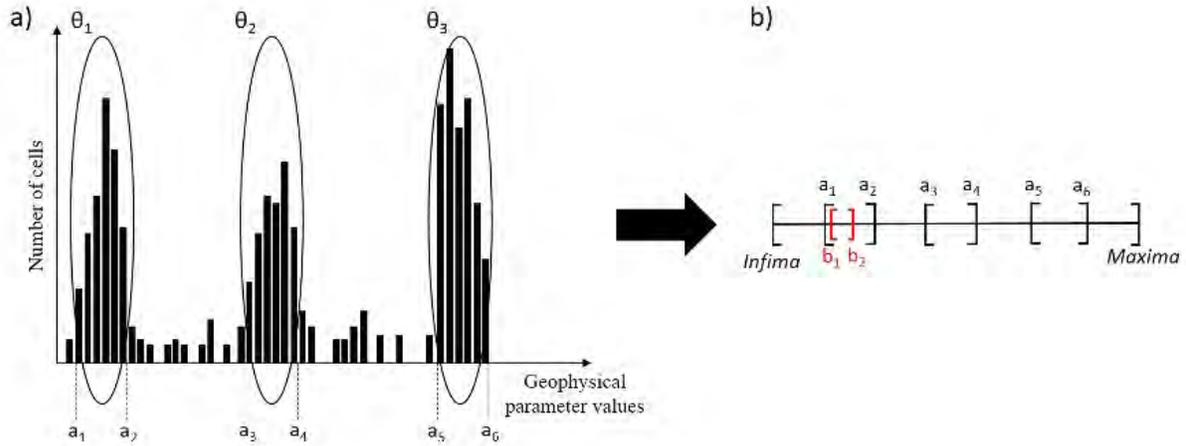


Fig. 1. a) Model classes' distribution of the geophysical parameter values from the considered subsoil section, allowing the selection of the geophysical classes in b). The red interval corresponds to an interval of inverted values, from one cell of a 2D section of subsoil, used for Wasserstein distances' calculation.

of the FoD and  $B$  being an interval of inverted values (Figure 1.b), Eq. (10):

$$d_{\text{Wass}}^2(A, B) = \left[ \frac{a_1 + a_2}{2} - \frac{b_1 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{a_2 - a_1}{2} - \frac{b_2 - b_1}{2} \right]^2 \quad (10)$$

This calculation estimates the distance between two intervals according to their size and the distance between them. The Wasserstein distances are calculated (using a logarithmic scale if the geophysical parameter requires it) between the inverted values with estimated inaccuracies, and the intervals associated with each event, chosen by the geophysicist. Each cell is finally associated with a standardized BBA respecting Eq. (3). This way, the more the distance of a geophysical interval resulting from the inversion is “close” to one event of the FoD, the more the mass of belief associated is important, and reciprocally.

### Geotechnical data

For the geotechnical part, the information proposed during an investigation campaign is spatially punctual (in the  $x$ - $z$  plane) and often contained in vertical soundings made from the surface. It is about associating masses of belief with the different events of the FoD for each cell of the considered vertical soundings. For this, the values proposed at each depth are considered with the associated inaccuracy, corresponding to the measurement error that could be attributed to the measuring device (Figure 2.a). Thus, as for the geophysical part, intervals of values are obtained.

The geotechnical mesh consisting of as many cells in depth as the number of geotechnical values (Figure 2.b) is generated. A mass of belief  $m_2(\cdot) = 1$  is assigned, in the drilling points,

to the events corresponding to the measured geotechnical parameter. A value of 1 is set since we are very confident in the information inside the boreholes unlike the spatialized geophysical information. A new mesh is then constructed (Figure 2.c), according to the size and depth of the boreholes. In order to characterize the entire section of the model, as does the geophysical method, and to associate mass values to each new cell (BBA), an exponential lateral decay of the belief mass is imposed, from the drilling point to the nearest borehole so that the decay rate is a function of the values proposed by the nearby borehole. So that, for a specific depth, Eq. (11):

$$BBA(x) = e^{-kC_v x} BBA(0), \quad (11)$$

with  $x$  being the distance from the considered cell to the reference borehole ( $x = 0$  in the borehole),  $k$  a decay factor fixed by the user to adjust the lateral decay rate,  $BBA(x)$  the belief mass values assigned to each event of the FoD for a position  $x$ , with  $BBA(0) = 1$ .  $C_v$  corresponds to the coefficient of variation expressed in Eq. (12), such as used in [26]:

$$C_v = \frac{\sqrt{\frac{1}{n_{\text{mesh}}-1} \sum_{i=1}^{n_{\text{mesh}}} (Q - Q_i)^2}}{Q}, \quad (12)$$

where  $Q$  is the geotechnical value of the reference cell in the considered borehole and  $Q_i$  the geotechnical value in the nearby borehole. For Figure 2.b,  $n_{\text{mesh}} = 3$  has been considered. If  $n_{\text{mesh}} = 5$  or  $7$ , the computation of the  $C_v$  will take into account 5 or 7 cells in the nearby borehole. Indeed, for two consecutive boreholes with similar values, at similar depth, the decay of the confidence is slower than for two consecutive boreholes presenting radically different

values. This decay of belief mass is carried out to the left and to the right, from each drilling.

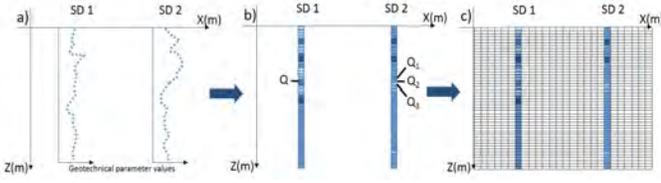


Fig. 2. Construction of a geotechnical discretization mesh from two vertical boreholes acquisition (SD1 and SD2). a) Representation of the geotechnical values for SD1 and SD2 according to the depth. b) The boreholes are divided in cells associated with belief mass equal to 1 for the considered event. c) Construction of a new mesh according to the size and depth of the boreholes.

If, between two boreholes, the mass of belief associated with a hypothesis  $A$  is less than 1 ( $m_2(A) < 1$ ), then the remainder of mass to be allocated to satisfy Eq. (3), is reported on the proposition “any type of material” represented by the union of all events, such as Eq. (13):

$$m_2(\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4) = 1 - m_2(A). \quad (13)$$

### C. Dimensioning of the mesh prior to the fusion

Each source of information imposes its own mesh but in order to combine the belief masses from the geophysical information source (source 1) and the geotechnical source (source 2), it is necessary to have a common mesh containing, for each cell, the geophysical and geotechnical BBAs. In order to not alter the quality of the information, no interpolation is carried out. It is decided to superimpose the geophysical discretization grid resulting from the 2D inversion to the geotechnical division, depending on the number and the borehole positions. Thus, an irregular mesh is obtained but without any approximation (Figure 3).

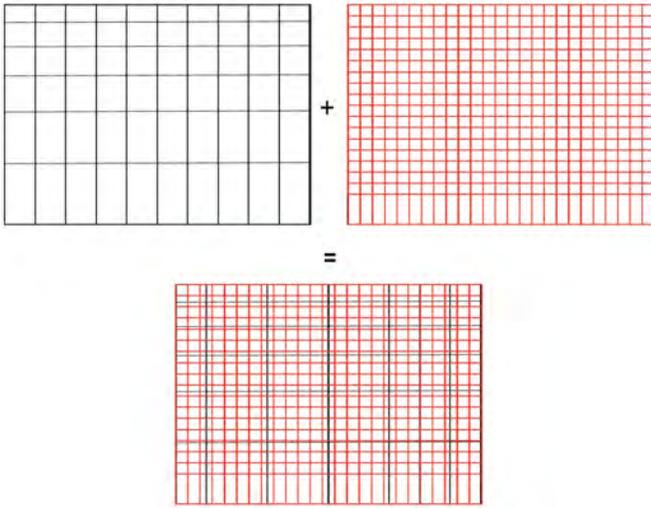


Fig. 3. Example of a geophysical mesh (in black) and a geotechnical mesh (in red) superimposed to propose a new irregular mesh to carry out the combination calculations and present the fusion results.

## III. SYNTHETIC STUDY

Below, a synthetic study based on artificial data is proposed in order to test this new proposed methodology. It is the opportunity to show the impact of different levels of noise on the geophysical information as well as the influence of the lateral decay factor  $k$  (Eq.11) on the results of the fusion in order to be able to choose a value for the use of such a methodology from real data.

### A. Considered methods

For this study, the electrical resistivity tomography (ERT) method stands for the geophysical information source and the Cone Penetrometer Test (CPT) method for the geotechnical information source.

The basic principle of DC-resistivity methods consist in injecting an electric current of known intensity [A] by means of two “current” electrodes and measuring a voltage [V] between two “potential” electrodes. Depending on the electrode layout, the topography, the properties of the materials and their distribution, apparent resistivity values can be computed. The depth of investigation depends on the spacing of the electrodes, the configuration of the electrodes and the nature of the soil [27]. By generalizing this principle, a two dimensional (2D) ERT consists in aligning a series of electrodes and acquiring a large number of measurements based on four electrodes configuration. The apparent resistivity data acquired are then inverted using an inversion code or software to reconstruct a complete 2D-section of electrical resistivity [ $\Omega \cdot m$ ]. Here the Res2Dinv software (ver 3.71.118) [27] has been used.

In order to obtain an artificial resistivity section of subsoil, a two steps procedure is followed. First, resistivity data are simulated using the Res2Dmod software [28], on the section that we want to consider. Second, apparent electrical resistivity values are inverted with Res2Dinv, considering a L1 norm [29] and an extended model discretization, to obtain the synthetic inverted section of electrical resistivity.

The CPT method consists in pushing rods into the soil, at a constant speed, with a conical tip at the end [30]. This test is often used for the determination of the soils mechanical resistance properties. The two measured parameters are the tip resistance  $q_c$  [MPa] and sleeve friction  $f_s$  [MPa]. Although the method uses two parameters, only  $q_c$  will be considered as the study parameter.

### B. FoD and considered model

For this synthetic study, a two-layer model is considered, composed of materials that can be likened to silts for the upper layer and clays for the underlying layer. The FoD therefore contains three material class hypotheses, such as:

$$\Theta = \{\theta_1, \theta_2, \theta_3\}, \quad (14)$$

with  $\theta_1$  the event corresponding to the clayey material,  $\theta_2$  to the silty material and  $\theta_3$  to unknown materials. The latter is associated with the union of the geophysical and geotechnical value ranges that do not correspond to those associated with  $\theta_1$  and  $\theta_2$ . This event  $\theta_3$  allows us in a certain way, to quantify

the lack of knowledge of the environment since it does not include the two first sets. The construction of the BBAs then consists in associating the data of the two considered sources to the events of the FoD. Figure 4 shows the two-layer model based on events from the FoD, used for this synthetic study.



Fig. 4. Representation of the events of the FoD in the imposed model of subsoil of the synthetic study.

### C. Construction of BBAs from geophysical and geotechnical data

#### Geophysical data

The electrical acquisition is simulated with a Wenner acquisition mode and with 96 electrodes interspaced from one meter. An electrical resistivity of  $100 \Omega\cdot\text{m}$  is considered for the upper material and a resistivity of  $30 \Omega\cdot\text{m}$  for the underlying one [31] (Figure 4). Electrical acquisitions are simulated with different noise levels (5, 10 and 15%). The results of this inversion (with 10% noise, figure 5.a) allow to highlight the presence of two layers but the interface between these two layers is not perfectly identified. A variation of thickness in the center of the model is visible. The interface is not straightforward and anomalies are present on the surface even though they are not part of the initial model.

From these inversion results, it is possible to define the ranges of electrical resistivities that will be associated with the different events considered for the fusion process. A distribution in modal classes is used to visualize the number of cells, in the discretized section of the 2D inversion, associated with specific range of resistivities (Figure 5.b). This distribution allows to highlight the two large material classes of the model. Thanks to it, the bounds of the considered events can thus be defined (in  $\Omega\cdot\text{m}$ ), so that the intervals have the same length (in logarithmic scale):

$$\begin{aligned}\theta_1 &= [25; 45], \\ \theta_2 &= [83; 149.4], \\ \theta_3 &= [13.89; 25 \cup 45; 83 \cup 149.4; 268.92].\end{aligned}\quad (15)$$

As explained in II.B., it is possible to associate belief masses with each cell of the mesh thanks to the values resulting from the inversion. As part of the construction of geophysical BBAs, the values presented Figure 6 are obtained. This figure highlights the association of the values of Figure 5.a with the events of the FoD, Eq. (15). The presence of a top layer ( $\theta_2$ ) and a base layer ( $\theta_1$ ) can be detected (Figure 6.a). It appears that there is a variation in the thickness of the layers in the center of the model, but the interface is not well characterized. Moreover, the intermediate values of electrical resistivity resulting from the inversion (Figure 5.a) between

$\theta_2$  and  $\theta_1$  layers induce the representation of a third material ( $\theta_3$ ) which has no reality in the model that has been fixed. The belief masses are maximum when the resistivity values correspond to the center of the resistivity classes set for each event (Eq. (15)).

#### Geotechnical data

Concerning the source of geotechnical information, the simulation of four vertical CPT soundings interspaced from 19 meters is proposed (Figure 7). 20 cm wide and up to 15 m deep boreholes are considered, and a value of  $q_c$  is recorded every 50 cm from the surface. An inaccuracy of  $10^{-2}$  MPa on the measurements is considered. For a fixed normalized friction ratio of 3%, a value of  $q_c$  of 20 MPa is considered for the upper silty material and a value of 0.2 MPa for the underlying clay material, as proposed in the Robertson diagram [32].

In order not to have uniform values of  $q_c$  for the materials and to try to represent the noisy reality of an acquisition in the field, values are drawn following a normal distribution defined for each event. Mean  $q_c$  values of 0.2 and 20 MPa are respectively used to define the normal distributions of the material classes. Standard deviation values equal to 10% of the mean values are associated, echoing the 10% noise used for the geophysical data. Keeping the minimum and maximum values, these random draws, make it possible to define the limits, in MPa, of the intervals associated with the elements (i.e. material classes) of the FoD:

$$\begin{aligned}\theta_1 &= [0.14; 0.27], \\ \theta_2 &= [13.5; 23.5], \\ \theta_3 &= [0.1; 0.14 \cup 0.27; 13.5 \cup 23.5; 100].\end{aligned}\quad (16)$$

The minimum and maximum values are fixed at 0.1 and 100 respectively because they are the minimum and maximum values in Robertson's diagram [32].

There are two types of sounding results according to their position (Figure 8). Once the values associated with the meshes of the sounding are obtained, it is possible to associate masses of belief to the whole section by extending the geotechnical information, as explained in II.B. In the framework of the construction of geotechnical BBAs and for  $k = 0.1$  (Eq.11), the obtained values are proposed in Figure 9. This figure highlights the fact that the confidence is maximum in the soundings. This method allows us to characterize the material  $\theta_2$  on the first 5 meters of the model and the material  $\theta_1$  from 10 to 15 meters deep.

The greater thickness of the material  $\theta_2$  in the center of the model is also well characterized. On the other hand, a great doubt appears in yellow (Figure 9.a) in certain areas, not allowing the determination of a specific material ( $\theta_1 \cup \theta_2 \cup \theta_3$ ). For the model base area, this can be explained by the fact that the soundings stop at 15 meters depth. Regarding the areas between 5 and 10 meters to the right and left of the first and last sounding, this is related to the fact that for these two soundings, the closest soundings propose different values at the same depths, the confidence attributed to the presence of

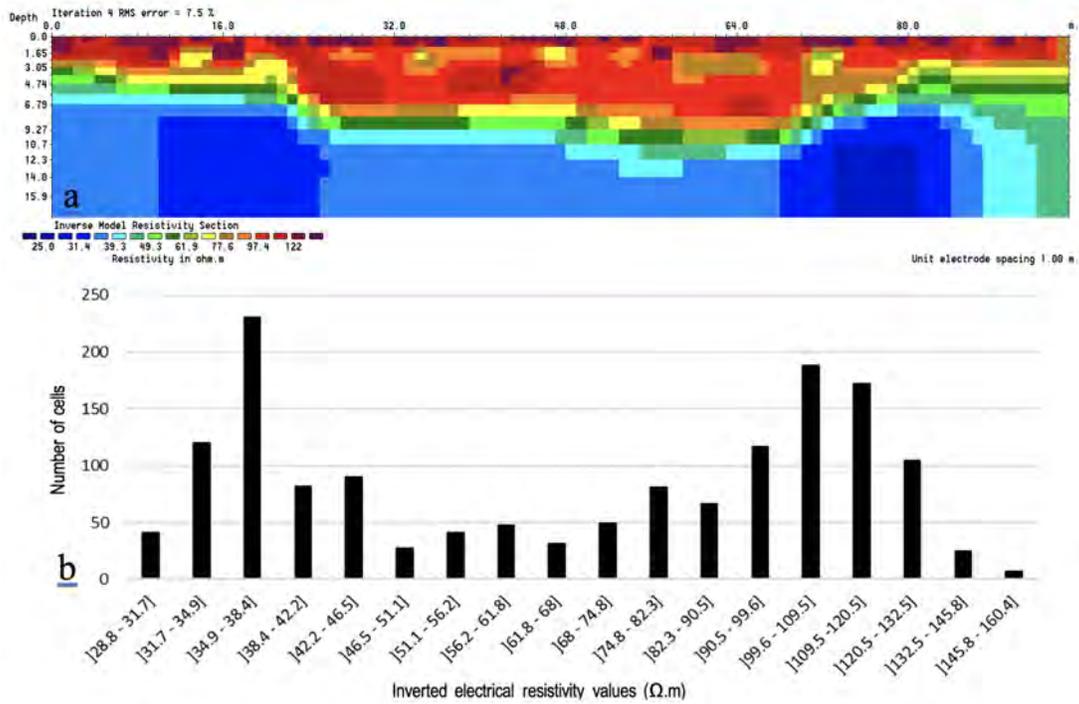


Fig. 5. a) Subsoil section displaying inverted electrical resistivity values from 10% noise data acquisition and b) model classes' distribution of the cells presented in a), according to the electrical resistivity values ( $\Omega.m$ ).

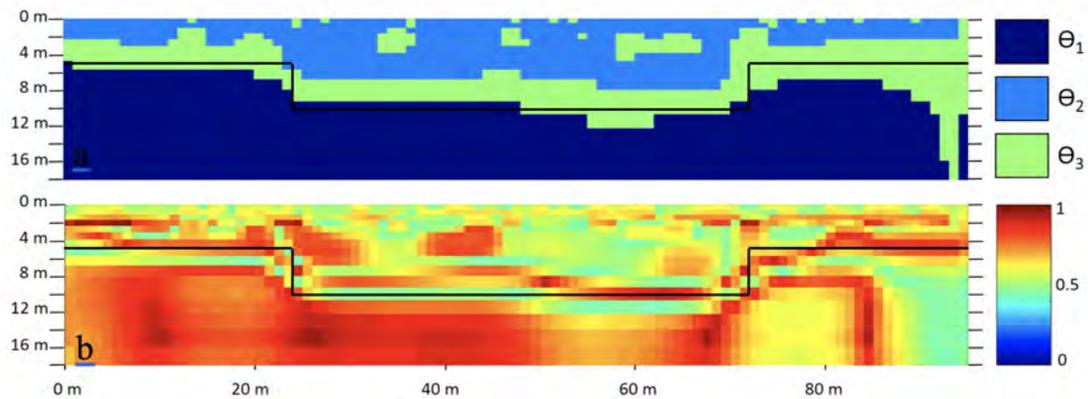


Fig. 6. a) Representation of the event having the highest belief mass according to the BBA construction from geophysical data and b) the associated belief mass values, considering a 10% noise. The black lines represent the position of the interface.

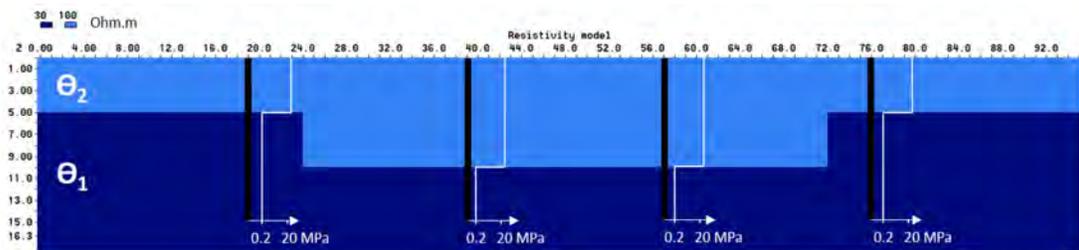


Fig. 7. 2D section of subsoil displaying true ER distribution with boreholes positions in black and associated tip resistance vertical profiles in white.

$\theta_1$  therefore decreases very quickly laterally. This decay is also high towards the edges of the model because no other

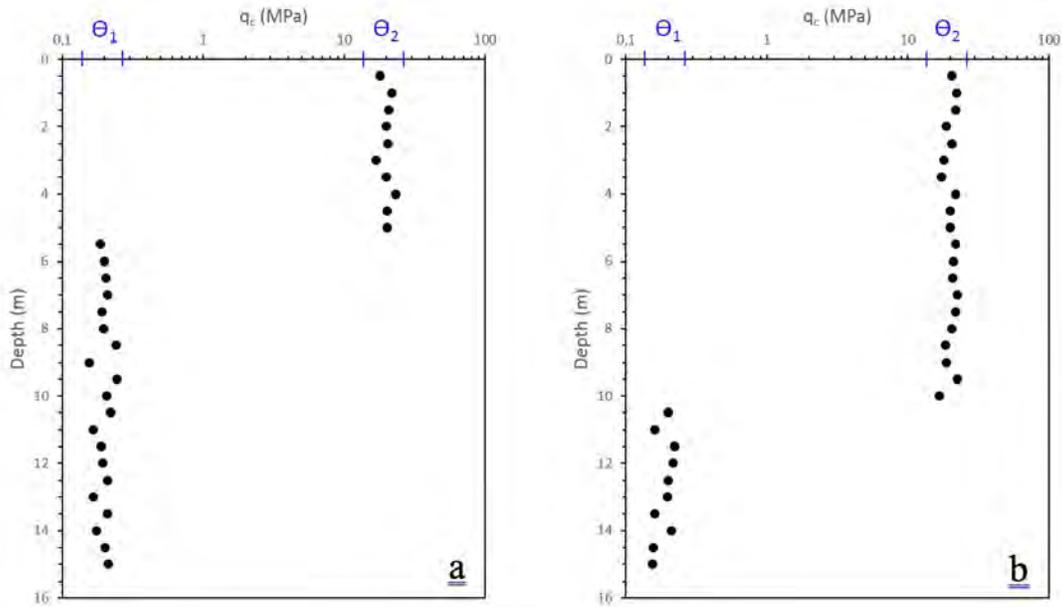


Fig. 8. Examples of the two types of simulated soundings with tip resistance values according to the investigation depth. a) corresponds to borehole # 1 and 4 on figure 7 while b) corresponds to borehole # 2 and 3 on figure 7.

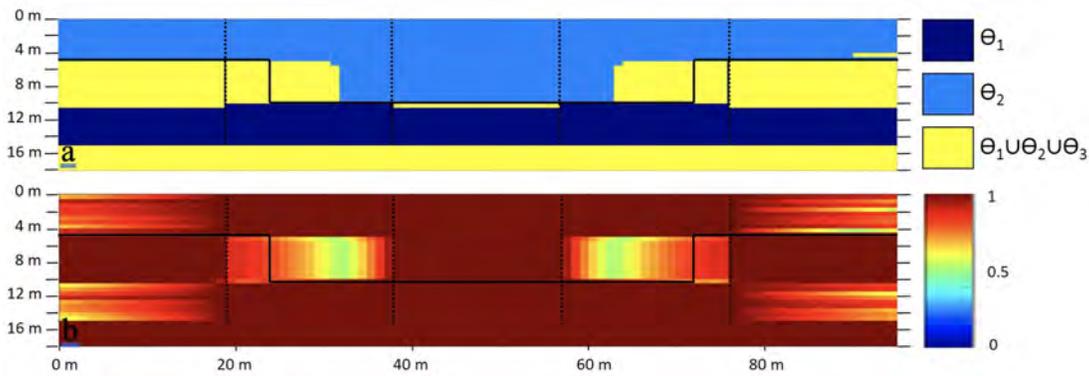


Fig. 9. a) Representation of the events having the highest belief mass according to the BBA construction from geotechnical data and b) the associated belief mass values. The borehole positions are in dashed lines while the black lines represent the position of the interface.

sounding is present at the ends to constrain the information.

*D. Effect of lateral decay factor and noise level on the fusion results*

We examine in this part the results of the fusion of belief masses established for the proposed synthetic model by varying the noise level of the geophysical information, as well as the value of the lateral decay factor  $k$  (Eq.11) influencing the lateral decay rate of geotechnical information.

Figure 10 shows the fusion results with different values of  $k$  ( $10^{-2}$ ,  $5 \cdot 10^{-2}$ ,  $10^{-1}$ ,  $5 \cdot 10^{-1}$  and 1) for a simulated noise of 10% on the acquired geophysical information. Noise was set at 10% since the electrical resistivity classes of the FoD were defined from the modal classes of the inverted 10% noise image, Figure 5.b. For each value of  $k$ , Figures 10.a and 10.b represent the results obtained by Smets fusion whereas Figures 10.c and 10.d represent the results obtained by PCR6

fusion. While Figures 10.a and 10.c show the material classes having the greatest mass of belief at the end of the fusion process, Figures 10.b and 10.d correspond to the values of these respective belief masses, between 0 and 1. These figures, display the events (materials) potentially present within the section, as well as their attached level of confidence.

The higher the value of  $k$  is, the higher the rate of confidence in geotechnical information is. This can be seen, for example, from the last borehole to the right end of the section (Figures 10.b) or between the 2nd and 3rd boreholes (Fig 10.d). The increase of  $k$  implies that for two soundings offering similar values at the same depth, the confidence associated with the corresponding type of material will tend to decrease. On the other hand, for two soundings proposing different values at the same depth, the increase of  $k$  will hardly have any impact on the belief masses associated with the selected events (e.g.

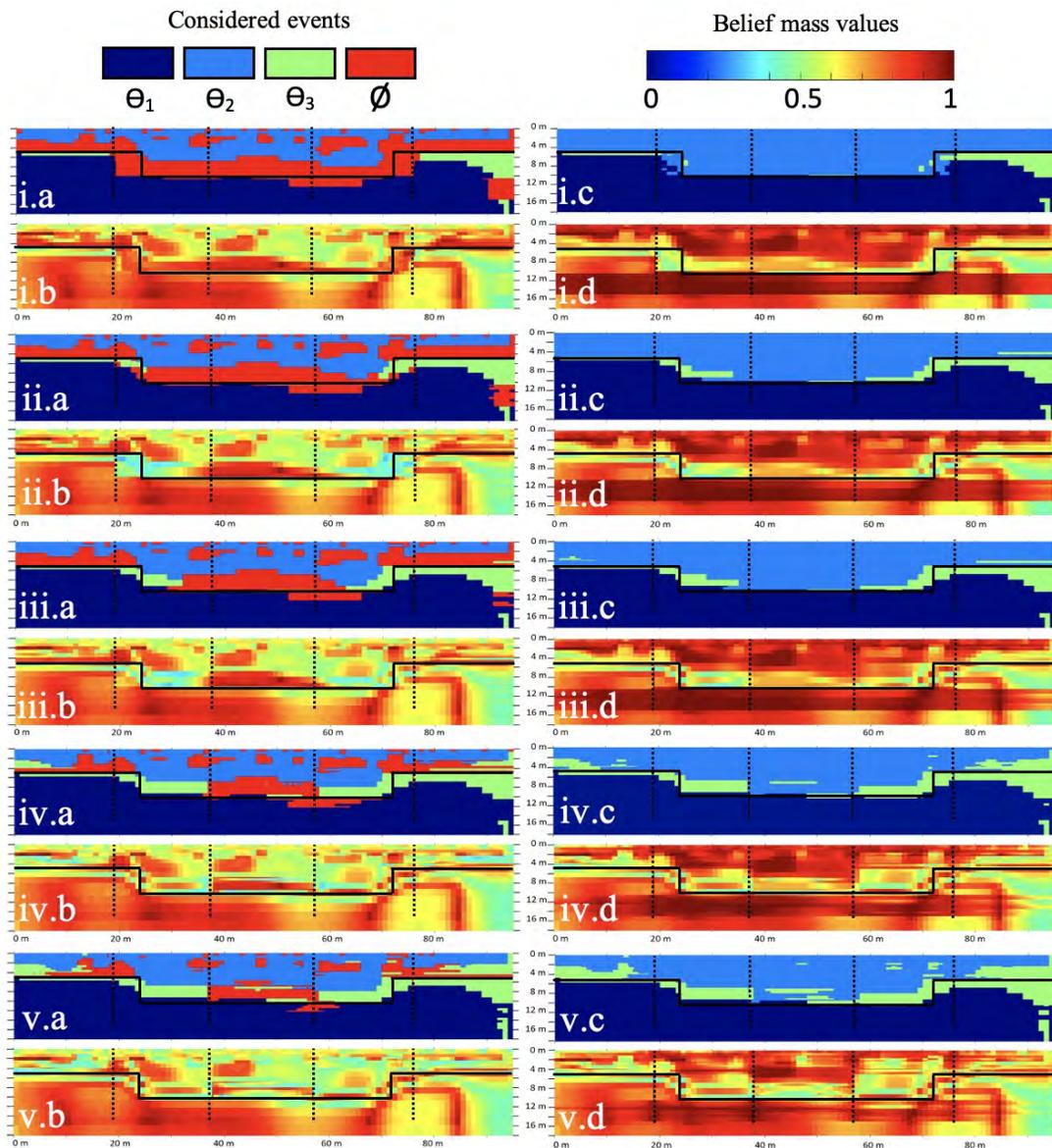


Fig. 10. Representation of the events having the highest belief mass in a) and c) and their associated mass values in b) and d), considering a 10% noise. In i)  $k = 0.01$ , ii)  $k = 0.05$ , iii)  $k = 0.1$ , iv)  $k = 0.5$  and v)  $k = 1$ . Figures on the left side are results for Smets fusion while figures on the right side are results for PCR6 fusion. The sounding positions are in dashed lines while the black lines represent the position of the interface.

between 5 and 10 m of depth between the boreholes 1 and 2, Figures 10.d).

With regard to the material classes identified after the fusion process, the more  $k$  increases, the more the quantity of conflict decreases (in red, Figures 10.a). This is explained by the fact that when there is little trust in the geotechnical data, there is little conflict with the geophysical data. In the meantime, an increase in the proportion of  $\theta_3$  is observed (Figures 10.c) close to the interface. This observation is explained by a larger mass attributed to the union of events and by geophysical data which propose intermediate values at the interface level.

In the following of this work, an intermediate value of  $k$  will be retained, equal to 0.1. With such parameter value, a good

confidence in information repeating between two successive soundings is obtained, but it also leaves room for doubt by having enough unknown material ( $\theta_3$ ) at the interfaces. The obtained fusion results with different noise levels added to the geophysical information (5, 10 and 15%) are shown in Figures 10.iii and 11 with  $k = 0.1$ .

The greater the amount of noise is, the less clear the interfaces proposed by the inversion are (Figures 5.a, 11.i.a, 11.ii.a). A greater number of anomalies are also present when the noise level increases. The noise level finally impacts the level of inaccuracy associated with the geophysical data used in the fusion process. Larger data inaccuracies induce wider value ranges considered for calculating Wasserstein distances,

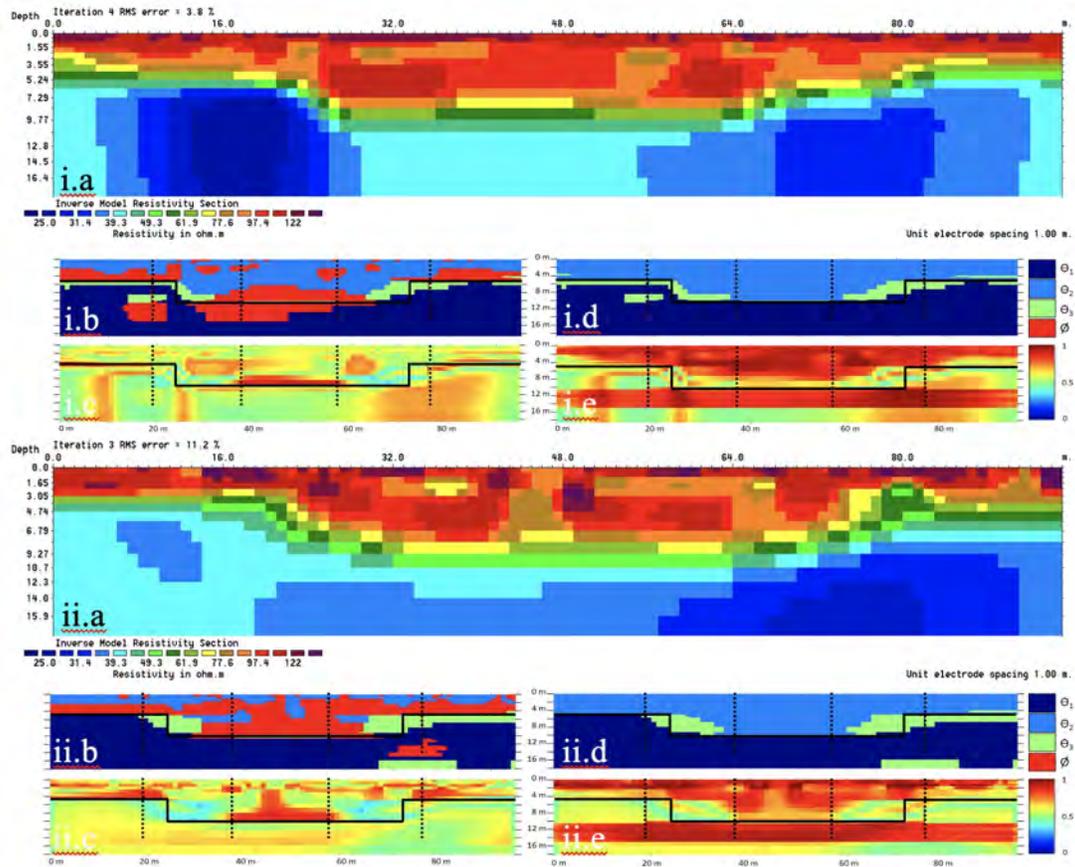


Fig. 11. a) Subsoil section displaying inverted electrical resistivity values from i) 5% noise and ii) 15% noise data acquisition. Representation of the events having the highest belief mass in b) and d) and their associated mass values in c) and e). Figures on the left side are results for Smets fusion while figures on the right side are results for PCR6 fusion. The sounding positions are in dashed lines while the black lines represent the position of the interface.

which in turn can bring to consider belief masses on more events of the FoD.

Since the classes associated with FoD elements were fixed from the values with 10% noise (section III.C and Figure 4.b), it is “reasonable” to have a higher confidence (higher belief masses) on these results than on the results with 5% and 15% noise (Figures 11.c, 11.e, 10.iii.b, 10.iii.d). The fusion process allows to override the noise effects, whether the noise level is 5 or 15%. This can be imputed to the computation of Wasserstein distances, taking into account the data inaccuracies and considering all geophysical classes.

#### IV. SETTING UP A TEST BENCH FOR REAL GEOTECHNICAL AND GEOPHYSICAL ACQUISITIONS

##### A. Materials

In order to be able to assess the validity of the developed fusion methodology, two methods of data acquisition were retained: (i) a mini-ERT device acting as the geophysical source of information and (ii) a laboratory penetration cone acting as the geotechnical source of information. Before setting up the test bench, it was necessary to select the materials that could be put in place in a tank in order to carry out the study. This selection implies that the materials used meet

several conditions in order to validate the methodology: they must have (i) distinct electrical resistivity ranges, (ii) distinct penetration depths and (iii) a certain homogeneity in the space to limit uncontrolled anomalous values.

1) *Mini ERT device*: Expressly for the purposes of this study, a mini ERT device (Figure 12) has been set up. This device consists of 48 electrodes of 6 mm length, positioned at regular intervals of one centimeter. It can be moved along the test bench to make multiple acquisitions and to cover a longer section.

2) *Laboratory penetration cone*: The laboratory penetration cone method is described in the French standard NF P 94-052-1 [33]. It consists in measuring a penetration depth of a cone, in millimeters, subjected to its own weight (Figure 13). The materials are tested individually, repeatedly, to determine an average value and a standard deviation of penetration depth for each material. These values can be used later in the study to simulate different drilling positions within the test bench. This method can be likened to the CPT method which is one of the most popular in situ geotechnical tests.

3) *Test bench and used materials*: For the validation of the methodology, we wanted to build a test bench that could be easily set up and controlled, with two or three layers and

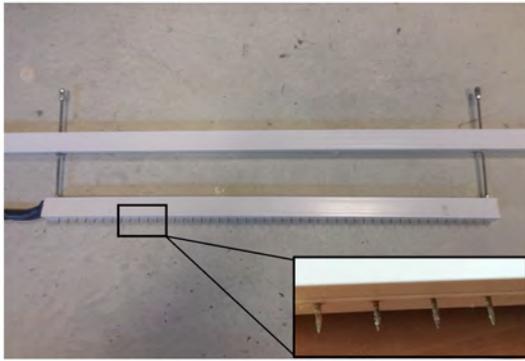


Fig. 12. Mini ERT device with 48 electrodes spaced 1 cm, adjustable height, used for electrical acquisitions in the test bench.



Fig. 13. Laboratory penetration cone.

variation of the interface positions. Fast-hardening natural fine-grained plaster as well as Hostun fine sand [34] are the retained constituents. These two materials meet the three conditions listed above. They were placed in a transparent PVC tank of  $100 \times 30 \times 17 \text{ cm}^3$  as shown in Figure 14 with an underlying layer of 5 cm of plaster (setting time = 69 h) overlaid by a layer of 2.5 cm of water saturated sand.

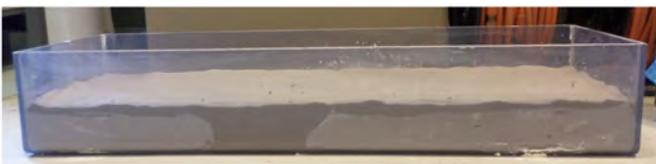


Fig. 14. Transparency view of the test bench.

A formwork was made during the placement of the plaster so that a 20 cm long anomaly could be inserted in. Saturated sand of 7.5 cm thickness is present instead of plaster. The contact between the materials and the bottom of the tank is at the origin of an interface that will be interesting to detect with the help of the methodology. 16 kg of plaster were mixed with 8 kg of water to obtain the material finally put in place. The electrical resistivity of the plaster was measured before and after the placement of the saturated sand to verify that the presence of water had a negligible impact on the electrical properties of the plaster.

TABLE I  
VALUES OF ELECTRICAL RESISTIVITY AND DEPTH OF PENETRATION OF THE MATERIALS SET UP WITHIN THE TEST BENCH.

	Plaster before pluviation	Saturated Hostun sands
<b>Electrical resistivity (<math>\Omega \cdot \text{m}</math>)</b>		
Mean	31.28	78.15
Standard deviation	3.23	11.18
Number of measures	12	52
<b>Penetration depth (mm)</b>		
Mean	0.11	17.31
Standard deviation	0.04	1.61
Number of measures	8	10

For the Hostun sand, 15.82 kg were pluviated in 5.8 kg of water, above the plaster to reach saturation. Trials had been carried out in advance to determine the proportions of water and sand required to achieve such a state as well as to validate the repeatability of such installation by pluviation. The values of electrical resistivities and penetration depths are displayed in Table I.

### B. FoD and BBA modeling

1) *FoD and target model*: A FoD consisting of four elements (material classes) is considered so that:

$$\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}, \quad (17)$$

with  $\theta_1$  the element corresponding to the plaster material;  $\theta_2$  corresponding to saturated sand;  $\theta_3$  corresponding to the hard and electrically insulating bottom of tank simulating a substrate and  $\theta_4$  corresponding to unknown materials, being the union of the ranges of values not corresponding to those associated with the 3 previously described materials. Figure 15 presents the target model in the form of events constituting the FoD, following the disposition of the materials within the test bench. Although the tank used is 1 m long, the ERT acquisition only covered a 83 cm long section, on the central line of the model, and allowed us to image up to 18 cm of depth.

2) *Construction of BBAs from geophysical and geotechnical data*: The electrical acquisition was carried out on 83 cm long, on the central line of the model, with a first acquisition on 47 cm, and three next acquisitions done after respective displacements of 12 cm (roll along method). The results obtained from the inversion of the acquired data are displayed in Figure 16.a. These results make it possible to highlight the existence of three distinct sets, at depths relatively close to the target model

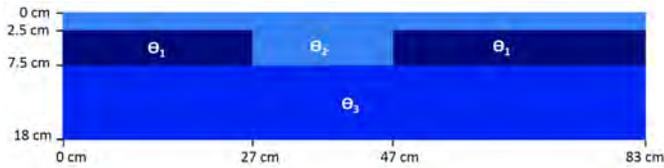


Fig. 15. Scheme of the idealized section model (with vertical exaggeration), including the FoD constituent events associated with the materials of the test bench.

(Figure 15) but presenting vertically slightly shifted interfaces, gradual rather than sharp. In addition, the variation in saturated sand thickness is poorly evaluated. Indeed, the anomalous zone is recognized but associated here, in its lower part, with values of electrical resistivities much larger than what they really are.

The proposed values, although in the same order of magnitude, do not exactly match the ranges of values measured on the materials independently (Table I). In order to characterize the events (materials) of the FoD, a distribution in modal classes (Figure 16.b) is used to visualize the number of cells of the discretized section for the 2D inversion, associated with their corresponding ranges of resistivities. This distribution makes it possible to highlight the three large sets of materials in the model. Thanks to it, the bounds of the events considered can thus be defined, in  $\Omega \cdot m$ , so that the intervals are the same length, as presented Eq. (18):

$$\begin{aligned} \theta_1 &= [10; 35], \\ \theta_2 &= [40; 140], \\ \theta_3 &= [9500; 33250], \\ \theta_4 &= [2.85; 10 \cup [35; 40 \cup [140; 9500 \cup [33250; 116375]. \end{aligned} \quad (18)$$

In contrast to information from the geophysical source, geotechnical data were obtained beforehand by laboratory penetration cone testing, and then numerically simulated prior to fusion. Several simulations proposing various positions of survey points were carried out. In order to simulate drilling points, the associated mean depth values (mm) and associated standard deviations (Table I) were used to draw values, following a normal distribution defined for each event. An average penetration depth value of 0 mm is used for  $\theta_3$  (bottom of tank) and an associated standard deviation of 0.01 mm, meaning that negative values may be drawn. These random draws, make it possible to define the limits, in mm, of the intervals associated with the events of the FoD as presented Eq. (19):

$$\begin{aligned} \theta_1 &= [0.04; 0.19], \\ \theta_2 &= [13; 21], \\ \theta_3 &= [-0.02; 0.02], \\ \theta_4 &= [-0.05; -0.02 \cup [0.02; 0.04 \cup [0.19; 13 \cup [21; 100]. \end{aligned} \quad (19)$$

Thus, 2 mm wide boreholes are simulated, down to 15 cm and acquiring every 5 mm with an associated inaccuracy of 0.01 mm. The values of penetration depth obtained can then be associated with the different materials of the model.

## V. TEST BENCH DATA FUSION RESULTS

The results of the geophysical and geotechnical information fusion, are proposed in Figure 17. The simulations were carried out according to four distinct vertical drill positioning configurations, represented in dashed lines in the figures and at regular intervals: i) 8 holes inter-spaced of 10 cm (Figure 17.i) ( $x = 10; 20; 30; 40; 50; 60; 70; 80$  cm), ii) 5 holes inter-spaced of 18 cm (Figure 17.ii) ( $x = 4, 22, 40, 58, 76$  cm), iii) 3 holes inter-spaced of 25 cm (Figure 17.iii) ( $x = 15, 40, 65$  cm), iv) 2 holes inter-spaced of 50 cm (Figure 17.iv) ( $x = 15, 65$  cm). The fusion results carried out are presented, respecting i) the hypothesis of Smets (Figures 17.a and 17.b), ii) the hypothesis of a closed-world (section II.1) with PCR6 rule (Figures 17.c and 17.d). Figures 17.b and 17.d represent the belief mass values associated with events having the largest mass, represented respectively in Figures 17.a and 17.c. The fusion results are analyzed and discussed in the next section.

## VI. FUSION RESULTS ANALYSIS AND DISCUSSION

### Different rules of combinations

Let us discuss and compare the results obtained by the 2 different combination rules used in an 8-boreholes simulation (Figure 17.i). In the framework of a model as rich in geotechnical information, the section proposed by the PCR6 method (Figure 17.i.c) is very close to the target model set up (Figure 15). The three sets are well characterized and the interfaces at 2.5 cm deep (sands-plaster) and at 7.5 cm deep (plaster-PVC tank and sand-PVC tank) are much better defined than by ERT alone (Figure 16.a). Moreover, thanks to this geotechnical information, the sand thickness anomaly could be correctly characterized as saturated sands (2) and not as a more resistive anomaly, in continuity with the insulating material from below, as suggested by the results of the inversion. The lateral extension of this anomaly is, moreover, well estimated (20 cm). The combination of Smets highlights the significant conflict existing between the two considered sources of information (Figure 17.i.a) concerning the two first layers.

Whatever method is used, the presence of a hypothesis 4 is found at the vertical and horizontal interfaces (Figures 17.i.a, i.c). This hypothesis does not correspond to any material set up in the test bench. The belief masses attributed to such a hypothesis, highlight the transition zones not conform to reality, proposed by the inversion of the electrical resistivity data (Figure 16.a). In comparison to the belief masses associated with the other hypotheses of the model, the belief masses associated with 4 are the lowest (Figures 17.i.b, i.d), showing that the confidence granted to such a material remains quite relative. An overall confidence drop is also observed from 15 cm depth. This corresponds to the maximum depth reached by the simulated boreholes. As confidence is extended laterally, the belief masses are constrained only by geophysical information to such a depth and therefore rely only on one source of information.

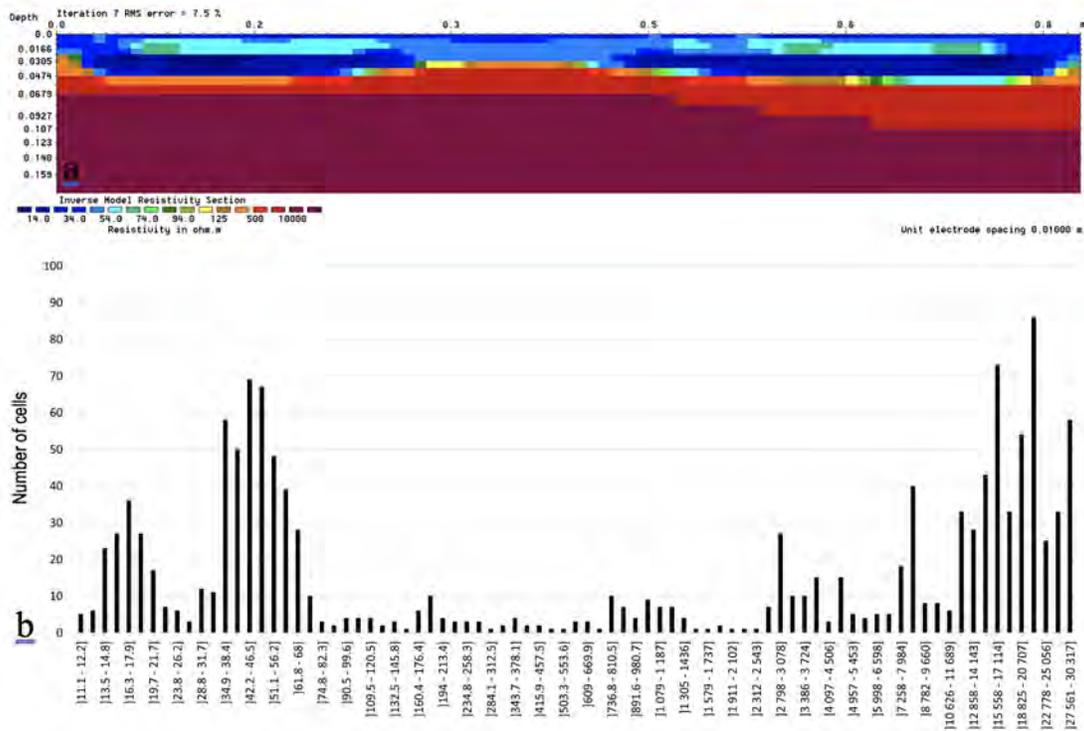


Fig. 16. a) inverse model resistivity section obtained by roll along acquisitions in the central line of the model and b) model classes' distribution of the cells presented in a), according to the electrical resistivity values ( $\Omega\cdot m$ ).

### Influence of the number of boreholes and positions

The first intuition would be to assume that the more the number of boreholes decreases, the more the method should be put in difficulty to properly characterize the section of the set up test bench. Although this is partly true, the quality of the results is not based as much on the number as on the positions of the drillings. Indeed, the anomaly of saturated sands contained between the two banks of plaster (Figure 15) is as well characterized in terms of lateral extension with three or five soundings (Figures 17.ii.c, iii.c). It also has an equivalent associated trust (Figures 17.ii.d, iii.d). It turns out that the belief masses associated with the event 1 (plaster) are even smaller for a fusion including three soundings (Figures 17.iii.d) than for a simulation of only two (Figures 17.iv.d).

The explanation of such results lies in the fact that being in the presence of consecutive boreholes, informing about the occurrence of different materials, at an equivalent depth, induces a rapid decrease in the confidence attributed to the boreholes. Therefore, more credibility is given to the geophysical information source, explaining the greater presence of 4, which reflects the gradual transitions in electrical resistivities. The masses associated with this event, however, remain relatively small. On the other hand, if two consecutive boreholes have the same geotechnical values, for a specific depth, the lateral decay rate will be low and no priority can be given to a different material existing between these two boreholes. That is why the sand anomaly in the center of the model does not appear in the results fusion with two soundings (Figures 17.iv.c) : no

borehole pass through the anomaly and the geophysical source is unable to characterize this material as saturated sand. The strength of these results is that they suggest the presence of 4 in this location, suggesting that the survey campaign should be reinforced (with a new borehole position for example).

The conflict presented by Smets combination (results in Figures 17.i.a, ii.a, iii.a and iv.a) is neither a function of the number of geotechnical soundings. In this study, the cases of fusion bringing the highest amount of conflict are in fact the ones with eight and two soundings. Nor is it to be confused with a lack of knowledge of the subsoil. Conflict zones highlight contradictory information between the two sources. These zones are generally between two consecutive boreholes providing the same information, but going against the available geophysical information. These are therefore potentially anomalous zones where the geophysical information must be considered carefully, in particular if the belief mass associated with the event retained after normalization is too low.

### Important considerations and potential in the application

It is important to consider that the effectiveness of this fusion methodology has been assessed by comparing the fusion results with a target model (Figure 15). However, this remains an idealized representation of the test bench set up and could be, in some places, quite far from reality (real interfaces not perfectly horizontal or vertical, materials not perfectly homogeneous, 3D effects neglected, ...). The approach is different from the one of the synthetic study (Figure 4) where the

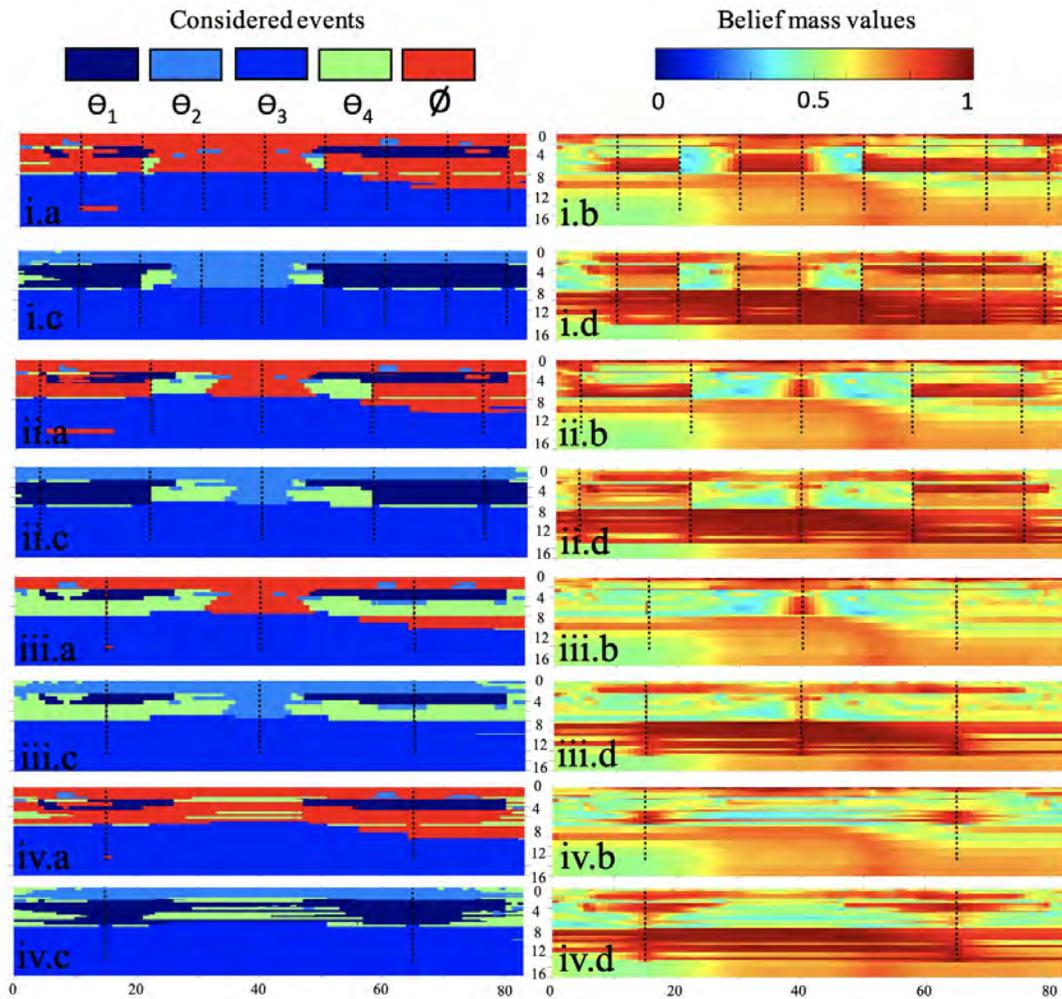


Fig. 17. Representation of the events having the highest belief mass in a) and c) and their associated mass values in b) and d). For i) 8 boreholes, ii) 5 boreholes, iii) 3 boreholes, iv) 2 boreholes are considered. For each case, (a,b) figures are results of Smets fusion, while (c,d) figures are results of PCR6 fusion. The borehole positions are in dashed lines.

model shown corresponds to the true model. In order to control the effectiveness of the fusion methodology, it was envisaged to carry out ex post verifications of the constituent materials. Unfortunately, for practical reasons, this could not be done (reworking of materials modifying their physical properties, interaction with water, delicate cutting and extraction, ...).

Regarding the fusion methodology developed, two aspects are debatable. First, the choice to set a mass of belief equal to 1 on the geotechnical information in the boreholes. Second, the effect of different random draw results on the fusion results. The choice of a maximum punctual confidence ( $m = 1$ ) in boreholes is defended in order to give a full and local confidence to geotechnical information as it is currently done during investigation campaigns. Excessive risks are not taken since the test bench is relatively well known and the synthetic model is perfectly well known. Thus, it is sure that simulated borehole values refer to the right materials. Furthermore, a value of  $m = 0.99$  instead of  $m = 1$ , for instance, does not significantly change the results and does

not change the interpretation and the resulting discussion. Regarding the effect of random draws, these draws were done following a normal distribution, the variations from one draw to another are minimal and the results of fusion differ little.

Such an information fusion algorithm, dedicated to the combination of data from geophysical and geotechnical sources, should prove useful for processing of data acquired during investigation campaigns for many different kinds of issues. It is possible to envisage its use with a larger number of materials, but also, and especially, with a larger number of data types from geophysical methods (seismics, ground penetrating radar) and geotechnical testing methods (penetration cone, core sampling with laboratory identification, permeability tests, ...) associated.

In the framework of a recognition campaign, the conflict zones, or zones with a low associated confidence, would make it possible to specify the locations where the investigation must be reinforced. The ultimate goal is to obtain a more robust and cost-effective diagnosis of the investigated structure, more

targeted for geotechnical investigation. This methodology has particularly shown its ability to correctly characterize interfaces, which corresponds to areas where the risk of instability is potentially the greatest. For a levee embankment issue, for example, the results from such a methodology could come to feed into models of breakage risks (ex: CARDigues [35]).

## VII. CONCLUSION

In this work, a new methodology has been presented, based on belief functions to take benefit and to combine two different and complementary kinds of information: geophysical and geotechnical. Each one having its own spatial distribution and related uncertainties and inaccuracies. A new representation of the information has been proposed, taking into consideration two different investigation methods, associated with degrees of belief. This representation is more informative than data superposition of different physical parameters.

In the first place, this new approach has been validated with a synthetic study, simulating data acquired by ERT and a CPT method, considering a 2D model with two layers and thickness variation. The results were obtained with different noise ratios applied to the geophysical data and different values of lateral decay coefficient for the geotechnical information. The most appropriate value to pick up for the coefficient has been pointed out and it has been showed that this approach was able to manage the noise ratio, thanks to the use of Wasserstein distances.

In order to address the problem of combining information acquired by geophysical and geotechnical methods during investigation campaigns, and to acquire values from real devices, a test bench composed of plaster and saturated sands was set up. The methods used to characterize such a physical model were the ERT method (geophysical) and the laboratory penetration cone method (geotechnical). While the data has been acquired by a dedicated small scale ERT device, on the surface and on the central line of the complete model, borehole were simulated respecting the penetration depth ranges previously established.

Fusion results were proposed following 2 combination rules (Smets and PCR6) as well as for four different simulations of number and positions of boreholes. The results highlighted the ability of this fusion approach to correctly characterize the test bench materials as well as to specify the positions of the interfaces (vertical and horizontal) between the materials. Moreover, for each result, thanks to a graphical representation, the associated confidence is proposed.

Further research should include cases of material mixtures and cases of different materials sharing common ranges of physical properties in order to test the ability of this methodology to differentiate them. We also wish to test this new methodology in real investigation campaigns in order to improve the available knowledge and strengthen the characterization. The level of confidence associated with the proposed results may be very relevant for decision support (eg models of failure hazards). The results of such a methodology should make it possible to propose the most relevant borehole

positions (that are a function of conflictual and anomalous areas), in order to make the quality of the information more cost-effective.

## ACKNOWLEDGMENTS

We would like to thank Quiterie Forquenot de la Fortelle and Gautier Gugole for their precious help throughout the experimental phase and also would like to thank the *Région Pays de la Loire* for their financial support.

## REFERENCES

- [1] J.P.T. Caris, T.W. Van Asch, *Geophysical, geotechnical and hydrological investigations of a small landslide in the French Alps*, Engineering Geology, Vol. 31(3-4), pp. 249–276, 1991.
- [2] A.J. Merritt, J.E. Chambers, W. Murphy, P.B. Wilkinson, L.J. West, D.A. Gunn, P.I. Meldrum, M. Kirkham, N. Dixon, *3D ground model development for an active landslide in Lias mudrocks using geophysical, remote sensing and geotechnical methods*, Landslides, Vol. 11(4), pp. 537–550, 2014.
- [3] J. Chambers, P. Meldrum, D. Gunn, P. Wilkinson, A. Merritt, W. Murphy, J. West, O. Kuras, E. Haslam, P. Hobbs, C. Pennington, C. Munro, *Geophysical-geotechnical sensor networks for landslide monitoring*, Landslide Science and Practice, pp. 289–294. Springer, Berlin, Heidelberg, 2013.
- [4] Z. Abidin, M. Hazreek, R. Saad, A. Fauziah, C. Wijeyesekera, T. Baharuddin, M. Faizal, *Integral analysis of geoelectrical (resistivity) and geotechnical (SPT) data in slope stability assessment*, Academic Journal of Science, Vol. 1(2), pp. 305–316, 2012.
- [5] N. James, T.G. Sitharam, G. Padmanabhan, C.S. Pillai, *Seismic microzonation of a nuclear power plant site with detailed geotechnical, geophysical and site effect studies*, Natural hazards, Vol. 71(1), pp. 419–462, 2014.
- [6] J.O. Coker, *Integration of Geophysical and Geotechnical Methods to Site Characterization for Construction Work at the School of Management Area, Lagos State Polytechnic, Ikorodu, Lagos, Nigeria*, International Journal of Energy Science and Engineering, Vol. 1(2), pp. 40–48, 2015.
- [7] M.T. Perri, J. Boaga, S. Bersan, G. Cassiani, S. Cola, R. Deiana, P. Simonini, S. Patti, *River embankment characterization: The joint use of geophysical and geotechnical techniques*, Journal of Applied Geophysics, Vol. 110, pp. 5–22, 2014.
- [8] P. Royet, S. Palma Lopes, C. Fauchard, P. Mériaux, L. Auriau, *Rapid and cost-effective dike condition assessment methods: geophysics*, Geography, Vol. 32(4), pp. 1–17, 2013.
- [9] F. Shaaban, A. Ismail, U. Massoud, H. Mesbah, A. Lethy, A.M. Abbas, *Geotechnical assessment of ground conditions around a tilted building in Cairo–Egypt using geophysical approaches*, Journal of the Association of Arab Universities for Basic and Applied Sciences, Vol. 13(1), pp. 63–72, 2013.
- [10] M. Foster, R. Fell, M. Spannagle, *The statistics of embankment dam failures and accidents*, Canadian Geotechnical Journal, Vol. 37(5), pp. 1000–1024, 2000.
- [11] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [12] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The annals of mathematical statistics, pp. 325–339, 1967.
- [13] E. Binaghi, L. Luzi, P. Madella, F. Pergalani, A. Rampini, *Slope instability zonation: a comparison between certainty factor and fuzzy Dempster-Shafer approaches*, Natural hazards, Vol. 17(1), pp. 77–97, 1998.
- [14] O.F. Althuwaynee, B. Pradhan, S. Lee, *Application of an evidential belief function model in landslide susceptibility mapping*, Computers & Geosciences, Vol. 44, pp. 120–135, 2012.
- [15] M.H. Tangestani, F. Moore, *The use of Dempster-Shafer model and GIS in integration of geoscientific data for porphyry copper potential mapping, north of Shahr-e-Babak, Iran*, International Journal of Applied Earth Observation and Geoinformation, Vol. 4(1), pp. 65–74, 2002.
- [16] K.A. Mogaji, H.S. Lim, K. Abdullah, *Regional prediction of groundwater potential mapping in a multifaceted geology terrain using GIS-based Dempster-Shafer model*, Arabian Journal of Geosciences, Vol.8(5), pp. 3235–3258, 2015.

- [17] M.S. Tehrani, L. Kumar, *The application of a Dempster-Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods*, Environmental Earth Sciences, Vol. 77(13), 490, 2018.
- [18] A. Martin, C. Osswald, J. Dezert, F. Smarandache, *General combination rules for qualitative and quantitative beliefs*, Journal of Advances in Information Fusion, Vol. 3(2), pp. 67–89, Dec. 2008.
- [19] R. Cooke, *Experts in uncertainty: opinion and subjective probability in science*, Oxford University Press on Demand, 1991.
- [20] P. Smets, *The combination of evidence in the transferable belief model*, IEEE Transactions on pattern analysis and machine intelligence, Vol. 12(5), pp. 447–458, 1990.
- [21] J. Dezert, P. Wang, A. Tchamova, *On The Validity of Dempster-Shafer Theory*, in Proc. of 15th Int. Conf on Information Fusion (FUSION 2012), pp. 655–660, Singapore, July 2012.
- [22] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion - Collected works*, American Research Press, Volume 3, 2009.
- [23] F. Smarandache, J. Dezert, *Information fusion based on new proportional conflict redistribution rules*, in Proc. of 7th Int. Conf. on Information Fusion (Fusion 2005, Philadelphia, PA, USA, July 25–29, 2005).
- [24] P. Kearey, M. Brooks, I. Hill, *An introduction to geophysical exploration*, John Wiley & Sons, 2013
- [25] L. Tran, L. Duckstein, *Comparison of fuzzy numbers using a fuzzy distance measure*, Fuzzy Sets and Systems, Vol. 130(3), pp. 331–341, 2002.
- [26] K.K. Phoon, F.H. Kulhawy, *Characterization of geotechnical variability*, Canadian Geotechnical Journal, Vol. 36(4), pp. 612–624, 1999.
- [27] M.H. Loke, *Tutorial 2-D and 3-D electrical imaging surveys*, Technical report, 2011.
- [28] M.H. Loke, *RES2DMOD ver. 3.01: Rapid 2D resistivity forward modelling using the finite difference and finite-element methods*, Software manual, 2002.
- [29] M.H. Loke, I. Acworth, T. Dahlin, *A comparison of smooth and blocky inversion methods in 2D electrical imaging surveys*, Exploration Geophysics, Vol. 34(3), pp. 182–187, 2003.
- [30] ISO 22476-1:2012(en) Geotechnical investigation and testing – Field testing – Part 1: Electrical cone and piezocone penetration test.
- [31] G.J. Palacky, *Resistivity characteristics of geologic targets*, Methods in Applied Geophysics Theory, Vol. 1, pp. 53–129, 1987.
- [32] P.K. Robertson, R.-G. Campanella, D. Gillespie, J. Grieg, *Use of piezometer cone data*, in Proc. of In-situ '86, AXE Specialty Conference, Blacksburg, VA, USA, 1986.
- [33] Normes nationales et documents normatifs nationaux, *Sols : reconnaissance et essais - Détermination des limites d'Atterberg - Partie 1 : limite de liquidité - Méthode du cône de pénétration*, NF P94-052-1, Nov. 1995.
- [34] E. Flavigny, J. Desrues, B. Palayer, *Note technique: le sable d'Hostun «RF»*, Revue française de géotechnique, Vol. 53, pp. 67–70, 1990.
- [35] H. Apel, A.H. Thieken, B. Merz, G. Blöschl, *Flood risk assessment and associated uncertainty*, Natural Hazards and Earth System Science, Vol. 4(2), pp. 295–308, 2004.

# Dual Iris Authentication System Using Dezert-Smarandache Theory

Kamel Ghanem Ghalem<sup>a</sup>, Fatiha Hendel<sup>b</sup>

<sup>a</sup>Ecole Supérieure En Génie Electrique et Energétique, BP 64 CH2 ACHABA HANIFI USTO 31000 Oran, Algérie.

<sup>b</sup>Univ. des Sciences et de la Technologie d'Oran Mohamed Boudiaf, USTO-MB, BP 1505, EL M'Naouer, 31000 Oran, Algérie.

Emails: ghalem\_kamel\_ghanem@esgee-oran.dz, fa\_hendel@yahoo.fr

Originally published as: K.G. Ghalem, F. Hendel, *Dual Iris Authentication System Using Dezert-Smarandache Theory*, International Journal of Electrical and Computer Engineering (IJECE) Vol. 9, No. 6, pp. 4703–4712, December 2019, and reprinted with permission.

**Abstract**—In this paper, a dual iris authentication using Dezert-Smarandache theory is presented. The proposed method consists of three main steps: In the first one, the iris images are segmented in order to extract only half iris disc that contains relevant information and is less affected by noise. For that, a Hough transform is used. The segmented images are normalized by Daugman rubber sheet model. In the second step, the normalized images are analyzed by a bench of two 1D Log-Gabor filters to extract the texture characteristics. The encoding is realized with a phase of quantization developed by J. Daugman to generate the binary iris template. For the authentication and the similarity measurement between both binary irises templates, the hamming distances are used with a previously calculated threshold. The score fusion is applied using DSMT combination rule. The proposed method has been tested on a subset of iris database CASIA-IrisV3-Interval. The obtained results give a satisfactory performance with accuracy of 99.96%, FAR of 0%, FRR of 3.89%, EER of 2% and processing time for one iris image of 12.36 s.

**Keywords:** Biometric, Iris, Authentication, Dezert-Smarandache theory.

## I. INTRODUCTION

When individuals log onto computers, or access an ATM, or pass through airport security, they have to reveal their identities. For this, individuals use passwords, ATM cards, and passports to prove their identities. However, passwords can be forgotten, and ATM cards or passports can be lost or stolen. In contrary, the biometric modalities (Fingerprint, face, iris, etc) speak to what and they also allow to prove our identity.

However, the unimodal biometric systems using one biometric modality for recognition cannot guarantee at present an excellent recognition rate. Furthermore, these systems suffer from limitations such as sensitivity to noise, data quality, non-universality, and spoof attacks. To overcome these problems, Multimodal biometric systems, which combine multiple biometric modalities, have been developed on purpose to achieve a better recognition rate.

The popular fusion method of the biometric traits can be done at two stages of recognition system:

### A. Fusion at feature extraction level

The data is acquired from each sensor is utilized to generate a feature vector. Then, the features are fused to form one feature vector.

### B. Fusion at matching score level

The matching score of each system is combined and compared with the stored template.

We use as a modality for recognition of individuals: iris, since their texture is

- Stable throughout the life of a person, unlike the fingerprint.
- Unique for each person, unlike a facial feature in identical twins.
- Unfalsifiable contrary to the characteristics of the voice.
- Iris is an internal organ well protected from the external environment, but nevertheless measurable, in a rather little invasive way, by simple image acquisition.

Daugman's algorithm [1] is one of the best iris algorithm known in biometrics. The algorithm consists of segment iris using Integro-Differential Operator and iris normalization is implemented using Daugman's polar representation. Then, iris encoding is applied using 2D Gabor filters to extract a binary code of 256 bytes. The Matching is processed by computing similarity between two iris codes using Hamming distance. The more Hamming distance is small, the more both codes are similar. A distance of 0 corresponds to a perfect match between both iris images, while two iris images of different person will have a Hamming distance close to 0.50. In 1997, Wildes [2] proposed a novel iris recognition system compared to Daugman algorithm [1]. The acquisition of iris is done by a CCD Camera in low luminosity. Then, the iris is segmented using Circular and Elliptic Hough transform and is normalized using a transformation function of pixels. After that, the iris is filtered by Laplacian of Gaussian filters with four different resolution levels. A normalized correlation is calculated for every resolution levels. The median of the values of correlations is computed for the filtered image. The fusion of four values is applied using Fisher's linear discriminant.

In 1998, W. Boles and B. Boashash [3], presented a new algorithm for recognition of individuals from iris images. The algorithm is insensitive to variation in the lighting conditions and noise levels. A Median filter is used for preprocessing. The advantage of this technique is to extract a features vector from 1D signals rather than 2D images analyzed in [1], [2] using zero-crossings of the dyadic wavelet transform at various resolution levels. Only a few selected intermediate resolutions

are used for matching. The matching is applied using different dissimilarity functions. Thus, make the algorithm faster and less sensitive to noise and quantification error.

In 2004, Ma et al. [4] presented an efficient algorithm for iris recognition. The iris is segmented by Canny filter and Hough transform. Then, the iris is normalized by histogram equalization. After that, A 1D Wavelet Transform is used to represent resulting 1D intensity signals. The position of local sharp variation points is registered as features. The matching is effectuated using the similarity function (exclusive or operation). This algorithm is efficient and faster than Daugman's algorithm [1].

In [5], the researchers proposed a modified Masek approach and a comparative study of the performance of the following methods: radial segmentation, Masek segmentation approach, modified Masek approach. The proposed method tested on Casia Iris Database V3 showed a good performance in terms of accuracy and processing time.

R. Biswas [6] has introduced an iris recognition system that includes different steps: segmentation, normalization, feature extraction, and classification. The segmentation of the pupil is performed by the Hough transform. Experimental results showed a recognition rate of 92%.

In [7], the authors proposed an iris recognition system based on "Fractal dimension of the box-counting method". First, the iris is segmented by Hough transform and is normalized by Daugman's rubber sheet model. Then, the feature extraction is processed by box counting. Finally, the matching is established using K-nearest Neighbor and Euclidean distance. Experiments tested on Casia Interval V4 database showed a good recognition rate equal to 92.63%.

D. Bobeldyk and A. Ross [8] have developed a method for predicting eye color from NIR iris images. Researchers have shown that a texture based approach based on the BSIF is more efficient than the intensity based approach based on raw pixel values. Experiments tested on the BioCOP database showed a good recognition rate of 90%. The BSIF distinguished "light color iris" and "dark color irides" using the SVM classifier.

The authors [9] presented a new method of classifying faked iris images of different patterns such as printed irises, contact lenses. The new classification method learns different characteristics of faked iris images by CNN and identifies legitimate and faked iris images using "Hierarchical Multi-Class". The tests carried out on the different databases: ND-contact, Casia-Iris-Interval and Casia-Iris-Syn, LivDet-Iris-2017-Warsaw showed a recognition rate equal to 100% and FAR = 0%, FRR = 0%.

H.G. Daway et al. [10] presented a new method for detecting the pupil. The method involves several steps, the most important of which depends on the difference in color and intensity between the pupil and its neighborhood. These characteristics are very important to locate and extract the pupil. Thus, the pupil is a region of very high intensity (color) compared to its neighborhood. The experimental results showed a recognition rate equal to 100%.

In purpose to improve overall performance in terms of recognition rate and mitigate errors, the researchers have used more than one biometric trait, and thus, the multibiometric systems have emerged. Numerous multi-biometric systems have been developed, which fusion is made at Matching score.

In [11], the researchers proposed a new approach for recognition using both irises. The iris is segmented using the Canny filter and Hough transform, then the segmented iris is normalized by J. Daugman's rubber sheet model. The iris feature extraction is carried out using convolution of the normalized iris with 1D Log-Gabor filters then the phase of filtered iris is quantized in order to generate a binary code. The Hamming distance is used for Matching. The Matching operation consists in comparing the two iris feature vectors of a person with the others; if the Hamming Distances are less than the threshold then the person is identified. The experimental results showed a good recognition rate equal to 99.92% with an FRR = 9.96%, while for unimodal systems (left iris and right iris) the recognition rate is equal to 99.87% with an FRR = 14.62% and FPR = 15.68%.

In [12], the authors presented the framework for multi-modal biometric fusion based on the uncertainty concept of Dempster-Shafer theory. A combination of quality measures and the accuracy of classifiers (equal error rate) are proposed to encode the uncertainty concept to improve the fusion. The proposed method revealed a good performance with an EER equal to 1%.

R. Dwivedi and S. Dey [13] proposed a cancelable multi-biometric system using score level fusion. The fusion of scores was applied by MC weighting at first level and RA weighting at the second level. The comparative analysis shows that the proposed fusion method outperforms the existing weighting approaches.

In this works, we extract only the interior half of the iris disc rather than the whole iris disc, which contains the most relevant information and it is less affected by noise. In addition, we combined two sources of information (left iris and right iris) with a high degree of conflict using Dezert-Smarandache Theory, which solves the problem of highly re-distributed masses conflicts arising under the Dempster Shafer theory.

The reminder paper is organized as follows: the research method is described in section II, Results and Analysis are presented in section III, conclusions are provided in section IV.

## II. RESEARCH METHOD

The key idea of our work is to extract only the discriminant information from iris texture and proposes Dezert-Smarandache Theory (DSMT) at score level fusion to operate under uncertainty in goal to achieve a good performance.

The proposed method is composed of four main stages: preprocessing, feature extraction, fusion, and matching.

### A. Preprocessing stage

First, the iris images require going through the preprocessing phase including segmentation and normalization.

1) *Iris segmentation*: The segmentation of iris is realized by commonly Edge detector method: Hough transform.

- HOUGH TRANSFORM ALGORITHM
- Generate edge map using the Canny filter.
- Canny parameters: the standard deviation of Gaussian smoothing filter:  $\sigma = 2$ ; weighting for vertical gradients = 0; weighting for horizontal gradients=1.
- Increase contrast in dark iris region. image gamma value: enhance the contrast of bright regions:  $\gamma = 1.9$ .
- Detect pixel corresponding to the local maxima Distance in pixel units to be looked at on each side of each pixel when determining, whether it is a local maximum or not:  $d = 1.5$ .
- Binarize iris image using Hysteresis thresholding. Low threshold  $T_1 = 0.19$ . High threshold  $T_2 = 0.20$ .

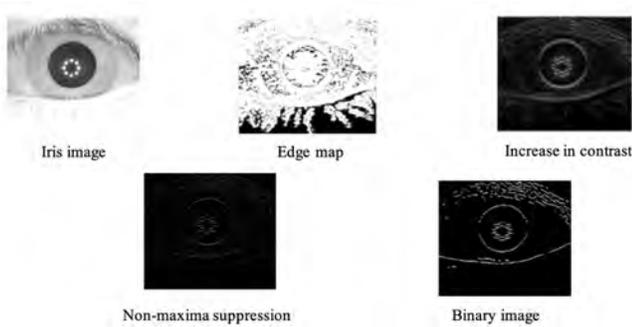


Figure 1. Different step of Hough Transform.

Then, a Circular Hough Transform detects at first the iris/sclera boundary and the iris/pupil boundary. The eyelashes are detected by global thresholding ( $T = 100$ ).

In this work, the objective is to extract only relevant information from iris, which is represented by the structural variation of the iris texture (high gradient areas), only the internal half of the iris disc is exploited rather than whole, because it contains the most relevant information [14] and it is less affected by the noise as shown in Figure 2. Indeed, the proposed technique decreases the complexity and the computation load without losing information (as shown in Table I).

$$r_{\text{half of iris disc}} = r_{\text{pupil}} + (r_{\text{iris}} - r_{\text{pupil}})/2 \quad (1)$$

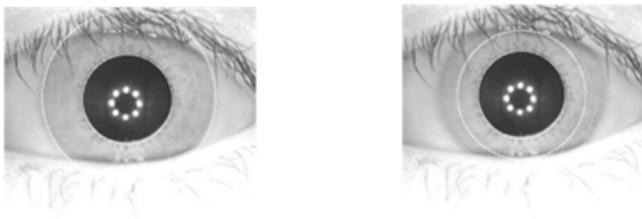


Figure 2. Delimitation of only the internal half of iris disc.

Table I  
COMPARISON: TOTAL IRIS DISC VS HALF IRIS DISC.

	Accuracy (%)	Processing time for one iris image (s)
Total iris disc	99.87	22.88
Half iris disc	99.96	12.36

From Table I, we denote that treatment using only half iris disc is more efficient with an accuracy of 99.96% and processing time for one iris image of 12.36 s than the treatment using a whole iris disc with accuracy 99.87% of and processing time of 22.88 s.

2) *Iris normalization*: The iris disc does not always have the same dimension, even for eye images of the same person; this is due to various problems as follows:

- a) Different acquisition conditions of the eye images. Dilation and contraction of the pupil due to the variation of the illumination level.
- b) The circles of iris and pupil are not concentric.

In order to overcome these problems, a stage of normalization is applied. It consists of transforming the region of the iris disc to rectify the dimensions of all the iris discs, by using the homogenous rubber sheet model proposed by Daugman [1]. It transforms each point in the iris area to the polar coordinates  $(r, \theta)$ , where  $r$  is on the interval  $[0, 1]$ , and  $\theta$  is an angle in  $[0, 2\pi]$ , as illustrated in Figure 3.

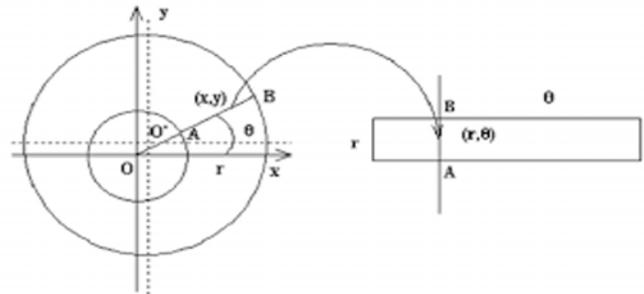


Figure 3. Daugman rubber sheet model [1].

In our system,  $(20 \times 240)$  points were used, but only  $(10 \times 240)$  points corresponding to the internal half of the iris disc that contains the most relevant information and which is less affected by noise, are retained for the next steps of the processing, as shown in Figure 4.



Figure 4. Normalization of the segmented iris.

### B. Feature extraction stage

After that, the feature extraction stage is applied in purpose to extract the most discrimination information present in the iris region. For this reason, a bench of two 1 D Log-Gabor filter is used.

### 1) 1D Log-Gabor filter:

- The Fast Fourier Transform is applied for each line of the normalized matrix image (FFT to 1D signals).
- Then, the Inverse Fast Fourier Transform IFFT is applied on the multiplication FFT (1D signals) by a 1D Log-Gabor Filter.
- The frequency response of a 1D Log-Gabor filter is given by:

$$G(f) = \exp\left(-\frac{\log(f/f_0)^2}{2 \times \log(\sigma/f_0)^2}\right) \quad (2)$$

#### Parameter setting:

- A bench of two 1D Log-Gabor filters is used.
- The standard deviation of the 1D Log-Gabor wavelet is given by  $\sigma = 2$ .
- The center frequency of the 1D Log-Gabor wavelet is given by  $f_0 = 0.05$ .

Indeed, the phase of a filtered image was quantized using four-quadrants of Daugman [1], when going from one quadrant to an adjacent quadrant, one bit is changed as shown in Figure 5.

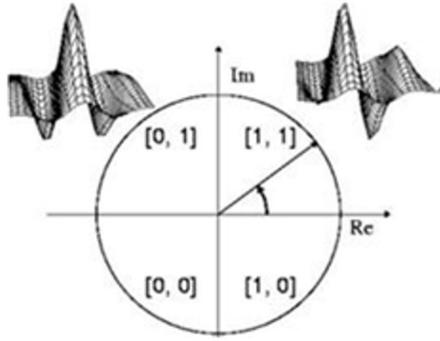


Figure 5. Quantization Phase [1].

The encoding process produces a bitwise template containing a number of information bits (as shown in Figure 6), the total number of bits in the template (9600 bits) will be the angular resolution (240) times the radial resolution (10), times 2, times the number of filters used (2).



Figure 6. Quantization Phase [1].

### C. Matching stage

The matching score comes before Fusion stage. It consists in comparing two iris code using Hamming distance. The Hamming Distance (HD) is defined by:

$$HD = \sum_{j=1}^N X_j \oplus Y_j \quad (3)$$

where  $X_j$  and  $Y_j$  are the two bitwise iris code,  $N$  is the number of bits in each iris code, and  $\oplus$  is xor operation.

Literally, the Hamming distance calculates the number of different and valid bits for the two iris code between  $X_j$  and  $Y_j$ .

The number of translation bits that compensates the rotation of the iris needs to be fixed. We applied a translation of the iris code in an interval  $[-3,+3]$  bits. We take into consideration the minimum Hamming distance.

### D. Fusion stage

In this stage, score level fusion using Dezert-Smarandache theory (DSMT) was applied on a goal to improve the performance of the dual iris system.

1) *Score level fusion:* Matching score level fusion combines the scores generated by multiple classifiers relating to the left and right iris to affirm the veracity of the claimed identity.

The Dezert-Smarandache theory operates under hyperpower set  $D^\Theta$ . Thus, DSMT is able to function properly not only with the unions but also with intersections. DSMT Classic has combination rule [15],[16] and [17]:

$$m(C) = \sum_{A \cap B = C} m_1(A)m_2(B); A, B \in D^\Theta, \forall C \in D^\Theta \quad (4)$$

#### Example

$$\Theta = \{S_{\text{left}}, S_{\text{right}}\},$$

$$\Theta = \{\emptyset, S_{\text{left}}, S_{\text{right}}, S_{\text{left}} \cup S_{\text{right}}, S_{\text{left}} \cap S_{\text{right}}\},$$

where

$\emptyset$ : Empty set;

$S_{\text{left}}$ : Hypothesis assuming that two individuals have same left iris;

$S_{\text{right}}$ : Hypothesis assuming that two individuals have a same right iris;

$S_{\text{left}} \cup S_{\text{right}}$ : Hypothesis assuming that two individuals have the same iris;

$S_{\text{left}} \cap S_{\text{right}}$ : Hypothesis assuming that two individuals have different iris.

### E. Decision

The decision is made by fixing a threshold. The two irises compared will be considered as belonging to the same person if the calculated score is inferior to a threshold.

## III. RESULTS AND ANALYSIS

### A. Simulation environment

The proposed method has been tested on a subset of iris database CASIA-IrisV3-Interval [18] in order to evaluate its performance in authentication mode. The subset contains 1180 eye images of 118 individuals (classes), and each individual has five iris samples for the left eye and five iris samples for the right eye.

**B. Performance metrics**

- False Reject Rate (FRR): also known as Type I error, is the measure of the probability that the biometric security system will incorrectly reject an access attempt by an authorized user;
- False Accept Rate (FAR): also known as Type II error, is the measure of the probability that the biometric security system will incorrectly accept an access attempt by an unauthorized user;
- EER (Equal Error Rate): The EER is the operating point for which the False Reject Rate (FRR) is equal to the False Accept Rate (FPR).

**C. Decidability**

Decidability [1] is the best metric which indeed takes into account the mean and standard deviation of the intra-class and inter-class distributions:

$$d' = \frac{|\mu_s - \mu_d|}{\sqrt{\frac{\sigma_s^2 + \sigma_d^2}{2}}} \tag{5}$$

Decidability  $d'$  is a distance in standard deviations calculated using (7), which is a function of the magnitude of the difference between the mean of the intra-class distribution  $\mu_s$ , and the mean of the inter-class distribution  $\mu_d$ , the standard deviation of the intra-class and inter-class distributions,  $\sigma_s$ , and  $\sigma_d$  respectively.

Table II  
DECIDABILITY TABLE FOR VARIOUS NUMBERS OF BIT-SHIFTS.

Numbers of shifts	$\mu_s$	$\sigma_s$	$\mu_d$	$\sigma_d$	$d'$
0	0.3300	0.0723	0.4914	0.0284	3.4314
1	0.3137	0.0697	0.4860	0.0279	3.8149
2	0.3072	0.0668	0.4812	0.0269	4.0264
3	0.3044	0.0653	0.4772	0.0258	4.0742
4	0.3032	0.0646	0.4738	0.0247	4.0431
5	0.3028	0.0642	0.4709	0.0238	3.9907
6	0.3025	0.0639	0.4684	0.0230	3.9362
7	0.0637	0.0642	0.0223	0.0216	3.8862
8	0.3023	0.0635	0.4645	0.0216	3.8303
9	0.3022	0.0634	0.4629	0.0211	3.7999
<b>10</b>	<b>0.2758</b>	<b>0.0639</b>	<b>0.4643</b>	<b>0.0201</b>	<b>4.3960</b>

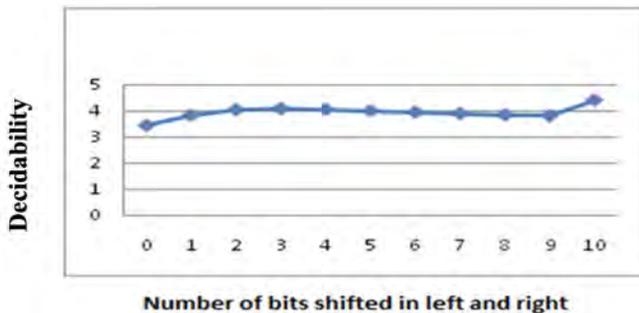


Figure 7. Decidability curve for various numbers of bit-shifts.

Using Equation (7), several different decidabilities are found out using 0-bit shift to 10-bit shift towards both left and right iris templates.

The higher decidability is equal to 4.3960 at 10 bit shift (as shown in Table II and Figure 7) that guarantees good separation of intra-class and inter-class distributions, which allows for more accurate recognition.

**D. Score level fusion**

In fact, we calculated the fusion score  $S_f$  using Hamming distances obtained by comparing the individuals from their iris

- $HD_L$ : Hamming distance obtained by comparing the individuals from their left iris;
- $S_L$ : Score obtained by comparing the individuals from their left iris;
- $HD_R$ : Hamming distance obtained by comparing the individuals from their right iris;
- $S_R$ : Score obtained by comparing the individuals from their right iris.

**Algorithm**

```

for each individual indiv
for each different iris, j :
% such as i, j belongs to iris set of indiv
Calculate the score S_L (i, j)=1-HD_(L ) (i, j)
Calculate the score S_R (i, j)=1-HD_R (i, j)
Calculate the fusion of score
S_f (i, j)=S_L (i, j)*S_R (i, j)
k=1
for s=0:0.05:1
if S_f(i, j) < s then
FN(k)=FN(k)+1 % false negative counter
k=k+1
end if
end
end
for each different individual indivi, indivj
for each different iris (i, j)
% such as i belong to iris set of indivi
% and j belong to iris set of indivj
Calculate the score S_L(i, j)=1-HD_L (i, j)
Calculate the score S_R(i, j)=1-HD_R(i, j)
Calculate the fusion of score
S_f (i, j)=S_L (i, j)*S_R (i, j)
k=1
for s=0:0.05:1
if S_f (i, j) >= s then
FP(k)=FP(k)+1 % false positive counter
k=k+1
end if
end
end
end
maxindv= number of individuals
    
```

```

nbtr=number of iris images per individual
nbinter=maxindv*(nbtrx(nbtr-1)/2)
nbintera=maxindv*(nbtrx(nbtr-1)/2)
TN=nbinter-FP % True Negative
TP=nbintera-FN % True Positive
TPR=100*(TP/nbintera) % True Positive Rate
TNR=100*(TN/nbinter) % True Negative Rate
FAR=100*(FP/nbinter) % False Accept Rate
FRR=100*(FN/nbintera) % False Reject Rate
Accuracy=100*((TP+TN)/(nbintera+nbinter))
    
```

where \* is the product operator of two numbers.

The dual iris system using DSMT at score level fusion reaches an accuracy rate of 99.96% and FAR of 0%, FRR of 3.89%, EER of 2% as shown in Figure 8 and Figure 9.

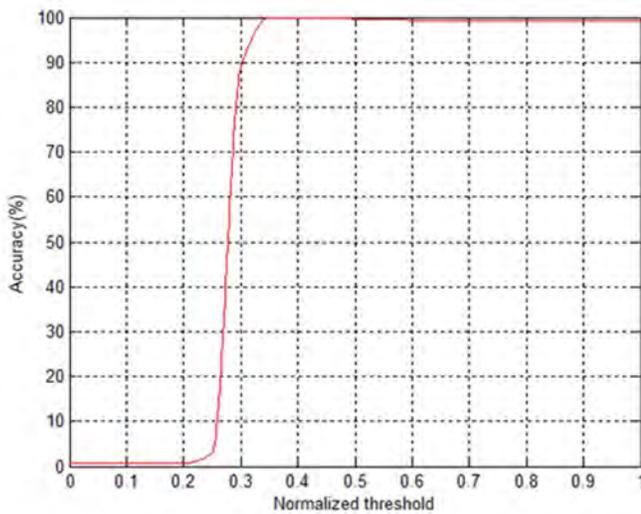


Figure 8. The accuracy of the dual iris system.

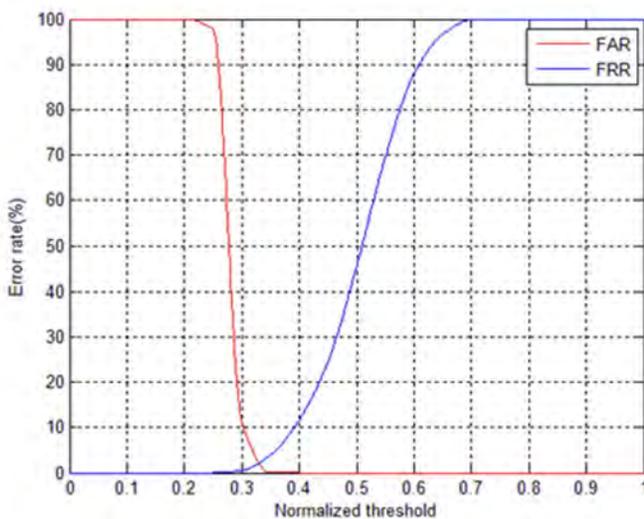


Figure 9. FRR and FAR of the dual iris system.

We conclude from Figure 10 that the ROC (Receiver Operating Characteristic) of dual iris system using DSMT at score level fusion fit the origin, which proves the performance of our method.

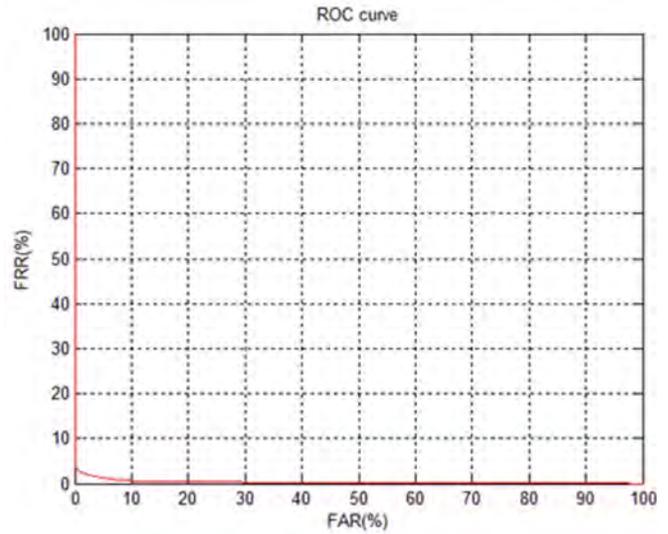


Figure 10. ROC of the dual iris system.

#### E. Comparison of various approaches

We denote from Table III, that the proposed dual iris authentication system gives a competitive performance with an accuracy of 99.96%, FAR of 0%, FRR of 3.89%, EER of 2% in comparison with other approaches. The problem of Dempster Shafer theory used in DST approach [12] (EER of 1%) that consists of combining two sources of information with a high degree of conflict is resolved. The proposed method is based on Dezert-Smarandache Classic rule (DsmC) that solves this problem. Iftakhar and al [11] used the fusion method based on the AND rule that gives an accuracy of 99.92%, FAR of 0%, FRR of 9.96%, which is more drastic and leads to improve the FAR. R. Dwindi and S. Dey obtained less performance in term of accuracy equal to 98.89%, EER of 0.69%, which used a score fusion methods likes MC weighting and RA weighting. These two methods using for optimization gives a little improvement to the performance of the system.

#### IV. CONCLUSION

The purpose of this work was to find out a dual iris authentication system that guarantees good performance and to make sure that there is no false acceptance rate, which promises useful security applications. The proposed method consists in segmenting, to normalizing, characterizing and encoding the iris. For the segmentation part, the detection of the iris/pupil circles was performed by Hough circular transform. Only the interior half of the iris disc containing the most relevant information and less affected by noise, which reduces time complexity was extracted. Iris normalization part was performed by the Daugman rubber sheet model with a

Table III  
DECIDABILITY TABLE FOR VARIOUS NUMBERS OF BIT-SHIFTS.

Iris Reco. System	Accuracy (%)	FAR (%)	FRR (%)	EER (%)
Iftakhar & Ashraful approach [11]	99.92	0	9.96	-
DST [12]	-	-	-	1
Dwivedi & Dey approach [13]	98.89	-	-	0.69
Proposed dual iris authentication system	99.96	0	3.89	2

resolution of  $10 \times 240$ . This stage was analyzed by the bench of two 1D Log-Gabor filters to extract the texture characteristics and the encoding was realized with a phase of quantization developed by J.Daugman to generate the binary iris template. For the authentication and the similarity measurement between both binary irises templates, the hamming distances are used with a previously calculated threshold. The score fusion is applied using Dezert-Smarandache Classic (DSmC) rule. The experiment tested on Casia-iris v3-interval shows that the proposed system gives a good performance compared to others approaches with an accuracy of 99.96%, FAR of 0%, FRR of 3.89%, EER of 2% and processing time for one iris image of 12.37 s.

- [16] F. Smarandache, J. Dezert, *Applications and Advances of DSMT for Information Fusion (Collected works)*, Volume 2, American Research Press (ARP), Rehoboth, 2006.
- [17] F. Smarandache, J. Dezert, *Applications and Advances of DSMT for Information Fusion (Collected works)*, Volume 3, American Research Press (ARP), Rehoboth, 2009.
- [18] Iris database CASIA-IrisV3, Chinese Academy of Sciences, Institute of Automation. Retrieved on Dec 2011.

## REFERENCES

- [1] J. Daugman, *How iris recognition works*, IEEE Trans.Circuits Syst. Video Techn., Vol. 14, pp. 21–30, 2004.
- [2] R.P. Wildes et al., *A machine-vision system for iris recognition*, Machine Vision and Applications, Vol. 9, pp. 1–8, 1996.
- [3] W.W. Boles, B. Boashash, *A human identification technique using images of the iris and wavelet transform*, IEEE Trans. on Signal Processing, Vol. 46, pp. 1185–1188, 1998.
- [4] L. Ma et al. *Efficient iris recognition by characterizing key local variations*, IEEE Trans. on Image Processing, Vol. 13: pp. 739–750, 2004.
- [5] W. Aydi, N. Masmoudi, L. Kamoun, *A Fast and Accurate Circular Segmentation Method for Iris Recognition Systems*, International Review on Computers and Software (IRECOS), Vol. 9 (3), pp. 468–477, 2014.
- [6] R. Biswas, J. Uddin, Md. J.Hasan, *A New Approach of Iris Detection and Recognition*, International Journal of Electrical and Computer Engineering (IJECE), Vol. 7(5), pp. 2530–2536, 2017.
- [7] C. Khotimah, D. Juniati, *Iris Recognition Using Feature Extraction of Box Counting Fractal Dimension*, J. Physics: Conf.Series, 2018.
- [8] D. Bobeldyk, A. Ross, *Predicting Eye Color from Near Infrared Iris Images*, International Conference on Biometrics, pp. 104–110, 2018.
- [9] Z. Yan, L. He, M. Zhang, Z. Sun, T. Tan, *Hierarcical Multiclass Iris Classification for Liveness detection*, International Conference on Biometrics, pp. 47-53, 2018.
- [10] H.G. Daway, H.H. Kareem, A.R. Hashim, *Pupil Detection Based on Color Difference and Circular Hough Transform*, International Journal of Electrical and Computer Engineering (IJECE), Vol. 8(5), pp. 3278–3284, 2018.
- [11] K.M. Iftakhar Hasan, M. Ashraful Amin, *Dual iris matching for biometric identification*, Signal Image and Video Processing, Vol. 8(8), pp. 1605–1611, 2014.
- [12] K. Nguyen, S. Denman, *Score-Level Multibiometric Fusion Based on Dempster-Shafer Theory Incorporating Uncertainty Factors*, IEEE Trans. on Human-Machine Systems, Vol. 45, pp. 132–140, 2015.
- [13] R. Dwivedi, S. Dey, *Score level fusion for concealable multi-biometric verification*, Pattern Recognition Letters, 2018.
- [14] K.G. Ghalem, F. Hendel, *Authentication and identification of individuals from the iris images*, Second International Workshop on Mathematics and Computer Science (IWMCS), 2014.
- [15] F. Smarandache, J. Dezert, *Applications and Advances of DSMT for Information Fusion (Collected works)*, Volume 1, American Research Press (ARP), Rehoboth, 2004.



# Method of Classification of Global Machine Conditions Based on Spectral Features of Infrared Images and Classifiers Fusion

Marek Fidali, Wojciech Jamrozik

Faculty of Mechanical Engineering, Silesian University of Technology,

Department of Fundamentals of Machinery Design, Gliwice, Poland.

Emails: marek.fidali@polsl.pl, wojciech.jamrozik@polsl.pl

Originally published as: M. Fidali, W. Jamrozik, *Method of Classification of Global Machine Conditions Based on Spectral Features of Infrared Images and Classifiers Fusion*, Quantitative InfraRed Thermography Journal, Vol. 16(1), pp. 129–145, 2019, and reprinted with permission.

**Abstract**—This paper describes an original method of global machine condition assessment for infrared condition monitoring and diagnostics systems. This method integrates two approaches: the first is processing and analysis of infrared images in frequency domain by the use of 2D Fourier transform and a set of  $F$ -image features, the second uses fusion of classification results obtained independently for  $F$ -image features. To find the best condition assessment solution the two different types of classifiers,  $k$ -nearest neighbours and support vector machine (SVM), as well as data fusion method based on Desert Smarandache theory have been investigated. This method has been verified using infrared images recorded during experiments performed on laboratory model of rotating machinery. The results obtained during the research confirm that the method could be successfully used for identification of operational conditions that are difficult to be recognized.

**Keywords:** classification, decision fusion, PCR6, infrared image analysis, Fourier Analysis, infrared thermography, condition base monitoring.

## I. INTRODUCTION

Infrared thermography is a modern and popular technique for thermal condition monitoring of machinery, apparatus and industrial processes [1].

Infrared cameras can be used in continuous condition monitoring systems for contactless detection and identification of object faults at its early stage, which is useful for planing object maintenance and overhauls.

Continuous condition monitoring system based on infrared device should include infrared image processing and recognition to classify the current operation condition of the object. Research connected with the application and development of infrared image processing and analysis, as well as artificial intelligence methods, to continuous thermographic objects monitoring and diagnostics has been carried out in several different academic and research centres [2], [3] and also by the authors [4]. In this article an original method of object condition identification, which can be used in continuous condition monitoring and diagnostics systems, has been proposed.

The method can be generalised to any diagnostic data acquired during continuous monitoring of different objects or industrial processes.

## II. METHOD

It has been assumed that the assessment of the general condition of an object could be determined on the basis of the analysis of infrared images that are acquired continuously by monitoring system during an object operation.

For a clear description of the method, let us assume that diagnosed object is a complex machinery containing several sub-assemblies (e.g. motor, couplings, journal bearing, pump, etc.).

Having acquired an infrared image of machinery in any moment of its operation, it is possible to define regions of interests (ROIs) containing only important parts of the diagnosed object. In such a way, the rest of the image content could be treated as an unwanted background that is not considered during the diagnostic process.

In the proposed method, whose brief algorithm is presented in (Fig. 1), each defined region of interest contained a sub-

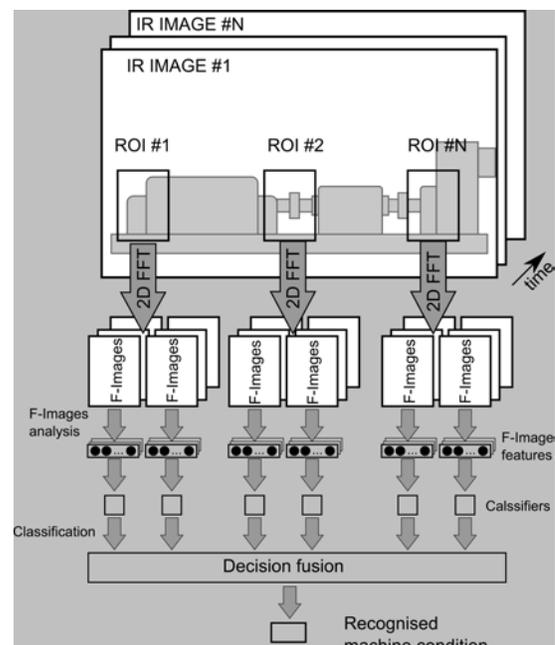


Figure 1: Idea of the method of identification of object conditions based on infrared images.

assembly of the machinery that could be treated as a kind of sub-image. Each sub-assembly in a different way reflects the machine's conditions, thus analysis of the sub-images of sub-assemblies allows us to acquire partial diagnostic information about global conditions of an object. Process of analysis of each sub-image gives sets of features that represent the condition of each machine sub-assembly at the moment of its operation corresponding to the time of infrared image acquisition. The local conditions of the sub-assemblies are related to the machine's global condition.

Having determined the feature vectors for infrared images acquired during machine operation in different conditions (including faults), it is possible to design a set of local classifiers that allow us to identify conditions of the machinery. At this stage, the classifiers could be treated as local experts.

Local diagnostic information provided by each classifier can be joined together to get information about global (overall) machinery condition. In the elaborated method, to aggregate diagnostic decisions and maximize final classification performance, application of decision fusion methods were used.

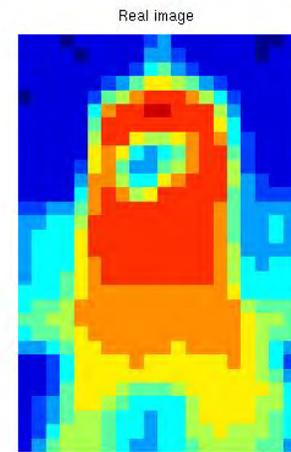
#### A. Processing and analysis of infrared images

The versatile nature of developed method allows us to apply different image processing and analysis methods to obtain a features set. For method verification purposes, the authors decided to use spectral representation of infrared images. Spectral representation of infrared images is obtained by use of the two-dimensional Fourier transform. One of the reasons of application of the 2D Fourier transform is a shift invariant property [5], which makes the method less sensitive to deviation in location of imaging device while observing an object. Spectral representation of infrared image could also emphasises diagnostic information that could be hidden in the real image.

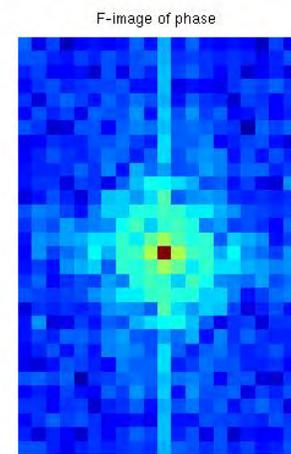
The result of Fourier transform of an infrared image is a two-dimensional spectrum, which could be represented by two images of magnitude and phase called also  $F$ -images. Frequency components on the  $F$ -images are distributed symmetrically and in many cases of the analysis it is enough to consider one quarter of the magnitudegrams and/or two adjacent quarters of the phasegrams. In most considered cases, the entire  $F$ -image is shown and analysed [4], [5]. This approach is most convenient for  $F$ -image interpretation purposes because frequency components generates specific symmetrically distributed patterns (similar to stars) (c.f. Fig. 2), whose shapes and locations depend on a content of the original infrared image.

To analyse the  $F$ -images, the three following features are defined:

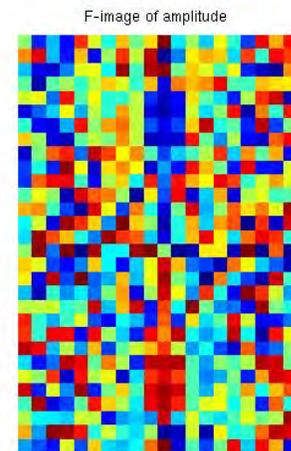
- HFP* Horizontal  $F$ -image Parameter,
- VFP* Vertical  $F$ -image Parameter,
- CFP* Circular  $F$ -image Parameter.



(a) Infrared image



(b)  $F$ -image of magnitude



(c)  $F$ -image of phase

Figure 2: An exemplary infrared image (a) and its  $F$ -images of magnitude (b) and phase (c) obtained on the basis of 2D Fourier analysis.

The features are mean values of  $F$ -images frequency components calculated over rectangular and circular areas, placed in the centers of the  $F$ -images in the way presented in Fig. 3. The dimensions of areas that were used to calculation of feature values were set experimentally (c.f. III-A)

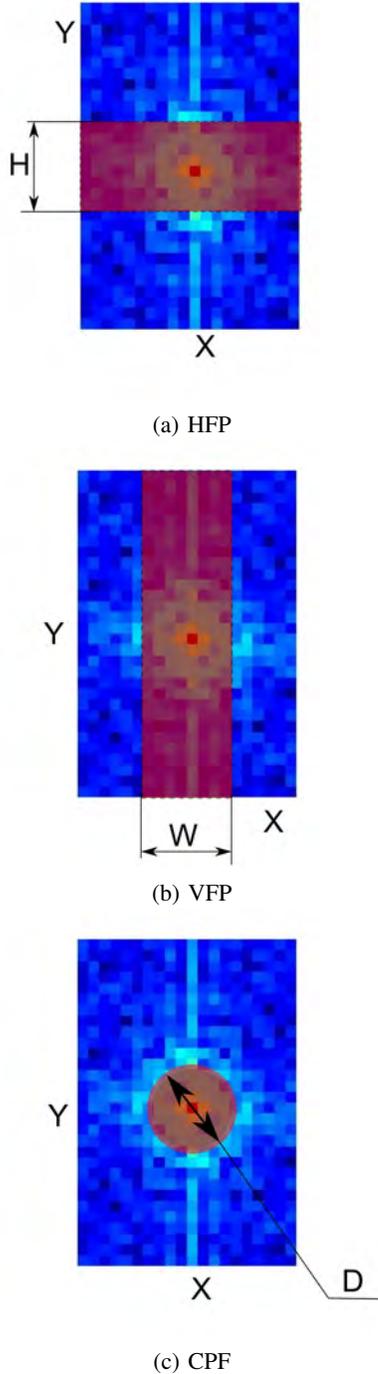


Figure 3: Graphical illustration of considered features of  $F$ -images.

### B. Classification of the machine's conditions

To classify machine operation conditions, a number of possible approaches could be chosen [6]. In practice, the choice of a classifier is a difficult problem and it is often based on a data specificity, as well as a researcher's experience.

The authors decided to apply two classifiers: a simple  $k$ -Nearest Neighbour ( $k$ -NN) classifier [7] and Support Vector Machine (SVM) [8], which is recognised as a very effective classification solution.

The author's intention was to show how to use the method and how the different classifiers behave.

To obtain a reliable and certain classification efficiency, the *leave-one-out* cross-validation (LOOCV) algorithm [9] has been applied.

The LOOCV validation method has a high variance but estimates of generalization error are comparable with other partitioning schemes used for classification efficiency evaluation [10].

The classifier accuracy measure that we used was the relative number of misclassification, which is calculated as follows:

$$err = N_e/N, \quad (1)$$

where  $N$  was the number of considered samples and  $N_e$  was the number of misclassified samples. On the basis of the  $err$  measure, the classifier efficiency was calculated in the following way:

$$eff = (1 - err) \cdot 100 \%. \quad (2)$$

### C. Decision fusion

In the elaborated method, joining of the classification results is proposed. There are some methods which allow treatment of the data jointly [11]. One of the interesting approaches is a decision fusion.

Decision fusion, which is also called classifier fusion, is the method that combines results of classification obtained from different classifiers trained over different types of data gathered from the same object. In this approach, classifiers are treated as "local experts", who make decision about the machine's condition.

The use of classifiers in technical diagnostic is connected with the uncertainty of the data on which those classifiers are trained. The sources of uncertainty could take the following form, for example, [12]: random events, measurement deviations, incompleteness of the set of considered diagnostic parameters and lack of knowledge about diagnosed object or process.

In general, most types of uncertainty could be characterised by the use of classical probability theory based on the Bayesian theorem [13], [14].

An alternative to the Bayesian methods is the Dempster-Shafer Theory (DST), also called the mathematical theory of evidence. The DST can deal with imprecise or incomplete data. In addition, DST can be interpreted as a generalisation of probability theory where probabilities are assigned to multiple

possible events (e.g. sets of events) as opposed to mutually exclusive singletons [15], [16].

The DST theory offers very important mechanisms of information aggregation coming from multiple sources by the use of rules for combining evidences. A lot of rules have been developed since establishing the DST.

Several interesting examples, including a detailed analysis of validity of Dempster's combination rule in different contexts, can be found in [17]–[19].

A generalisation and in some points an extension of the Dempster-Shafer evidence theory is The Dezert-Smarandache Theory (DSmT) [20] of plausible and paradoxical reasoning. DSmT overcomes some limitations of DST [20], [21] because it allows us to formally combine any kind of information. DSmT bases itself on similar terms as DST. The DSmT introduces the generalised frame of discernment  $\Theta$ , which contains  $n$  exhaustive elements  $(\theta_1, \dots, \theta_n)$ . In the classification case, elements of  $\Theta$  are all considered classes (class labels).

On the basis of the generalised frame of discernment, hyper-power set  $D^\Theta$  can be created of all single class labels but also of allowed class labels logical combinations. This means that classification can not only be made for single classes but the tested sample can also be assigned simultaneously to several classes ( $\theta_i \cap \theta_k \neq \emptyset$ ) or there can be some uncertainty in the reasoning process and the same test sample can be member of one or other class ( $\theta_i \cup \theta_k \neq \emptyset$ ). Each of these combinations is called focal element. For each element of  $D^\Theta$ , a Generalized Basic Belief Assignment (GBBA) is possible. In other words, as the result of classification some belief is assigned for test sample  $x$  that is a member of certain classes, several classes or there is some doubt to which class it should be assigned. From the formal site:  $m(\cdot) : D^\Theta \rightarrow [0, 1]$  so GBBA can take values from 0 to 1, and if  $m_x(A) = 1$  there is 100% belief that test element  $x$  belongs to class  $A$ . In contrast, for empty set - e.g.. unknown class  $m(\emptyset) = 0$ . Belief assigned to all elements of  $D^\Theta$  should sum up to 1:  $\sum_{A \in D^\Theta} m(A) = 1$ . This means that in the frame of discernment,  $D^\Theta$  tested elements are for sure member of one of classes or class combinations defined by  $D^\Theta$ ; so, no other unknown classes are allowed.

Similarly to DST, the DSmT also allows to aggregate information with the use of combination rules. For this purpose, many combination rules have been elaborated [20], [22]. During the research, a PCR6 rule was used. The key idea of the PCR6 rule is to transfer the partial conflicting Basic Belief Assignment *BBA* proportionally to the individual *BBA* of non-empty elements involved in the conflict [23].

#### D. GBBA calculation

The calculation of evidence is crucial for classifier fusion based on the methods demanding the *BBA* or the *GBBA* for each class [24], [25].

A simple method, which is ideal for research at the preliminary stage, has been developed for the evidence calculation from *k*-NN classifiers [26]. To obtain the output for a given sample, a set of distance measures to a number of known samples is calculated and it can be regarded as a class

distribution. Identification of  $k$  nearest neighbours of a element  $x$  irrespective of class label is made. Then, the number of neighbours  $k_i$  supporting assignment of element  $x$  to class  $C_i$  is calculated. Accordingly, the *GBBA* function of class  $C_i$  is calculated as follows [26]:

$$m(\{C_i\}) = k_i/k \quad (3)$$

In case of SVM classifier, which unfortunately gives only class labels, the probabilities of class distribution were obtained applying extension introduced by Wu [27]. In the presented research, we deal with only one occurring condition at the time, therefore probabilities are very useful. It can be assumed that SVM classifier outputs are degrees of support for each class representing identified machine conditions. These outputs can be directly transformed into mass assignments:  $p_i \rightarrow m(i)$ , where  $p_i$  is the probability of condition  $i$  occurrence and  $m(i)$  is the belief that condition  $i$  occurred provided by single SVM classifier on the basis of available evidence (in a form of feature space).

### III. METHOD VERIFICATION

Our method verification considers several different aspects of the method's application. First of all, verification should confirm that method can be useful in condition monitoring of machinery. The second important task of verification was to indicate what kind of classifier should be used and what is the best way to perform data fusion. To do this, we use two earlier described classifiers and compare results of classification obtained by use of single classifiers with results of classifier fusion, as well as results of classification obtained for multidimensional space of features. Investigation helps us to find the best solution to an answer the question: is fusion of simple classifiers a better solution than application of the classifier to single or multidimensional space of features?

The method was verified on the basis of digital infrared images taken during diagnostic experiments. All of our computations were performed using Matlab 2007b software.

#### A. Considered experimental data

The experiments have been performed using a laboratory stand that consists of a laboratory model of rotating machinery and an infrared imaging system (Fig. 4).

During the experiment, a sequence of 840 infrared images of resolution  $320 \times 480$  pixels has been recorded. The thermographic images have been taken every 30 s. The images have represented the machine operating in the conditions presented in the table I.

For reference, condition S1 decided to record two times more images to make it easier to recognise by classifiers. It should be pointed out that conditions S2, S3 and S4 are difficult to distinguish and have been simulated intentionally to check whether it was possible to notice a small change in operational condition. Such small changes were also desirable for testing the ability of the classifiers to recognise nearly indistinguishable changes in the machine's condition.

Table I: Description of conditions simulated during the experiment.

Condition Id	Description of fault	No of acquired images
S1	machine without faults	240
S2	50% throttling of air pump	120
S3	90% throttling of air pump	120
S4	90% throttling of air pump and clearance of second bearing mounting	120
S5	load of disk brake	120
S6	faulty bearing no 2	120

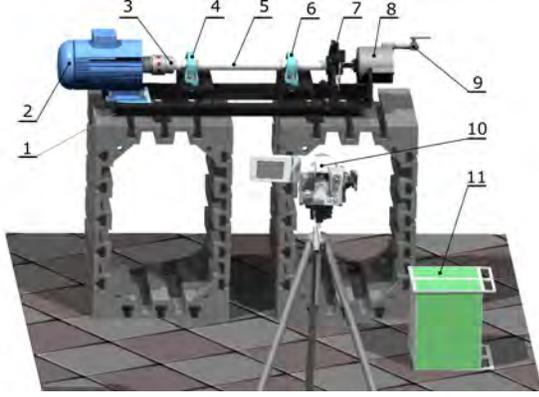


Figure 4: Visualisation of the laboratory stand. 1-frame, 2-motor (1.5 kW, 2500 rpm), 3-coupling, 4-bearings set no 1, 5-shaft, 6-bearings set no 2, 7-break set, 8-air pump, 9-throttle valve, 10-infrared camera connected to PC, 11-motor controller.

The infrared images acquired during the experiment have been pre-processed. The first step of the pre-processing was the selection of two Regions of Interest of size  $20 \times 30$  pixels (ROI1 and ROI2) (Fig. 5). These ROIs represented the bearing housings. It was expected that changes in the machine's condition would affect changes of bearing temperature and should be revealed in the infrared images.

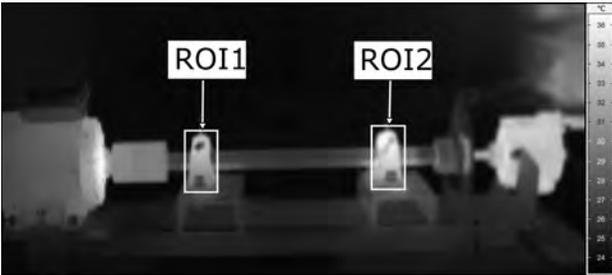


Figure 5: Infrared image of the operating laboratory stand, with marked ROIs of the first (left, ROI1) and the second (right, ROI2) bearings.

According to the proposed method (c.f. II), sub-images corresponding to regions of interest (ROI1 and ROI2) were transformed to frequency domain using Fast Fourier Transform (FFT) algorithm.  $F$ -images (magnitude and phase) obtained after transformation was analysed and image features were

calculated. Each infrared image was represented by 12 features, whose names were coded in the following way:

EstimatorId\_FImageType\_ROIID

e.g. HFP\_P\_R1 means that the value of the feature HFP was calculated for  $F$ -image of phase determined in ROI1.

It is obvious that values of presented  $F$ -image features depend on dimensions and content of the region of interest (ROI), as well as type of  $F$ -image (magnitude and phase). To consider a variety in content of each type of  $F$ -images, each of the proposed feature could be fitted to the image content by setting a value of the feature parameter  $W$ ,  $H$ , and  $D$ .

To find the optimal values of  $F$ -image feature parameters  $W$ ,  $H$ , and  $D$  an exhaustive search of feature space based on criterion of the maximum machine conditions classifier performance has been performed. Features have been calculated for each acceptable value of the feature parameters (from 1 to the maximal value  $H_{\max} = 30$ ,  $W_{\max} = 20$ , and  $D_{\max} = 20$ ). Constrains followed from the size ( $20 \times 30$  pixels) of the considered  $F$ -images.

For optimisation purposes, a k-Nearest Neighbour ( $k$ -NN) classifier was used. A number of nearest neighbours parameter was set to  $k = 10$  according to recommendations presented in [28]. Classification efficiency was calculated in the way presented in the theoretical background (c.f. II-B) and *leave-one-out* cross-validation (LOOCV) algorithm was used. Optimal values of feature parameters are presented in Tab. II.

Table II: Optimal values of feature parameters and basic statistics of classification efficiencies.

feat. num.	feat. name	estimator parameter name	estimator parameter value	mean eff [%]
1	ROI1_VFP_A	H	20	59.6
2	ROI2_VFP_A	H	18	80.6
3	ROI1_VFP_P	H	2	23.9
4	ROI2_VFP_P	H	7	25.2
5	ROI1_HFP_A	W	29	56.9
6	ROI2_HFP_A	W	26	80.8
7	ROI1_HFP_P	W	9	28.3
8	ROI2_HFP_P	W	9	24.3
9	ROI1_CFP_A	D	14	62.3
10	ROI2_CFP_A	D	20	75.6
11	ROI1_CFP_P	D	6	37.1
12	ROI2_CFP_P	D	6	43.4

The feature values, calculated using determined optimal parameters of  $F$ -image features, were data source for classification of machine conditions.

*B. Classification results for one and multidimensional feature space*

The first step of the method verification was assessment of application of one and multidimensional *F*-images feature space for purposes of classification of machine condition. As mentioned earlier, k-NN and SVM classifiers were applied. In case of k-NN classifier, a  $k = 10$  neighbours was used. The Euclidean distance function was used as a distance metric in k-NN classifier. In case of SVM classifier, one-against-all strategy is implemented for multi-class classification. A Gaussian kernel was applied. Mean classifier efficiencies of considered machine conditions as a function of feature space dimension were shown in Fig. 6. As one can expect, classification efficiency increase with size of feature space and for almost all conditions reach efficiency above 80% for size of feature space equal to 4 and more.

A detailed analysis of maximal classifier efficiencies is presented in Table III. The results show that in case of conditions S1, S3, S4, S5 and S6 maximum efficiencies could be achieved for one dimensional space of feature vales for both types of applied classifiers. Values of maximal classification efficiencies are given in bold. The highest classification effi-

ciency values have been obtained on the basis of CFP feature, which indicates its usefulness in analysing the *F*-images. The greatest number of maximum classification efficiency (100%) was obtained using the SVM classifier. SVM gave the best results for features of *F*-images of phase whereas k-NN gave good results for *F*-images of magnitude within the region of interest ROI2. Region ROI2 covered more load bearing support, which affected its highest temperature and thus intensive infrared radiation.

A plot of classification efficiencies for condition S2 presented in Fig. 6 and values in Table III clearly show that condition S2 is poor recognizable. Analysis of classification efficiencies for condition S2 shows that application of one dimensional feature space allowed to obtain maximal efficiency equals 58.3% with application of k-NN classifier. The SVM classifier was unable to correctly recognize condition S2, where SVM allowed to obtain maximal efficiency equal 8.3%.

Looking at feature values distribution for condition S2 presented in Figure 7, it is clear that SVM was unable to find proper global decision boundaries. Exemplary decision boundary for condition S2 vs all other conditions can be

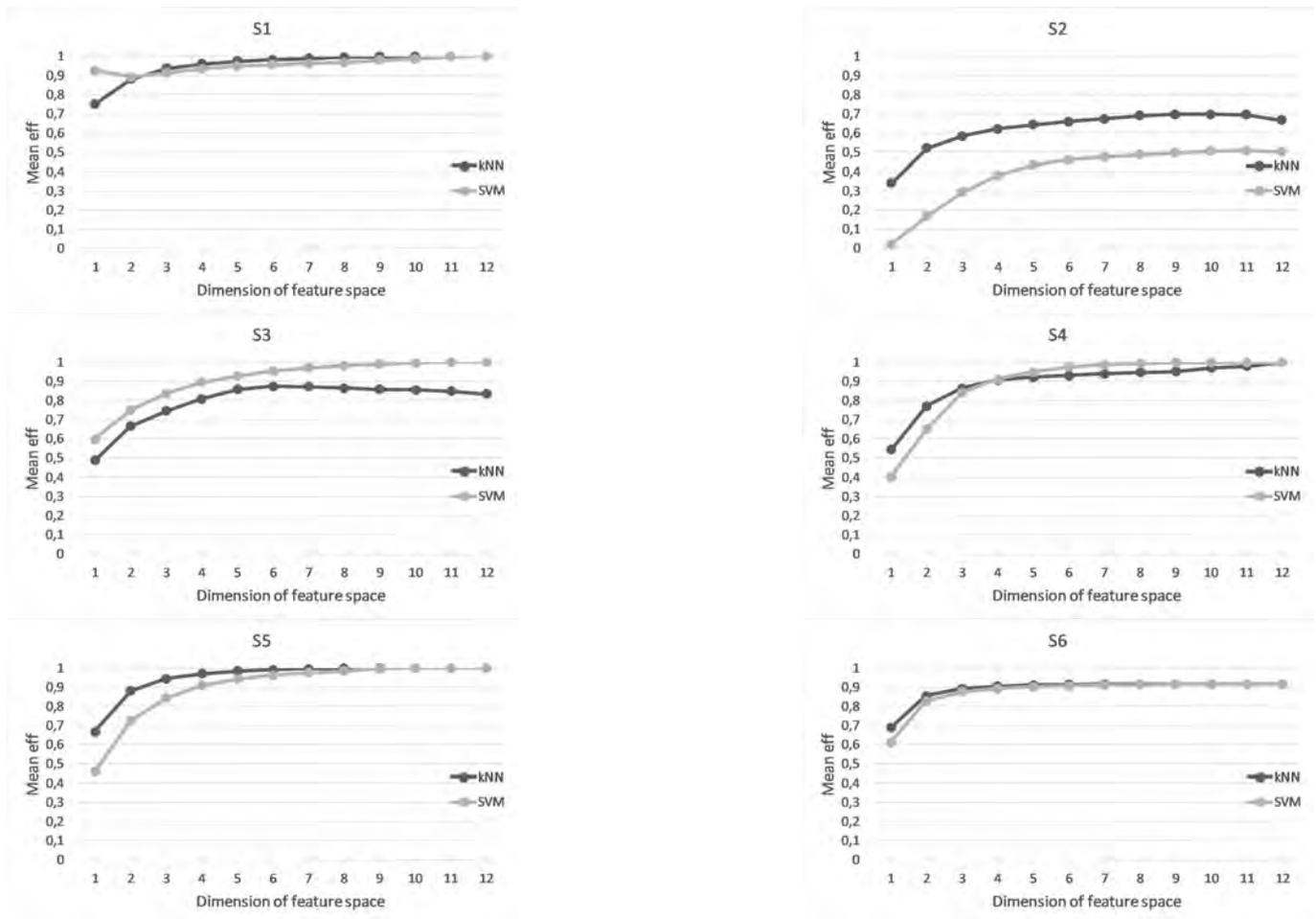


Figure 6: Evolution of the belief in a cell crossed by an obstacle observed by a sensor.

Table III: Classification efficiencies obtained for individual  $F$ -image features

Feature space	Simulated machine conditions					
	S1	S2	S3	S4	S5	S6
kNN(HFP_A_R1)	62.5	16.7	91.7	33.3	91.7	91.7
SVM(HFP_A_R1)	87.5	0.0	<b>100.0</b>	0.0	83.3	91.7
kNN(HFP_P_R1)	54.2	16.7	8.3	25.0	25.0	16.7
SVM(HFP_P_R1)	87.5	0.0	<b>100.0</b>	0.0	83.3	91.7
kNN(HFP_A_R2)	91.7	58.3	41.7	83.3	<b>100.0</b>	91.7
SVM(HFP_A_R2)	83.3	8.3	<b>100.0</b>	83.3	<b>100.0</b>	83.3
kNN(HFP_P_R2)	50.0	16.7	8.3	8.3	16.7	8.3
SVM(HFP_P_R2)	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
kNN(VFP_A_R1)	58.3	8.3	<b>100.0</b>	58.3	91.7	91.7
SVM(VFP_A_R1)	79.2	0.0	<b>100.0</b>	16.7	75.0	91.7
kNN(VFP_P_R1)	50.0	0.0	0.0	8.3	16.7	16.7
SVM(VFP_P_R1)	<b>100.0</b>	0.0	0.0	0.0	0.0	0.0
kNN(VFP_A_R2)	91.7	58.3	25.0	83.3	<b>100.0</b>	91.7
SVM(VFP_A_R2)	83.3	8.3	<b>100.0</b>	83.3	<b>100.0</b>	83.3
kNN(VFP_P_R2)	62.5	0.0	16.7	25.0	8.3	16.7
SVM(VFP_P_R2)	95.8	0.0	0.0	0.0	25.0	16.7
kNN(CFP_A_R1)	87.5	16.7	91.7	0.0	91.7	83.3
SVM(CFP_A_R1)	<b>100.0</b>	8.3	<b>100.0</b>	25.0	75.0	83.3
kNN(CFP_P_R1)	62.5	0.0	0.0	8.3	25.0	16.7
SVM(CFP_P_R1)	<b>100.0</b>	0.0	<b>100.0</b>	75.0	0.0	<b>100.0</b>
kNN(CFP_A_R2)	91.7	58.3	75.0	91.7	<b>100.0</b>	91.7
SVM(CFP_A_R2)	83.3	0.0	<b>100.0</b>	91.7	91.7	83.3
kNN(CFP_P_R2)	62.5	50.0	16.7	33.3	8.3	41.7
SVM(CFP_P_R2)	<b>100.0</b>	0.0	16.7	<b>100.0</b>	0.0	58.3

seen in Fig. 7. Taking into consideration the distribution of feature values for condition S2, the strategy of classification using SVM classifier with linear boundaries is insufficient to distinguish between S2 and other classes. Application of feature spaces dimensionality of 3-8 increase maximal classification performance of condition S2 which was, respectively, 83% for k-NN and 75% for SVM classifiers. Minimal space giving maximal classification performance with use of k-NN classifier was constructed with use of the following two sets of features:

$VFP\_P\_R2, HFP\_A\_R2, HFP\_P\_R2$  and  $VFP\_A\_R1, HFP\_P\_R2, CFP\_A\_R1$

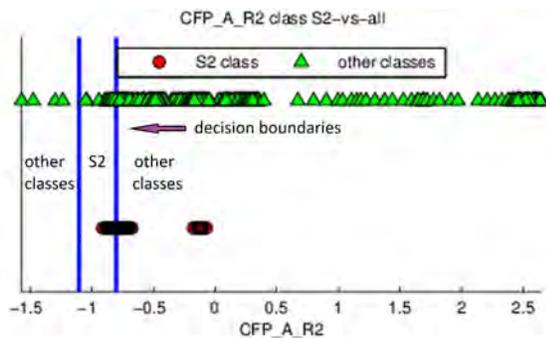


Figure 7: Distribution of CFP\_A\_R2 feature for condition S2.

To assess which classes are most similar a confusion matrix was prepared (Fig. 8). In each column there is percentage fraction of each class that was assigned to various predicted classes. Taking into consideration only single kNN classifiers, that were trained over 1D data set it can be seen, that conditions S2 and S3 are most difficult to distinguish. It is connected

with the way in which those conditions were simulated, when only degree throttling of air pump was changed.

Predicted class	Actual class					
	1	2	3	4	5	6
1	92.3	0.0	0.0	0.0	0.0	0.0
2	7.7	58.4	30.0	0.0	0.0	0.0
3	0.0	41.3	70.0	0.0	0.0	0.0
4	0.0	0.0	0.0	100.0	0.0	0.0
5	0.0	0.0	0.0	0.0	85.7	0.0
6	0.0	0.0	0.0	0.0	14.3	100.0

Figure 8: Normalized confusion matrix for single kNN classifier.

#### IV. CLASSIFIER FUSION RESULTS

Results of the classification of machine conditions shown in Section II-B, (Tab. III) indicate that the proposed features of the  $F$ -images are useful for assessing machine conditions. For the majority of concerned machine conditions, it was possible to obtain the maximum classification efficiency on the basis of selected individual features of  $F$ -images. However, for condition S2, reliable condition assessment was not possible. To increase classification efficiency, fusion of classifiers was applied. We carried out an exhaustive computation considered all combinations of two, three and four k-NN and SVM classifiers of all considered  $F$ -image features.

Mean classification efficiencies after classifier fusion as a function of number of fused classifier for all considered conditions are presented in Fig. 9. Table IV presents the highest classification efficiencies obtained for all condition after fusion of two, three and four combinations of different individual k-NN and SVM classifiers. Our results show that fusion of two classifiers is sufficient to obtain maximal classification almost for all conditions. Classifier fusion also allowed us to raise the highest classification efficiency for condition S2 by the 8.3% (from 58.3% to 66.7%) in comparison to the results obtained for the individual classifiers. The maximum efficiency of the classification was obtained as a result of the fusion of k-NN classifiers only. Fusion of the SVM classifiers does not ensure an increase of the classification efficiency for this class.

The most interesting observation made after the analysis of classification performances is the lack of an increase of the efficiency for the condition S2 according to the number of fused single classifiers. This is caused by the presence of the classifiers that assign the high degree of belief to the wrong states. Fusing more than two classifiers did not cause an increase of the relative number of classifier combinations giving maximal performance. These results find confirmation

Table IV: Comparison of maximal classification efficiencies for all conditions after fusion different numbers of single kNN and SVM classifiers using PCR6 rule.

Fuzz. Class. #	Class. Type	Simulated machine conditions					
		S1	S2	S3	S4	S5	S6
2	kNN(.)	100	66.7	100	100	100	100
	SVM(.)	100	8.3	100	100	100	100
3	kNN(.)	100	66.7	100	100	100	100
	SVM(.)	100	8.3	100	100	100	100
4	kNN(.)	100	66.7	100	100	100	100
	SVM(.)	100	8.3	100	100	100	100

in [29], which showed that adding additional experts at some point leads to obtaining totally conflicted and useless classifier combinations. Analysis of classifiers combinations giving the highest performances indicates that they are composed from complementary rather than individually best performing classifiers. Taking into account the obtained results, it can be concluded that the fusion of two selected classifiers is sufficient. In case of the considered data, a pair of classifiers assuring highest efficiency 66.7% was *HFP\_A\_R2, CFP\_A\_R1*.

Taking into consideration the very good results of

classification obtained for multidimensional feature spaces decided to perform fusion of kNN classifiers calculated for two dimensional feature spaces. As could have been expected, the results were very good (Table V). Maximal classification efficiency for condition S2 was increased to 83.3% for four following combination of classifiers and feature spaces:

- PCR6{kNN{VFP\_A\_R2,CFP\_A\_R2}},
- kNN{HFP\_P\_R2,CFP\_A\_R1}},
- PCR6{kNN{HFP\_P\_R2,CFP\_A\_R1}},
- kNN{HFP\_P\_R2,CFP\_A\_R2}},
- PCR6{kNN{HFP\_P\_R2,CFP\_A\_R1}},
- kNN{HFP\_P\_R2,CFP\_P\_R2}},
- PCR6{kNN{HFP\_P\_R1,CFP\_P\_R1}},
- kNN{HFP\_P\_R2,CFP\_A\_R1}}.

Its worth mentioning that maximal classification efficiency using single k-NN classifier for condition S2 with the use of three and four dimensional (3D and 4D) space of feature was also 83.3%. The presented results confirm the ability of decision fusion algorithms to identify machinery conditions which are difficult to be recognised. In contrast, the SVM classifier results for the 2D feature space was maximally

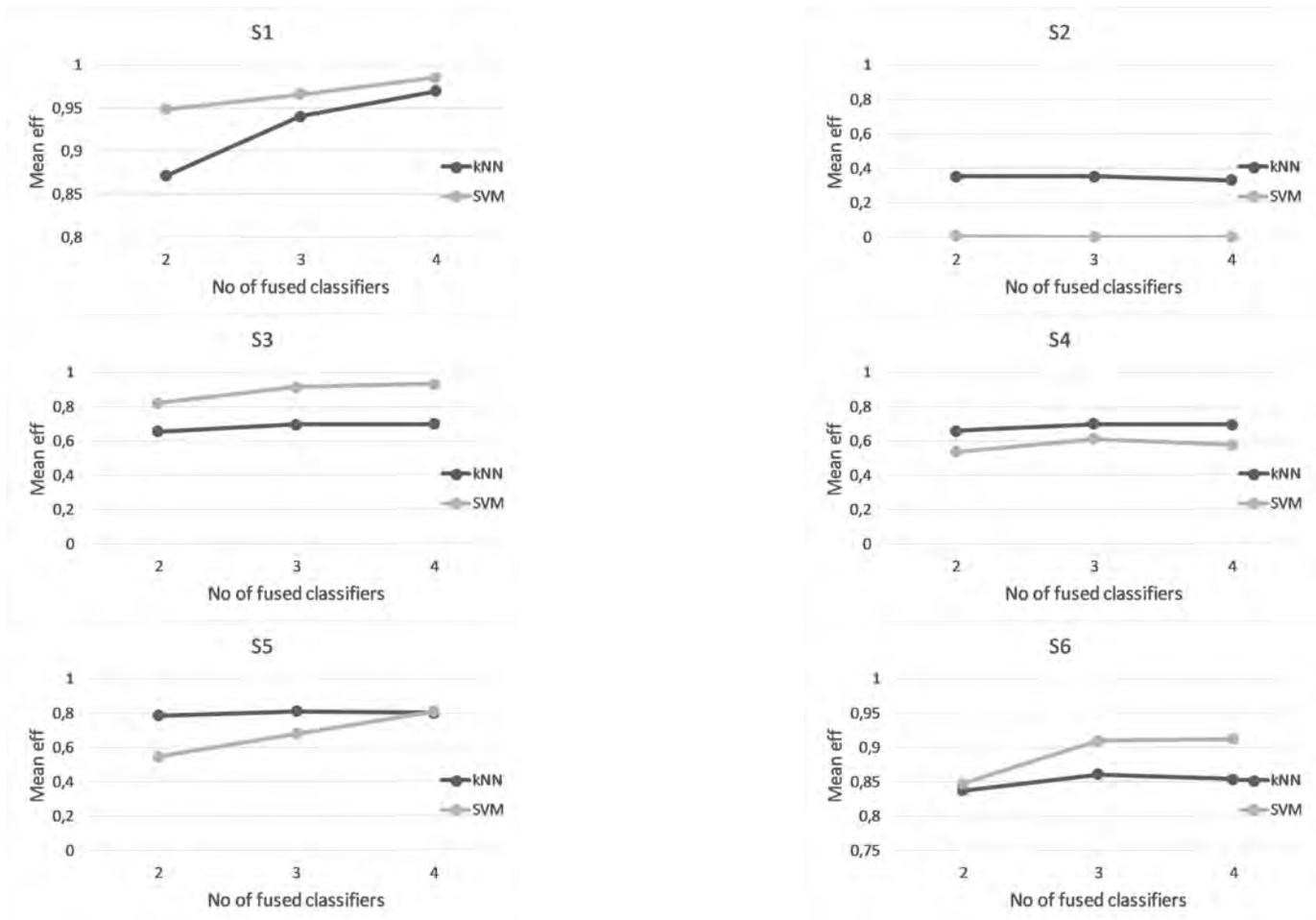


Figure 9: Plots of mean classification efficiencies as a function of different number of fused classifiers for considered machine conditions.

62.5%. Accordingly, the increase of classification performance in comparison to single feature space is visible and in this the kNN classifier was proven to be better than the SVM classifier.

Table V: Comparison of maximal classification efficiencies for all conditions fusion of 2 classifiers trained over 2D feature space.

Fuzz. Class. #	Class. Type	Simulated machine conditions					
		S1	S2	S3	S4	S5	S6
2	kNN(.)	100	83.3	100	100	100	100
	SVM(.)	100	62.5	100	100	100	100

## V. CONCLUSIONS

In this paper, the method of object condition assessment using multiple classifiers fusion approach based on the generalised evidence theory is proposed. Fused classifiers have been trained over the data represented by three parametric spectral features of  $F$ -images. The  $F$ -images were the result of the 2D Fourier transform of infrared images acquired during object observation. During the research, optimal parameters of the features were evaluated and  $F$ -image features were computed. Based on the spectral features of the infrared images the classification process was performed. For comparison purposes k-NN and SVM classifiers were used. The results of the classification have shown that the proposed features of an  $F$ -image of thermograms could be useful for the evaluation of a machine's condition. Circular Fourier Power (CFP) seemed to be suitable enough for the estimation of magnitude, as well as phase  $F$ -images.

The proposed approach of classifier fusion is suitable for the assessment of machine global condition on the basis of pre-selected features of spectral infrared images. Classification efficiencies obtained using classifier fusion are higher than those calculated taking into consideration a single classifier. It must be mentioned that features chosen for the member classifiers in fusion process should be heterogeneous to assure high classification efficiency. Moreover, the increase of the number of considered ROIs should entail a reduction of the uncertainty of the information, which is used in the decision making about the machine's global condition. Although the connection between diversity of features and the classification performance is not always straightforward, the analysis of the obtained results leads to the statement that in the considered case, the influence of feature heterogeneity degree on the fusion results is quite noticeable.

The problem of high homogeneity of data could be resolved by classification of multidimensional space of homogeneous feature values and the next application of the fusion of such a classifier. This strategy was verified during the presented research and the obtained results confirmed the ability of classifier fusion to increase classification efficiency of condition S2, which was difficult to recognise.

It can be expected that conclusions made from the research could be generalised to data represented by other infrared image features and diagnostic signals. However, it needs further investigation.

## REFERENCES

- [1] R.A. Thomas, *The thermography monitoring handbook*, Oxford, UK: Coxmoor Pub. (Machine & systems condition monitoring series), 1999.
- [2] A.M. Younus, B.S. Yang, *Intelligent fault diagnosis of rotating machinery using infrared thermal image*, Expert Syst. Appl., Vol. 39, pp. 2082–2091, 2012.
- [3] A. Widodo, D. Satrijo, T. Prahasto, et al., *Confirmation of thermal images and vibration signals for intelligent machine fault diagnostics*, Int. J. Rotating Mach., Vol. 2012, pp. 1–10, 2012.
- [4] M. Fidali, *Methodology of thermographical diagnostics of technical object*, (in polish), Gliwice: Wydawnictwo Naukowe Instytutu Technologi Eksploatacji - PIB, 2013.
- [5] S.M. Nixon, S. Aguado, *Feature extraction and image processing*, Oxford, Boston: Newnes, 2002.
- [6] A.K. Jain, R.P. Duin, J. Mao, *Statistical pattern recognition: a review*, IEEE Trans. on Pattern Anal. Mach. Intell., Vol. 22(1), pp. 4–37, 2000.
- [7] P. Cunningham, S.K. Delany, *K-nearest neighbour classifiers*, Technical report UCD-CSI-2007-4, 2007.
- [8] V.N. Vapnik, *The nature of statistical learning theory*, Berlin, Heidelberg: Springer-Verlag, 1995.
- [9] T. Evgeniou, E.M. Pontil, A. Elisseeff, *Leave-one-out error, stability, and generalization of voting combinations of classifiers*, Mach. Learning, Vol. 55(1), pp. 71–97, 2004.
- [10] R. Kohavi, *A study of cross-validation and bootstrap for accuracy estimation and model selection*, in Proc. of the 14th International joint conference on Artificial intelligence (IJCAI'95), pp. 1137–1145, Montreal: Morgan Kaufmann, 1995.
- [11] M. Fidali, *Method of machine technical state assessment on the basis of joint signal analysis*, Mech. Syst. Signal Process, Vol. 25(3), pp. 871–883, 2011.
- [12] J. Korbicz, J. Kościelny, Z. Kowalczyk, et al. (Editors), *Fault diagnosis. Models, artificial intelligence, applications*, Berlin: Springer, 2004.
- [13] M. Bednarski, W. Cholewa, W. Frid, *Identification of sensitivities in bayesian networks*, Eng. Appl. Artif. Intell, Vol. 17(4), pp. 327–335, 2004.
- [14] N. Subrahmanya, Y.C. Shin, P.H. Meckl, *A bayesian machine learning method for sensor selection and fusion with application to on-board fault diagnostics*, Mech. Syst. Signal Process, Vol. 24(1), pp. 182–192, 2010.
- [15] G. Shafer, *A mathematical theory of evidence*, Mercer County, NJ: Princeton University Press, 1976.
- [16] A. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. Math. Statist., Vol. 38, pp. 325–339, 1967.
- [17] P. Wang, *A defect in Dempster-Shafer theory*, in Proc. of the 10th Conf. on uncertainty in artificial intelligence, Ramon Lopez de Mantaras & David Poole (Editors), Morgan Kaufmann Publishers, pp. 560–566, 1994.
- [18] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of 15th Int. Conf. on information fusion (Fusion 2012), pp. 655–660, Singapore, July 2012.
- [19] F. Smarandache, J. Dezert, *Modified PCR rules of combination with degrees of intersections*, in Proc. of 18th Int. Conf. on information fusion (Fusion 2015), p. 2100–2107, Washington DC, USA, July 2015.
- [20] J. Dezert, F. Smarandache, *DSMT: A New Paradigm Shift for Information Fusion*, in Proc. of Cogis'06 Conference, pp. 15–17, Paris, France, March 2006.
- [21] L. Zadeh, *On the validity of Dempster's rule of combination*, Memo UCB/ERL M79/24, University of California, Berkeley, CA, 1979.
- [22] P. Smets, R. Kennes, *The transferable belief model*, Artif. Intell., Vol. 66, pp.191–234, 1994.
- [23] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of the 16th Int. Conf. on Information Fusion (Fusion 2013), pp. 1119–1126, Istanbul, Turkey, July, 2013.
- [24] L. Xu, A. Krzyzak, C. Suen, *Methods of combining multiple classifiers and their applications to handwriting recognition*, IEEE Trans. on Syst. Man and Cybernetics, Vol. 22(3), pp. 418–435, 1992.
- [25] B. Zhang, S. Srihari, *Class-wise multi-classifier combination based on Dempster-Shafer theory*, in Proc. of 7th Int. Conf. on control, automation, robotics and vision (ICARCV), Vol. 2, pp. 698–703, Marina Mandarin, Singapore, December 2002.

- [26] D. Han, C. Han, Y. Yang, *Multiple k-nn classifiers fusion based on evidence theory*, in Proc. of IEEE Int. Conf. on automation and logistics, pp. 2155–2159, August, Jinan, China, 2007.
- [27] T.F. Wu, C.J. Lin, R.C. Weng, *Probability estimates for multi-class classification by pairwise coupling*, J. Mach. Learn., Vol. 5, pp. 975–1005, 2004.
- [28] G.E.A.P.A. Batista, M.C. Monard, *A study of K-nearest neighbour as an imputation method*, in Proc. of Second Int. Conf. on hybrid intelligent systems, pp. 251–260, Maebashi City, Japan, December 2002.
- [29] A. Martin, A.-L. Jousselme, C. Osswald, *Conflict measure for the discounting operation on belief functions*, in Proc. of 11th Int. Conf. on information fusion (Fusion 2008), p. 1–8, Cologne, Germany, July 2008.

# Inter-Criteria Analysis Based on Belief Functions for GPS Surveying Problems

Stefka Fidanova<sup>a</sup>, Jean Dezert<sup>b</sup>, Albena Tchamova<sup>a</sup>

<sup>a</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

Emails: stefka@parallel.bas.bg, jean.dezert@onera.fr, tchamova@bas.bg

Originally published as: S. Fidanova, J. Dezert, A. Tchamova, *Inter-Criteria Analysis Based on Belief Functions for GPS Surveying Problems*, in Proc. of INISTA 2019, Sofia, Bulgaria, July 3–5, 2019 (Best Paper Awards), and reprinted with permission.

**Abstract**—In this paper we present an application of a new Belief Function-based Inter-Criteria Analysis (BF-ICrA) approach for Global Positioning System (GPS) Surveying Problems (GSP). GPS surveying is an NP-hard problem. For designing Global Positioning System surveying network, a given set of earth points must be observed consecutively. The survey cost is the sum of the distances to go from one point to another one. This kind of problems is hard to be solved with traditional numerical methods. In this paper we use BF-ICrA to analyze an Ant Colony Optimization (ACO) algorithm developed to provide near-optimal solutions for Global Positioning System surveying problem.

**Keywords:** Inter-Criteria Analysis, BF-ICrA, GPS surveying, PCR6, belief functions.

## I. INTRODUCTION

In our previous work [1] we did apply classical Atanassov's Inter-Criteria Analysis (ICrA) to examine some relations between considered GSP's and ACO algorithm performance. In this paper we consider a recent improved version of ICrA based on belief functions [2] and show how to apply it in same GSP problematic to revise and refine our previous analysis.

After a short presentation of GSP problematic and ACO in the next section, and brief basics of BF in section III, we recall the classical Atanassov's ICrA method in section IV and we present the new ICrA method based on Belief Functions, called BF-ICrA, in section V. In section VI, we show how to apply BF-ICrA for GSP problematic. Concluding remarks are given in Section VII.

## II. PRESENTATION OF ACO AND GSP PROBLEMATIC

### A. GPS surveying problem description

GPS satellites continuously transmit radio signals to the Earth while orbiting it. A receiver, with unknown position on Earth, has to detect and convert the signals received from all of the satellites into useful measurements. These measurements would allow a user to compute a three-dimensional coordinate position: location of the receiver. Any GPS observation is proven to have biases, hence, in order to survey an appropriate combination of measurement processing strategies must be used to minimize their effect on the positioning results. Differencing data collected simultaneously from two or more GPS receivers to several GPS satellites allows to eliminate or significantly reduce most of the biases. The GPS network

can be defined as set of stations  $(a_1, a_2, \dots, a_n)$ , which are co-ordinated by placing receivers  $(X_1, X_2, \dots)$  on them to determine sessions  $(a_1a_2, a_1a_3, a_2a_3, \dots)$  among them. The problem is to search for the best order in which these sessions can be organized to give the best schedule. Thus, the schedule can be defined as a sequence of sessions to be observed consecutively. The solution is represented by linear graph with weighted edges. The nodes represent the stations and the edges represent the moving cost. The objective function of the problem is the cost of the solution which is the sum of the costs (time) to move from one point to another one,  $C(V) = \sum C(a_i, a_j)$ , where  $a_i a_j$  is a session in solution  $V$ . For example if the number of points (stations) is 4, a possible solution is  $V = (a_1, a_3, a_2, a_4)$  and it can be represented by linear graph  $a_1 \rightarrow a_3 \rightarrow a_2 \rightarrow a_4$ . The moving costs are as follows:  $C(a_1, a_3), C(a_3, a_2), C(a_2, a_4)$ . Thus the cost of the solution is  $C(V) = C(a_1, a_3) + C(a_3, a_2) + C(a_2, a_4)$ . In practice, determining how each GPS receiver should be moved between stations to be surveyed in an efficient manner taking into account some important factors such as time, cost etc. The problem is to search for the best order, with respect to the time, in which these sessions can be observed to give the cheapest schedule or to minimize  $C(V)$ . The initial data is a cost matrix, which represents the cost (time, or distance) of moving a receiver from one point to another. Solving such problems - GSPs - to optimality requires a very high computational time. Therefore, meta-heuristic methods are used to provide near-optimal solutions for large networks within acceptable amount of computational effort. In this paper, we consider the Max-Min Ant System (MMAS) meta-heuristic [3] and we present it briefly in the next subsection.

### B. Ant colony optimization for GPS surveying problem

Real ants foraging for food lay down quantities of pheromone (chemical cues) marking the path that they follow. An isolated ant moves essentially at random but an ant encountering a previously laid pheromone will detect it and decide to follow it with high probability and thereby reinforce it with a further quantity of pheromone. The repetition of the above mechanism represents the auto-catalytic behavior of real ant colony where the more the ants follow a trail, the more attractive that trail becomes.

The ACO algorithm uses a colony of artificial ants that behave as cooperative agents in a mathematics space where they are allowed to search and reinforce pathways (solutions) in order to find the optimal ones. The problem is represented by graph and the ants walk on the graph to construct solutions. The solution is represented by path in the graph. After initialization of the pheromone trails, ants construct feasible solutions, starting from random nodes, then the pheromone trails are updated. At each step ants compute a set of feasible moves and select the best one (according to some probabilistic rules) to carry out the rest of the tour. The transition probability  $p_{ij}$ , to choose the node  $j$  when the current node is  $i$ , is based on the heuristic information  $\eta_{ij}$  and pheromone trail level  $\tau_{ij}$  of the move, where  $i, j = 1, \dots, n$ .

$$p_{ij} = \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum_{k \in U_{\text{unused}}} \tau_{ik}^\alpha \eta_{ik}^\beta}. \quad (1)$$

The higher value of the pheromone and the heuristic information, the more profitable is to select this move and resume the search. In the beginning, the initial pheromone level is set to a small positive constant value  $\tau_0$  and then ants update this value after completing the construction stage. ACO algorithms adopt different criteria to update the pheromone level.

In our implementation we use MAX-MIN Ant System (MMAS) [3], [4], which is one of the best ant approaches. In MMAS the main is using fixed upper bound  $\tau_{max}$  and lower bound  $\tau_{min}$  of the pheromone trails. Thus accumulation of big amount of pheromone by part of the possible movements and repetition of same solutions is partially prevented. The main features of MMAS are:

The aim of using only one solution is to make solution elements, which frequently occur in the best found solutions, get large reinforcement. Pheromone trail update is given by:

$$\tau_{ij} \leftarrow \rho \tau_{ij} + \Delta \tau_{ij}, \quad (2)$$

where

$$\Delta \tau_{ij} = \begin{cases} 1/C(V_{best}) & \text{if } (i, j) \in \text{best solution,} \\ 0 & \text{otherwise,} \end{cases}$$

and  $V_{best}$  is the iteration best solution and  $i, j = 1, \dots, n$ .

To avoid stagnation of the search, the range of possible pheromone value on each movement is limited to an interval  $[\tau_{min}, \tau_{max}]$ .  $\tau_{max}$  is an asymptotic maximum of  $\tau_{ij}$  and  $\tau_{max} = 1/(1 - \rho)C(V^*)$ , while  $\tau_{min} = 0.087\tau_{max}$ . Where  $V^*$  is the optimal solution, but it is unknown, therefore we use  $V_{best}$  instead of  $V^*$ .

When all ants have completed their solutions, the pheromone level is updated by applying the global update rule. Only the pheromone corresponding to the best found solution is increased by the similar to the MMAS way. The global update rule is intended to provide a greater amount of pheromone on the paths of the best solution. It is a kind of intensification of the search around the best found solution.

We use heuristic information equals to one over the cost of the session.

### III. BASICS OF THE THEORY OF BELIEF FUNCTIONS

Let consider a finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where  $\theta_i \cap \theta_j = \emptyset$  for  $i \neq j$ . The power-set of  $\Theta$  (i.e. the set of all subsets of  $\Theta$ ) is denoted  $2^\Theta$ . A basic belief assignment (BBA) associated with a given source of evidence is defined [5] as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are usually interpreted respectively as lower and upper bounds of unknown (possibly subjective) probability measure [6]. They are defined by<sup>1</sup>

$$Bel(A) \triangleq \sum_{B \subseteq A, B \in 2^\Theta} m(B), \quad \text{and} \quad Pl(A) \triangleq 1 - Bel(\bar{A}). \quad (3)$$

If  $m(A) > 0$ ,  $A$  is called a focal element of  $m(\cdot)$ . When all focal elements are singletons then  $m(\cdot)$  is called a *Bayesian BBA* and its corresponding  $Bel(\cdot)$  function is homogeneous to a probability measure. Historically the combination of BBAs is accomplished by Dempster's rule in Dempster-Shafer Theory (DST) [5]. Because of serious problems of Dempster's rule<sup>2</sup>, we recommend the Proportional Conflict Redistribution rule no. 6 (PCR6) proposed by Martin and Osswald in [10] (Vol. 3) which remains the most appealing alternative rule for BBA combination so far.

### IV. ATANASSOV'S INTER-CRITERIA ANALYSIS (ICRA)

Based on Intuitionistic Fuzzy Sets (IFS) [11], the Inter-Criteria Analysis (ICrA) has been introduced in 2014 by Atanassov et al. in [12], and then improved in [13], [15]. ICrA aims to identify the possible links between the criteria involved in a process of evaluation of multiple objects against multiple criteria. The aim of ICrA is to discover any existing correlations between the criteria themselves. Such analysis can permit (when possible) to reduce the complexity of large multiple criteria decision-making (MCDM) problems [2]. Until now the classical<sup>3</sup> ICrA has been applied in different fields: medicine [16], [17], optimization [18]–[21], workforce planning [22], competitiveness analysis [23], radar detection [24], ranking [25]–[27], etc. In this section we just recall the basic principles of classical ICrA.

Let consider a set of alternatives (or objects)  $\mathbf{A} \triangleq \{A_1, A_2, \dots, A_M\}$  ( $M > 2$ ), and a set of criteria  $\mathbf{C} \triangleq \{C_1, C_2, \dots, C_N\}$  ( $N \geq 1$ ). The available information is expressed by a  $M \times N$  score matrix<sup>4</sup>  $\mathbf{S} \triangleq [S_{ij} = C_j(A_i)]$ , and (eventually) the importance factor  $w_j \in [0, 1]$  of each criterion  $C_j$  with  $\sum_{j=1}^N w_j = 1$ . The ICrA method consists to build an  $N \times N$  Inter-Criteria (IC) matrix  $\mathbf{K}$  from the score matrix  $\mathbf{S}$ . The elements of the IC matrix  $\mathbf{K}$  consist of all Intuitionistic Fuzzy (IF) pairs  $(\mu_{jj'}, \nu_{jj'}) \in [0, 1] \times [0, 1]$  whose

<sup>1</sup>In the notations, the symbol  $\triangleq$  means *equal by definition*.

<sup>2</sup>that is: 1) insensitivity to the level of conflict between sources in some cases and dictatorial behavior [7], [8], and 2) inconsistency of Shafer's belief conditioning [9] with bounds of conditional probabilities.

<sup>3</sup>We refer Atanassov's ICrA as the classical approach in the sequel.

<sup>4</sup>also called benefit or payoff matrix in Multi-Criteria Decision-Making framework.

components express respectively the degree of agreement and the degree of disagreement between criteria  $C_j$  and  $C_{j'}$  for  $j, j' \in \{1, 2, \dots, N\}$ . For a given column  $j$  (i.e. criterion  $C_j$ ), it is always possible to compare with  $>$ ,  $<$  and  $=$  operators all the scores  $S_{ij}$  for  $i = 1, 2, \dots, M$  because the scores of each column are expressed in same unit. The construction of IC matrix  $\mathbf{K}$  can be used to search relations between the criteria because the method compares homogeneous data relatively to a same column. Atanassov in [14] prescribes<sup>5</sup> the normalization of the element  $S_{ij}$  of score matrix  $\mathbf{S}$  by taking

$$S_{ij}^{\text{norm}} = (S_{ij} - S_j^{\min}) / (S_j^{\max} - S_j^{\min}) \in [0, 1], \quad (4)$$

where

$$\begin{cases} S_j^{\min} = \min\{S_{1j}, \dots, S_{Mj}\}, \\ S_j^{\max} = \max\{S_{1j}, \dots, S_{Mj}\}. \end{cases} \quad (5)$$

The construction of the  $N \times N$  IC matrix  $\mathbf{K}$  is based on the pairwise comparisons between every two criteria along all evaluated alternatives. More precisely in [14] the degree of agreement between criteria  $C_j$  and  $C_{j'}$   $\mu_{jj'}$ , and their degree of disagreement  $\nu_{jj'}$  are calculated by

$$\mu_{jj'} \triangleq \frac{2K_{jj'}^{\mu}}{M(M-1)}, \quad \text{and} \quad \nu_{jj'} \triangleq \frac{2K_{jj'}^{\nu}}{M(M-1)}, \quad (6)$$

where  $K_{jj'}^{\mu}$  be the number of cases in which the inequalities  $S_{ij} > S_{i'j}$  and  $S_{i'j'} > S_{ij'}$  hold simultaneously, and  $K_{jj'}^{\nu}$  be the number of cases in which the inequalities  $S_{ij} > S_{i'j}$  and  $S_{i'j'} < S_{ij'}$  hold simultaneously.

By construction the IC matrix  $\mathbf{K}$  is always a symmetric matrix. Atanassov provides explicit formulas in [14] for  $K_{jj'}^{\mu}$  and  $K_{jj'}^{\nu}$ , which depend on a particular choice of the signum function. Because of this the results of  $K_{jj'}^{\mu}$  and  $K_{jj'}^{\nu}$  are disputable and that is why some authors [22], [28] propose other methods to calculate  $K_{jj'}^{\mu}$  and  $K_{jj'}^{\nu}$  values for making the Inter-Criteria Analysis.

Once the IC matrix  $\mathbf{K} = [K_{jj'}]$  of intuitionistic fuzzy pairs is calculated one needs to analyze it to decide which criteria  $C_j$  and  $C_{j'}$  are in strong agreement (or positive consonance) reflecting the correlation between  $C_j$  and  $C_{j'}$ , in strong disagreement (or negative consonance) reflecting non correlation between  $C_j$  and  $C_{j'}$ , or in dissonance reflecting the uncertainty situation where nothing can be said about the non correlation or the correlation between  $C_j$  and  $C_{j'}$ .

At the beginning of ICRA development it was not very clear how these intuitionistic fuzzy (IF) pairs  $(\mu_{jj'}, \nu_{jj'})$  had to be used and that is why Atanassova [29], [30] proposed to handle both components of the IF pair. For this, she interpreted pairs  $(\mu_{jj'}, \nu_{jj'})$  as points located in the elementary  $TFU$  triangle, where the point  $T$  of coordinate  $(1, 0)$  represents the maximal positive consonance (i.e. the true consonance), the point  $F$  with coordinate  $(0, 1)$  represents the maximal negative consonance (i.e. the falsity), and the point  $U$  with coordinates  $(0, 0)$  represents the maximal dissonance (i.e. the uncertainty). From this interpretation it becomes quite easy to identify the

top of consonant IF pairs  $(\mu_{jj'}, \nu_{jj'})$  that fall in bottom right corner of  $(TFU)$  triangle limited by vertical line from  $x$ -axis  $x = \alpha$ , and horizontal line from  $y$ -axis  $y = \beta$ , where  $\alpha$  and  $\beta$  are two ad-hoc threshold values in  $[0, 1]$ . The set of consonant IF pairs are then ranked according to their (Euclidean) distance  $d_{C_j C_{j'}}^T$  with respect to  $T$  point of coordinate  $(1, 0)$  defined by

$$d_{C_j C_{j'}}^T = d((1, 0), (\mu_{jj'}, \nu_{jj'})) = \sqrt{(1 - \mu_{jj'})^2 + \nu_{jj'}^2}. \quad (7)$$

It is worth noting that  $\mu_{jj'}$  and  $\nu_{jj'}$  values are in fact linked with belief function through the following formulas

$$Bel_{jj'}(\theta) = \mu_{jj'} \quad (8)$$

$$Pl_{jj'}(\theta) = 1 - \nu_{jj'}, \quad (9)$$

$$U_{jj'}(\theta) = Pl_{jj'}(\theta) - Bel_{jj'}(\theta) = 1 - \nu_{jj'} - \mu_{jj'}, \quad (10)$$

where  $\theta$  means: the criteria  $C_j$  and  $C_{j'}$  are totally positively consonant (i.e. totally correlated), whereas  $\bar{\theta}$  means: the criteria  $C_j$  and  $C_{j'}$  are totally negatively consonant (uncorrelated). The FoD is defined as  $\Theta \triangleq \{\theta, \bar{\theta}\}$ .  $U_{jj'}(\theta)$  represents the dissonance (the uncertainty about the correlation) of the criteria  $C_j$  and  $C_{j'}$ . From this, one can easily define any BBA  $m_{jj'}(\theta)$ ,  $m_{jj'}(\bar{\theta})$  and  $m_{jj'}(\theta \cup \bar{\theta})$  of  $2^\Theta$  by taking

$$m_{jj'}(\theta) = \mu_{jj'}, \quad (11)$$

$$m_{jj'}(\bar{\theta}) = \nu_{jj'}, \quad (12)$$

$$m_{jj'}(\theta \cup \bar{\theta}) = 1 - \mu_{jj'} - \nu_{jj'}. \quad (13)$$

**Remark 1:** The construction of the Inter-Criteria Matrix  $\mathbf{K}$  is not unique and depends on the choice of algorithm of construction of  $\mu_{jj'}$  and  $\nu_{jj'}$  (and the choice of the signum function) as reported in [28]. This can yield different ICRA results in general.

**Remark 2:** The construction of  $\mu_{jj'}$  and  $\nu_{jj'}$  appears to be only a crude approximation of true values because they are only based on counting the valid " $>$ " or " $<$ " inequalities. In fact, their calculations do not exploit how bigger and how smaller the scores values are in each comparison done. So it yields a lack of precision on estimation of  $\mu_{jj'}$  and  $\nu_{jj'}$  values.

ICRA can be very useful for verification of algorithm correctness. When the optimization problem have a lot of constraints with ICRA we can find if some of the constrain is subconstrain of some other and to exclude it. With the help of ICRA we can divide constraints to two or more groups, more sensitive and less sensitive and to solve problem first according more sensitive constraints and later to less sensitive ones. To circumvent the aforementioned drawbacks, we present succinctly in the next section a new ICRA approach based on belief functions which is presented in more details with examples in [2].

## V. A NEW ICRA METHOD BASED ON BELIEF FUNCTIONS

The new Belief Function based ICRA method, called BF-ICRA for short, presented in this section improves Atanassov's ICRA. It provides a more precise construction of  $\mu_{jj'}$  and  $\nu_{jj'}$  values because it exploits all available information included in the score matrix.

<sup>5</sup>Although this normalization is not very necessary in fact for ICRA making.

BF-ICrA starts with the construction of an  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  from the score matrix  $\mathbf{S} = [S_{ij}]$ . The elements  $m_{ij}$  of the BBA matrix  $\mathbf{M}$  are obtained as follows - see [31] for details and justification.

$$m_{ij}(A_i) = Bel_{ij}(A_i), \quad (14)$$

$$m_{ij}(\bar{A}_i) = Bel_{ij}(\bar{A}_i) = 1 - Pl_{ij}(A_i), \quad (15)$$

$$m_{ij}(A_i \cup \bar{A}_i) = Pl_{ij}(A_i) - Bel_{ij}(A_i), \quad (16)$$

where<sup>6</sup>

$$Bel_{ij}(A_i) \triangleq Sup_j(A_i)/A_{\max}^j, \quad (17)$$

$$Bel_{ij}(\bar{A}_i) \triangleq Inf_j(A_i)/A_{\min}^j, \quad (18)$$

with

$$Sup_j(A_i) \triangleq \sum_{k \in \{1, \dots, M\} | S_{kj} \leq S_{ij}} |S_{ij} - S_{kj}|, \quad (19)$$

$$Inf_j(A_i) \triangleq - \sum_{k \in \{1, \dots, M\} | S_{kj} \geq S_{ij}} |S_{ij} - S_{kj}|, \quad (20)$$

and

$$A_{\max}^j \triangleq \max_i Sup_j(A_i), \quad (21)$$

$$A_{\min}^j \triangleq \min_i Inf_j(A_i). \quad (22)$$

For another criterion  $C_{j'}$  and the  $j'$ -th column of the score matrix we will obtain another set of BBA values  $m_{ij'}(\cdot)$ . Applying this method for each column of the score matrix we are able to compute the BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  whose each component is in fact a triplet  $(m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i))$  of BBA values in  $[0, 1]$  such that  $m_{ij}(A_i) + m_{ij}(\bar{A}_i) + m_{ij}(A_i \cup \bar{A}_i) = 1$  for all  $i = 1, \dots, M$  and  $j = 1, \dots, N$ .

The next step of BF-ICrA approach is the construction of the  $N \times N$  Inter-Criteria Matrix  $\mathbf{K} = [K_{jj'}]$  from  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  where elements  $K_{jj'}$  corresponds to the BBA  $(m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  about positive consonance  $\theta$ , negative consonance  $\bar{\theta}$  and uncertainty between criteria  $C_j$  and  $C_{j'}$  respectively. The construction of the triplet  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  is based on two steps:

- **Step 1 (BBA construction):** Getting  $m_{jj'}^i(\cdot)$ .

For each alternative  $A_i$  for  $i = 1, \dots, M$ , we first compute the BBA  $(m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  for any two criteria  $j, j' \in \{1, 2, \dots, N\}$ . For this, we consider two sources of evidences (SoE) indexed by  $j$  and  $j'$  providing the BBA  $m_{ij}$  and  $m_{ij'}$  defined on the simple FoD  $\{A_i, \bar{A}_i\}$  and denoted  $m_{ij} = [m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i)]$  and  $m_{ij'} = [m_{ij'}(A_i), m_{ij'}(\bar{A}_i), m_{ij'}(A_i \cup \bar{A}_i)]$ . We also denote  $\Theta = \{\theta, \bar{\theta}\}$  the FoD about the relative state of the two SoE, where  $\theta$  means that the two SoE agree,  $\bar{\theta}$  means that they disagree and  $\theta \cup \bar{\theta}$  means that we don't know. Hence, two SoE are in total agreement if both commit their maximum belief mass to the same element  $A_i$  or to

the same element  $\bar{A}_i$ . Similarly, two SoE are in total disagreement if each one commits its maximum mass of belief to one element and the other to its opposite, that is if one has  $m_{ij}(A_i) = 1$  and  $m_{ij'}(\bar{A}_i) = 1$ , or if  $m_{ij}(\bar{A}_i) = 1$  and  $m_{ij'}(A_i) = 1$ . Based on this very simple and natural principle, one can now compute the belief masses as follows:

$$m_{jj'}^i(\theta) = m_{ij}(A_i)m_{ij'}(A_i) + m_{ij}(\bar{A}_i)m_{ij'}(\bar{A}_i), \quad (23)$$

$$m_{jj'}^i(\bar{\theta}) = m_{ij}(A_i)m_{ij'}(\bar{A}_i) + m_{ij}(\bar{A}_i)m_{ij'}(A_i), \quad (24)$$

$$m_{jj'}^i(\theta \cup \bar{\theta}) = 1 - m_{jj'}^i(\theta) - m_{jj'}^i(\bar{\theta}). \quad (25)$$

$m_{jj'}^i(\theta)$  represents the degree of agreement between the BBA  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$  for the alternative  $A_i$ ,  $m_{jj'}^i(\bar{\theta})$  represents the degree of disagreement of the two BBAs and  $m_{jj'}^i(\theta \cup \bar{\theta})$  the level of uncertainty (i.e. how much we don't know if they agree or disagree). By construction  $m_{jj'}^i(\cdot) = m_{jj'}^i(\cdot), m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}) \in [0, 1]$  and  $m_{jj'}^i(\theta) + m_{jj'}^i(\bar{\theta}) + m_{jj'}^i(\theta \cup \bar{\theta}) = 1$ . This BBA modeling permits to build a set of  $M$  symmetrical Inter-Criteria Belief Matrices (ICBM)  $\mathbf{K}^i = [K_{jj'}^i]$  of dimension  $N \times N$  relative to each alternative  $A_i$  whose components  $K_{jj'}^i$  correspond to the triplet of BBA values  $m_{jj'}^i = (m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  modeling the belief of agreement and of disagreement between  $C_j$  and  $C_{j'}$  based on  $A_i$ .

- **Step 2 (fusion):** Getting  $\mathbf{m}_{jj'}(\cdot)$ .

In this step, one needs to combine the BBAs  $\mathbf{m}_{jj'}^i(\cdot)$  for  $i = 1, \dots, M$  altogether to get the component  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  of the Inter-Criteria Belief matrix (ICBM)  $\mathbf{K} = [K_{jj'}]$ . For this, we recommend to use the PCR6 fusion rule [10] (Vol. 3) because of known deficiencies of Dempster's rule. Because of computational complexity of PCR6 fusion rule when  $M$  becomes large, one may prefer to approximate the fusion result by using the simple averaging rule. Simple Matlab™code for PCR6 rule can be found in [32] for convenience.

The computational complexity of BF-ICrA is of course higher than the complexity of ICrA because it makes a more precise evaluation of local and global inter-criteria belief matrices with respect to IF inter-criteria matrices of ICrA. The overall reduction of the computational burden of the original MCDM problem thanks to BF-ICrA depends highly on the problem under concern, the complexity and cost to evaluate each criteria involved in it, as well as the number of redundant criteria identified by BF-ICrA method.

Once the global Inter-Criteria Belief Matrix (ICBM)  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$  is calculated, we can identify the criteria that are in strong agreement, in strong disagreement, and those on which we are uncertain. For identifying the criteria that are in strong agreement, we evaluate the distance of each component of  $K_{jj'}$  with the BBA representing the best agreement state and characterized

<sup>6</sup>assuming that  $A_{\max}^j \neq 0$  and  $A_{\min}^j \neq 0$ . If  $A_{\max}^j = 0$  then  $Bel_{ij}(A_i) = 0$ , and if  $A_{\min}^j = 0$  then  $Pl_{ij}(A_i) = 1$ .

by the specific BBA<sup>7</sup>  $m_T(\theta) = 1$ . From a similar approach we can also identify, if we want, the criteria that are in very strong disagreement using the distance of  $m_{jj'}(\cdot)$  with respect to the BBA representing the best disagreement state characterized by the specific BBA  $m_F(\bar{\theta}) = 1$ . We use the  $d_{BI}(\cdot, \cdot)$  distance presented in [33] for measuring the distance  $d(m_1, m_2)$  between the two BBAs<sup>8</sup>  $m_1(\cdot)$  and  $m_2(\cdot)$  over the same FoD. It is defined by

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in \mathcal{E}^\Theta} d_W^2(BI_1(X), BI_2(X))}, \quad (26)$$

where the *Belief-Intervals* are defined by  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)]$  and computed from  $m_1(\cdot)$  and  $m_2(\cdot)$  thanks to formula (3).  $d_W(BI_1(X), BI_2(X))$  is Wasserstein's distance between intervals calculated by

$$d_W([a_1, b_1], [a_2, b_2]) = \sqrt{\left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2},$$

and  $N_c = 1/2^{|\Theta|-1}$  is a normalization factor to get  $d_{BI}(m_1, m_2) \in [0, 1]$ .

## VI. APPLICATION OF BF-ICRA TO GSP

In this section, we analyze the experimental results obtained using MMAS algorithm described in the previous section. For this, we use real data from Malta and Seychelles GPS networks composed of 38 sessions and 71 sessions respectively denoted GSP1 and GSP2. We use also 6 larger test problems range from 100 to 443 sessions denoted GSP3, ..., GSP8. The results are obtained by performing 30 independent runs, for every experiment. The details of our MMAS implementation are given in [1]. So in our GSP example we consider 8 GSP criteria  $C_i = GSP_i$ ,  $i = 1, \dots, 8$  and six average costs as results  $A_1, \dots, A_6$ , where  $A_1$  is the cost average for the first 5 runs,  $A_2$  the cost average for the first 10 runs,  $A_3$  for the first 15 runs, ... and finally  $A_6$  for all the 30 runs. Table I shows the values of averaged costs obtained for this problem. It corresponds to the transpose of the score matrix  $\mathbf{S}$ .

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$
$C_1 = GSP1$	899.00	898.00	898.33	898.50	899.40	899.50
$C_2 = GSP2$	916.40	915.60	922.47	924.80	924.72	922.07
$C_3 = GSP3$	41336.40	41052.40	40991.93	40935.90	40832.20	40910.60
$C_4 = GSP4$	3244.80	3303.30	3327.00	3344.55	3345.60	3341.93
$C_5 = GSP5$	1656.20	1660.80	1663.93	1664.95	1666.96	1665.90
$C_6 = GSP6$	1673.60	1683.50	1690.73	1688.75	1690.24	1692.67
$C_7 = GSP7$	3420.00	3430.70	3433.13	3426.85	3429.44	3428.57
$C_8 = GSP8$	3758.20	3755.70	3758.73	3760.50	3760.80	3765.80

Table I

TRANSPOSE OF THE SCORE MATRIX  $\mathbf{S} = [S_{ij}]$  OF GSP PROBLEM.

<sup>7</sup>We use the index  $T$  in the notation  $m_T(\cdot)$  to refer that the agreement is true, and  $F$  in  $m_F(\cdot)$  to specify that the agreement is false.

<sup>8</sup>Here we will take  $m_1(\cdot) = m_{jj'}(\cdot)$  and  $m_2(\cdot) = m_T(\cdot)$ , or  $m_2(\cdot) = m_F(\cdot)$

Hence in this problem  $M = 6$  and  $N = 8$ , and  $\mathbf{S} = [S_{ij}]$  is a  $6 \times 8$  score matrix. Based on classical ICRA approach, one gets the following IC matrices<sup>9</sup>

$$\mathbf{K}^\mu = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 1 & 0.60 & 0.27 & 0.67 & 0.73 & 0.67 & 0.33 & 0.87 \\ 0.60 & 1 & 0.27 & 0.80 & 0.73 & 0.53 & 0.47 & 0.73 \\ 0.27 & 0.27 & 1 & 0.07 & 0 & 0.20 & 0.40 & 0.13 \\ 0.67 & 0.80 & 0.07 & 1 & 0.93 & 0.73 & 0.53 & 0.80 \\ 0.73 & 0.73 & 0 & 0.93 & 1 & 0.80 & 0.60 & 0.87 \\ 0.67 & 0.53 & 0.20 & 0.73 & 0.80 & 1 & 0.67 & 0.80 \\ 0.33 & 0.47 & 0.40 & 0.53 & 0.60 & 0.67 & 1 & 0.47 \\ 0.87 & 0.73 & 0.13 & 0.80 & 0.87 & 0.80 & 0.47 & 1 \end{bmatrix} \end{matrix}$$

$$\mathbf{K}^\nu = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 0 & 0.40 & 0.73 & 0.33 & 0.27 & 0.33 & 0.67 & 0.13 \\ 0.40 & 0 & 0.73 & 0.20 & 0.27 & 0.47 & 0.53 & 0.27 \\ 0.73 & 0.73 & 0 & 0.93 & 1 & 0.80 & 0.60 & 0.87 \\ 0.33 & 0.20 & 0.93 & 0 & 0.07 & 0.27 & 0.47 & 0.20 \\ 0.27 & 0.27 & 1 & 0.07 & 0 & 0.20 & 0.40 & 0.13 \\ 0.33 & 0.47 & 0.80 & 0.27 & 0.20 & 0 & 0.33 & 0.20 \\ 0.67 & 0.53 & 0.60 & 0.47 & 0.40 & 0.33 & 0 & 0.53 \\ 0.13 & 0.27 & 0.87 & 0.20 & 0.13 & 0.20 & 0.53 & 0 \end{bmatrix} \end{matrix}$$

The element  $K_{jj'}^\mu$  of matrix  $\mathbf{K}^\mu$  expresses the *degree of agreement* between criteria  $C_j = GSP_j$  and  $C_{j'} = GSP_{j'}$ , whereas the element  $K_{jj'}^\nu$  of matrix  $\mathbf{K}^\nu$  expresses the *degree of disagreement* between  $C_j = GSP_j$  and  $C_{j'} = GSP_{j'}$ . Based on these results, one sees that ACO algorithm performs similarly for  $GSP_2$ ,  $GSP_4$ ,  $GSP_5$  and  $GSP_8$  because they are all in high agreement. Indeed  $\mu_{jj'}$  values for  $j, j' \in \{2, 4, 5, 8\}$  are quite high (greater than 70%). They are GPS networks with different numbers of sessions, but may have a similar structure, therefore, the value of agreement is high. For other networks, we can conclude that they have very different structure. What is worth noting is that there appears also a strong agreement of GSP1 with GSP8 because  $\mu_{18} = 0.87$ . But because GSP8 is also in strong agreement with GSP2, GSP4, GSP5 and with GSP1 it is logically expected that GSP1 should be also in agreement with GSP2, GSP4, GSP5, which is unfortunately not the case based on this classical ICRA. This example points out some inconsistency of ICRA result because of the too crude method of estimation of the degree of agreement and disagreement between criteria based on IFS.

Now if we consider the same example with the same score matrix  $\mathbf{S}$  (built from Table I), we obtain the following IC Belief matrices<sup>10</sup>

$$\mathbf{K}(\theta) = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 0.9098 & 0.6732 & 0.1791 & 0.5968 & 0.6106 & 0.5620 & 0.1659 & 0.7789 \\ 0.6732 & 0.9546 & 0.0364 & 0.8983 & 0.8783 & 0.8341 & 0.5532 & 0.7016 \\ 0.1791 & 0.0364 & 0.8722 & 0.0172 & 0.0154 & 0.0178 & 0.0366 & 0.1137 \\ 0.5968 & 0.8983 & 0.0172 & 0.9552 & 0.9146 & 0.9163 & 0.7395 & 0.6092 \\ 0.6106 & 0.8783 & 0.0154 & 0.9146 & 0.8917 & 0.8778 & 0.6922 & 0.6315 \\ 0.5620 & 0.8341 & 0.0178 & 0.9163 & 0.8778 & 0.9060 & 0.7630 & 0.6441 \\ 0.1659 & 0.5532 & 0.0366 & 0.7395 & 0.6922 & 0.7630 & 0.8587 & 0.2484 \\ 0.7789 & 0.7016 & 0.1137 & 0.6092 & 0.6315 & 0.6441 & 0.2484 & 0.8508 \end{bmatrix} \end{matrix}$$

$$\mathbf{K}(\bar{\theta}) = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 0.0207 & 0.1941 & 0.5385 & 0.2578 & 0.1757 & 0.2117 & 0.5335 & 0.0399 \\ 0.1941 & 0.0166 & 0.8323 & 0.0486 & 0.0298 & 0.0513 & 0.1808 & 0.0682 \\ 0.5385 & 0.8323 & 0.0117 & 0.9002 & 0.8754 & 0.8548 & 0.7062 & 0.5486 \\ 0.2578 & 0.0486 & 0.9002 & 0.0187 & 0.0216 & 0.0204 & 0.0606 & 0.1193 \\ 0.1757 & 0.0298 & 0.8754 & 0.0216 & 0.0170 & 0.0201 & 0.0558 & 0.0832 \\ 0.2117 & 0.0513 & 0.8548 & 0.0204 & 0.0201 & 0.0154 & 0.0390 & 0.0726 \\ 0.5335 & 0.1808 & 0.7062 & 0.0606 & 0.0558 & 0.0390 & 0.0110 & 0.3495 \\ 0.0399 & 0.0682 & 0.5486 & 0.1193 & 0.0832 & 0.0726 & 0.3495 & 0.0100 \end{bmatrix} \end{matrix}$$

From ICBM  $\mathbf{K}(\theta)$  and  $\mathbf{K}(\bar{\theta})$  we compute the matrix  $\mathbf{D}(\theta)$  of distance of  $m_{jj'}(\cdot)$  to the full agreement state with BBA

<sup>9</sup>For presentation convenience and due to typesetting column width, we decompose et present the IC matrix  $\mathbf{K} = [K_{jj'} = (K_{jj'}^\mu, K_{jj'}^\nu)]$  into two distinct matrices  $\mathbf{K}^\mu = [K_{jj'}^\mu]$  and  $\mathbf{K}^\nu = [K_{jj'}^\nu]$ .

<sup>10</sup>For presentation convenience, the ICBM  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$  is decomposed into three matrices  $\mathbf{K}(\theta) = [K_{jj'}^\theta = m_{jj'}(\theta)]$ ,  $\mathbf{K}(\bar{\theta}) = [K_{jj'}^{\bar{\theta}} = m_{jj'}(\bar{\theta})]$  and  $\mathbf{K}(\theta \cup \bar{\theta}) = [K_{jj'}^{\theta \cup \bar{\theta}} = 1 - m_{jj'}(\theta) - m_{jj'}(\bar{\theta})]$ .

$m_F(\theta) = 1$  based on  $d_{BI}(\cdot)$  distance. We get the following distances to full agreement

$$\mathbf{D}(\theta) = [D_{jj'} = d_{BI}(m_{jj'}, m_T)]$$

	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$
$C_1$	0.0590	0.2633	0.6845	0.3331	0.2892	0.3314	0.6893	0.1406
$C_2$	0.2633	0.0321	0.8987	<b>0.0767</b>	<b>0.0803</b>	0.1135	0.3230	0.1950
$C_3$	0.6845	0.8987	0.0774	0.9418	0.9306	0.9192	0.8381	0.7241
$C_4$	0.3331	<b>0.0767</b>	0.9418	0.0326	<b>0.0566</b>	0.0552	0.1706	0.2668
$C_5$	0.2892	<b>0.0803</b>	0.9306	<b>0.0566</b>	0.0679	0.0770	0.1958	0.2404
$C_6$	0.3314	0.1135	0.9192	0.0552	0.0770	0.0592	0.1494	0.2293
$C_7$	0.6893	0.3230	0.8381	0.1706	0.1958	0.1494	0.0849	0.5626
$C_8$	0.1406	0.1950	0.7241	0.2668	0.2404	0.2293	0.5626	0.0892

The element  $D_{jj'}$  represents the agreement distance between  $C_j$  and  $C_{j'}$ , the lower the better. From the values of elements of  $\mathbf{D}(\theta)$  matrix one sees clearly that ACO performs similarly for GSP2, GSP4 and GSP5 because distances  $D_{24}$ ,  $D_{25}$ , and  $D_{45}$  are very small. Also we see that GSP6 is also in good agreement with GSP4 and GSP5 but is relatively less in agreement with GSP2 because  $D_{26} = 0.1135$ . As we see there is no inconsistency in this new BF-ICrA method with respect to what provides classical ICrA because with BF-ICrA we have a much better and precise estimation of degrees of agreement and disagreement between criteria for making the analysis thanks to a proper belief functions modeling.

## VII. CONCLUSION

The GPS surveying problem and a new InterCriteria Analysis based on belief functions were addressed in this paper to overcome the potential inconsistencies of the results generated by the classical ICrA method. This technique proposes a more precise and refined method for estimating the degree of agreement and disagreement between criteria which use the whole information available in the data. Instances containing from 38 to 443 sessions have been solved using MMAS algorithm and we did compare the performance of ACO algorithms applied to eight GPS networks. Our results shows that ACO can provide fast near-optimal solution for observing GPS networks, and could help to improve the services based on GPS networks. From this new Inter-Criteria Analysis we are able to identify some relations and dependences between the considered eight GSPs and MMAS algorithm performance.

## ACKNOWLEDGEMENTS

This paper is partially supported by the National Scientific Fund of Bulgaria under grant DFNI DN12/5 and by Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program and co-financed by the European Union through the European structural and Investment funds.

## REFERENCES

- [1] S. Fidanova, V. Atanassova, O. Roeva, *Ant Colony Optimization Application to GPS Surveying Problems: InterCriteria Analysis*, in Atanassov K. et al. (Eds) Uncertainty and Imprecision in Decision Making and Decision Support: Cross-Fertilization, New Models and Applications, IWIFSGN 2016. Advances in Intelligent Systems and Computing, Vol. 559. Springer, Cham, 2018.
- [2] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Simplification of Multi-Criteria Decision-Making Using Inter-Criteria Analysis and Belief Functions*, in Proc. of Fusion 2019 Int. Conf, Ottawa, Canada, July 2–5, 2019.
- [3] T. Stutzle, H.H. Hoos, *MAX-MIN Ant System*, in Dorigo, M., Stutzle, T., Di Caro, G. (Eds.) Future Generation Computer Systems, Vol. 16, pp. 889–914, 2000.
- [4] M. Dorigo, L.M. Gambardella, *Ant colony system: a cooperative learning approach to the traveling salesman problem*, IEEE Trans. Evol. Comput., Vol. 1, pp. 5–66, 1997.
- [5] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Press, 1976.
- [6] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., Vol. 38, pp. 325–339, 1967.
- [7] J. Dezert, P. Wang, A. Tchamova, *On The Validity of Dempster-Shafer Theory*, in Proc. of Fusion 2012, Singapore, July 2012.
- [8] A. Tchamova, J. Dezert, *On the Behavior of Dempster's Rule of Combination and the Foundations of Dempster-Shafer Theory*, IEEE IS-2012, Sofia, Bulgaria, Sept. 6–8, 2012.
- [9] J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Conditional Belief Functions*, International Journal of Intelligent Systems, pp. 1-27, 2018.
- [10] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion - Collected works*, American Research Press, Vol. 1–4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [11] K. Atanassov, *On Intuitionistic Fuzzy Sets Theory*, Springer, 2012.
- [12] K. Atanassov, D. Mavrov, V. Atanassova, *InterCriteria decision making: a new approach for multicriteria decision making, based on index matrices and intuitionistic fuzzy sets*, Issues IFSs GNs 11, pp. 1–8, 2014.
- [13] K. Atanassov, V. Atanassova, G. Gluhchev, *InterCriteria Analysis: Ideas and problems*, Notes on IFS, Vol. 21, No. 1, pp. 81–88, 2015.
- [14] K. Atanassov et al., *InterCriteria analysis over normalized data*, Intelligent Systems (IS), 2016 IEEE 8th Int. Conference on, IEEE, 2016.
- [15] K. Atanassov et al., *An approach to a constructive simplification of multiagent multicriteria decision making problems via interCriteria analysis*, C.R. de l'Acad. Bulgare des Sci., Vol. 70, No. 8, 2017.
- [16] S. Krumova, et al., *InterCriteria analysis of calorimetric data of blood serum proteome*, in Bioch. et Biophys. Acta, Gen. Subjects, 1861, 2017.
- [17] B. Zaharieva et al., *InterCriteria decision making approach for Beherev's disease analysis*, in Int. J. of Bioautom, Vol. 22(2), 2018.
- [18] T. Pencheva et al., *InterCriteria Analysis of Genetic Algorithm Parameters in Parameter Identification*, Notes on IFS, Vol. 21(2), 2015.
- [19] S. Sotirov, et al., *Application of the Intuitionistic Fuzzy InterCriteria Analysis Method to a Neural Network Preprocessing Procedure*, in Proc of 9th EUSFLAT, pp. 1559–1564, Gijon, Spain, June 30–July 7th, 2015.
- [20] M. Angelova, O. Roeva, T. Pencheva, *InterCriteria analysis of crossover and mutation rates relations in simple genetic algorithm*, in Proc. of Conf on Computer Sci. and Inf. Syst, Vol. 5, pp. 419–424, 2015.
- [21] O. Roeva, S. Fidanova, M. Paprzycki, *InterCriteria analysis of ACO and GA hybrid algorithms*, Stud. Comput. Intell., Vol. 610, pp. 107–126, 2016.
- [22] O. Roeva et al., *InterCriteria Analysis of ACO Performance for Workforce Planning Problem*, Studies in Comp. Intell., Vol. 795, Springer 2019.
- [23] V. Atanassova et al., *Discussion on the threshold values in the Inter-Criteria decision making approach*, Notes on Intuitionistic Fuzzy Sets, Vol. 20(2), pp. 94–99, 2014.
- [24] L. Doukovska, V. Atanassova, *InterCriteria Analysis approach in radar detection threshold analysis*, Notes on IFS, Vol. 21(4), 2015.
- [25] V. Bureva et al., *Application of the InterCriteria decision making method to Bulgarian universities ranking*, Int. Workshop on IFSs, Bulgaria, 2015.
- [26] V. Bureva, E. Sotirova, H. Panayotov, *The InterCriteria decision making method to Bulgarian university ranking system*, Annual of Informatics Section, Vol. 8, pp. 54–70, 2015–2016.
- [27] M. Krawczak et al., *Application of the InterCriteria Decision Making Method to Universities Ranking*, Adv. in Intell. Syst. and Comp., Vol. 401, pp. 365–372, Springer, 2016.
- [28] N. Ikonov, P. Vassilev, O. Roeva, *ICrAData software for InterCriteria analysis*, Int. J. BioAutomation, Vol. 22, No. 2, 2018.
- [29] V. Atanassova, *Interpretation in the Intuitionistic Fuzzy Triangle of the Results Obtained by the InterCriteria Analysis*, Proc. of 16th World Congr. of IFSA, Atlantis Press, 2015.
- [30] V. Atanassova et al., *Traversing and ranking of elements of an intuitionistic fuzzy set in the intuitionistic fuzzy interpretation triangle*, Adv. in Intell. Syst. and Comp., Vol. 401, pp. 161–174, 2016.

- [31] J. Dezert, D Han, H. Yin, *A New Belief Function Based Approach for Multi-Criteria Decision-Making Support*, in Proc. of Fusion 2016 Conf., Heidelberg, Germany, July 2016
- [32] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, Proc.of Fusion Conf 2010, Edinburgh, Scotland, UK, 26–29 July 2010.
- [33] D. Han, J. Dezert, Y. Yang, *New Distance Measures of Evidence based on Belief Intervals*, in Proc. of Belief 2014, Oxford, UK, Sept. 2014.



# Simplification of Multi-Criteria Decision-Making Using Inter-Criteria Analysis and Belief Functions

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>, Jean-Marc Tacnet<sup>d</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

<sup>d</sup>University Grenoble Alpes, Irstea, ETNA, Grenoble, France.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn, jean-marc.tacnet@irstea.fr

Originally published as: J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Simplification of Multi-Criteria Decision-Making Using Inter-Criteria Analysis and Belief Functions*, in Proc. of Fusion 2019 Int. Conf., Ottawa, Canada, July 2–5, 2019, and reprinted with permission.

**Abstract**—In this paper we propose a new Belief Function-based Inter-Criteria Analysis (BF-ICrA) for the assessment of the degree of redundancy of criteria involved in a multicriteria decision making (MCDM) problem. This BF-ICrA method allows to simplify the original MCDM problem by withdrawing all redundant criteria and thus diminish the complexity of MCDM problem. This is of prime importance for solving large MCDM problems whose solution requires the fusion of many belief functions. We provide simple examples to show how this new BF-ICrA works.

**Keywords:** Inter-Criteria Analysis, ICrA-BF, MultiCriteria Decision Making, MCDM, belief functions, information fusion.

## I. INTRODUCTION

In a Multi-Criteria Decision-Making (MCDM) problem we consider a set of alternatives (or objects)  $\mathbf{A} \triangleq \{A_1, A_2, \dots, A_M\}$  ( $M > 2$ ), and a set of criteria  $\mathbf{C} \triangleq \{C_1, C_2, \dots, C_N\}$  ( $N \geq 1$ ). We search for the best alternative  $A^*$  given the available information expressed by a  $M \times N$  score matrix (also called benefit or payoff matrix)  $\mathbf{S} \triangleq [S_{ij} = C_j(A_i)]$ , and (eventually) the importance factor  $w_j \in [0, 1]$  of each criterion  $C_j$  with  $\sum_{j=1}^N w_j = 1$ . The set of normalized weighting factors is denoted by  $\mathbf{w} \triangleq \{w_1, w_2, \dots, w_N\}$ . Depending on the context of the MCDM problem, the score  $S_{ij}$  of each alternative  $A_i$  with respect to each criteria  $C_j$  can be interpreted either as a cost (i.e. an expense), or as a reward (i.e. a benefit). By convention and without loss of generality<sup>1</sup> we will always interpret the score as a reward having monotonically increasing preference. Thus, the best alternative  $A_j^*$  for a given criteria  $C_j$  will be the one providing the highest reward/benefit.

The MCDM problem is not easy to solve because the scores are usually expressed in different (physical) units and different scales. This necessitates a choice of score/data normalization yielding rank reversal problems [1], [2]. Usually there is no same best alternative choice  $A^*$  for all criteria, so a compromise must be established to provide a reasonable

and acceptable solution of the MCDM problem for decision-making support.

Many MCDM methods exist, see references in [3]. Most popular methods are AHP<sup>2</sup> [4], ELECTRE<sup>3</sup> [5], TOPSIS<sup>4</sup> [6], [7]. In 2016 and 2017, we did develop BF-TOPSIS methods [3], [8] based on Belief Functions (BF) to improve the original TOPSIS approach to avoid data normalization and to deal also with imprecise score values as well. It appears however that the complexity of these new BF-TOPSIS methods can become a bottleneck for their use in large MCDM problems because of the fusion step of basic belief assignments required for the implementation of the BF-TOPSIS. That is why a simplification of the MCDM problem (if possible) is very welcome in order to save computational time and resources. This is the motivation of the present work.

For this aim we propose a new Inter-Criteria Analysis (ICrA) based on belief functions for identifying and estimating the possible degree of agreement (i.e. redundancy) between some criteria driven from the data (score values). This permits to remove all redundant criteria of the original MCDM problem and thus solving a simplified (almost) equivalent MCDM problem faster and at lower computational cost. ICrA has been developed originally by Atanassov et al. [9]–[11] based on Intuitionistic Fuzzy Sets [12], and it has been applied in different fields like medicine [13]–[15], optimization [16]–[20], workforce planning [21], competitiveness analysis [22], radar detection [23], ranking [24]–[27], etc. In this paper we improve ICrA approach thanks to belief functions introduced by Shafer in [28] from original Dempster's works [29]. We will refer it as BF-ICrA method in the sequel.

After a short presentation of basics of belief functions in section II, we present Atanassov's ICrA method in section III and discuss its limitations. In Section IV we present the new BF-ICrA approach based on a new construction of Basic Belief Assignment (BBA) matrix from the score matrix and a new establishment of Inter-Criteria belief matrix. In section

<sup>2</sup>Analytic Hierarchy Process

<sup>3</sup>ELimination Et Choix Taduisant la REalité

<sup>4</sup>Technique for Order Preference by Similarity to Ideal Solution

<sup>1</sup>because it suffices to multiply the scores values by  $-1$  to reverse the preference ordering.

V a method of simplification of MCDM using BF-ICrA is proposed. Examples are given in VI with concluding remarks in Section VII.

## II. BASICS OF THE THEORY OF BELIEF FUNCTIONS

To follow classical notations of the theory of belief functions, also called Dempster-Shafer Theory (DST) [28], we assume that the answer (i.e. the solution, or the decision to take) of the problem under concern belongs to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are exclusive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A BBA (or mass function) associated with a given source of evidence is defined [28] as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are usually interpreted respectively as lower and upper bounds of unknown (possibly subjective) probability measure [29]. They are defined by<sup>5</sup>

$$\text{Bel}(A) \triangleq \sum_{B \subseteq A, B \in 2^\Theta} m(B), \quad \text{and} \quad \text{Pl}(A) \triangleq 1 - \text{Bel}(\bar{A}). \quad (1)$$

If  $m(A) > 0$ ,  $A$  is called a focal element of  $m(\cdot)$ . When all focal elements are singletons then  $m(\cdot)$  is called a *Bayesian BBA* [28] and its corresponding  $\text{Bel}(\cdot)$  function is homogeneous to a probability measure. The vacuous BBA, or VBBA for short, representing a totally ignorant source is defined as  $m_v(\Theta) = 1$ . The main challenge of the decision-maker consists to combine efficiently the possible multiple BBAs  $m_s(\cdot)$  given by  $s > 1$  distinct sources of evidence to obtain a global (combined) BBA, and to make a final decision from it. Historically the combination of BBAs is accomplished by Dempster's rule proposed by Shafer in DST. Because Dempster's rule presents several serious problems (insensitivity to the level of conflict between sources in some cases, inconsistency with bounds of conditional probabilities when used for belief conditioning, dictatorial behavior, counter-intuitive results), many fusion rules have been proposed in the literature as alternative to Dempster's rule, see [30], Vol. 2 for a detailed list of fusion rules. We will not detail here all the possible combination rules but just mention that the Proportional Conflict Redistribution rule no. 6 (PCR6) proposed by Martin and Osswald in [30] (Vol. 3) is one of the most serious alternative rule for BBA combination available so far.

## III. ATANASSOV'S INTER-CRITERIA ANALYSIS (ICrA)

Atanassov's Inter-Criteria Analysis (ICrA) approach is based on a  $M \times N$  score matrix<sup>6</sup>  $\mathbf{S} \triangleq [S_{ij} = C_j(A_i), i = 1, \dots, M, j = 1, \dots, N]$ , and intuitionistic fuzzy pairs [12] including two membership functions  $\mu(\cdot)$  and  $\nu(\cdot)$ . Mathematically, an intuitionistic fuzzy set (IFS)  $A$  is denoted by  $A \triangleq \{(x, \mu_A(x), \nu_A(x)) | x \in E\}$ , where  $E$  is the set of possible values of  $x$ ,  $\mu_A(x) \in [0, 1]$  defines the membership of

$x$  to the set  $A$ , and  $\nu_A(x) \in [0, 1]$  defines the non-membership of  $x$  to the set  $A$ , with the restriction  $0 \leq \mu_A(x) + \nu_A(x) \leq 1$ . The ICrA method consists to build an  $N \times N$  Inter-Criteria (IC) matrix from the score matrix  $\mathbf{S}$ . The elements of the IC matrix consist of all intuitionistic fuzzy pairs  $(\mu_{jj'}, \nu_{jj'})$  whose components express respectively the degree of agreement and the degree of disagreement between criteria  $C_j$  and  $C_{j'}$  for  $j, j' \in \{1, 2, \dots, N\}$ . For a given column  $j$  (i.e. criterion  $C_j$ ), it is always possible to compare with  $>$ ,  $<$  and  $=$  operators all the scores  $S_{ij}$  for  $i = 1, 2, \dots, M$  because the scores of each column are expressed in same unit. The construction of IC matrix can be used to search relations between the criteria because the method compares homogeneous data relatively to a same column. In [32] Atanassov prescribes to normalize the score matrix before applying ICrA as follows

$$S_{ij}^{\text{norm}} = (S_{ij} - S_j^{\text{min}}) / (S_j^{\text{max}} - S_j^{\text{min}}), \quad (2)$$

if one wants to apply it in the dual manner for the search of InterObjects analysis (IOBa).

Because we focus on ICrA only, we don't need to apply a score matrix normalization because each column of the score matrix represents the values of a same criterion for different alternatives, and the criterion values are expressed with the same unit (e.g.  $m$ ,  $m^2$ ,  $\text{sec}$ ,  $Kg$ , or  $\text{€}$ , etc).

### A. Construction of Inter-Criteria matrix

The construction of the  $N \times N$  IC matrix, denoted<sup>7</sup>  $\mathbf{K}$ , is based on the pairwise comparisons between every two criteria along all evaluated alternatives. Let  $K_{jj'}^\mu$  be the number of cases in which the inequalities  $S_{ij} > S_{i'j}$  and  $S_{ij'} > S_{i'j'}$  hold simultaneously, and let  $K_{jj'}^\nu$  be the number of cases in which the inequalities  $S_{ij} > S_{i'j}$  and  $S_{ij'} < S_{i'j'}$  hold simultaneously. Because the total number of comparisons between the alternatives is  $M(M-1)/2$  then one always has necessarily

$$0 \leq K_{jj'}^\mu + K_{jj'}^\nu \leq \frac{M(M-1)}{2}, \quad (3)$$

or equivalently after the division by  $\frac{M(M-1)}{2} > 0$

$$0 \leq \frac{2K_{jj'}^\mu}{M(M-1)} + \frac{2K_{jj'}^\nu}{M(M-1)} \leq 1. \quad (4)$$

This inequality permits to define the elements of  $N \times N$  IC matrix  $\mathbf{K} = [K_{jj'}]$  as intuitionistic fuzzy (IF) pairs  $K_{jj'} = (\mu_{jj'}, \nu_{jj'})$  where

$$\mu_{jj'} \triangleq \frac{2K_{jj'}^\mu}{M(M-1)}, \quad \text{and} \quad \nu_{jj'} \triangleq \frac{2K_{jj'}^\nu}{M(M-1)}. \quad (5)$$

$\mu_{jj'}$  measures the degree of agreement between criteria  $C_j$  and  $C_{j'}$ , and  $\nu_{jj'}$  measures their degree of disagreement. By construction the IC matrix  $\mathbf{K}$  is always a symmetric matrix.

<sup>5</sup>where the symbol  $\triangleq$  means *equal by definition*.

<sup>6</sup>called index matrix by Atanassov in [31].

<sup>7</sup>We use  $\mathbf{K}$  because it corresponds to the first letter of word Kriterium, meaning criteria in German. The letter  $C$  is being already in use.

The computation of  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  can be done explicitly thanks to Atanassov's formulas [32]

$$K_{jj'}^\mu = \sum_{i=1}^{M-1} \sum_{i'=i+1}^M [sgn(S_{ij} - S_{i'j})sgn(S_{ij'} - S_{i'j'}) + sgn(S_{i'j} - S_{ij})sgn(S_{i'j'} - S_{ij'})], \quad (6)$$

and

$$K_{jj'}^\nu = \sum_{i=1}^{M-1} \sum_{i'=i+1}^M [sgn(S_{ij} - S_{i'j})sgn(S_{i'j'} - S_{ij'}) + sgn(S_{i'j} - S_{ij})sgn(S_{ij'} - S_{i'j'})], \quad (7)$$

where the signum function  $sgn(\cdot)$  used by Atanassov is defined as follows

$$sgn(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases} \quad (8)$$

Actually the values of  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  depend on the choice of  $sgn(x)$  function<sup>8</sup>. That is why in [21], [33], the authors propose different algorithms implemented under Java in an ICRA software yielding different  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  values for making the analysis and to reduce the dimension (complexity) of the original MCDM problem.

### B. Inter-criteria analysis

Once the Inter-Criteria matrix  $\mathbf{K} = [K_{jj'}]$  of intuitionistic fuzzy pairs is calculated one needs to analyze it to decide which criteria  $C_j$  and  $C_{j'}$  are in strong agreement (or positive consonance) reflecting the correlation between  $C_j$  and  $C_{j'}$ , in strong disagreement (or negative consonance) reflecting non correlation between  $C_j$  and  $C_{j'}$ , or in dissonance reflecting the uncertainty situation where nothing can be said about the non correlation or the correlation between  $C_j$  and  $C_{j'}$ . If one wants to identify the set of criteria  $C_{j'}$  for  $j' \neq j$  that are strongly correlated with  $C_j$  then we can sort  $\mu_{jj'}$  values in descending order to identify those in strong positive consonance with  $C_j$ . In [25], [26], the authors propose a qualitative scale to refine the levels of consonance and dissonance and for helping the decision making procedure. A dual approach based on  $\nu_{jj'}$  values can be made to determine the set of criteria that are not correlated with  $C_j$ . An other approach [10], [27] proposes to define two thresholds  $\alpha, \beta \in [0; 1]$  for the positive and negative consonance respectively against which the components  $\mu_{jj'}$  and  $\nu_{jj'}$  of  $K_{jj'} = (\mu_{jj'}, \nu_{jj'})$  will be compared. The correlations between the criteria  $C_j$  and  $C_{j'}$  are called "positive consonance", "negative consonance" or "dissonance" depending on their  $\mu_{jj'}$  and  $\nu_{jj'}$  values with respect to chosen thresholds  $\alpha$  and  $\beta$ , see [22] for details. More precisely,  $C_j$  and  $C_{j'}$  are in

- $(\alpha, \beta)$  positive consonance (i.e. correlated):  
If  $\mu_{jj'} > \alpha$  and  $\nu_{jj'} < \beta$ .
- $(\alpha, \beta)$  negative consonance (i.e. no correlated):

<sup>8</sup>for instance if we use  $sgn(x) = 1$  if  $x \geq 0$  and  $sgn(x) = 0$  if  $x < 0$ , we will obtain, in general, other  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  values.

If  $\mu_{jj'} < \beta$  and  $\nu_{jj'} > \alpha$ .

- $(\alpha, \beta)$  dissonance (i.e. full uncertainty): Otherwise.

At the beginning of ICRA development it was not very clear how these intuitionistic fuzzy (IF) pairs  $(\mu_{jj'}, \nu_{jj'})$  had to be used and that is why Atanassova [34], [35] proposed to handle both components of the IF pair. For this, she interpreted pairs  $(\mu_{jj'}, \nu_{jj'})$  as points located in the elementary  $TFU$  triangle, where the point  $T$  of coordinate  $(1, 0)$  represents the maximal positive consonance (i.e. the true consonance), the point  $F$  with coordinate  $(0, 1)$  represents the maximal negative consonance (i.e. the falsity), and the point  $U$  with coordinates  $(0, 0)$  represents the maximal dissonance (i.e. the uncertainty). From this interpretation it becomes easy to identify the top of consonant IF pairs  $(\mu_{jj'}, \nu_{jj'})$  that fall in bottom right corner of  $(TFU)$  triangle limited by vertical line from  $x$ -axis  $x = \alpha$ , and horizontal line from  $y$ -axis  $y = \beta$ . The set of consonant IF pairs are then ranked according to their Euclidean distance  $d_{C_j C_{j'}}^T$  with respect to  $T$  point of coordinate  $(1, 0)$  defined by

$$d_{C_j C_{j'}}^T = d((1, 0), (\mu_{jj'}, \nu_{jj'})) = \sqrt{(1 - \mu_{jj'})^2 + \nu_{jj'}^2}. \quad (9)$$

In the MCDM context only the criteria that are negatively consonant (or uncorrelated) must be kept for solving MCDM and saving computational resources because they have no (or only very low) dependency with each other, so that each uncorrelated criterion provides useful information. The set of criteria that are positively consonant (if any), called the consonant set, indicates somehow a redundancy of information between the criteria belonging to it in term of decisional behavior. Therefore all these positively consonant criteria must be represented by only one representative criterion that will be kept in the MCDM analysis to simplify MCDM problem. Also all the criteria that are deemed strongly dissonant (if any) could be taken out of the original MCDM problem because they only introduce uncertainty in the decision-making.

### C. General comments on ICRA

Although appealing at the first glance, the classical ICRA approach induces the following comments:

- 1) The IF values  $\mu_{jj'}$  and  $\nu_{jj'}$  can be easily interpreted in the belief function framework. Indeed, the belief and plausibility of (positive) consonance between criteria  $C_j$  and  $C_{j'}$  can be directly linked to the values  $\mu_{jj'}$  and  $\nu_{jj'}$  by taking  $Bel_{jj'}(\theta) = \mu_{jj'}$  and  $Pl_{jj'}(\theta) = 1 - \nu_{jj'}$ . Moreover  $U_{jj'}(\theta) = Pl_{jj'}(\theta) - Bel_{jj'}(\theta) = 1 - \nu_{jj'} - \mu_{jj'}$  represents the dissonance (the uncertainty about the correlation) of the criteria  $C_j$  and  $C_{j'}$ . Here the proposition  $\theta$  means: the criteria  $C_j$  and  $C_{j'}$  are totally positively consonant (i.e. totally correlated) and the frame of discernment is defined as  $\Theta \triangleq \{\theta, \bar{\theta}\}$ , where  $\bar{\theta}$  means: the criteria  $C_j$  and  $C_{j'}$  are totally negatively

consonant (uncorrelated). From this, one can define any BBA  $m_{jj'}(\theta)$ ,  $m_{jj'}(\bar{\theta})$  and  $m_{jj'}(\theta \cup \bar{\theta})$  of  $2^\Theta$  by

$$m_{jj'}(\theta) = \mu_{jj'}, \quad (10)$$

$$m_{jj'}(\bar{\theta}) = \nu_{jj'}, \quad (11)$$

$$m_{jj'}(\theta \cup \bar{\theta}) = 1 - \mu_{jj'} - \nu_{jj'}. \quad (12)$$

- 2) The construction of  $\mu_{jj'}$  and  $\nu_{jj'}$  proposed in the classical ICRA is disputable because it is only based on counting the valid " $>$ " or " $<$ " inequalities but it doesn't exploit how bigger and how smaller the scores values are in each comparison done in the construction of the Inter-Criteria Matrix  $\mathbf{K}$ . Therefore the construction of  $\mu_{jj'}$  and  $\nu_{jj'}$  is actually only a very crude method to estimate IF pairs.
- 3) The construction of the Inter-Criteria Matrix  $\mathbf{K}$  is in fact not unique as reported in [33]. This will yield different results in general.
- 4) The exploitation of the ICRA method depends on the choice of  $\alpha$  and  $\beta$  thresholds that will impact the final result.
- 5) The classical ICRA method cannot deal directly with imprecise or missing score values.

#### IV. A NEW ICRA METHOD BASED ON BELIEF FUNCTIONS

In this paper we propose a new ICRA method, called BF-ICRA for short, based on belief functions that circumvents most of the aforementioned drawbacks of classical ICRA. Here we show how to get more precisely the Inter-Criteria Belief Matrix and how to exploit it for MCDM simplification.

##### A. Construction of BBA matrix from the score matrix

From any non-zero score matrix  $\mathbf{S} = [S_{ij}]$ , we can construct the  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  as follows

$$m_{ij}(A_i) = Bel_{ij}(A_i), \quad (13)$$

$$m_{ij}(\bar{A}_i) = Bel_{ij}(\bar{A}_i) = 1 - Pl_{ij}(A_i), \quad (14)$$

$$m_{ij}(A_i \cup \bar{A}_i) = Pl_{ij}(A_i) - Bel_{ij}(A_i). \quad (15)$$

Assuming  $A_{\max}^j \neq 0$  and  $A_{\min}^j \neq 0$ , we take<sup>9</sup>

$$Bel_{ij}(A_i) \triangleq Sup_j(A_i)/A_{\max}^j, \quad (16)$$

$$Bel_{ij}(\bar{A}_i) \triangleq Inf_j(A_i)/A_{\min}^j, \quad (17)$$

where  $A_{\max}^j \triangleq \max_i Sup_j(A_i)$  and  $A_{\min}^j \triangleq \min_i Inf_j(A_i)$  and with

$$Sup_j(A_i) \triangleq \sum_{k \in \{1, \dots, M\} | S_{kj} \leq S_{ij}} |S_{ij} - S_{kj}|, \quad (18)$$

$$Inf_j(A_i) \triangleq - \sum_{k \in \{1, \dots, M\} | S_{kj} \geq S_{ij}} |S_{ij} - S_{kj}|. \quad (19)$$

The entire justification of these formulas can be found in our previous works [3]. For example, consider the  $j$ -th column corresponding to a criterion  $C_j$  of a score matrix  $\mathbf{S} = [S_{ij}]$  with seven rows given by  $\mathbf{s}_j = [10, 20, -5, 0, 100, -11, 0]^T$ ,

where  $T$  denotes the transpose. Then based on above formula we get the BBA values listed in Table I.

For another criterion  $C_{j'}$  and the  $j'$ -th column of the score matrix we will obtain another set of BBA values  $m_{ij'}(\cdot)$ . Applying this method for each column of the score matrix we are able to compute the BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  whose each component is in fact a triplet  $(m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i))$  of BBA values in  $[0, 1]$  such that  $m_{ij}(A_i) + m_{ij}(\bar{A}_i) + m_{ij}(A_i \cup \bar{A}_i) = 1$  for all  $i = 1, \dots, M$  and  $j = 1, \dots, N$ .

##### B. Construction of Inter-Criteria Matrix from BBA matrix

The next step of BF-ICRA approach is the construction of the  $N \times N$  Inter-Criteria Matrix  $\mathbf{K} = [K_{jj'}]$  from  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  where elements  $K_{jj'}$  corresponds to the BBA  $(m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  about positive consonance  $\theta$ , negative consonance  $\bar{\theta}$  and uncertainty between criteria  $C_j$  and  $C_{j'}$  respectively. The principle of construction of the triplet  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  is based on two steps that will be detailed in the sequel:

- Step 1: For each alternative  $A_i$ , we first compute the BBA  $(m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  for any two criteria  $j, j' \in \{1, 2, \dots, N\}$ .
- Step 2: The BBA  $(m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  is then obtained by the combinations of the  $M$  BBA  $m_{jj'}^i(\cdot)$ .

##### Construction of BBA $m_{jj'}^i(\cdot)$

The mass of belief  $m_{jj'}^i(\theta)$  represents the degree of agreement between the BBA  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$  for the alternative  $A_i$ , and  $m_{jj'}^i(\bar{\theta})$  represents the degree of disagreement between  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$ . The mass  $m_{jj'}^i(\theta \cup \bar{\theta})$  is the degree of uncertainty about the agreement (or disagreement) between  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$  for the alternative  $A_i$ . The calculation of  $m_{jj'}^i(\theta)$  could be envisaged in several manners.

The first manner would consist to consider the degree of conflict [28]  $k_{jj'}^i \triangleq \sum_{X, Y \subseteq \Theta | X \cap Y = \emptyset} m_{ij}(X) m_{ij'}(Y)$  and consider the Bayesian BBA  $m_{jj'}^i(\theta) = 1 - k_{jj'}^i$ ,  $m_{jj'}^i(\bar{\theta}) = k_{jj'}^i$  and  $m_{jj'}^i(\theta \cup \bar{\theta}) = 0$ . Instead of using Shafer's conflict, the second manner would consist to use a normalized distance  $d_{jj'}^i = d(m_{ij}, m_{ij'})$  to measure the closeness between  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$ , and then consider the Bayesian BBA modeling defined by  $m_{jj'}^i(\theta) = 1 - d_{jj'}^i$ ,  $m_{jj'}^i(\bar{\theta}) = d_{jj'}^i$  and  $m_{jj'}^i(\theta \cup \bar{\theta}) = 0$ . These two manners however are not very satisfying because they always set to zero the degree of uncertainty between the

<sup>9</sup>If  $A_{\max}^j = 0$  then  $Bel_{ij}(A_i) = 0$ , and if  $A_{\min}^j = 0$  then  $Pl_{ij}(A_i) = 1$ .

Table I  
BBAS CONSTRUCTED FROM SCORE VALUES.

	$m_{ij}(A_i)$	$m_{ij}(\bar{A}_i)$	$m_{ij}(A_i \cup \bar{A}_i)$
$A_1$	0.0955	0.5236	0.3809
$A_2$	0.1809	0.4188	0.4003
$A_3$	0.0102	0.8115	0.1783
$A_4$	0.0273	0.6806	0.2921
$A_5$	1.0000	0	0
$A_6$	0	1.0000	0
$A_7$	0.0273	0.6806	0.2921

agreement and disagreement of the BBA, and the second manner depends also on the choice of the distance metric. So, we propose a more appealing third manner of the BBA modeling of  $m_{jj'}^i(\theta)$ ,  $m_{jj'}^i(\bar{\theta})$ , and  $m_{jj'}^i(\theta \cup \bar{\theta})$ . For this, we consider two sources of evidences (SoE) indexed by  $j$  and  $j'$  providing the BBA  $m_{ij}$  and  $m_{ij'}$  defined on the simple FoD  $\{A_i, \bar{A}_i\}$  and denoted  $m_{ij} = [m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i)]$  and  $m_{ij'} = [m_{ij'}(A_i), m_{ij'}(\bar{A}_i), m_{ij'}(A_i \cup \bar{A}_i)]$ . We also denote  $\Theta = \{\theta, \bar{\theta}\}$  the FoD about the relative state of the two SoE, where  $\theta$  means that the two SoE agree,  $\bar{\theta}$  means that they disagree and  $\theta \cup \bar{\theta}$  means that we don't know. Then the BBA modeling is based on the important remarks

- Two SoE are in total agreement if both commit their maximum belief mass to the element  $A_i$  or to element  $\bar{A}_i$ . So they perfectly agree if  $m_{ij}(A_i) = m_{ij'}(A_i) = 1$ , or if  $m_{ij}(\bar{A}_i) = m_{ij'}(\bar{A}_i) = 1$ . Therefore the pure degree of agreement<sup>10</sup> between two sources is modeled by

$$m_{jj'}^i(\theta) = m_{ij}(A_i)m_{ij'}(A_i) + m_{ij}(\bar{A}_i)m_{ij'}(\bar{A}_i). \quad (20)$$

- Two SoE are in total disagreement if each one commits its maximum mass of belief to one element and the other to its opposite, that is if one has  $m_{ij}(A_i) = 1$  and  $m_{ij'}(\bar{A}_i) = 1$ , or if  $m_{ij}(\bar{A}_i) = 1$  and  $m_{ij'}(A_i) = 1$ . Hence the pure degree of disagreement<sup>11</sup> between two sources is modeled by

$$m_{jj'}^i(\bar{\theta}) = m_{ij}(A_i)m_{ij'}(\bar{A}_i) + m_{ij}(\bar{A}_i)m_{ij'}(A_i). \quad (21)$$

- All possible remaining products between components of  $m_{ij}$  and  $m_{ij'}$  reflect the part of uncertainty we have about the SoE (i.e. we don't know if they agree or disagree). Hence the degree of uncertainty between the two sources is modeled by

$$\begin{aligned} m_{jj'}^i(\theta \cup \bar{\theta}) &= m_{ij}(A_i)m_{ij'}(A_i \cup \bar{A}_i) + m_{ij}(\bar{A}_i)m_{ij'}(A_i \cup \bar{A}_i) \\ &+ m_{ij}(A_i \cup \bar{A}_i)m_{ij'}(A_i) + m_{ij}(A_i \cup \bar{A}_i)m_{ij'}(\bar{A}_i) \\ &+ m_{ij}(A_i \cup \bar{A}_i)m_{ij'}(A_i \cup \bar{A}_i). \end{aligned} \quad (22)$$

By construction  $m_{jj'}^i(\cdot) = m_{jj'}^i(\cdot)$ , hence this BBA modeling permits to build a set of  $M$  symmetrical Inter-Criteria Belief Matrices (ICBM)  $\mathbf{K}^i = [K_{jj'}^i]$  of dimension  $N \times N$  relative to each alternative  $A_i$  whose components  $K_{jj'}^i$  correspond to the triplet of BBA values  $m_{jj'}^i = (m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  modeling the belief of agreement and of disagreement between  $C_j$  and  $C_{j'}$  based on  $A_i$ . One has also<sup>12</sup>  $m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}) \in [0, 1]$  and  $m_{jj'}^i(\theta) + m_{jj'}^i(\bar{\theta}) + m_{jj'}^i(\theta \cup \bar{\theta}) = 1$ . This BBA construction can be easily extended

for modeling the agreement, disagreement and uncertainty of  $n > 2$  criteria  $C_{j_1}, \dots, C_{j_n}$  altogether if needed by taking

$$\begin{aligned} m_{j_1 \dots j_n}^i(\theta) &= \prod_{k=1}^n m_{ij_k}(A_i) + \prod_{k=1}^n m_{ij_k}(\bar{A}_i), \\ m_{j_1 \dots j_n}^i(\bar{\theta}) &= \sum_{\substack{X_{j_1}, \dots, X_{j_n} \in \{A_i, \bar{A}_i\} \\ X_{j_1} \cap \dots \cap X_{j_n} = \emptyset}} \prod_{k=1}^n m_{ij_k}(X_{j_k}), \\ m_{j_1 \dots j_n}^i(\theta \cup \bar{\theta}) &= 1 - m_{j_1 \dots j_n}^i(\theta) - m_{j_1 \dots j_n}^i(\bar{\theta}). \end{aligned}$$

### Construction of BBA $m_{jj'}^i(\cdot)$

Once all the BBAs  $m_{jj'}^i(\cdot)$  ( $i = 1, \dots, M$ ) are calculated one combines them to get the component  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  of the Inter-Criteria Belief matrix (ICBM)  $\mathbf{K} = [K_{jj'}]$ . This fusion step can be done in many ways depending on the combination rule chosen by the user. If the number of alternatives  $M$  is not too large we recommend to combine the BBAs  $m_{jj'}^i(\cdot)$  with PCR6 fusion rule [30] (Vol. 3) because of known deficiencies of Dempster's rule. If  $M$  is too large to prevent PCR6 working on computer, we can just use the simple averaging rule of combination in these high dimensional MCDM problems.

### V. SIMPLIFICATION OF ORIGINAL MCDM

Once the global Inter-Criteria Belief Matrix  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$  is calculated, we need to identify and cluster the criteria that are in strong agreement, in strong disagreement, and those on which we are uncertain. For identifying the criteria that are in very strong agreement, we evaluate the distance of each component of  $K_{jj'}$  with the BBA representing the best agreement state and characterized by the specific BBA<sup>13</sup>  $m_T(\theta) = 1$ . From a similar approach we can also identify, if we want, the criteria that are in very strong disagreement using the distance of  $m_{jj'}(\cdot)$  with respect to the BBA representing the best disagreement state characterized by the specific BBA  $m_F(\bar{\theta}) = 1$ . As alternative of Jusselme's distance [37], we use the  $d_{BI}(\cdot, \cdot)$  distance based on belief interval [36] because it is a good method for measuring the distance  $d(m_1, m_2)$  between the two BBAs<sup>14</sup>  $m_1(\cdot)$  and  $m_2(\cdot)$  over the same FoD. It is defined by

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^{\Theta}} d_W^2(BI_1(X), BI_2(X))}, \quad (23)$$

where the *Belief-Intervals* are defined by  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)]$  and computed from  $m_1(\cdot)$  and  $m_2(\cdot)$  thanks to formula (1).

<sup>10</sup>or positive consonance according Atanassov's terminology.

<sup>11</sup>or negative consonance according Atanassov's terminology.

<sup>12</sup>because  $(m_{ij}(A_i) + m_{ij}(\bar{A}_i) + m_{ij}(A_i \cup \bar{A}_i))(m_{ij'}(A_i) + m_{ij'}(\bar{A}_i) + m_{ij'}(A_i \cup \bar{A}_i)) = 1 \cdot 1 = 1$ .

<sup>13</sup>We use the index  $T$  in the notation  $m_T(\cdot)$  to refer that the agreement is true, and  $F$  in  $m_F(\cdot)$  to specify that the agreement is false.

<sup>14</sup>Here  $m_1(\cdot) = m_{jj'}(\cdot)$ , and  $m_2(\cdot) = m_T(\cdot)$  or  $m_2(\cdot) = m_F(\cdot)$

$d_W(BI_1(X), BI_2(X))$  is Wassertein's distance between intervals calculated by

$$d_W([a_1, b_1], [a_2, b_2]) = \sqrt{\left[\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2}\right]^2 + \frac{1}{3}\left[\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2}\right]^2},$$

and  $N_c = 1/2^{|\Theta|-1}$  is a factor to get  $d_{BI}(m_1, m_2) \in [0, 1]$ .

Because all criteria that are in strong agreement somehow contain redundant (correlated) information and behave similarly from decision-making standpoint, we propose to simplify the original MCDM problem by keeping in the MCDM only criteria that are non redundant. The remaining criteria can be eventually weighted by their degree of importance reflecting the number of different criteria that are in agreement through this BF-ICrA approach.

For instance, if one has a seven criteria MCDM problem and if criteria  $C_1, C_2$  and  $C_3$  are in strong agreement we will only select one remaining criterion among  $\{C_1, C_2, C_3\}$  and we give it a weight of  $w_1 + w_2 + w_3$ . Moreover if  $C_4$  and  $C_5$  are in strong agreement also we will only select one remaining criterion among  $\{C_4, C_5\}$  and we give it a weight of  $w_4 + w_5$ , and we will use the weight  $w_6$  for  $C_6$ , and  $w_7$  for  $C_7$ . Hence the original MCDM problem will reduce to a four simplified MCDM problem that can be solved using BF-TOPSIS method already presented in details in [3] and in [8], or with AHP [4] if one prefers, or with any other chosen method that the system-designer may prefer.

The strategy for selecting the most representative criterion among a set of redundant criteria is not unique and depends mainly on the cost necessary (i.e. human efforts, data mining, computational resources, etc) for getting the values of the score matrix of the problem under concern. The least costly criteria may be a good option of selection. In the next section we provide simple examples for BF-ICrA and, for simplicity, we will select the representative criterion as being the one with smallest index. So in the aforementioned example the simplified MCDM problem will reduce to a  $M \times 4$  MCDM problem involving only four criteria  $C_1, C_4, C_6$  and  $C_7$ .

The BF-ICrA method proposed in this work allows also, in principle, to make a refined analysis (if necessary) based on IC matrices  $\mathbf{K}_{jj'}^i$  about the origin of disagreement between criteria with respect to each alternative  $A_i$  in order to identify the potential inconsistencies in original MCDM problem. This aspect is not developed in this paper and has been left for future investigations. It is worth mentioning that the analysis of the number of redundant criteria versus time improvements that could be proposed as an effective measure of performance of this approach depends highly of the application under consideration and the difficulty (and cost) to get the value of each criteria. For convenience the Figure 1 shows the flow chart of BF-ICrA to help the reader to have a better understanding of this new proposed method.

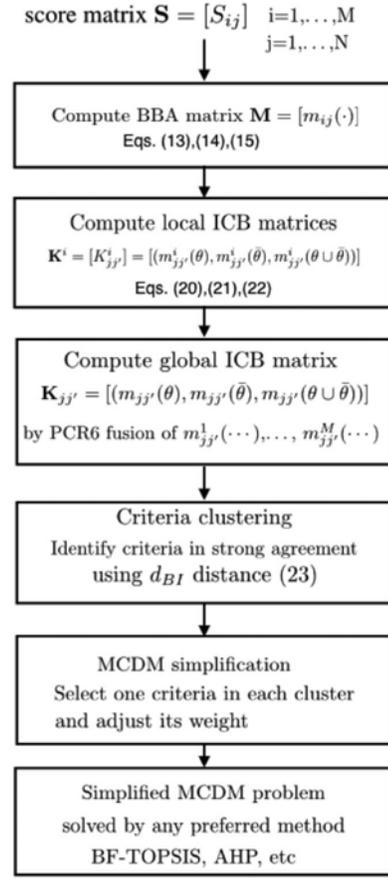


Figure 1. Flow chart of BF-ICrA method.

## VI. EXAMPLES

### A. Example 1 (Comparison of $\mathbf{K}$ matrices)

Here we compare the construction of the global IC matrix  $\mathbf{K}$  based on Atanassov ICrA and our new BF-ICrA approach. For this, we use the  $5 \times 4$  MCDM example given in [33] based on the following score matrix (called sample data matrix in [33]). Each row of  $\mathbf{S}$  corresponds to an alternative, and each column to a criterion. In [33], the authors use rows for criteria and columns for alternatives so they work with  $\mathbf{S}^T$ .

$$\mathbf{S} = [S_{ij}] = \begin{bmatrix} 6 & 7 & 4 & 4 \\ 5 & 7 & 3 & 5 \\ 3 & 8 & 5 & 6 \\ 7 & 1 & 9 & 7 \\ 6 & 3 & 1 & 8 \end{bmatrix}.$$

Based on Atanassov's ICrA method (using unbiased algorithm presented in details in [33]) we will get the following  $4 \times 4$  global Inter-Criteria  $\mathbf{K}^\mu$  and  $\mathbf{K}^\nu$  matrices

$$\mathbf{K}^\mu = [K_{jj'}^\mu] = \begin{bmatrix} 0.9 & 0 & 0.5 & 0.5 \\ 0 & 0.9 & 0.5 & 0.3 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.3 & 0.5 & 1 \end{bmatrix},$$

$$\mathbf{K}^\nu = [K_{jj'}^\nu] = \begin{bmatrix} 0 & 0.8 & 0.4 & 0.4 \\ 0.8 & 0 & 0.4 & 0.6 \\ 0.4 & 0.4 & 0 & 0.5 \\ 0.4 & 0.6 & 0.5 & 0 \end{bmatrix}.$$

Regrouping these two matrices into one matrix  $\mathbf{K} = [K_{jj'}]$  with components  $K_{jj'} = (K_{jj'}^\mu, K_{jj'}^\nu, 1 - K_{jj'}^\mu - K_{jj'}^\nu)$ , one gets the following global Inter-Criteria matrix  $\mathbf{K}$

$$\mathbf{K} = \begin{bmatrix} (0.9, 0, 0.1) & (0, 0.8, 0.2) & (0.5, 0.4, 0.1) & (0.5, 0.4, 0.1) \\ (0, 0.8, 0.2) & (0.9, 0, 0.1) & (0.5, 0.4, 0.1) & (0.3, 0.6, 0.1) \\ (0.5, 0.4, 0.1) & (0.5, 0.4, 0.1) & (1, 0, 0) & (0.5, 0.5, 0) \\ (0.5, 0.4, 0.1) & (0.3, 0.6, 0.1) & (0.5, 0.5, 0) & (1, 0, 0) \end{bmatrix}.$$

According to this  $\mathbf{K}$  matrix it appears intuitively that none of the criterion is in strong agreement with others. We observe that criteria  $C_1$  and  $C_2$  are in relatively strong disagreement because  $K_{12}^\nu = K_{21}^\nu = 0.8$  which is quite close to one. Criteria  $C_2$  and  $C_4$  are in relatively medium disagreement because  $K_{24}^\nu = K_{42}^\nu = 0.6$ . In this example no MCDM simplification is prescribed based on Atanassov's ICRA. To get a more precise evaluation of degree of agreement between criteria based on Atanassov's ICRA we apply formula (23) to get the  $\mathbf{D}_{BI}^\theta$  distance matrix from each component of  $\mathbf{K}$  to the total agreement state  $m_T = [m(\theta), m(\bar{\theta}), m(\theta \cup \bar{\theta})] = [1, 0, 0]$ . Hence we get

$$\mathbf{D}_{BI}^\theta = \begin{bmatrix} 0.0577 & \mathbf{0.9018} & 0.4509 & 0.4509 \\ \mathbf{0.9018} & 0.0577 & 0.4509 & 0.6506 \\ 0.4509 & 0.4509 & 0 & 0.5000 \\ 0.4509 & 0.6506 & 0.5000 & 0 \end{bmatrix}.$$

As we see from this  $\mathbf{D}_{BI}^\theta$  matrix, the distance of the inter-criteria BBA for  $C_1$  and  $C_2$  with respect to the total agreement state  $m_T(\theta) = 1$  is very large (i.e. 0.9018) which means that  $C_1$  and  $C_2$  strongly disagree in this example as we expect from a more intuitive reasoning based on  $K_{12}^\nu = 0.8$  value. Similar analyses can be done for all (non diagonal) elements of  $\mathbf{D}_{BI}^\theta$  to identify which criteria are in strong agreement, or not (if any).

Based on our new BF-ICRA method we first compute the  $5 \times 4$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  from the score matrix  $\mathbf{S}$  based on formulas (13)-(15). We get (all the values of results have been rounded at their second digit)

$$\mathbf{M} \approx \begin{bmatrix} (0.5, 0.08, 0.42) & (0.71, 0.05, 0.24) & (0.18, 0.35, 0.47) & (0, 1, 0) \\ (0.25, 0.33, 0.42) & (0.71, 0.05, 0.24) & (0.09, 0.53, 0.38) & (0.1, 0.6, 0.3) \\ (0, 1, 0) & (1, 0, 0) & (0.30, 0.24, 0.46) & (0.3, 0.3, 0.4) \\ (1, 0, 0) & (0, 1, 0) & (1, 0, 0) & (0.6, 0.1, 0.3) \\ (0.5, 0.09, 0.41) & (0.14, 0.62, 0.24) & (0, 1, 0) & (1, 0, 0) \end{bmatrix}.$$

The construction of Inter-Criteria Matrices  $\mathbf{K}^i = [K_{jj'}^i]$  (for  $i = 1, \dots, 5$ ) from the BBA matrix  $\mathbf{M}$  based on formulas (20)-

(22) yields the following five matrices

$$\mathbf{K}^1 \approx \begin{bmatrix} (0.26, 0.08, 0.66) & (0.36, 0.08, 0.56) & (0.12, 0.19, 0.69) & (0.08, 0.5, 0.42) \\ (0.36, 0.08, 0.56) & (0.51, 0.07, 0.42) & (0.14, 0.26, 0.6) & (0.05, 0.71, 0.24) \\ (0.12, 0.19, 0.69) & (0.14, 0.26, 0.6) & (0.15, 0.13, 0.72) & (0.35, 0.18, 0.47) \\ (0.08, 0.5, 0.42) & (0.05, 0.71, 0.24) & (0.35, 0.18, 0.47) & (1, 0, 0) \end{bmatrix}$$

$$\mathbf{K}^2 \approx \begin{bmatrix} (0.17, 0.17, 0.66) & (0.19, 0.25, 0.56) & (0.20, 0.16, 0.64) & (0.23, 0.18, 0.59) \\ (0.19, 0.25, 0.56) & (0.51, 0.07, 0.42) & (0.09, 0.38, 0.53) & (0.10, 0.43, 0.47) \\ (0.20, 0.16, 0.64) & (0.09, 0.38, 0.53) & (0.29, 0.09, 0.62) & (0.33, 0.10, 0.57) \\ (0.23, 0.18, 0.59) & (0.10, 0.43, 0.47) & (0.33, 0.10, 0.57) & (0.37, 0.12, 0.51) \end{bmatrix}$$

$$\mathbf{K}^3 \approx \begin{bmatrix} (1, 0, 0) & (0, 1, 0) & (0.24, 0.3, 0.46) & (0.3, 0.3, 0.4) \\ (0, 1, 0) & (1, 0, 0) & (0.30, 0.24, 0.46) & (0.3, 0.3, 0.4) \\ (0.24, 0.3, 0.46) & (0.30, 0.24, 0.46) & (0.15, 0.14, 0.71) & (0.16, 0.16, 0.68) \\ (0.3, 0.3, 0.4) & (0.3, 0.3, 0.4) & (0.16, 0.16, 0.68) & (0.18, 0.18, 0.64) \end{bmatrix}$$

$$\mathbf{K}^4 \approx \begin{bmatrix} (1, 0, 0) & (0, 1, 0) & (1, 0, 0) & (0.6, 0.1, 0.3) \\ (0, 1, 0) & (1, 0, 0) & (0, 1, 0) & (0.1, 0.6, 0.3) \\ (1, 0, 0) & (0, 1, 0) & (1, 0, 0) & (0.6, 0.1, 0.3) \\ (0.6, 0.1, 0.3) & (0.1, 0.6, 0.3) & (0.6, 0.1, 0.3) & (0.370, 0.12, 0.51) \end{bmatrix}$$

$$\mathbf{K}^5 \approx \begin{bmatrix} (0.26, 0.08, 0.66) & (0.12, 0.32, 0.56) & (0.08, 0.5, 0.42) & (0.5, 0.08, 0.42) \\ (0.12, 0.32, 0.56) & (0.40, 0.18, 0.42) & (0.62, 0.14, 0.24) & (0.14, 0.62, 0.24) \\ (0.08, 0.5, 0.42) & (0.62, 0.14, 0.24) & (1, 0, 0) & (0, 1, 0) \\ (0.5, 0.08, 0.42) & (0.14, 0.62, 0.24) & (0, 1, 0) & (1, 0, 0) \end{bmatrix}$$

The componentwise PCR6 fusion of all five  $\mathbf{K}^i$  matrices provides the following global Inter-Criteria matrix  $\mathbf{K}_{PCR6}$

$$\mathbf{K}_{PCR6} \approx \begin{bmatrix} (0.90, 0.02, 0.08) & (0.06, 0.83, 0.11) & (0.55, 0.15, 0.30) & (0.49, 0.25, 0.26) \\ (0.06, 0.83, 0.11) & (0.95, 0.01, 0.04) & (0.18, 0.58, 0.24) & (0.08, 0.80, 0.12) \\ (0.55, 0.15, 0.30) & (0.18, 0.58, 0.24) & (0.89, 0.02, 0.09) & (0.22, 0.48, 0.30) \\ (0.49, 0.25, 0.26) & (0.08, 0.80, 0.12) & (0.22, 0.48, 0.30) & (0.90, 0.02, 0.08) \end{bmatrix}$$

Applying formula (23) we get the following  $\mathbf{D}_{BI}^\theta$  distance matrix from each component of  $\mathbf{K}_{PCR6}$  to the total agreement state  $m_T = [m(\theta), m(\bar{\theta}), m(\theta \cup \bar{\theta})] = [1, 0, 0]$

$$\mathbf{D}_{BI}^\theta = \begin{bmatrix} 0.0601 & 0.8845 & 0.3124 & 0.3909 \\ 0.8845 & 0.0327 & 0.7037 & 0.8618 \\ 0.3124 & 0.7037 & 0.0668 & 0.6355 \\ 0.3909 & 0.8618 & 0.6355 & 0.0622 \end{bmatrix}. \quad (24)$$

We see that  $\mathbf{K}$  and  $\mathbf{K}_{PCR6}$  are different specially  $K_{23} = K_{32} = (0.5, 0.4, 0.1)$  with respect to  $K_{23}^{PCR6} = K_{32}^{PCR6} = (0.18, 0.58, 0.24)$ . Based on  $\mathbf{D}_{BI}^\theta$  matrix (24) it is obvious that no criteria strongly agree in this example so that no judicious MCDM simplification is recommended according to BF-ICRA.

### B. Example 2 (MCDM simplification)

Here we consider a more interesting example showing how an MCDM simplification is possible. We consider a  $6 \times 5$  MCDM problem with the following score matrix.

$$\mathbf{S} = [S_{ij}] = \begin{bmatrix} 7.5914 & 18.1828 & 18.3221 & 95.6739 & 4.5674 \\ 8.7753 & 20.5506 & 20.8240 & 48.0229 & -0.1977 \\ -1.3492 & 0.3017 & 0.7804 & 79.8283 & 2.9828 \\ 8.8739 & 20.7478 & 21.2302 & 13.3305 & -3.6669 \\ 5.2207 & 13.4413 & 13.5201 & 41.5979 & -0.8402 \\ -1.7320 & -0.4639 & 0.0213 & 91.4893 & 4.1489 \end{bmatrix}.$$

It is not very obvious to identify the closeness of these criteria (if any) to know if there is some underlying relationship between them. For the analysis, we apply the BF-ICRA approach proposed in this work. After applying all derivations (similarly to those presented in Example 1), we finally get the following  $\mathbf{D}_{BI}^\theta$  distance matrix from each component of  $\mathbf{K}_{PCR6}$  to the total agreement state  $m_T = [m(\theta), m(\bar{\theta}), m(\theta \cup \bar{\theta})] = [1, 0, 0]$

$$\mathbf{D}_{BI}^\theta = \begin{bmatrix} 0.0239 & \mathbf{0.0239} & \mathbf{0.0250} & 0.7512 & 0.7512 \\ \mathbf{0.0239} & 0.0239 & \mathbf{0.0250} & 0.7512 & 0.7512 \\ \mathbf{0.0250} & \mathbf{0.0250} & 0.0262 & 0.7595 & 0.7595 \\ 0.7512 & 0.7512 & 0.7595 & 0.0568 & \mathbf{0.0568} \\ 0.7512 & 0.7512 & 0.7595 & \mathbf{0.0568} & 0.0568 \end{bmatrix}.$$

From the analysis of upper off-diagonal components of  $\mathbf{D}_{BI}^\theta$  (put in boldface for convenience) it is clear that criteria  $C_1$ ,  $C_2$  and  $C_3$  are in almost total agreement because their distance is close to zero. Also we can observe from  $\mathbf{D}_{BI}^\theta$  that criteria  $C_4$  and  $C_5$  are also very close. So the original  $6 \times 5$  MCDM problem in this example can be simplified into a  $6 \times 2$  MCDM problem considering only the simplified score matrix involving only  $C_1$  and  $C_4$  because  $C_2$  and  $C_3$  behave similarly to  $C_1$  for decision-making, and  $C_5$  behaves similarly to  $C_4$ . Then the simplified MCDM will have to be solved by any preferred technique.

Does the BF-ICrA make sense in this example? The answer is positive because it suffices to remark that the columns of the score matrix are not totally independent because  $C_2(A_i) = 2 \cdot C_1(A_i) + 3$ ,  $C_3(A_i) = C_2(A_i) + \epsilon$  ( $\epsilon$  being a small contamination noise), and  $C_5(A_i) = 0.1 \cdot C_4(A_i) - 5$ . Hence the decision based either on  $C_1$ ,  $C_2$  or  $C_3$  will be very close, as well as the decision based on  $C_4$  or  $C_5$ . Therefore the result of BF-ICrA makes sense and the expected simplification of MCDM is well obtained from BF-ICrA. If we apply AHP, which is nothing but the weighted arithmetic average and we use the normalized score matrix based on (2), or BF-TOPSIS methods to solve original MCDM (assuming equal importance of criteria), or if we apply simplified MCDM based on BF-ICrA, we will get same preference order  $A_1 \succ A_2 \succ A_4 \succ A_5 \succ A_6 \succ A_3$ . So, the best decision to make is  $A_1$  in this example.

## VII. CONCLUSION

In this paper we have proposed a new method called BF-ICrA to simplify (when it is possible) Multi-Criteria Decision-Making problems based on inter-criteria analysis and belief functions. This method is in the spirit of Atanassov's method but proposes a better construction of Inter-Criteria Matrix that fully exploits all information of the score matrix, and the closeness measure of agreement between criteria based on belief interval distance. This BF-ICrA approach for simplifying MCDM could deal also with imprecise or missing score values using the technique presented in [8]. An application of BF-ICrA for GPS surveying problem is presented in [38], and applications of BF-ICrA for simplifying and solving real MCDM problems for the prevention of natural risks in mountains will be the object of forthcoming investigations.

## REFERENCES

- [1] J. Barzilai, B. Golany, *AHP rank reversal, normalization and aggregation rules*, INFOR Vol. 32, No. 2, pp. 57–63, 1994.
- [2] D. Pavlicic, *Normalization affects the results of MADM methods*, Yugoslav J. of Operations Research, Vol. 11, No. 2, pp. 251–265, 2001.
- [3] J. Dezert, D. Han, H. Yin, *A New Belief Function Based Approach for Multi-Criteria Decision-Making Support*, in Proc. of Fusion Conf., 2016.
- [4] T. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, 1980.
- [5] X. Wang, E. Triantaphyllou, *Ranking irregularities when evaluating alternatives by using some ELECTRE methods*, Omega, Vol. 36(1), 2008.
- [6] C.L. Hwang, K. Yoon, *Multiple Attribute Decision Making*, in Lecture Notes in Economics and Math. Syst. 186, Springer-Verlag, Berlin, 1981.
- [7] Y.J. Lai, T.Y. Liu, C.L. Hwang, C.L., *TOPSIS for MODM*, European Journal of Operational Research, Vol. 76, No. 3, pp. 486–500, 1994.
- [8] J. Dezert, D. Han, J.-M. Tacnet, *Multi-Criteria Decision-Making with Imprecise Scores and BF-TOPSIS*, in Proc. of Fusion Conf., Xi'an, 2017.
- [9] K. Atanassov, D. Mavrov, V. Atanassova, *InterCriteria Decision Making. A New Approach for Multicriteria Decision Making, Based on Index Matrices and Intuitionistic Fuzzy Sets*, Issues in IFS and Generalized Nets, Vol. 11, pp. 1–8, 2014.
- [10] K. Atanassov, V. Atanassova, G. Gluhchev, *InterCriteria Analysis: Ideas and problems*, Notes on IFS, Vol. 21, No. 1, pp. 81–88, 2015.
- [11] K. Atanassov et al., *An approach to a constructive simplification of multiagent multicriteria decision making problems via intercriteria analysis*, C.R. de l'Acad. Bulgare des Sci., Vol. 70, No. 8, 2017.
- [12] K. Atanassov, *On Intuitionistic Fuzzy Sets Theory*, Springer, 2012.
- [13] S. Todinova et al., *Blood Plasma Thermograms Dataset Analysis by Means of InterCriteria and Correlation Analyses for the Case of Colorectal Cancer*, Int. J. of BIO Autom., Vol. 20(1), pp 115–124, 2016.
- [14] S. Krumova, et al., *InterCriteria analysis of calorimetric data of blood serum proteome*, in Bioch. et Biophys. Acta, Gen. Subjects, 1861, 2017.
- [15] B. Zaharieva et al., *InterCriteria decision making approach for Beherev's disease analysis*, in Int. J. of Bioautom, Vol. 22(2), 2018.
- [16] T. Pencheva et al., *InterCriteria Analysis of Genetic Algorithm Parameters in Parameter Identification*, Notes on IFS, Vol. 21(2), 2015.
- [17] S. Sotirov, et al., *Application of the Intuitionistic Fuzzy InterCriteria Analysis Method to a Neural Network Preprocessing Procedure*, Proc of 9th EUSFLAT, pp. 1559–1564, Gijon, Spain, June 30–July 7th, 2015.
- [18] O. Roeva et al., *InterCriteria analysis of a model parameters identification using genetic algorithm*, Proc. of the Federated Conf. on Computer Science and Information Systems, Vol. 5, pp. 501–506, 2015.
- [19] M. Angelova, O. Roeva, T. Pencheva, *InterCriteria analysis of crossover and mutation rates relations in simple genetic algorithm*, Proc. of Conf on Computer Sci. and Inf. Syst, Vol. 5, pp. 419–424, 2015.
- [20] O. Roeva, S. Fidanova, M. Paprzycki, *InterCriteria analysis of ACO and GA hybrid algorithms*, Stud. Comput. Intell., 610, pp. 107–126, 2016.
- [21] O. Roeva et al., *InterCriteria Analysis of ACO Performance for Workforce Planning Problem*, Studies in Comp. Intell., Vol. 795, Springer 2019.
- [22] V. Atanassova et al., *Discussion on the threshold values in the InterCriteria decision making approach*, Notes on Intuitionistic Fuzzy Sets, Vol. 20(2), pp. 94–99, 2014.
- [23] L. Doukovska, V. Atanassova, *InterCriteria Analysis approach in radar detection threshold analysis*, Notes on IFS, Vol. 21(4), 2015.
- [24] L. Doukovska et al., *InterCriteria Analysis Applied to EU Micro, Small, Medium and Large Enterprises*, in Proc. 5th Int. Symp. on BMSD, 2015.
- [25] V. Bureva et al., *Application of the InterCriteria decision making method to Bulgarian universities ranking*, Int. Workshop on IFSs, Bulgaria, 2015.
- [26] V. Bureva, E. Sotirova, H. Panayotov, *The InterCriteria decision making method to Bulgarian university ranking system*, Annual of Informatics Section, Vol. 8, pp. 54–70, 2015–2016.
- [27] M. Krawczak et al., *Application of the InterCriteria Decision Making Method to Universities Ranking*, Adv. in Intell. Syst. and Comp., Vol. 401, pp. 365–372, Springer, 2016.
- [28] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Press, 1976.
- [29] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., Vol. 38, pp. 325–339, 1967.
- [30] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, American Research Press, Vol. 1–4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [31] K. Atanassov, *Index Matrices: Towards an Augmented Matrix Calculus*, Springer, Cham, 2014.
- [32] K. Atanassov et al., *InterCriteria analysis over normalized data*, Proc. of 8th IEEE Int. Conf. on Intelligent Syst., 2016.
- [33] N. Ikononov, P. Vassilev, O. Roeva, *ICrAData software for InterCriteria analysis*, Int. J. BioAutomation, Vol. 22, No. 2, 2018.
- [34] V. Atanassova, *Interpretation in the Intuitionistic Fuzzy Triangle of the Results Obtained by the InterCriteria Analysis*, Proc. of 16th World Congr. of IFSA, Atlantis Press, 2015.
- [35] V. Atanassova et al., *Traversing and ranking of elements of an intuitionistic fuzzy set in the intuitionistic fuzzy interpretation triangle*, Adv. in Intell. Syst. and Comp., Vol. 401, pp. 161–174, 2016.
- [36] D. Han, J. Dezert, Y. Yang, *New Distance Measures of Evidence based on Belief Intervals*, Proc. of Belief 2014, Oxford, UK, Sept. 2014.
- [37] A.-L. Jousselme, D. Grenier, É. Bossé, *A new distance between two bodies of evidence*, Information Fusion, Vol. 2, No. 2, pp. 91–101, 2001.
- [38] S. Fidanova, J. Dezert, A. Tchamova, *Inter-Criteria Analysis Based on Belief Functions for GPS Surveying Problems*, Proc. of Int. Symposium on INnovations in Intelligent SysTems and Applications (INISTA), Sofia, Bulgaria, July 3–5, 2019.

# DSmT-based Group DEMATEL Method With Reaching Consensus

Yuan-Wei Du<sup>a, b</sup>, Wen Zhou<sup>b</sup>

<sup>a</sup>Management College, Ocean University of China, Qingdao 266100, People's Republic of China.

<sup>b</sup>Marine Development Studies Institute of OUC,

Key Research Institute of Humanities and Social Sciences at Universities,  
Ministry of Education, Qingdao 266100, People's Republic of China.

Emails: duyuanwei@foxmail.com, zwzhouwen9572@163.com

Originally published as: Y.-W. Du, W. Zhou, *DSmT-Based Group DEMATEL Method with Reaching Consensus, Group Decision and Negotiation*, Vol. 28, pp. 1201–1230, 2019, and reprinted with permission.

**Abstract**—The decision-making trial and evaluation laboratory (DEMATEL) method employs expert assessments expressed by crisp values to construct a group initial direct-relation (IDR) matrix. However, it tends to be a low-precision expression, especially in complex practical problems. Although significant efforts have been made to improve the DEMATEL method, these improvements tend to neglect individual characteristics and group consensus, resulting in unconvincing decision results. This study provides a Dezert-Smarandache theory (DSmT)-based group DEMATEL method with reaching consensus. In order to reasonably determine the group IDR matrix, basic belief assignment (BBA) function is employed to extract expert assessments and the proportional conflict redistribution rule no.5 (PCR5) of DSmT is employed to make fusion to derive the temporary group IDR matrix. Moreover, the consensus measures at both expert level and pair-factors level are calculated to determine whether the acceptable consensus level has been reached or not. If the required consensus level is not reached, a feedback mechanism will be activated to help experts reach a consensus. A consensus group IDR matrix for the group DEMATEL can be obtained with the help of feedback mechanism, based on which an algorithm is summarized for the proposed method to identify major factors in a complex system. Finally, numerical comparison and discussion are introduced to verify the effectiveness and applicability of the proposed method and algorithm.

**Keywords:** DEMATEL, group decision making; Dezert-Smarandache theory (DSmT), consensus reaching, evidence distance, expert weight.

## I. INTRODUCTION

Between 1972 and 1976, the Science and Human Affairs Program of Battelle Memorial Institute of Geneva developed the decision-making trial and evaluation laboratory (DEMATEL) method. This method aimed to describe the basic concept of contextual relations and identify cause-effect chain factors for a complex decision problem in an understandable manner by addressing the influence relations among factors given by experts [1], [2]. It was considered to be a credible decision-making method.

The DEMATEL method has been extensively used to solve complex decision problems because of its simplicity and effectiveness, including problems pertaining to hospital service quality [3], decision making [4], sustainable supply chain

management [5], [6], etc. In special, in order to determine the weights of factors by considering their relations, the DEMATEL method is also extended into decision making fields, such as analytic hierarchy process (AHP) [7], [8], analytic network process (ANP) [9]–[14], and technique for order preference by similarity to ideal solution (TOPSIS) [15], [16]. In the DEMATEL method, the initial decision information is always subjectively given by experts in the form of crisp values and calculated to obtain an individual or group initial direct-relation (IDR) matrix by simple operations (e.g., weighted sum). However, such descriptions and operations are considered to hardly reflect the vagueness of the real world [17]. Therefore, scholars have carried out some research to improve the DEMATEL combined with fuzzy theory [18]. Several fuzzy DEMATEL methods have been introduced. For examples, Abdullah [19] introduced interval-valued intuitionistic fuzzy numbers to improve the judgement of DEMATEL in a group decision-making (GDM) environment. Addae [20] used a two-step fuzzy DEMATEL method to solve a practical problem. Asan [21] proposed a new interval-valued hesitant fuzzy approach to DEMATEL to explicitly deal with hesitation in expert assessments and offered a better representation of uncertainty, etc [22]–[25].

Obviously, all these extensions have made great contributions to the DEMATEL method, and its ability of dealing with complex problems can be strengthened to some extent. As currently defined, instead of one expert, a panel gives the assessments on the influence relations of factors, and multiple experts arrive at an acceptable result [26]. This process makes it necessary to extend the traditional DEMATEL method to a group DEMATEL method, which belongs to GDM problems. Although some literature considers the DEMATEL from the perspective of GDM, we believe that these group DEMATEL methods have a lot of room for further improvement in both expert assessment extraction and group IDR matrix construction.

Firstly, the expert assessment extraction of group DEMATEL method should be improved to obtain accurate decision information in accordance with experts' cognitive competence. As introduced earlier, DEMATEL is a GDM method totally

based on expert assessments to conduct later computation and analyzation. Considering the complexity of reality and individual characteristics of experts, there is no doubt that not every expert is proficient in all areas. In other words, experts may give incomplete or uncertain assessments to the influence relations among factors for a specific complex practical problem. However, all of the existing group DEMATEL methods default that each expert can give a definite assessment for every pair of factors by means of crisp values or fuzzy numbers, and neglect the problem that the assessments given by experts may be incomplete in practice. If experts who cannot give definite assessments with crisp values or fuzzy numbers are required to give the assessments in those forms, the final decision of DEMATEL resulted from the ineffective assessments may be erroneous. Therefore, understanding how to depict and fuse incomplete information from experts is of great importance to improve the properties of the group DEMATEL method. Fortunately, the basic belief assignment (BBA) function, as a key concept in the Dezert-Smarandache theory (DSMT) of evidence [27], can directly express the uncertainty by assigning probability to the subsets of a set composed of multiple objects rather than to each individual object [28]. The BBA functions generated from different evidence sources (experts) could be well fused by the proportional conflict redistribution rule no. 5 (PCR5). All these features exactly meet the requirements of the group DEMATEL method. Accordingly, DSMT is used to extract and fuse expert assessments to derive the group IDR matrix in this paper. Moreover, the differences in experts' knowledge backgrounds and cognitive abilities on a particular problem are reflected by expert weights [29]. Expert weights are reflected as discounting parameters to reflect one's relative importance in a group during fusion process. In this paper, expert weights are calculated based on similarity functions of expert assessments.

Secondly, the group IDR matrix construction of group DEMATEL method should be improved to obtain the acceptable decision results in accordance with experts' satisfaction. In the GDM, "group" refers to not only the number of experts merely, but also the experts who have common interests in reaching a consensus for the ultimate satisfactory results despite individual differences. This principle helps to reduce biased evaluations and inherent partiality in GDM processes [30]. Unfortunately, the group IDR matrix in the existing group DEMATEL methods is frequently constructed by making arithmetic average values for individual IDR matrices, while whether experts agree with the group results is scarcely concerned. The group IDR matrix plays a fundamental role in the entire DEMATEL processes and it has a significant influence on the effectiveness of final results. If strong inconsistency and conflict exist among experts, the group IDR matrix may not be able to precisely describe the real influence relations of factors. Therefore, it is particularly important to construct a group IDR matrix according to the consensus rules, that is, to construct a group consensus IDR matrix by reaching general or widespread agreement among the experts involved in the GDM processes [31]. Fortunately, Herrera-Viedma proposed

a rational consensus model in GDM composed of a selection process and a consensus-reaching process (CRP), which had become a hot issue in the recent GDM area. The CRP has been successfully introduced to make GDM with different situations, such as hesitant fuzzy preference relations, Delphi processes, multi-attribute large-scale GDM, sentiment analysis, and virtual reality industry [3]–[6], [32]. Traditionally, unanimous agreement of all experts in CRP is required. However, the desired result can hardly be achieved because of the diversity of opinions, knowledge, and experiences of experts. Therefore, the concept of "soft consensus" has been provided, in which, "soft" means better reflecting all possible levels of agreement by setting an acceptable consensus level (CL) threshold value (such as 0.8 rather than 1) and guiding the CRP until a high-level agreement is achieved among the individuals. Soft consensus can be reached through an iterative dynamic process with several collection and adjustment rounds [33], [34]. Hence, in this paper, we consider a soft CRP in the group IDR matrix construction, and transform the original static group DEMATEL problems into dynamic ones.

The motivation of this paper is to improve group DEMATEL method according to the following three aspects. Firstly, the BBA function is used to extract expert assessments to accurately express uncertainty and incompleteness, thus reducing the loss of accuracy. Secondly, the initial assessments are discounted with expert weights by using Shafer's discounting method. Moreover, the PCR5 of DSMT is used to fuse the discounted assessments to overcome the defects in the intuitional paradox of Dempster's combination rule. Thirdly, we apply a soft CRP to help reach an acceptable CL in the construction of group IDR matrix to ensure the consistency and satisfaction among experts. This paper is organized as follows. Section II briefly introduces the basic knowledge of DEMATEL and DSMT. In Section III, the DSMT-based group DEMATEL method with reaching consensus is proposed, and the corresponding algorithm is summarized. In Section IV, the numerical comparison and discussion are provided to demonstrate the performances of the proposed method and algorithm. In Section V, the conclusion is drawn and the future research directions are briefly discussed.

## II. PRELIMINARIES

In order to facilitate the later formulation, some basic concepts of DEMATEL and the Dezert-Smarandache theory (DSMT) are given in this section.

### A. DEMATEL method

The DEMATEL method is an effective way to analyze the influence relations among factors of a system. Through an analysis of the total influence relation of factors by the DEMATEL method, we can obtain a better understanding of structural relations and an ideal way to solve complicated system problems. Consider that a group of experts are invited to assess the influence relations for a set of factors  $F = \{f_1, \dots, f_L\}$  with a set of grade levels  $\{0, 1, 2\}$ , where the expert set is  $E = \{e_1, \dots, e_K\}$ , and the degree of influence to which he

or she believes factor  $f_i$  has an effect on factor  $f_j$  (denoted by  $f_i \rightarrow f_j$ ) is assessed by expert  $e_k$  and denoted as  $g_{ij}^k, \forall i, j, k$ . The meaning of each element in the set of grade levels is that 0 denotes “no influence”, 1 denotes a “low influence” and 2 denotes a “very high influence”. The central concepts in the DEMATEL are defined as follows.

**Definition 1** [35]. Suppose a pairwise comparison of influence degree from the  $i$ -th to the  $j$ -th factor given by the  $k$ -th expert  $e_k$  is denoted as  $g_{ij}^k$  with 0–2 grade levels, and the grade levels given by each expert form a  $L \times L$  non-negative answer matrix  $G^k = [g_{ij}^k]_{L \times L}$ ,  $k = 1, \dots, K$ . The group IDR matrix, which represents the initial direct relation between each pair of factors derived from experts, is obtained by calculating the average values of all experts’ answer matrices as  $G = [g_{ij}]_{L \times L}$ , where  $g_{ij} = \sum_{k=1}^K g_{ij}^k / K$ , for  $i, j = 1, \dots, L$ .

In Definition 1, the 0, 1, 2 grade levels mean “no influence”, “medium influence”, and “high influence”, respectively. Note that, the diagonal elements of each answer matrix  $G^k$  are all set to zero, which means that the factors do not influence themselves.

**Definition 2** [35]. The maximal row-wise and column-wise sum of matrix  $G$  is calculated by  $g' = \max(\max_{1 \leq i \leq L} \sum_{j=1}^L g_{ij}, \max_{1 \leq j \leq L} \sum_{i=1}^L g_{ij})$ ; then the normalized IDR matrix  $D = [d_{ij}]_{L \times L}$  can be computed according to (1).

$$D = G/g'. \quad (1)$$

**Definition 3** [35]. Suppose the direct and indirect relations among several factors are represented by the total relation matrix, and it is defined as in (2).

$$A = \lim_{N \rightarrow \infty} (D + D^2 + \dots + D^N) = D(I - D)^{-1}. \quad (2)$$

Some kinds of extensions are further discussed to strengthen the original DEMATEL. One kind of extensions is used to overcome the drawback that raising the normalized IDR matrix to the power of infinity may not converge to zero, and hence, the total relation matrix may not converge (see (2)). A very small positive number  $\mu$  (e.g.,  $\mu = 10^{-5}$ ) is introduced in the maximal row-wise and column-wise sum of matrix  $G$  as  $g'' = \max(\max_{1 \leq i \leq L} \sum_{j=1}^L g_{ij}, \mu + \max_{1 \leq j \leq L} \sum_{i=1}^L g_{ij})$ . Other steps remained unchanged as in the original DEMATEL. The revised DEMATEL guarantees that the normalized IDR matrix to infinite power will converge to zero and that the total relation matrix can be obtained smoothly [36].

**Definition 4** [35]. Suppose  $r$  and  $c$  represent the sum of rows and the sum of columns of the total relation matrix  $A$ . According to  $A = [a_{ij}]_{L \times L}$ ,  $r$  and  $c$  can be defined as follows:

$$\begin{aligned} r &= [r_i]_{L \times 1} = \left( \sum_{j=1}^L a_{ij} \right)_{L \times 1}, \\ c &= [c_i]_{1 \times L} = \left( \sum_{j=1}^L a_{ji} \right)_{1 \times L}. \end{aligned} \quad (3)$$

where  $r_i$  shows the total influence, both direct and indirect, given by the factor  $f_i$  to other factors;  $c_i$  shows the total influence, both direct and indirect, received by the factor  $f_i$  from other factors;  $r_i + c_i$  is defined as the prominence, showing the degree of the important role that the factor  $f_i$  plays in the complex system; and  $r_i - c_i$  shows the net influence that the factor  $f_i$  contributes to the complex system. Note that, if  $r_i - c_i$  is positive, the factor  $f_i$  is a net causer; if  $r_i - c_i$  is negative, the factor  $f_i$  is a net receiver.

### B. Dezert-Smarandache theory

DSMT, jointly proposed by Dezert and Smarandache [37], can be used to obtain more accurate fusion results of BBA functions especially in high conflicting information cases. It has a series of proportional conflict redistribution rules to make fusion for evidences [38], among which, PCR5 is the most widely used one with the advantages in dealing with conflict belief functions. For example, it provides the appropriate redistribution of conflict beliefs and can produce a reasonable fusion result even in highly conflicting cases. These attractive features motivate the use of DSMT in GDM problems, such as map reconstruction of robot [39], decision making support [40], [41], target type tracking [42], image processing [43], data classification [44]–[48], clustering [47], [49], [50], and so on.

In DSMT framework, the frame  $\Theta = \{\theta_1, \dots, \theta_Y\}$  is a finite set of  $Y$  exhaustive propositions that are not necessarily mutually exclusive. The hyper-power set  $D^\Theta$  is defined as the set of all composite propositions built from elements of  $\Theta$  with  $\cup$  and  $\cap$  operators, such that [51]:

- (i)  $\emptyset, \theta_1, \dots, \theta_Y \in D^\Theta$ ;
- (ii) if  $\theta_y, \theta_{y'} \in D^\Theta$ , then  $\theta_y \cup \theta_{y'} \in D^\Theta$  and  $\theta_y \cap \theta_{y'} \in D^\Theta$ ;
- (iii) No other elements belong to  $D^\Theta$ , except those obtained by using rules (i) or (ii).

**Definition 5** [37]. Suppose  $\Theta = \{\theta_1, \dots, \theta_Y\}$  is a set of exhaustive propositions; then the basic belief assignment is defined over the hyper-power set  $D^\Theta$ . If the mapping function  $m : D^\Theta \rightarrow [0, 1]$  could fulfill the following:

$$m(\emptyset) = 0, \quad \sum_{\theta \in D^\Theta} m(\theta) = 1, \quad (4)$$

then  $m(\cdot)$  is called the BBA function. If  $m(\theta) > 0$ ,  $\theta$  is called a focal element.

In DSMT framework, PCR5 for making fusion for two pieces of evidence is introduced as follows.

**Definition 6** [37]. Suppose the BBA functions of two pieces of evidence are  $m_1$  and  $m_2$  on  $D^\Theta$ ; then, PCR5 to fuse  $m_1$  and  $m_2$  can be defined as follows for all  $\theta \in D^\Theta$ :

$$m_{PCR5}(\theta) = \begin{cases} 0, & \text{if } \theta = \emptyset, \\ \sum_{\substack{\theta' \cap \theta'' = \theta \\ \theta', \theta'' \subseteq D^\Theta}} m_1(\theta') m_2(\theta'') \\ \quad + \sum_{\substack{\theta''' \in D^\Theta \\ \theta \cap \theta''' = \emptyset}} \left[ \frac{m_1(\theta)^2 m_2(\theta''')}{m_1(\theta)^2 + m_2(\theta''')} \right. \\ \quad \left. + \frac{m_2(\theta)^2 m_1(\theta''')}{m_2(\theta)^2 + m_1(\theta''')} \right], & \text{if } \theta \neq \emptyset. \end{cases} \quad (5)$$

### III. THE PROPOSED METHOD

In the proposed method, we construct the group IDR matrix in the DEMATEL by following two steps. Firstly, we develop an expert assessments extraction and fusion mechanism with DSMT to obtain a temporary group IDR matrix (see III-A). Secondly, we activate the soft CRP based on the current individual and group IDR matrices to reach a soft consensus among experts (see III-B).

#### A. Expert Assessment Extraction and Fusion

In traditional DEMATEL, experts are advised to give their assessments of the influence relations among factors by means of crisp values to construct the group IDR matrix  $G = [g_{ij}]_{L \times L}$  (see Definition 1), which is a rough extraction method with low precision. When restricted by expert experiences and knowledge, the assessments given by experts may be ignorant and partially credible. Taking the influence degree for  $f_i \rightarrow f_j$  and expert  $e_k$  as an example, the expert may well know this problem and he or she can give a definite result by using one of the given grade levels  $\{0, 1, 2\}$ . On the contrary, the expert may not know this problem at all, in which case he or she cannot give any assessment information about the influence degree. In most instances, the expert knows this problem to a certain extent and he or she may point out that the influence degree belongs to several grade levels but may not be sure which is the best one. Obviously, experts cannot give their assessments by means of crisp values in the latter two situations—that is, the original extraction method is not practical and does not consider experts' personality. Thus, we employed the BBA function to extract expert assessments in this paper as shown in Fig. 1.

As shown in Fig. 1, the expert assessment extraction and fusion mechanism includes three major steps. (1) Extracting expert assessments with BBA function; (2) Discounting these assessments with Shafer's discounting method; (3) Fusing the discounted assessments with PCR5. Next, we will give a detailed definition and explanation of the procedures involved in the above processes.

Let the frame of discernment be  $\Theta = \{\theta_1, \theta_2, \theta_3\} = \{0, 1, 2\}$ , which can be seen as the discernment frame of DSMT. Expert  $e_k$  is asked to assess the influence degree for  $f_i \rightarrow f_j$  and his/her assessment is allowed to be expressed by the BBA function as given in Definition 5 and shown in (6).

$$b_{ij}^k = \{(\theta, b_{ij}^k(\theta)) \mid \sum_{\theta \subseteq \Theta} b_{ij}^k(\theta) = 1; b_{ij}^k(\theta) \geq 0, \theta \subseteq \Theta, \forall i, j, k\}. \quad (6)$$

All of the BBA functions make up the individual IDR matrix for expert  $e_k$ , denoted as  $B^k = [b_{ij}^k]_{L \times L}$ , with  $i, j = 1, \dots, L$  and  $k = 1, \dots, K$ . Note that the basic beliefs in  $b_{ij}^k$  can be assigned not only to singleton grade levels but also to any subsets of  $\Theta$ , thereby it is allowed such an assessment (also called a piece of evidence) to be profiled by a BBA defined on the hyper-power set  $D^\Theta$ . It is capable of reflecting ignorance in expert assessments, and the basic beliefs in  $b_{ij}^k$

can be given to  $\Theta$  (global ignorance) or to  $\theta \subset \Theta$  (local ignorance) according to the unknown and partial assessments [52]. Expert assessments could be expressed precisely through BBA functions as discussed, laying a great foundation for later DEMATEL procedures.

**Example 1.** Assume expert  $e_1$  points out the influence degree for  $f_i \rightarrow f_j$  has 30% probabilities belonging to  $\theta_1$  and 70% probabilities belonging to  $\theta_2$ . Thus, his or her assessment is described as  $b_{ij}^1(\theta) = \{(\theta_1, 0.30), (\theta_2, 0.70)\}$ . Expert  $e_2$  points out the influence degree has 20% probabilities belonging to  $\theta_1$  and has 80% probabilities belonging to  $\theta_2 \cup \theta_3$  but is not sure which is the best one. Therefore, his or her assessment is described as  $b_{ij}^2(\theta) = \{(\theta_1, 0.20), (\theta_2 \cup \theta_3, 0.80)\}$ . Expert  $e_3$  points out the influence degree for  $f_i \rightarrow f_j$  has 100% probabilities belonging to  $\theta_2 \cap \theta_3$ . Thus, his or her assessment is described as  $b_{ij}^3(\theta) = \{(\theta_2 \cap \theta_3, 1.00)\}$ . Expert  $e_4$  cannot give any information about the influence degree. Therefore, his or her assessment is described as  $b_{ij}^4(\theta) = \{(\Theta, 1.00)\}$ .

Normally, expert weight is subjective and relative to reflect the importance of one's assessments in a group, and it is usually denoted by  $w$  in  $[0, 1]$ , with 0 and 1 respectively standing for not important at all and the most important [53], [54]. It is obvious to find that the expert weight can be determined by AHP [55], ANP [56] or Delphi [57], and it can also be determined subjectively according to the requirements of actual issues [58]. As discussed, expert weight is used to account for one's relative importance among all experts—that is, the closer one's assessments are to others' assessments, the more important the expert is likely to be [40]. Hence, in our opinion, expert weight could be calculated indirectly based on the similarity between one expert and other experts—that is, the higher the similarity of assessments between the expert and others, the larger the weight of the expert. Expert weight is directly proportional to the similarity between expert assessments and others' assessments. Thus, we use evidential distance to depict the similarity of expert assessments, based on which we could calculate expert weight in simple ways. The Euclidean evidential distance and Euclidean evidential similarity, which have little computation complexity and fast convergence speed, are defined as follows.

**Definition 7** [59]. Suppose  $m_1$  and  $m_2$  are two BBA functions for the same frame of discernment  $\Theta$ ,  $\theta_n$  is the  $n$ -th element of  $D^\Theta$ , and  $|D^\Theta|$  is the cardinality of  $D^\Theta$ . The distance between  $m_1$  and  $m_2$  is defined as follows.

$$Dist_E(m_1, m_2) = \frac{1}{\sqrt{2}} \sqrt{\sum_{n=1}^{|D^\Theta|} [m_1(\theta_n) - m_2(\theta_n)]^2}. \quad (7)$$

**Definition 8** [59]. Suppose  $m_1$  and  $m_2$  are two BBA functions for the same frame of discernment  $\Theta$ ,  $\theta_n$  is the  $n$ -th element of  $D^\Theta$ , and  $|D^\Theta|$  is the cardinality of  $D^\Theta$ . The Euclidean similarity function  $Sim_E(m_1, m_2)$  is defined based on the Euclidean evidential distance as follows.

$$Sim_E(m_1, m_2) = 1 - \frac{1}{\sqrt{2}} \sqrt{\sum_{n=1}^{|D^\Theta|} [m_1(\theta_n) - m_2(\theta_n)]^2}. \quad (8)$$

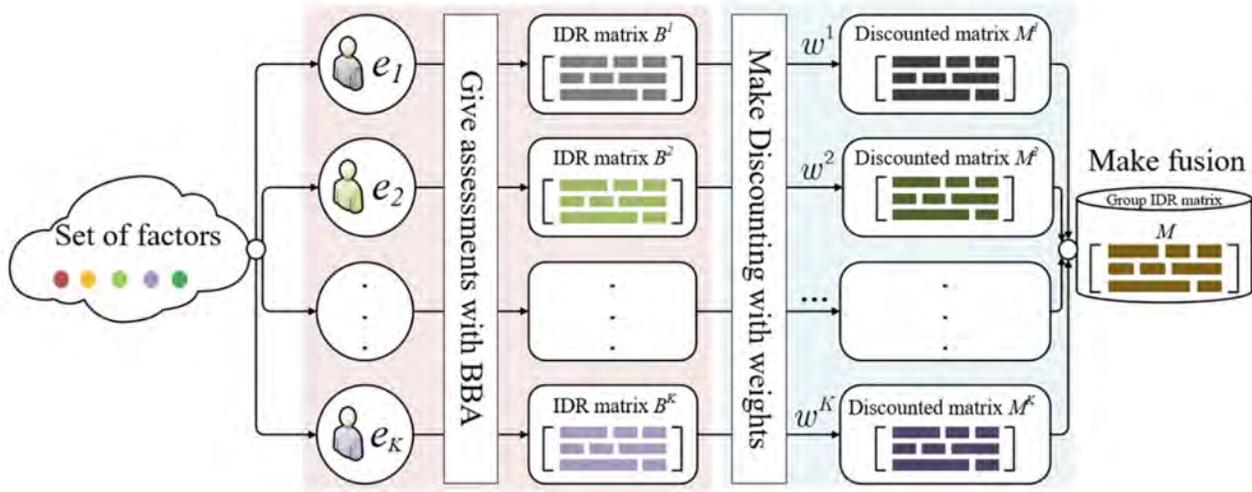


Figure 1. Expert assessments extraction and fusion mechanism.

**Example 2.** Assume two pieces of evidence,  $b_1 = \{(\theta_2, 1)\}$  and  $b_2 = \{(\theta_1, 0.4), (\theta_2, 0.6)\}$ , and that the distance between them could be computed by (7):  $Dist_E(m_1, m_2) = \frac{1}{\sqrt{2}}\sqrt{(0 - 0.4)^2 + (1 - 0.6)^2} = 0.4$ . The similarity between them is computed by (8):  $Sim_E(m_1, m_2) = 1 - 0.4 = 0.6$ .

On the basis of the previous definitions, the similarity between experts can be directly calculated. Thus, we define the following expert weight calculation method.

**Definition 9** [60]. Suppose the individual IDR matrix consisting of all influence relations among factors given by expert  $e_k$  is  $B^k = [b_{ij}^k]_{L \times L}$ , with  $i, j = 1, \dots, L$  and  $k = 1, \dots, K$ . The similarity between experts  $e_k$  and  $e_{k'}$  on each pair of factors  $f_i \rightarrow f_j$  is  $Sim_E(b_{ij}^k, b_{ij}^{k'})$ , and the similarity between any two experts is computed as  $s^{kk'} = \sum_{i,j=1, i \neq j}^L Sim_E(b_{ij}^k, b_{ij}^{k'}) / L^2$ , where  $L^2$  denotes the quantity of factor pairs, making up the similarity matrix as follows.

$$S = [s^{kk'}]_{K \times K} = \begin{bmatrix} 1 & \dots & s^{1k} & \dots & s^{1K} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s^{k1} & \dots & 1 & \dots & s^{kK} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s^{K1} & \dots & s^{Kk} & \dots & 1 \end{bmatrix}. \quad (9)$$

The support of  $e_k$  can be obtained by adding all of the elements in the similarity matrix  $S$  that are related to expert  $e_k$  except for self-similarity, i.e.,  $Sup(e_k) = \sum_{k'=1, k' \neq k}^K s^{kk'}$ ,  $k = 1, \dots, K$  where  $Sup(e_k)$  represents the support degree of expert  $e_k$  received from other experts. By normalizing them, we could obtain the expert credibility  $Crd(e_k)$  which is generally regarded as expert weight  $w^k$  as follows.

$$w^k = Crd(e_k) = Sup(e_k) / \sum_{k=1}^K Sup(e_k), \quad k = 1, \dots, K. \quad (10)$$

**Example 3.** Assume experts  $e_1, e_2, e_3$  give their assessments for the influence relations among factor set  $F = \{F_1, F_2, F_3\}$ , respectively, as follows:

$$B^1 = \begin{bmatrix} \{(\theta_1, 1)\} & \{(\theta_2, 1)\} & \{(\theta_3, 1)\} \\ \{(\theta_2, 0.7), (\theta_3, 0.3)\} & \{(\theta_1, 1)\} & \{(\theta_1, 0.4), (\theta_3, 0.6)\} \\ \{(\theta_1, 1)\} & \{(\theta_2, 1)\} & \{(\theta_1, 1)\} \end{bmatrix},$$

$$B^2 = \begin{bmatrix} \{(\theta_1, 1)\} & \{(\theta_2, 1)\} & \{(\theta_3, 1)\} \\ \{(\theta_1, 0.8), (\theta_2, 0.2)\} & \{(\theta_1, 1)\} & \{(\theta_1, 1)\} \\ \{(\theta_1, 1)\} & \{(\theta_1, 0.5), (\theta_2, 0.5)\} & \{(\theta_1, 1)\} \end{bmatrix},$$

$$B^3 = \begin{bmatrix} \{(\theta_1, 1)\} & \{(\theta_2, 1)\} & \{(\theta_1, 0.2), (\theta_2 \cup \theta_3, 0.8)\} \\ \{(\theta_2 \cup \theta_3, 1)\} & \{(\theta_1, 1)\} & \{(\theta_1, 1)\} \\ \{(\theta_1, 1)\} & \{(\theta_1, 1)\} & \{(\theta_1, 1)\} \end{bmatrix}.$$

The similarity matrix can be calculated by Definition 9 as follows:

$$S = \begin{bmatrix} 1.00 & 0.47 & 0.29 \\ 0.47 & 1.00 & 0.41 \\ 0.29 & 0.41 & 1.00 \end{bmatrix}.$$

We calculate the support degree of experts as  $Sup(e_1) = 0.47 + 0.29 = 0.76$ ,  $Sup(e_2) = 0.47 + 0.41 = 0.88$ ,  $Sup(e_3) = 0.29 + 0.41 = 0.70$ . Then we compute the expert weight as  $w^1 = Crd(e_1) = 0.76 / (0.76 + 0.88 + 0.70) \approx \{(\theta_1, P_{ij}^1), (\theta_2, P_{ij}^2), (\theta_3, P_{ij}^3)\}$ ,  $w^2 = 0.87 / 2.34 \approx 0.37$ ,  $w^3 = 0.70 / 2.34 \approx 0.30$ .

As mentioned above, expert weight should be considered through Shafer's discounting method, which multiplies the masses of focal elements by the expert weight and transfers all of the remaining discounted mass to the full ignorance  $\Theta$ . Mathematically, Shafer's discounting method can be given as in Definition 10.

**Definition 10** [61]. Suppose the BBA function is  $b$  as in (4), and  $w$  is a parameter to discount the evidence,  $0 \leq w \leq 1$ . Then, Shafer’s discounting method is defined as follows:

$$m(\theta) = \begin{cases} 0, & \theta = \emptyset, \\ w \cdot b(\theta), \theta \subseteq \Theta & \theta \neq \emptyset, \\ w \cdot b(\theta) + 1 - w, & \theta = \Theta. \end{cases} \quad (11)$$

If the sum of weights is equal to 1, then the discounting parameter is usually derived by standardizing the weights as  $\bar{w}^k = w^k / \max(w^k | k = 1, \dots, K)$ . The discounted expert assessments could be described as follows.

$$m_{ij}^k = \{(\theta, m_{ij}^k(\theta)) | \sum_{\theta \subseteq \Theta} m_{ij}^k(\theta) = 1; m_{ij}^k(\theta) \geq 0, \theta \subseteq \Theta, \forall i, j, k\}. \quad (12)$$

The discounted individual IDR matrix is denoted as  $M^k = [m_{ij}^k]_{L \times L}$ , with  $i, j = 1, \dots, L$  and  $k = 1, \dots, K$ .

**Example 4.** Assume expert weight is  $W = \{w_1 = 0.40, w_2 = 0.20, w_3 = 0.20, w_4 = 0.20\}$ . The assessments given by four experts are the same as in Example 1. The discounting parameters for the four experts can be standardized as  $\bar{w}^1 = w_1 / \max(w_1, \dots, w_4) = 0.40/0.40 = 1.00$ , and similarly we get  $\bar{w}^2 = \bar{w}^3 = \bar{w}^4 = 0.20/0.40 = 0.50$ . Taking them into (11), the discounted assessments are as follows:  $m_{ij}^1(\theta) = \{(\theta_1, 0.30), (\hat{a}\theta_2, 0.70)\}$ ,  $m_{ij}^2(\theta) = \{(\theta_1, 0.10), (\theta_2 \cup \theta_3, 0.70), (\Theta, 0.50)\}$ ,  $m_{ij}^3(\theta) = \{(\theta_2 \cap \theta_3, 0.50), (\Theta, 0.50)\}$ , and  $m_{ij}^4(\theta) = \{(\Theta, 1.00)\}$ .

The DSMT framework with PCR5 rule can be used to make fusion for individual discounted assessments as in (12) and the fusion result can be described as follows:

$$m_{ij} = \{(\theta, m_{ij}(\theta)) | \sum_{\theta \subseteq \Theta} m_{ij}(\theta) = 1; m_{ij}(\theta) \geq 0, \theta \subseteq \Theta, \forall i, j\}. \quad (13)$$

**Example 5.** Assume two pieces of evidence are:  $m_{ij}^1(\theta) = \{(\theta_1, 0.20), (\theta_2, 0.30), (\Theta, 0.50)\}$ , and  $m_{ij}^2(\theta) = \{(\theta_2, 0.70), (\Theta, 0.30)\}$ . Taking them into (5), the fusion results are obtained as  $m_{ij}(\theta) = \{(\theta_1, 0.09), (\theta_2, 0.76), (\Theta, 0.15)\}$ .

**B. The Soft CRP for Group DEMATEL**

Soft CRP can be designed with the aim of supporting experts until a group consensus is reached by following several discussion and adjustment rounds. As discussed in Section I, situations in which all of the experts agree with each other unanimously are rare or not desirable in the decision making process. Since the “unanimous consensus” of conflict tolerance, which has the ability to satisfy all pairs of BBA functions, rarely exists, the choice of a “soft consensus” is largely subjective and application oriented [62]. That is, setting an acceptable CL threshold value (suppose that the value is 0.8 there) to guide the whole process to reach group consensus. Basically, this process relies on making assumptions about experts’ willingness to change their opinion or preferences

[63], [64]. Initially, consensus measures are computed based on individual and group IDR matrices to determine whether an acceptable CL has been reached or not. If so ( $CL \geq 0.8$ ), the soft CRP is finished and a consensus group IDR matrix can be obtained. Otherwise ( $CL < 0.8$ ), the feedback mechanism will be activated, and the experts who are not contributing to the consensus are identified and the advice about how to alter their assessments is generated [65].

This consensus measure can indicate the current consensus situation throughout the soft CRP. According to the characteristics of group DEMATEL, the consensus measure is categorized into two levels, i.e., pair-factors level and expert level. The pair-factors level is the most basic level and reflects the original conflict degree. At this level, experts make some adjustments to increase consistency among the group. The expert level is the highest level, which reflects the conflict degree between a specific expert and the group as a whole for the collected results on all pairs of factors [66]. We employ these two levels of consensus measures to identify inconsistencies between experts and group.

The calculation methods for consensus measures can be derived based on the similarities between these assessments. At the pair-factors level, the consensus measure values can be calculated based on the Euclidean similarity function as in Definition 11. At the expert level, the consensus measure values can be calculated by adding the consensus measure values on all of pair-factors for an expert as in Definition 12. Obviously, the order for calculating the two levels of consensus measure is pair-factors level at first and then expert level. When determining the inconsistency assessments that need to be modified, the order is the highest level at first and then the most basic level. Experts only adjust the conflicting assessments based on the recommendations at the pair-factors level.

**Definition 11.** In the  $t$ -th round, suppose the assessments of the pair of factors  $f_i \rightarrow f_j$  given/calculated by expert  $e_k$  and group are  $b_{ij}^{k,t}$  and  $m_{ij}^t$ . Then, the consensus measure values can be calculated by the Euclidean similarity function as follows.

$$c_{ij}^{k,t} = 1 - \frac{1}{\sqrt{2}} \sqrt{\sum_{n=1}^{|D^\Theta|} [b_{ij}^{k,t}(\theta_n) - m_{ij}^t(\theta_n)]^2}. \quad (14)$$

**Definition 12.** In the  $t$ -th round, suppose the consensus measure value of expert  $e_k$  for  $f_i \rightarrow f_j$  at pair-factors level is  $c_{ij}^{k,t}$ ,  $\forall i, j, k$ , then the consensus measure value of expert  $e_k$  at expert level in this round can be calculated as follows.

$$c^{k,t} = \sum_{i,j=1, i \neq j}^L c_{ij}^{k,t} / L^2, \quad (15)$$

where  $L^2$  denotes the quantity of factor pairs.

The closer the value  $c^{k,t}$  is to 0, the greater the conflict degree; the closer the value is to 1, the smaller the conflict. Obviously,  $c^{k,t} = 0$  indicates complete conflict between expert  $e_k$  and group, and  $c^{k,t} = 1$  indicates no conflict between them.

If each expert's consensus measure  $c^{k,t}$  is larger than the acceptable CL threshold value, indicating that existing conflict degree can be accepted by all experts and group consensus is reached, then the soft CRP is finished and the current collected group IDR matrix  $M^t = [m_{ij}^t]_{L \times L}$  is taken as a consensus result. Otherwise, if one's consensus measure  $c^{k,t}$  is lower than the acceptable CL threshold value, indicating that strong conflict degree among experts, then the feedback mechanism is carried out to help the inconsistent experts adjust their assessments to enhance group consensus, and a new round ( $t + 1$ ) of group IDR matrix construction should be initiated.

Note that, expert weights may be not fixed during several of the rounds for the reason that expert assessments may be changed in the context of not reaching the acceptable CL. Thus, in each new round, we need to recalculate expert weights based on the latest assessments. In order to prevent the collective assessments from failing to converge after several discussion rounds, we incorporate a maximum number of rounds ( $t_{MAX}$ ) in the soft CRP to develop. It can be ensured that the feedback mechanism will not be carried out when and the current collected group IDR matrix will be taken as the final result even if the acceptable CL has not been reached yet.

Suppose the acceptable CL threshold values at expert level and pair-factors level are  $\varepsilon_e$  and  $\varepsilon_f$ . Then the processes of feedback mechanism can be divided into the following two steps.

- Step 1: Experts whose consensus measure values at expert level are lower than the threshold value  $\varepsilon_e$  in the  $t$ -th round are identified as follows.

$$\tilde{E}_t = \{k | c^{k,t} \leq \varepsilon_e\}. \quad (16)$$

- Step 2: For the identified experts in step 1, their assessments for such pairs of factors that the consensus measure values are lower than the threshold value are identified as follows.

$$\tilde{F}_t = \{k, i \rightarrow j | c_{ij}^{k,t} \leq \varepsilon_f \wedge k \in \tilde{E}_t\}. \quad (17)$$

When  $\tilde{E}_t$  and  $\tilde{F}_t$  have been identified, personal advices for experts will be generated to reduce the conflict caused by the moderator. Since pair-factors level is the most basic level and contains the original assessments of experts, only the experts whose consensus measure values  $c^{k,t}$  are lower than the threshold value  $\varepsilon_e$  may obtain the adjusted advices. After the identified experts all finish their adjustments, a new round of group IDR matrix construction will be carried out to determine the consensus situation. It is obvious to find that the above procedure is very similar to the Delphi method.

### C. DSMT-Based Group DEMATEL Method and Algorithm

The construction of a group IDR matrix includes two main processes. The first is collection process, which focuses on expert assessment extraction and fusion; and the second is soft consensus reaching process, which aims at reaching an acceptable CL among experts. After the above two processes,

the final group IDR matrix  $M$  can be constructed. Obviously,  $M$  is different from the original IDR matrix  $G$  as given in Definition 1 for two reasons. (1)  $M$  is a consensus group IDR matrix that satisfies all experts, while  $G$  does not consider consensus and satisfaction of experts. (2)  $M$  is made up of BBA functions that is capable of reflecting local or global ignorance in experts' mind, while  $G$  is made up of crisp values. Note that, the IDR matrix included in the DEMATEL is required to be exact influence degrees rather than BBA functions. Therefore, we apply the generalized pignistic probability as in Definition 13, which is an extension of the original pignistic probability [67], to reassign the local and global ignorance in the BBA functions to singleton grade levels. Then, we calculate the expected value to derive the exact influence degrees.

**Definition 13** [68]. Suppose the frame of discernment in the DSMT framework is  $\Theta = \{\theta_1, \dots, \theta_Y\}$ , then the generalized pignistic probability for all  $A \in D^\Theta$  can be calculated by (18).

$$P(A) = \sum_{X \in D^\Theta} \frac{|X \cap A|}{|X|} m(X). \quad (18)$$

where  $|X|$  denotes the cardinal of proposition  $X$ .

Taking  $m_{ij}$  in the final group IDR matrix  $M$  into (18), the transformation results can be obtained and recorded as  $\{(\theta_1, P_{ij}^1), \dots, (\theta_Y, P_{ij}^Y)\}$ , where  $P_{ij}^y$  denotes the probability of influence degree  $\theta_y$ ,  $y = 1, \dots, Y$ . The subscript  $Y$  is equal to 3 for the reason that the frame of discernment is defined as  $\Theta = \{\theta_1, \theta_2, \theta_3\} = \{0, 1, 2\}$  in this paper. Which influence grade level or influence degree is attached to the relationship of  $f_i \rightarrow f_j$ ? We follow two kinds of principles to solve this problem: (1) Highest probability principle. The grade level with the highest probability  $\tilde{g}_{ij} = \theta_*$  is chosen as the final influence grade level, where  $\{\theta_* | P_{ij}^* = \max(P_{ij}^1, \dots, P_{ij}^Y)\}$ ; (2) Expected value principle. The expected value  $\tilde{g}_{ij} = \sum_{y=1}^Y \theta_y \times P_{ij}^y$ ,  $\forall i, j$  is calculated as the final influence degree. According to the real situation, we obtain the final influence degree for  $f_i \rightarrow f_j$  by using one of the two principles, and then the final group IDR matrix  $\tilde{G} = [\tilde{g}_{ij}]_{L \times L}$  can be constructed.

The process of DSMT-based group DEMATEL is illustrated as in Fig. 2. As shown in Fig.2, the complete steps of DSMT-based group DEMATEL can be summarized as follows.

- Step 1: *Define group DEMATEL problem and soft CRP parameters.* Suppose the set of factors is  $F = \{f_1, \dots, f_L\}$ , the set of experts is  $E = \{e_1, \dots, e_K\}$ , the set of grade levels is  $\Theta = \{0, 1, 2\}$ , and the threshold value to filter out major factors is  $\eta$ . Then, the acceptable CL threshold values at expert level and pair-factors level are  $\varepsilon_e$  and  $\varepsilon_f$ , the round counter is  $t$  ( $t$  is set to one at first), and the maximum number of rounds is  $t_{MAX}$ . A moderator is invited to participate in the group decision making process and responsible for managing the whole process such as consensus measure value calculation, inconsistent expert identification, the maximum number of round judgement.
- Step 2: *Expert assessments extraction and fusion.* Experts are advised to indicate the influence degree to which he or

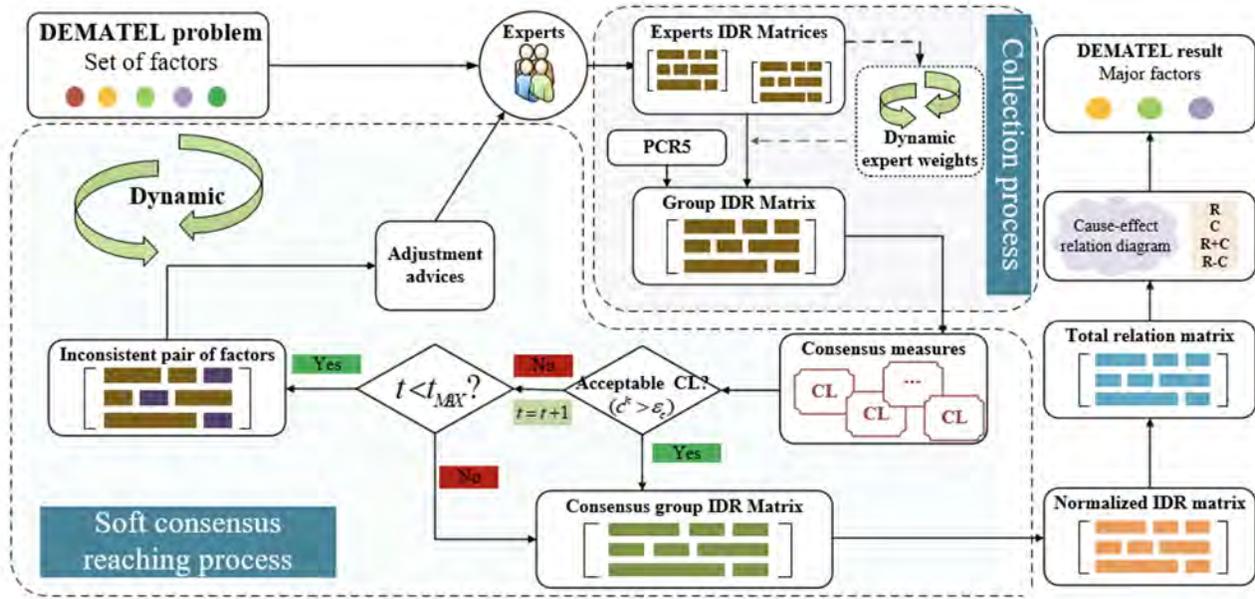


Figure 2. The process of proposed method.

she believes factor  $f_i$  has an effect on factor  $f_j$  (denoted by  $f_i \rightarrow f_j$ ) under the framework of DSMT. The BBA functions assessed for each pair of factors  $f_i \rightarrow f_j$  by expert  $e_k$  are described as  $b_{ij}^k$  (as in (6)), and make up an individual IDR matrix  $B^k = [b_{ij}^k]_{L \times L}$ . On the basis of expert assessments, expert weights  $W = [w^1, \dots, w^K]$  are calculated by Definition 8. Initial assessments  $b_{ij}^k$  are discounted by Shafer's discounting method with expert weight as in Definition 10 to obtain the discounted assessments  $m_{ij}^k$ . Then PCR5 is applied to fuse those BBA functions to obtain group assessments  $m_{ij}$ , making up group IDR matrix  $M = [m_{ij}]_{L \times L}$ .

- Step 3: *Soft consensus reaching process*. The moderator calculates the consensus measure values  $c_{ij}^{k,t}$  and  $c_{ij}$  by means of expert and group IDR matrices to identify whether an acceptable CL is reached in the current round. If so, the soft CRP is finished and a consensus group IDR matrix  $M$  has been obtained. Generalized pignistic probability is introduced to deal with global ignorance and local ignorance as in (18) and the influence degree for  $f_i \rightarrow f_j$  ( $\forall i, j$ ) is determined by one of the two principles as mentioned above to construct  $\tilde{G} = [\tilde{g}_{ij}]_{L \times L}$ , and proceeding to step 4. Otherwise, let  $t = t + 1$  and carry out the feedback mechanism. The moderator identifies the experts who differed strongly from the group by (16) and provides advices for them by (17). Then, return to step 2 and follow the same two processes. Note that the moderator also should detect whether the maximum number of rounds (means by  $t = t_{MAX}$ ) has been reached before carrying out the feedback mechanism. If so, the mechanism will be ceased, taking the current group IDR matrix  $M$  as the final result and applying the generalized

pignistic probability to obtain  $\tilde{G} = [\tilde{g}_{ij}]_{L \times L}$ . Then proceed to step 4.

- Step 4: *Calculate the normalized IDR matrix*. To guarantee that the normalized IDR matrix to infinite power will converge to zero and that the total relation matrix can be smoothly obtained,  $\tilde{g}_{ij}'' = \max(\max_{1 \leq i \leq L} \sum_{j=1}^L \tilde{g}_{ij}, \mu + \max_{1 \leq j \leq L} \sum_{i=1}^L \tilde{g}_{ij})$  is introduced as shown in (1) to calculate the normalized IDR matrix  $D = [d_{ij}]_{L \times L}$ , where  $d_{ij} = \tilde{g}_{ij} / \tilde{q}_{ij}'', \forall i, j$ .
- Step 5: *Compute total relation matrix*. The total relation matrix is calculated by  $A = D(I - D)^{-1}$ . The influencing degree of  $f_i$  is computed by  $r_i = \sum_{j=1}^L a_{ij}$ , the influenced degree of  $f_i$  is computed by  $c_i = \sum_{j=1}^L a_{ji}$ , the prominence degree of  $f_i$  is computed by  $r_i + c_i$ , and the net influence degree of  $f_i$  is computed by  $r_i - c_i$ .
- Step 6: *Obtain the major factors with the threshold value*. To simplify the complexity of a system to a manageable level, negligible factors should be filtered out. Only those factors whose prominence degrees are greater than the threshold value ( $r_i + c_i \geq \eta$ ) should be chosen and considered in the complex system. Generally, the cause-effect relation diagram can be made for the major factors based on their prominence degrees and net influence degrees.

Through the analysis of total relation of the factors by DEMATEL, a better understanding of the structural relation and an ideal way to solve complicate system problems can be obtained. The algorithm can be summarized to help moderator manage the whole processes.

---

**Algorithm 1** The algorithm for DSMT-based group DEMATEL method with reaching consensus

---

**Inputs:** Set of factors  $F = \{f_1, \dots, f_L\}$ , set of grade levels  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ , set of experts  $E = \{e_1, \dots, e_K\}$ , DEMATEL threshold value  $\eta$ , threshold values at expert level and pair-factors level  $\varepsilon_e$  and  $\varepsilon_f$ , round counter  $t$ , accepted maximum number of rounds  $t_{MAX}$ .

**Outputs:** Set of major factors  $\hat{F} = \{f_1, \dots, f_L\}$ .

**Begin**

$t=1$ ,  $\tilde{F}_i = \{k, i \rightarrow j \mid \forall k, \forall i, \forall j\}$ ,  $b_{ij}^{k,0} = \{(\theta_1, 1)\}$  for  $\forall i, j, k$

**Step 1: Collection process**

For  $i=1$  to  $L$

  For  $j=1$  to  $L$

    If  $i \neq j$

      For  $k=1$  to  $K$

        If  $\{k, i \rightarrow j\} \in \tilde{F}_i$  Then

          Give/modify the influence degree  $b_{ij}^{k,t}$  by expert  $e_k$

        Else

$b_{ij}^{k,t} = b_{ij}^{k,t-1}$

        EndIf

      EndFor

      Calculate expert weight  $W_i = \{w^{k,j} \mid k=1, \dots, K\}$  by Eqs.(8)-(10)

      Calculate  $m_{ij}^{k,t}$  in Eq.(12) by discounting  $b_{ij}^{k,t}$  with  $w^{k,j}$

      Make fusion to get  $m_{ij}^t = m_{ij}^{1,t} \oplus \dots \oplus m_{ij}^{K,t}$  with Eq.(5)

    Else

      Let  $m_{ij}^t = \{(\theta_1, 1)\}$

    EndIf

  EndFor

EndFor

Get temporary group IDR matrix in the  $t$ th round  $M^t = [m_{ij}^t]_{L \times L}$

**Step 2: Soft consensus reaching process**

For  $i=1$  to  $L$

  For  $j=1$  to  $L$

    Calculate consensus measure value at pair-factors level by  $c_{ij}^{k,t} = 1 - \sqrt{\sum_{n=1}^{|\Theta|} [b_{ij}^{k,t}(\theta_n) - m_{ij}^t(\theta_n)]^2} / \sqrt{2}$

  EndFor

EndFor

Calculate consensus measure value at expert level by  $c^{k,j} = \sum_{i=1}^L c_{ij}^{k,t} / L^2$

Derive inconsistent expert set by Eq.(16) and derive  $\tilde{E}_i = \{k \mid c^{k,i} < \varepsilon_e\}$

Derive inconsistent pair of factors set by Eq.(17) and derive  $\tilde{F}_i = \{k, i \rightarrow j \mid c_{ij}^{k,t} < \varepsilon_f \wedge k \in \tilde{E}_i\}$

If  $\tilde{F}_i \neq \emptyset$  and  $t \leq t_{MAX}$  Then

  For each  $\{k, i \rightarrow j\} \in \tilde{F}_i$

    Give the advices that expert  $e_k$  is suggested to make corrections for  $b_{ij}^{k,t}$

  EndFor

  Let  $t=t+1$

  Return to Step 1

Else

  Get final BBA-formed group IDR matrix  $M^t = [m_{ij}^t]_{L \times L}$

  Turn to Step 3

EndIf

**Step 3: Group DEMATEL process**

For  $i=1$  to  $L$

  For  $j=1$  to  $L$

    Take  $m_{ij}^t$  into Eq.(18) and get its generalized pignistic probability  $\{(\theta_1, P_{ij}^1), (\theta_2, P_{ij}^2), (\theta_3, P_{ij}^3)\}$

    Use the highest probability or the expected value principle to obtain the final influence degree  $\tilde{g}_{ij}$

  EndFor

EndFor

Construct final group IDR matrix  $\tilde{G} = [\tilde{g}_{ij}]_{L \times L}$

Calculate normalized IDR matrix by  $D = \tilde{G} / \max(\max_{1 \leq i \leq L} \sum_{j=1}^L \tilde{g}_{ij}, \mu + \max_{1 \leq j \leq L} \sum_{i=1}^L \tilde{g}_{ij})$

Calculate total-relation matrix by  $A = (a_{ij})_{L \times L} = D(I - D)^{-1}$

Initialize set of major factors by  $\hat{F} = \emptyset$

For  $i=1$  to  $L$

  Calculate the important degree of  $f_i$  by  $r_i + c_i = \sum_{j=1}^L a_{ij} + \sum_{j=1}^L a_{ji}$

  If  $r_i + c_i \geq \eta$

    Add  $f_i$  into the set of major factors by  $\hat{F} = \hat{F} \cup f_i$

  EndIf

EndFor

**End**

---

Table I  
ASSESSMENTS OF EXPERT  $e_1$  IN THE FIRST ROUND.

	$f_1 \rightarrow f_2$	$f_1 \rightarrow f_3$	$f_1 \rightarrow f_4$	$f_1 \rightarrow f_5$	$f_2 \rightarrow f_3$	$f_2 \rightarrow f_4$	$f_3 \rightarrow f_4$	$f_4 \rightarrow f_5$	$f_5 \rightarrow f_2$	$f_5 \rightarrow f_3$
$\theta_1$	0.00	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
$\theta_2$	0.00	0.50	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00
$\theta_1 \cup \theta_2$	0.30	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.80	0.80
$\theta_3$	0.70	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00
$\theta_1 \cup \theta_3$	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00
$\theta_2 \cup \theta_3$	0.00	0.00	0.00	0.20	0.50	0.20	0.00	0.00	0.00	0.00
$\theta_1 \cup \theta_2 \cup \theta_3$	0.00	0.00	1.00	0.80	0.00	0.60	0.00	1.00	0.20	0.20

IV. NUMERICAL COMPARISON AND DISCUSSION

In this section, we apply the proposed method and the fuzzy DEMATEL method to conduct a numerical simulation case, in which the key factors will be selected by two methods respectively. Afterwards, the results are compared and discussed.

A. The Proposed Method

- Step 1: Define DEMATEL problem and soft CRP parameters.

Similar to the examples in the literatures about DSmt [69], here the group IDR matrix construction of the given example works on the classical power set  $2^\Theta$ , not on the hyper-power set  $D\Theta$ . In the numerical simulation case, suppose the set of factors is  $F = \{f_1, f_2, f_3, f_4, f_5\}$ , the set of experts is  $E = \{f_1, e_2, e_3, e_4, e_5\}$ , the set of grade levels is  $\Theta = \{\theta_1, \theta_2, \theta_3\} = \{0, 1, 2\}$ , the threshold value is  $\eta = 0.35$ , and the acceptable CL threshold values at expert level is  $\varepsilon_e = 0.55$ , and at pair-factors level is  $\varepsilon_f = 0.50$ , the round counter is  $t$ , and the maximum number of rounds is  $t_{MAX} = 5$ .

- Step 2: Expert assessments extraction and fusion.

Experts are advised to indicate the influence degree to which he or she believes factor  $f_i$  has an effect on factor  $f_j$  (denoted by  $f_i \rightarrow f_j$ ) under the framework of DSmt. As an example, the BBA functions on each pair of factors  $f_i \rightarrow f_j$  from expert  $e_k$  are shown in Table I. Due to limited space, other experts' assessments are not given in this paper.

On the basis of the experts' assessments, we calculate the similarity matrix by the Euclidean similarity function as given in Definition 8 and shown in (19).

$$S = \begin{bmatrix} 1.00 & 0.20 & 0.16 & 0.13 & 0.18 \\ 0.20 & 1.00 & 0.15 & 0.17 & 0.19 \\ 0.16 & 0.15 & 1.00 & 0.14 & 0.18 \\ 0.13 & 0.17 & 0.14 & 1.00 & 0.18 \\ 0.18 & 0.19 & 0.18 & 0.18 & 1.00 \end{bmatrix}. \quad (19)$$

Hence,  $Sup(e_1)^1 = 0.20 + 0.16 + 0.13 + 0.18 = 0.68$ ,  $Sup(e_2)^1 = 0.72$ ,  $Sup(e_3)^1 = 0.64$ ,  $Sup(e_4)^1 = 0.62$ , and  $Sup(e_5)^1 = 0.73$ ; the expert weight is computed as  $w^{1,1} = Sup(e_1)^1 / \sum_{k=1}^5 Sup(e_k)^1 = 0.68/3.40 = 0.20$ ,  $w^{2,1} = 0.21$ ,  $w^{3,1} = 0.19$ ,  $w^{4,1} = 0.18$ , and  $w^{5,1} = 0.22$ ; the discounting parameters are derived by  $\bar{w}^{1,1} = w^{1,1} / \max\{w^{k,1} | k = 1, \dots, 5\} = 0.20/0.22 = 0.93$ ,  $\bar{w}^{2,1} = 0.98$ ,  $\bar{w}^{3,1} = 0.87$ ,  $\bar{w}^{4,1} = 0.82$ , and  $\bar{w}^{5,1} = 1.00$ .

Taking  $b_{ij}^{k,1}$  and  $\bar{w}^{k,1}$  into Shafer's discounting method (as in (11)), we derive the discounted assessments  $m_{ij}^{k,1}$ ,  $i, j = 1, \dots, 5$ , and  $k = 1, 2, 3, 4, 5$ . With the help of the MATLAB code for PCR5, we fuse the BBA functions and derive a holism result for the influence degrees between any two factors, as shown in Table II.

- Step 3: Soft consensus reaching process.

The moderator calculates the consensus measure values  $c_{ik}^{k,1}$  and  $c^{k,1}$  by means of  $b_{ik}^{k,1}$  and  $m_{ik}^1$  to identify whether an acceptable CL has been reached. Taking expert and fusion assessment BBA functions into (13) and (14), we calculate the consensus measure values at two levels, as shown in Tables III and IV. The bold values of data in Table III indicate the pair of factors that do not reach the set consensus level. The bold values of data in Table IV indicate the experts that do not reach the set consensus level. According to the selected threshold value  $\varepsilon_e = 0.55$  and (16), we find that the acceptable CL is not reached in this round for  $\tilde{F}_t \neq \emptyset$ . Then, the moderator lets  $t = 1 + 1 = 2$ , which is lower than  $t_{MAX}$ , and carries out the feedback mechanism. The order from expert level to pair-factors level should be followed to determine the inconsistent experts and the pairs of factors that need to be modified (for those values less than  $\varepsilon_f = 0.50$ ). Tables III and IV show that expert  $e_3$  greatly differs from the group results on the pairs of factors  $f_1 \rightarrow f_4$ ,  $f_2 \rightarrow f_4$ , and  $f_4 \rightarrow f_5$ , and  $e_4$  greatly differs from the group results on the pairs of factors  $f_1 \rightarrow f_2$ ,  $f_2 \rightarrow f_3$ ,  $f_2 \rightarrow f_4$ , and  $f_4 \rightarrow f_5$  (see those underlined data in Tables III and IV). Experts  $e_3$  and  $e_4$  need to modify those assessments on the mentioned pairs of factors. V and VI.

Because some of the expert assessments are changed. In the second round, expert weights are recalculated to be  $w^{1,2} = 0.19$ ,  $w^{2,2} = 0.20$ ,  $w^{3,2} = 0.20$ ,  $w^{4,2} = 0.21$ , and  $w^{5,2} = 0.21$ , and the corresponding discounting parameters are standardized as  $\bar{w}^{1,2} = 0.89$ ,  $\bar{w}^{2,2} = 0.93$ ,  $\bar{w}^{3,2} = 0.96$ ,  $\bar{w}^{4,2} = 1.00$ , and  $\bar{w}^{5,2} = 1.00$ . Taking  $b_{ik}^{k,2}$  and  $\bar{w}^{k,2}$  into (11), we obtain the discounted assessments  $m_{ij}^{k,2}$ , ( $\forall i, j$ ) and fuse these results by PCR5 to construct the group IDR matrix in the second round. The moderator calculates the new consensus measure values at two levels. Fusion results and consensus measures at expert level in the second round are shown in Tables VII and VIII.

Table II  
FUSION RESULTS OF EXPERTS' ASSESSMENTS IN THE FIRST ROUND.

	$f_1 \rightarrow f_2$	$f_1 \rightarrow f_3$	$f_1 \rightarrow f_4$	$f_1 \rightarrow f_5$	$f_2 \rightarrow f_3$	$f_2 \rightarrow f_4$	$f_3 \rightarrow f_4$	$f_4 \rightarrow f_5$	$f_5 \rightarrow f_2$	$f_5 \rightarrow f_3$
$\theta_1$	0.00	0.22	0.00	0.13	0.00	0.05	0.44	0.00	0.07	0.05
$\theta_2$	0.00	0.21	0.38	0.26	0.34	0.19	0.17	0.15	0.25	0.05
$\theta_1 \cup \theta_2$	0.08	0.16	0.00	0.00	0.03	0.05	0.08	0.00	0.29	0.32
$\theta_3$	0.66	0.22	0.19	0.27	0.20	0.56	0.12	0.56	0.22	0.42
$\theta_1 \cup \theta_3$	0.02	0.06	0.00	0.00	0.00	0.05	0.16	0.06	0.01	0.06
$\theta_2 \cup \theta_3$	0.02	0.00	0.17	0.07	0.20	0.04	0.00	0.05	0.07	0.04
$\theta_1 \cup \theta_2 \cup \theta_3$	0.22	0.13	0.26	0.27	0.23	0.06	0.03	0.18	0.09	0.06

Table III  
THE CONSENSUS MEASURE VALUES AT PAIR-FACTORS LEVEL IN THE FIRST ROUND.

	$f_1 \rightarrow f_2$	$f_1 \rightarrow f_3$	$f_1 \rightarrow f_4$	$f_1 \rightarrow f_5$	$f_2 \rightarrow f_3$	$f_2 \rightarrow f_4$	$f_3 \rightarrow f_4$	$f_4 \rightarrow f_5$	$f_5 \rightarrow f_2$	$f_5 \rightarrow f_3$
$e_1$	0.78	0.65	0.38	0.53	0.58	0.40	0.80	0.29	0.56	0.53
$e_2$	0.74	0.61	0.50	0.53	0.35	0.42	0.78	0.79	0.56	0.79
$e_3$	0.78	0.68	<b>0.32</b>	0.57	0.57	<b>0.20</b>	0.56	<b>0.45</b>	0.52	0.55
$e_4$	<b>0.27</b>	0.57	0.53	0.68	<b>0.37</b>	<b>0.35</b>	0.80	<b>0.38</b>	0.66	0.70
$e_5$	0.78	0.67	0.44	0.57	0.49	0.67	0.57	0.77	0.66	0.60

Table IV  
THE CONSENSUS MEASURE VALUES AT EXPERT LEVEL IN THE FIRST ROUND.

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
$c^{k,1}$	0.55	0.61	<b>0.52</b>	<b>0.53</b>	0.62

Table V  
ADJUSTED ASSESSMENTS OF EXPERT  $e_3$  IN THE SECOND ROUND.

	$f_1 \rightarrow f_2$	$f_1 \rightarrow f_3$	$f_1 \rightarrow f_4$	$f_1 \rightarrow f_5$	$f_2 \rightarrow f_3$	$f_2 \rightarrow f_4$	$f_3 \rightarrow f_4$	$f_4 \rightarrow f_5$	$f_5 \rightarrow f_2$	$f_5 \rightarrow f_3$
$\theta_1$	0.00	0.00	0.00	0.50	0.00	0.04	0.00	0.00	0.00	0.00
$\theta_2$	0.00	0.00	0.22	0.50	0.80	0.16	0.00	0.28	0.00	0.00
$\theta_1 \cup \theta_2$	0.30	0.30	0.00	0.00	0.20	0.04	0.20	0.00	0.00	0.70
$\theta_3$	0.70	0.30	0.37	0.00	0.00	0.45	0.40	0.28	0.00	0.00
$\theta_1 \cup \theta_3$	0.00	0.00	0.00	0.00	0.00	0.04	0.40	0.07	0.20	0.30
$\theta_2 \cup \theta_3$	0.00	0.00	0.07	0.00	0.00	0.03	0.00	0.07	0.50	0.00
$\theta_1 \cup \theta_2 \cup \theta_3$	0.00	0.40	0.34	0.00	0.00	0.24	0.00	0.30	0.30	0.00

Table VI  
ADJUSTED ASSESSMENTS OF EXPERT  $e_4$  IN THE SECOND ROUND.

	$f_1 \rightarrow f_2$	$f_1 \rightarrow f_3$	$f_1 \rightarrow f_4$	$f_1 \rightarrow f_5$	$f_2 \rightarrow f_3$	$f_2 \rightarrow f_4$	$f_3 \rightarrow f_4$	$f_4 \rightarrow f_5$	$f_5 \rightarrow f_2$	$f_5 \rightarrow f_3$
$\theta_1$	0.00	0.00	0.00	0.00	0.00	0.03	0.50	0.00	0.00	0.00
$\theta_2$	0.00	0.00	0.00	0.50	0.22	0.14	0.30	0.08	0.00	0.00
$\theta_1 \cup \theta_2$	0.06	0.30	0.00	0.00	0.02	0.02	0.20	0.00	0.40	0.00
$\theta_3$	0.47	0.70	0.50	0.50	0.13	0.38	0.00	0.36	0.60	0.60
$\theta_1 \cup \theta_3$	0.02	0.00	0.00	0.00	0.00	0.05	0.00	0.14	0.00	0.20
$\theta_2 \cup \theta_3$	0.02	0.00	0.50	0.00	0.13	0.12	0.00	0.12	0.00	0.20
$\theta_1 \cup \theta_2 \cup \theta_3$	0.43	0.00	0.00	0.00	0.50	0.26	0.00	0.30	0.00	0.00

Table VII  
FUSION RESULTS OF EXPERTS' ASSESSMENTS IN THE SECOND ROUND.

	$f_1 \rightarrow f_2$	$f_1 \rightarrow f_3$	$f_1 \rightarrow f_4$	$f_1 \rightarrow f_5$	$f_2 \rightarrow f_3$	$f_2 \rightarrow f_4$	$f_3 \rightarrow f_4$	$f_4 \rightarrow f_5$	$f_5 \rightarrow f_2$	$f_5 \rightarrow f_3$
$\theta_1$	0.00	0.20	0.00	0.12	0.00	0.02	0.43	0.00	0.06	0.06
$\theta_2$	0.00	0.20	0.47	0.26	0.41	0.10	0.17	0.06	0.22	0.05
$\theta_1 \cup \theta_2$	0.16	0.17	0.00	0.00	0.05	0.11	0.08	0.00	0.26	0.29
$\theta_3$	0.69	0.26	0.13	0.23	0.20	0.51	0.14	0.74	0.27	0.42
$\theta_1 \cup \theta_3$	0.04	0.06	0.00	0.00	0.00	0.05	0.17	0.04	0.02	0.07
$\theta_2 \cup \theta_3$	0.04	0.00	0.16	0.06	0.26	0.08	0.00	0.02	0.08	0.05
$\theta_1 \cup \theta_2 \cup \theta_3$	0.07	0.11	0.24	0.33	0.08	0.13	0.01	0.15	0.09	0.06

Table VIII  
THE CONSENSUS MEASURE VALUES AT EXPERT LEVEL IN THE SECOND ROUND.

	$e_1$	$e_2$	$e_3$	$e_4$	$e_5$
$c^{k;2}$	0.55	0.63	0.66	0.68	0.62

As shown in Table VIII, the acceptable CL has been reached in the second round, which means that the current group IDR matrix is a consensus result and can be applied to the group DEMATEL method. Then we derive pignistic probability transformation for each element in final BBA-formed group IDR matrix  $M^2 = [m_{ij}^2]_{5 \times 5}$ . Take the pair of factors  $f_1 \rightarrow f_2$  as an example,  $m_{12}(\theta_1) = m(\theta_1) + \frac{1}{2}m(\theta_1 \cup \theta_2) + \frac{1}{2}m(\theta_1 \cup \theta_3) + \frac{1}{3}m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.12$ ,  $m_{12}(\theta_2) = 0.12$ ,  $m_{12}(\theta_3) = 0.76$ , and  $\tilde{g}_{12} = 0 \times 0.12 + 1 \times 0.12 + 2 \times 0.76 = 1.64$ . After pignistic probability transformation, the final group IDR matrix is constructed as in (20).

$$\tilde{G} = \begin{bmatrix} 0.00 & 1.64 & 0.99 & 1.22 & 1.13 \\ 0.00 & 0.00 & 1.33 & 1.53 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.67 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.76 \\ 0.00 & 1.12 & 1.25 & 0.00 & 0.00 \end{bmatrix}. \quad (20)$$

- Step 4: Calculate the normalized IDR matrix.

Let consider  $\mu = 0.00001$ . We obtain

$$\tilde{g}_{ij}'' = \max(\max_{1 \leq i \leq L} \sum_{j=1}^L \tilde{g}_{ij}, \mu + \max_{1 \leq j \leq L} \sum_{i=1}^L \tilde{g}_{ij}) \approx 4.98,$$

and the normalized IDR matrix can be calculated by (1). The result is shown as in (21).

$$D = \begin{bmatrix} 0.00 & 0.33 & 0.20 & 0.24 & 0.23 \\ 0.00 & 0.00 & 0.27 & 0.31 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.13 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.35 \\ 0.00 & 0.22 & 0.25 & 0.00 & 0.00 \end{bmatrix}. \quad (21)$$

- Step 5: Compute the total relation matrix.

The total relation matrix can be calculated by (2). The result is shown as

$$A = \begin{bmatrix} 0.00 & 0.14 & 0.08 & 0.10 & 0.09 \\ 0.00 & 0.00 & 0.08 & 0.11 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.02 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.13 \\ 0.00 & 0.05 & 0.08 & 0.00 & 0.00 \end{bmatrix}. \quad (22)$$

According to  $A = [a_{ij}]_{5 \times 5}$  and (3), we derive the following parameters: the total influence  $r_i = \sum_{j=1}^5 a_{ij}$  given by the factor  $f_i$  to other factors, the total influence  $c_i = \sum_{j=1}^5 a_{ji}$  received by the factor  $f_i$  from other factors, the degree of the important role  $r_i + c_i$ , and the net influence  $r_i - c_i$ . These parameters are listed in Table IX. Additionally, we construct the cause-effect relation diagram of factors with the horizontal axis  $r + c$  and the vertical axis  $r - c$ , as shown in Fig. 3. According to Definition 4 and Fig. 3,  $f_1$  is known as net causer,

Table IX  
THE ATTRIBUTE PARAMETERS OF FACTORS.

	$r$	$c$	$r + c$	$r - c$
$f_1$	0.41	0.00	0.41	0.41
$f_2$	0.19	0.19	0.38	0.00
$f_3$	0.02	0.24	0.26	-0.22
$f_4$	0.13	0.23	0.36	-0.10
$f_5$	0.13	0.22	0.35	-0.09

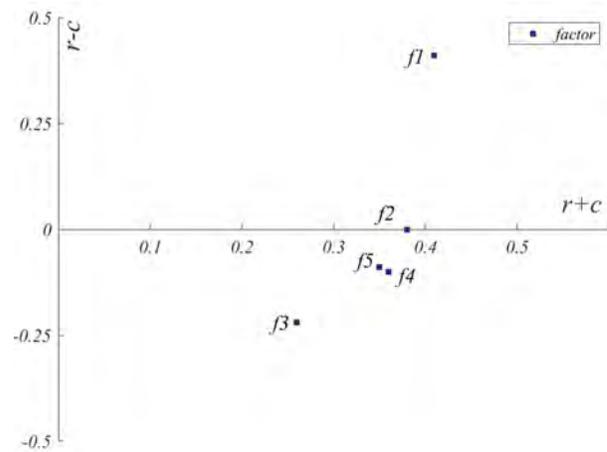


Figure 3. The cause-effect relation diagram derived by the proposed method.

whereas  $f_3$ ,  $f_4$  and  $f_5$  are net receivers. Because  $r - c = 0$  for  $f_2$ ,  $f_2$  is neither net causer, or net receiver.

- Step 6: Obtain major factors with the threshold value. Because of the threshold value  $\eta = 0.35$ , we select the factors whose degrees of the important role are greater than the threshold ( $r_i + c_i \geq \eta$ ) as major factors. As a result, the major factors in the complex system are  $\hat{F} = \{f_1, f_2, f_4, f_5\}$ .

We find that the Algorithm 1 can be programmed easily. Thus, the proposed method in this study is valid and applicable for solving group DEMATEL problems.

### B. The Fuzzy DEMATEL Method

The fuzzy DEMATEL method proposed by Wu and Lee [18] is a vital and significant improvement of DEMATEL. Since the fuzzy DEMATEL method is not only a major extension of DEMATEL but also extracts information with fuzzy linguistic terms, here we employ it to make a comparison with the proposed method. To ensure the results of two kinds of methods can be compared with each other, the set of factors  $F = \{f_1, f_2, f_3, f_4, f_5\}$ , the set of experts  $E = \{e_1, e_2, e_3, e_4, e_5\}$  and the threshold value is  $\eta = 0.35$  are all the same as in Subsection IV-A. Besides, the inputs of fuzzy DEMATEL method should be generated from the

initial assessments given by experts in the proposed method to ensure its comparability. According to the procedure of fuzzy DEMATEL method, its computation processes can be summarized as follows.

- Step 1: *Define DEMATEL problem and fuzzy linguistic scale.*

The DEMATEL problem is defined as the same as in Subsection IV-A and the fuzzy linguistic scale is defined as in Table X, where {No, L, H} are equal to  $\{\theta_1, \theta_2, \theta_3\}$  as defined in this paper.

Table X  
THE FUZZY LINGUISTIC SCALE.

Linguistic terms	Triangular fuzzy numbers
High influence (H)	(0.75,1.00,1.00)
Low influence (L)	(0.25,0.50,0.75)
No influence (No)	(0.00,0.00,0.25)

- Step 2: *Exact and fuse expert assessments..*

In order to ensure the comparability, the inputs of fuzzy DEMATEL method should be generated from the initial assessments given by experts in the proposed method. Because the assessments given by experts in Subsection 4.1 are in the form of BBA functions, how to make a transformation from the BBA functions to the inputs of fuzzy DEMATEL method is quite important. It is reasonable and logical to choose the grade level with the highest probability as expert assessments in the fuzzy DEMATEL method. Following this thought, the inputs of fuzzy DEMATEL method are transformed from the experts' assessments in Subsection IV-A. For example, the transformed assessments of expert are shown in Table XI. As introduced in fuzzy DEMATEL, the Converting

Table XI  
THE TRANSFORMED ASSESSMENTS OF EXPERT  $e_1$ .

	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$
$f_1$	No	H	L	No	No
$f_2$	No	No	H	H	No
$f_3$	No	No	No	L	No
$f_4$	No	No	No	No	H
$f_5$	No	No	L	No	No

Fuzzy data into Crisp Scores (CFCS) defuzzification method is applied to aggregate these assessments by five experts. Due to limited space, the detailed CFCS steps can be referred to [18] and are not repeated here. The IDR matrix  $G' = [g'_{ij}]_{5 \times 5}$  is produced as in (23).

$$G' = \begin{bmatrix} 0.04 & 0.19 & 0.87 & 0.19 & 0.50 \\ 0.19 & 0.04 & 0.96 & 0.96 & 0.04 \\ 0.04 & 0.04 & 0.04 & 0.41 & 0.19 \\ 0.04 & 0.04 & 0.04 & 0.04 & 0.87 \\ 0.04 & 0.50 & 0.59 & 0.04 & 0.04 \end{bmatrix}. \quad (23)$$

- Step 3: *Calculate the normalized IDR matrix.*

The normalizing method for IDR matrix in fuzzy DEMATEL is the same as in traditional DEMATEL. The

normalized IDR matrix is calculated by taking (23) into (1) and it is shown as in (24).

$$D' = \begin{bmatrix} 0.02 & 0.08 & 0.35 & 0.08 & 0.20 \\ 0.08 & 0.02 & 0.38 & 0.38 & 0.02 \\ 0.02 & 0.02 & 0.02 & 0.16 & 0.08 \\ 0.02 & 0.02 & 0.02 & 0.02 & 0.35 \\ 0.02 & 0.20 & 0.24 & 0.02 & 0.02 \end{bmatrix}. \quad (24)$$

- Step 4: *Compute the total relation matrix.*

The computing method for total-relation matrix in fuzzy DEMATEL is also the same as in traditional DEMATEL. The total-relation IDR matrix is calculated by taking (24) into (2) and it is shown as in (25).

$$A' = \begin{bmatrix} 0.02 & 0.01 & 0.18 & 0.02 & 0.07 \\ 0.01 & 0.02 & 0.21 & 0.20 & 0.00 \\ 0.00 & 0.00 & 0.02 & 0.03 & 0.01 \\ 0.00 & 0.00 & 0.00 & 0.02 & 0.14 \\ 0.00 & 0.05 & 0.09 & 0.00 & 0.02 \end{bmatrix}. \quad (25)$$

According to (3), we derive the following parameters as shown in Table XII and the cause-effect relation diagram as shown in Fig. 4. According to Definition 4 and Fig. 4, it is obvious to find that  $f_1$  and  $f_2$  are net causers, while  $f_3$ ,  $f_4$  and  $f_5$  are net receivers.

Table XII  
THE FACTORS' ATTRIBUTES PARAMETERS.

	$r'$	$c'$	$r' + c'$	$r' - c'$
$f_1$	0.30	0.03	0.33	0.27
$f_2$	0.44	0.08	0.52	0.36
$f_3$	0.07	0.50	0.57	-0.43
$f_4$	0.17	0.28	0.44	-0.11
$f_5$	0.16	0.25	0.41	-0.09

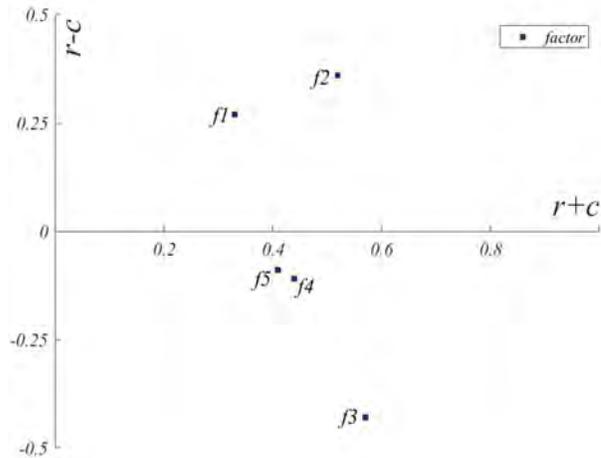


Figure 4. The cause-effect relation diagram derived by fuzzy DEMATEL method.

- Step 5: *Set a threshold value and obtain the major factors.* Because of the threshold value  $\eta = 0.35$ , only the factors with  $r' + c' \geq \eta$  should be chosen as the major factors. As a result, the major factors in the complex system are  $\tilde{F}' = \{f_2, f_3, f_4, f_5\}$ .

### C. Discussion

It is obvious to find that the major factors in the complex system determined by the proposed method are  $\hat{F} = \{f_1, f_2, f_4, f_5\}$ , while those determined by the fuzzy DEMATEL method are  $\hat{F}' = \{f_2, f_3, f_4, f_5\}$ . The results of two kinds of methods are different from each other. Which one is more reasonable? Now we make discussions from the following three aspects.

(1) Expert assessment extraction mechanism. The assessments of experts are used as the fundamental inputs of DEMATEL whether in the proposed method or in the fuzzy DEMATEL method. The proposed method allows experts to give assessments with BBA functions, while the fuzzy DEMATEL method employs the fuzzy linguistic scale to express their assessments. The local or global ignorance in experts' minds can be well reflected in the proposed method (e.g.,  $b_{12}^{1,1} = \{((\theta_1, \theta_2), 0.30), (\theta_3, 0.70)\}$ ), but it is unfortunate to find that such the ignorance cannot included in the fuzzy DEMATEL method (it seems that only the grade level/fuzzy linguistic scale with the highest probability may be allowed to express assessments, e.g.,  $b_{12}^{1,1} = \{L\} = \{\theta_3\}$ ). Consequently, it is believed that the proposed method is more feasible than the fuzzy DEMATEL method in the aspect of expert assessment extraction mechanism.

(2) The importance roles of experts. As discussed in subsection III-A, the importance roles of experts are reflected by expert weights that are calculated based on expert assessments information, hence, the weight parameters could effectively express the relative importance of experts in the group and they have a significant impact on the group DEMATEL decision results. If the importance roles of experts are neglected in the process of decision making, the decision results may lose effectiveness. The proposed method calculates expert weights based on expert assessments with the aid of evidence distance and employs Shafer's discounting method to modify the subjective assessments. Unfortunately, the importance roles of experts are unconsidered in the whole processes of fuzzy DEMATEL method. Consequently, it is believed that the proposed method is more accurate than the fuzzy DEMATEL method in the aspect of reflecting the importance roles of experts.

(3) Group consensus reaching. Group consensus reaching is a key problem in the GDM field. Only when the assessments given by experts reach an acceptable CL, the GDM results are seen to be valuable and credible. In other words, the GDM results that lack of consensus may be ineffective and have few reference value for decision-making. From subsection IV-B, we know that the fuzzy DEMATEL method is a static method without taking the group consensus reaching into consideration. In the proposed method, we apply the soft CRP into the construction of group IDR matrix to help the experts group reach a high CL, among which the feedback and modification mechanism are introduced. Consequently, it is believed that the proposed method is more reasonable than the fuzzy DEMATEL in the aspect of reaching group consensus.

### V. CONCLUSIONS

In the present study, the DSMT is used to extract and fuse expert assessments, and a soft CRP is introduced to construct the group IDR matrix for the DEMATEL. The DSMT-based group DEMATEL method and the corresponding algorithm are proposed. Moreover, a numerical comparison is performed to discuss the applicability of the proposed method and algorithm. The main contributions of the present study can be summarized into three aspects.

Firstly, an expert assessment extraction and fusion mechanism is established on the basis of DSMT. Expert assessments on the influence relations among factors are extracted by BBA functions, which can help experts to express uncertainty and incompleteness assessments. The PCR5 of DSMT, which can overcome the defects of intuitional paradox in Dempster's combination rule, is employed to make fusion for the individual BBA functions discounted by Shafer's discounting method. Secondly, expert weights are defined and introduced to reflect importance roles of experts in a group. Following the principle of pairwise comparisons, expert weights are calculated by the Euclidean similarity function to reflect experts' relative importance in the group. Expert weights are not fixed during the whole processes of group IDR matrix construction and they may vary dynamically with different expert assessments in each round. Shafer's discounting method is used to discount expert assessments in each round so as to reflect the importance roles of experts in the group dynamically. The above processes are beneficial to obtain an accurate group IDR matrix for the DEMATEL.

Thirdly, a soft CRP is established to construct the group IDR matrix with consensus and an algorithm is summarized for the group DSMT-based DEMATEL. The consensus measures at expert and pair-factors levels are defined to help establish the feedback/modification mechanism, based on which a soft CRP is established for the construction of consensus group IDR matrix, so that experts can reach a high CL. An algorithm for DSMT-based group DEMATEL method with reaching consensus is proposed to identify major factors in a complex system. The proposed algorithm can be programmed easily and is valid and applicable for solving group DEMATEL problems.

The proposed method forms an expert assessment extraction mechanism on the basis of the BBA function. However, the belief degrees or probabilities in BBA functions given by experts may hardly be assessed with exact values in more complex situations. Therefore, investigating how to deal with the group DEMATEL with the interval BBA function may be a good direction for future research.

### ACKNOWLEDGMENTS

This research was supported by the Major Program of National Social Science Foundation of China under Grant No. 18ZDA055, the Key Program of National Social Science Fund of China under Grant No. 16AJL007, the National Natural Science Foundation of China (NSFC) under Grant Nos. 71874167, 71804170, 71901199 and 71462022, and

the Special Funds of Taishan Scholars Project of Shandong Province under Grant No. tsqn20171205.

## REFERENCES

- [1] E.G. Fontela, *Structural analysis of the world problematique: (Methods)*, Volume 2 of DEMATEL Report Series, Battelle-Geneva Research Centre, 1974.
- [2] A. Gabus, *Communicating with those bearing collective responsibility*, Battelle-Geneva Research Centre, 1973.
- [3] J.I. Shieh, H.H. Wu, K.K. Huang, *A DEMATEL method in identifying key success factors of hospital service quality*, Knowledge-Based Systems, Vol. 23, pp. 277–282, 2010. <https://doi.org/10.1016/j.knsys.2010.01.013>
- [4] J. Michnik, *Weighted influence non-linear gauge system (WINGS) - An analysis method for the systems of interrelated components*, European Journal of Operational Research, Vol. 228, pp. 536–544, 2013. <https://doi.org/10.1016/j.ejor.2013.02.007>
- [5] H. Liang, J. Ren, Z. Gao, et al., *Identification of critical success factors for sustainable development of biofuel industry in China based on grey decision-making trial and evaluation laboratory (DEMATEL)*, Journal of Cleaner Production, Vol. 131, pp. 500–508, 2016. <https://doi.org/10.1016/j.jclepro.2016.04.151>
- [6] J. Ren, A. Manzardo, S. Toniolo, A. Scipioni, *Sustainability of hydrogen supply chain. Part I: Identification of critical criteria and cause-effect analysis for enhancing the sustainability using DEMATEL*, International Journal of Hydrogen Energy, Vol. 38, pp. 14159–14171, 2013. <https://doi.org/10.1016/j.ijhydene.2013.08.126>
- [7] J. Sara, R.M. Stikkelman, P.M. Herder, *Assessing relative importance and mutual influence of barriers for CCS deployment of the ROAD project using AHP and DEMATEL methods*, International Journal of Greenhouse Gas Control, Vol. 41, pp. 336–357, 2015. <https://doi.org/10.1016/j.ijggc.2015.07.008>
- [8] S. Balsara, P.K. Jain, A. Ramesh, *An integrated approach using AHP and DEMATEL for evaluating climate change mitigation strategies of the Indian cement manufacturing industry*, Environmental Pollution, Vol. 252, pp. 863–878, 2019. <https://doi.org/10.1016/j.envpol.2019.05.059>
- [9] I. Gölcük, A. Baykasoğlu, *An analysis of DEMATEL approaches for criteria interaction handling within ANP*, Expert Systems with Applications, Vol. 46, pp. 346–366, 2016. <https://doi.org/10.1016/j.eswa.2015.10.041>
- [10] S. Lan, R.Y. Zhong, *An Evaluation Model for Financial Reporting Supply Chain Using DEMATEL-ANP*, Procedia CIRP, Vol. 56, pp. 516–519, 2016. <https://doi.org/10.1016/j.procir.2016.10.101>
- [11] T. Ghaemi Rad, A. Sadeghi-Niaraki, A. Abbasi, S.M. Choi, *A methodological framework for assessment of ubiquitous cities using ANP and DEMATEL methods*, Sustainable Cities and Society, Vol. 37, pp. 608–618, 2018. <https://doi.org/10.1016/j.scs.2017.11.024>
- [12] L.E. Quezada, H.A. López-Ospina, P.I. Palominos, A.M. Oddershede, *Identifying causal relationships in strategy maps using ANP and DEMATEL*, Computers and Industrial Engineering, Vol. 118, pp. 170–179, 2018. <https://doi.org/10.1016/j.cie.2018.02.020>
- [13] Z. Chen, X. Ming, X. Zhang, et al., *A rough-fuzzy DEMATEL-ANP method for evaluating sustainable value requirement of product service system*, Journal of Cleaner Production, Vol. 228, pp. 485–508, 2019. <https://doi.org/10.1016/j.jclepro.2019.04.145>
- [14] H. Dinçer, S. Yüksel, L. MartíÁñez, *Interval type 2-based hybrid fuzzy evaluation of financial services in E7 economies with DEMATEL-ANP and MOORA methods*, Applied Soft Computing Journal, Vol. 79, pp. 186–202, 2019. <https://doi.org/10.1016/j.asoc.2019.03.018>
- [15] A. Baykasoğlu, I. Gölcük, *Development of an interval type-2 fuzzy sets based hierarchical MADM model by combining DEMATEL and TOPSIS*, Expert Systems with Applications, Vol. 70, pp. 37–51, 2017. <https://doi.org/10.1016/j.eswa.2016.11.001>
- [16] A. Baykasoğlu, V. Kaplanoglu, Z.D.U. Durmuşoglu, C. Şahin, *Integrating fuzzy DEMATEL and fuzzy hierarchical TOPSIS methods for truck selection*, Expert Systems with Applications, Vol. 40, pp. 899–907, 2013. <https://doi.org/10.1016/j.eswa.2012.05.046>
- [17] S.J. Chen, C.L. Hwang, *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, Springer-Verlag, New York. In collaboration with Frank P. Hwang, 1992.
- [18] W.W. Wu, Y.T. Lee, *Developing global managers' competencies using the fuzzy DEMATEL method*, Expert Systems with Applications, Vol. 32, pp. 499–507, 2007. <https://doi.org/10.1016/j.eswa.2005.12.005>
- [19] L. Abdullah, N. Zulkifli, H. Liao, et al., *An interval-valued intuitionistic fuzzy DEMATEL method combined with Choquet integral for sustainable solid waste management*, Engineering Applications of Artificial Intelligence, Vol. 82, pp. 207–215, 2019. <https://doi.org/10.1016/j.engappai.2019.04.005>
- [20] B.A. Addae L. Zhang, P. Zhou, F. Wang, *Analyzing barriers of Smart Energy City in Accra with two-step fuzzy DEMATEL*, Cities, Vol. 89, pp. 218–227, June 2019. <https://doi.org/10.1016/j.cities.2019.01.043>
- [21] U. Asan, C. Kadaifci, E. Bozdog, et al., *A new approach to DEMATEL based on interval-valued hesitant fuzzy sets*, Applied Soft Computing Journal, Vol. 66, pp. 34–49, May 2018. <https://doi.org/10.1016/j.asoc.2018.01.018>
- [22] M.S. Bhatia, R.K. Srivastava, *Analysis of external barriers to remanufacturing using grey-DEMATEL approach: An Indian perspective*, Resources, Conservation and Recycling, Vol. 136, pp. 79–87, 2018. <https://doi.org/10.1016/j.resconrec.2018.03.021>
- [23] F. Acuña-Carvajal, L. Pinto-Tarazona L, H. López-Ospina et al., *An integrated method to plan, structure and validate a business strategy using fuzzy DEMATEL and the balanced scorecard*, Expert Systems with Applications, Vol. 122, pp. 351–368, 2019. <https://doi.org/10.1016/j.eswa.2019.01.030>
- [24] L. Cui, H.K. Chan, Y. Zhou, et al., *Exploring critical factors of green business failure based on Grey-Decision Making Trial and Evaluation Laboratory (DEMATEL)*, Journal of Business Research, Vol. 98, pp. 450–461, 2019. <https://doi.org/10.1016/j.jbusres.2018.03.031>
- [25] Y.W. Du, W. Zhou, *New improved DEMATEL method based on both subjective experience and objective data*, Engineering Applications of Artificial Intelligence, Vol. 83, pp. 57–71, 2019. <https://doi.org/10.1016/j.engappai.2019.05.001>
- [26] C.H. Cheng, Y. Lin, *Evaluating the best main battle tank using fuzzy decision theory*, European Journal of Operational Research, Vol. 142, pp. 174–186, 2002. [https://doi.org/10.1016/S0377-2217\(01\)00280-6](https://doi.org/10.1016/S0377-2217(01)00280-6)
- [27] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion (Collected Works)*, Volume 1, American Research Press (ARP), Rehoboth, USA, 2004.
- [28] Y.W. Du, N. Yang, J. Ning, *IFS/ER-based large-scale multiattribute group decision-making method by considering expert knowledge structure*, Knowledge-Based Systems, Vol. 162, pp. 124–135, 2018. <https://doi.org/10.1016/j.knsys.2018.07.034>
- [29] Y.W. Du, Y.M. Wang, M. Qin, *New evidential reasoning rule with both weight and reliability for evidence combination*, Computers and Industrial Engineering, Vol. 124, pp. 493–508, 2018. <https://doi.org/10.1016/j.cie.2018.07.037>
- [30] G. Büyüközkan, G. Çifçi, *Evaluation of the green supply chain management practices: A fuzzy ANP approach*, Production Planning and Control, Vol. 23, pp. 405–418, 2012. doi: 10.1080/09537287.2011.561814
- [31] I.J. Pérez, F.J. Cabrerizo, S. Alonso, et al., *On dynamic consensus processes in group decision making problems*, Information Sciences, Vol. 459, pp. 20–35, 2018. <https://doi.org/10.1016/j.ins.2018.05.017>
- [32] F. Herrera, E. Herrera-Viedma, J.L. Verdegay, *A rational consensus model in group decision making using linguistic assessments*, Fuzzy Sets and Systems, Vol. 88, pp. 31–49, 1997.
- [33] J. Kacprzyk, S. Zadrozny, *Soft computing and web intelligence for supporting consensus reaching*, Soft Computing, Vol. 14, pp. 833–846, 2010. <https://doi.org/10.1007/s00500-009-0475-4>
- [34] F.J. Cabrerizo, R. Al-Hmouz, A. Morfeq, et al., *Soft consensus measures in group decision making using unbalanced fuzzy linguistic information*, Soft Computing, Vol. 21, pp. 3037–3050, 2017. <https://doi.org/10.1007/s00500-015-1989-6>
- [35] D. Singhal, S. Tripathy, S. Kumar Jena, *DEMATEL approach for analyzing the critical factors in remanufacturing process*, Materials Today: Proceedings 5, pp. 18568–18573, 2018. <https://doi.org/10.1016/j.matpr.2018.06.200>
- [36] H.S. Lee, G.H. Tzeng, W. Yeih, et al., *Revised DEMATEL: Resolving the infeasibility of DEMATEL*, Applied Mathematical Modelling, Vol. 37, pp. 6746–6757, 2013. <https://doi.org/10.1016/j.apm.2013.01.016>
- [37] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion (Collected Works)*, Volume 2, American Research Press (ARP), Rehoboth, USA, 2006.
- [38] Y.W. Du, S.S. Wang, Y.M. Wang, *Group fuzzy comprehensive evaluation method under ignorance*, Expert Systems with Applications, Vol. 126, pp. 92–111, 2019. <https://doi.org/10.1016/j.eswa.2019.02.006>
- [39] R. Singh, M. Vatsa, A. Noore, *Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face*

- recognition, *Pattern Recognition*, Vol. 41, pp. 880–893, 2008. <https://doi.org/10.1016/j.patcog.2007.06.022>
- [40] S. Huang, X. Su, Y. Hu, et al., *A new decision-making method by incomplete preferences based on evidence distance*, *Knowledge-Based Systems*, Vol. 56, pp. 264–272, 2014. <https://doi.org/10.1016/j.knsys.2013.11.019>
- [41] Z.-G. Liu, J. Dezert, Q. Pan, G. Mercier, *Combination of sources of evidence with different discounting factors based on a new dissimilarity measure*, *Decision Support Systems*, Vol. 52(1), pp. 133–141, Dec. 2011. <https://doi.org/10.1016/j.dss.2011.06.002>
- [42] F. Faux, F. Luthon, *Theory of evidence for face detection and tracking*, *International Journal of Approximate Reasoning*, Vol. 53(5), pp. 728–746, 2012. <https://doi.org/10.1016/j.ijar.2012.02.002>
- [43] Z.-G. Liu, J. Dezert, G. Mercier, Q. Pan, *Dynamic evidential reasoning for change detection in remote sensing images*, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50, pp. 1955–1967, 2012. <https://doi.org/10.1109/TGRS.2011.2169075>
- [44] C. Lian, S. Ruan, T. Denœux, *An evidential classifier based on feature selection and two-step classification strategy*, *Pattern Recognition*, Vol. 48, pp. 2318–2327, 2015. <https://doi.org/10.1016/j.patcog.2015.01.019>
- [45] Z.-G. Liu, Q. Pan, J. Dezert, *Evidential classifier for imprecise data based on belief functions*, *Knowledge-Based Systems*, Vol. 52, pp. 246–257, 2013. <https://doi.org/10.1016/j.knsys.2013.08.005>
- [46] Z.-G. Liu, Q. Pan, J. Dezert, G. Mercier, *Credal c-means clustering method based on belief functions*, *Knowledge-Based Systems*, Vol. 74, pp. 119–132, 2015. <https://doi.org/10.1016/j.knsys.2014.11.013>
- [47] Z.-G. Liu, Q. Pan, G. Mercier, J. Dezert, *A New Incomplete Pattern Classification Method Based on Evidential Reasoning*, *IEEE Transactions on Cybernetics*, Vol. 45, pp. 635–646, 2015. <https://doi.org/10.1109/TCYB.2014.2332037>
- [48] Z.-G. Liu, Q. Pan, J. Dezert, A. Martin, *Adaptive imputation of missing values for incomplete pattern classification*, *Pattern Recognition*, Vol. 52, pp. 85–95, 2016. <https://doi.org/10.1016/j.patcog.2015.10.001>
- [49] T. Denœux, O. Kanjanatarakul, S. Sriboonchitta, *EK-NNclus: A clustering procedure based on the evidential K-nearest neighbor rule*, *Knowledge-Based Systems*, Vol. 88, pp. 57–69, 2015. <https://doi.org/10.1016/j.knsys.2015.08.007>
- [50] Z.-G. Liu, J. Dezert, G. Mercier, Q. Pan, *Belief C-Means: An extension of Fuzzy C-Means algorithm in belief functions framework*, *Pattern Recognition Letters*, Vol. 33, pp. 291–300, 2012. <https://doi.org/10.1016/j.patrec.2011.10.011>
- [51] Q. Guo, Y. He, T. Jian, et al., *An evidence clustering DSMT approximate reasoning method for more than two sources*, *Digital Signal Processing: A Review Journal*, Vol. 56, pp. 79–92, 2016. <https://doi.org/10.1016/j.dsp.2016.05.007>
- [52] Y.W. Du, W.M. Xu, *Multiattribute group decision making based on interval-valued intuitionistic fuzzy sets and analytically evidential reasoning methodology*, *Journal of Intelligent and Fuzzy Systems*, Vol. 33, pp. 2953–2960, 2017. <https://doi.org/10.3233/JIFS-169346>
- [53] Y.W. Du, Y.M. Wang, *Evidence combination rule with contrary support in the evidential reasoning approach*, *Expert Systems with Applications*, Vol. 88, pp. 193–204, 2017. <https://doi.org/10.1016/j.eswa.2017.06.045>
- [54] Y.W. Du, N. Yang, W. Zhou, C.X. Li, *A Reliability-Based Consensus Model for Multiattribute Group Decision-Making with Analytically Evidential Reasoning Approach*, *Mathematical Problems in Engineering*, Vol. 2018, pp. 1–14, 2018. <https://doi.org/10.1155/2018/1651857>
- [55] T.L. Saaty, *Decision-making with the AHP: Why is the principal eigenvector necessary*, *European Journal of Operational Research*, Vol. 145, pp. 85–91, 2003. [https://doi.org/10.1016/S0377-2217\(02\)00227-8](https://doi.org/10.1016/S0377-2217(02)00227-8)
- [56] T.L. Saaty, *Time dependent decision-making; dynamic priorities in the AHP/ANP: Generalizing from points to functions and from real to complex variables*, *Mathematical and Computer Modelling*, Vol. 46(7–8), pp. 860–891, Oct. 2007. <https://doi.org/10.1016/j.mcm.2007.03.028>
- [57] N. Stebler, G. Schuepbach-Regula, P. Braam, L.C. Falzon, *Use of a modified Delphi panel to identify and weight criteria for prioritization of zoonotic diseases in Switzerland*, *Preventive Veterinary Medicine*, Vol. 121, pp. 165–169, 2015. <https://doi.org/10.1016/j.prevetmed.2015.05.006>
- [58] W. Li, W. Liang, L. Zhang, Q. Tang, *Performance assessment system of health, safety and environment based on experts' weights and fuzzy comprehensive evaluation*, *Journal of Loss Prevention in the Process Industries*, Vol. 35, pp. 95–103, 2015. <https://doi.org/10.1016/j.jlp.2015.04.007>
- [59] X. Li, J. Dezert, F. Smarandache, X. Huang, *Evidence supporting measure of similarity for reducing the complexity in information fusion*, *Information Sciences*, Vol. 181, pp. 1818–1835, 2011. <https://doi.org/10.1016/j.ins.2010.10.025>
- [60] F.B. Yang, X.X. Wang, *Conflict Evidence Composition Method of D-S Evidence Theory*, National Defense Industry Press, Beijing, China, 2010.
- [61] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [62] W. Liu, *Analyzing the degree of conflict among belief functions*, *Artificial Intelligence*, Vol. 170, pp. 909–924, 2006. <https://doi.org/10.1016/j.artint.2006.05.002>
- [63] E. Herrera-Viedma, S. Alonso, F. Chiclana, F. Herrera, *A Consensus Model for Group Decision Making With Incomplete Fuzzy Preference Relations*, *IEEE Trans. on Fuzzy Systems*, Vol. 15(5), pp. 863–877, Oct. 2007.
- [64] I.J. Pérez, F.J. Cabrerizo, E. Herrera-Viedma, *Group decision making problems in a linguistic and dynamic context*, *Expert Systems with Applications*, Vol. 38, pp. 1675–1688, 2011. <https://doi.org/10.1016/j.eswa.2010.07.092>
- [65] J. Xu, Z. Wu, *A discrete consensus support model for multiple attribute group decision making*, *Knowledge-Based Systems*, Vol. 24, pp. 1196–1202. <https://doi.org/10.1016/j.knsys.2011.05.007>
- [66] J. Wu, F. Chiclana, H. Fujita, E. Herrera-Viedma, *A visual interaction consensus model for social network group decision making with trust propagation*, *Knowledge-Based Systems*, Vol. 122, pp. 39–50, 2017. <https://doi.org/10.1016/j.knsys.2017.01.031>
- [67] P. Smets, *Decision making in the TBM: The necessity of the pignistic transformation*, *International Journal of Approximate Reasoning*, Vol. 38, pp. 133–147, 2005. <https://doi.org/10.1016/j.ijar.2004.05.003>
- [68] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion (Collected Works)*, Volume 3, American Research Press (ARP), Rehoboth, USA, 2009.
- [69] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in *Proc. of the 13th Int. Conf. on Information Fusion (Fusion 2010)*, Edinburgh, Scotland, UK, 26–29 July 2010. <https://doi.org/10.1109/ICIF.2010.5712071>

# A Novel Multi-Criteria Discounting Combination Approach for Multi-Sensor Fusion

Yilin Dong<sup>a</sup>, Xinde Li<sup>a,b</sup>, Jean Dezert<sup>c</sup>, Shuzhi Sam Ge<sup>d,e</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing, China.

<sup>b</sup>The School of Cyber Science and Engineering, Southeast University, Nanjing, China.

<sup>c</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>d</sup>Social Robotics Laboratory, ECE Dept., Interactive Digital Media Institute, National Univ. of Singapore, Singapore.

<sup>e</sup>Institute for Future (IFF), Qingdao University, China.

Emails: yldong@shmtu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr, samge@nus.edu.sg

Originally published as: Y. Dong, X. Li, J. Dezert, S.S. Ge, *A Novel Multi-Criteria Discounting Combination Approach for Multi-Sensor Fusion*, IEEE Sensors Journal, Vol. 19(20), pp. 9411–9421, October 2019, and reprinted with permission.

**Abstract**—Belief function theory manages uncertain information and offers useful combination rules for multi-sensor fusion. However, when sensor readings are in conflict or even unreliable, the quality of the fusion result is significantly affected. Recently, many discounting approaches have been proposed to combine unreliable sensor readings. The discounting factors involved in these methods are often determined based on a single criterion which is not sufficient in general to obtain a precise assessment of the reliability degrees of the sources to combine. In this work, that is why we propose a novel discounting combination approach, in which the reliability factors are obtained by using the multi-criteria strategy. Our discounting combination method includes two main steps. The first step to assess the sensor's reliability is based on belief function-based technique for order preference by similarity to ideal solution (BF-TOPSIS). The second step is to discount and global combine all involved sensor readings according to their degree of reliability with proportional conflict redistribution no. 6 (PCR6) rule. Several simulations and comprehensive comparisons with classical approaches are given to show the efficiency of our proposed method.

**Keywords:** Belief function theory, multi-sensor fusion, conflict measure, multi-criteria, sensor reliability.

## I. INTRODUCTION

In order to achieve the accurate and complete description of an environment, multi-sensor fusion technology is applied to combine data from multi-sources. In view of their good applicability, multi-sensor systems play an essential role in real-world applications which include wireless sensor networks [1], [2], image processing [3], [4], target tracking [5], health-related areas [6], environmental monitoring [7] and so on [8], [9]. Nevertheless, depending on environmental and working conditions, like sensor failure, deterioration of energy supply, adverse weather conditions etc., the corresponding sensor readings can be incorrect, imprecise, conflicting or even unreliable so that it may yield a wrong decision. Thus, in order to avoid a degradation of the multi-sensor fusion system performances [10], the reliability degree of sensor reading needs to be estimated in the process of combination.

Before evaluating the reliability degrees of sensors, the imprecision or uncertain information of sensor readings should be mathematically described. Many methods can be used to handle uncertainty information, such as maximum entropy

[11], [12], Bayesian theory [13]–[15] and Belief Functions (BF) theory [16]–[18]. BF allows to model uncertainty and fuse sensors' measurements [19], and in this paper we focus our discussions on BF theory. Several classical combination rules are provided by BF theory to fuse the pieces of sensor readings. Among all available combination rules, Dempster's rule proposed by Shafer in [20] is the most well-known rule still used in many application even if it remains very controversial. Indeed, if there exists the high or even low conflict between sensor readings during the combination, a counter-intuitive result may occur. To circumvent the problem of Dempster's rule, usually the system designer discounts original sensor readings before applying the combination [21]–[24].

In the discounting approach, the primary challenge is to accurately estimate the reliability degrees of the sensors. Several researchers in [22], [23], [25] estimated the reliability degree of each sensor reading according to the single criterion. However, classical methods based on mono-criterion strategies are not good enough to assess the reliability factor. Recently, Frikha [26], [27] presented two multi-criteria strategies to compute the discounting factors. In [26], authors proposed a novel method to evaluate the degree of imprecision of sensor readings and conflict between pieces of evidence according to six criteria by using PROMETHEE II. Similarly, Frikha [27] suggested another way to estimate the reliability of the sensors based on Analytical Hierarchy Process (AHP). Based on Frikha's works, Sarabi-Jamab in [28] also followed the multi-criteria line and proposed a new selective multi-criteria method based on AHP. Different from the mentioned methods in [26] and [27], all involved criteria are further evaluated and the most discriminative criteria are selected. Some appealing findings have been revealed in [26], [27] and [28] when compared to those mono-criterion approaches currently available in the literature. However, all the multi-criteria methods used in the aforementioned references require a data normalization step, which affects the process of the precise evaluation how much better or worse to some extent a sensor is with respect to the others.

In this paper, a novel multi-criteria discounting combination method is proposed. The discounting factors associated with sensor readings are evaluated by using Belief Function based Technique for order preference by Similarity to Ideal Solution (BF-TOPSIS) [29]. We take two classes of criteria into account in the process of calculating the weights of sensor readings. The first class is the conflict between the sensors. The second class is the imprecision of the information provided by each sensor. Once all the weights are calculated, the involved sensor readings are discounted and global combined with Proportional Conflict Redistribution no. 6 (PCR6) rule. The main contributions of this paper are summarized as follows:

- A novel multi-criteria discounting combination approach is given. In our method, the recent proposed BF-TOPSIS is first applied to evaluate the reliability of the involved sensors. We also propose a new method to calculate weights of criteria involved in BF-TOPSIS. Moreover, the comprehensive comparisons between classical discounting combination methods are also illustrated in detail;
- The global fusion with PCR6 rule in Dezert-Smarandache Theory (DSMT) is applied for combining sensor readings. Unlike the conventional discounting combination approaches [26], [27] and [28], which combine all evidences with Dempster's rule, we herein use PCR6 to combine all discounted sensor readings. The advantage of this new approach is that it yields reasonable results particularly when the sensor readings are in high conflict.

This paper is organized as follows: we introduce the basic concepts of BF theory in section II. The Section III describes the discounting strategies in BF theory and the involved criteria applied in our proposed method. In section IV, our proposed discounting combination rule based on the BF-TOPSIS is described in detail. Then, the simulation results and discussions are given in section V. Finally, we conclude and give some perspectives in section VI.

## II. BASICS OF BELIEF FUNCTIONS

Belief function assigns mass of belief to the subsets of Frame of Discernment (FoD). In general, a mass function  $m(\cdot)$  is a mapping defined as follows [20] and for  $X \subseteq 2^\Theta$ :

$$m : 2^\Theta \rightarrow [0, 1], \sum_{X \subseteq 2^\Theta} m(X) = 1, \quad (1)$$

$$m(\emptyset) = 0, m(X) > 0. \quad (2)$$

where  $\Theta$  represents FoD which includes a set of  $p$  hypotheses.  $m(\cdot)$  is a mapping function and this function is also called Basic Belief Assignment (BBA). When  $m(X) > 0$ , the element  $X$  is called Focal Element (FE) of  $m$ . The set of focal elements of a BBA  $m(\cdot)$  is denoted  $F(m)$ .

In BF theory, the combination of two independent Body of Evidences (BoEs) by Dempster's rule is denoted  $m_1 \oplus m_2$ . For  $\forall X \subseteq 2^\Theta, X \neq \emptyset$ , the belief of  $X$  is given by [20]:

$$(m_1 \oplus m_2)(X) = \frac{1}{1-K} \cdot \sum_{Y, Z \subseteq 2^\Theta, Y \cap Z = X} m_1(Y)m_2(Z). \quad (3)$$

where  $K$  represents the degree of conflict between  $m_1$  and  $m_2$  as:

$$K = \sum_{Y, Z \subseteq 2^\Theta, Y \cap Z = \emptyset} m_1(Y)m_2(Z). \quad (4)$$

To palliate the drawbacks of Dempster's rule, Martin and Osswald [30] proposed a very interesting combination rule: PCR6. Due to its good performance, it is widely applied in recent applications. The combination of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  by the PCR6 rule is given as follows: for  $m_{PCR6}(\emptyset) = 0$  and  $\forall X \in 2^\Theta$

$$m_{PCR6}(X) = m_{12}(X) + \sum_{Y \in 2^\Theta \setminus \{X\}, X \cap Y = \emptyset} \left[ \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right]. \quad (5)$$

where  $m_{12}(X) = \sum_{Y, Z \subseteq 2^\Theta, Y \cap Z = X} m_1(Y)m_2(Z)$  is the conjunctive operator, and each element  $X$  and  $Y$  are expressed in their disjunctive normal form. If the denominator involved in the fraction is zero, then this fraction is discarded.

We recall that the PCR6 formula for the combination of two BBAs coincides with PCR5 formula originally developed by Smarandache and Dezert in [31]. The combination of more than two BBAs altogether with PCR5 and with PCR6 fusion rule provides in general different results. The choice of PCR6 with respect to PCR5 has been justified at first by Martin and Osswald in [30] from on a specific application, and then theoretically by Smarandache and Dezert in [32]. The general formula of PCR6 for combining more than two BBAs is given in details in [30] with examples.

To make a final decision in BF from a BBA  $m(\cdot)$ , Smets [20] suggests to transform  $m(\cdot)$  into pignistic probability by using function BetP in pignistic level. For  $\forall X \subseteq \Theta$ ,  $Bet P(\cdot)$  is defined as:

$$Bet P(X) = \sum_{Y \subseteq 2^\Theta} \frac{|X \cap Y|}{|Y|} \cdot m(Y). \quad (6)$$

where  $|Y|$  refers to the cardinality of a subset  $Y$ .

## III. DISCOUNTING PROCEDURE AND ASSESSMENT CRITERIA

### A. Discounting Procedure

The discounting operations are frequently conducted by using the discounting factor  $\omega$  with each sensor reading. Firstly introduced by Shafer [20], this factor  $\omega$  is evaluated and regarded as the reliability of the sensor reading. In this paper, the discounting factors are determined by multi-criteria strategy: BF-TOPSIS. In general, the parameter  $\omega$  varies in an interval:  $[0, 1]$ : that is to say, if the value of  $\omega$  is closer to 1, the greater the reliability of sensor reading is. The discounting steps are given as follows and for  $\forall X \in 2^\Theta \setminus \{\emptyset\}$ :

$$\begin{cases} m^\omega(X) = \omega \cdot m(X), \\ m^\omega(\emptyset) = \omega \cdot m(\emptyset) + (1 - \omega). \end{cases} \quad (7)$$

where  $X$  refers to the FE of  $m(\cdot)$  and  $\omega \in [0, 1]$ .

## B. Assessment Criteria

1) *Imprecision*: The degree of non-specificity of a sensor reading is also regarded as a degree of imprecision. With the belief function, the imperfection of BoE is mainly caused by two factors: the first one is contradiction (strife:  $C_1$ ) proposed by Vejnarová and Klir [33] and imprecision (non-specificity:  $C_2$ ) proposed by Dubois and Prade [34]. The measure of contradiction (i.e. the strife  $C_1$ ) is defined as:

$$\begin{aligned} C_1(m) &= St(m) \\ &= - \sum_{X \in F(m)} m(X) \log_2 \left( \sum_{Y \in F(m)} \frac{|X \cap Y|}{|X|} m(Y) \right), \end{aligned} \quad (8)$$

where  $|X \cap Y|$  and  $|X|$  refers to the cardinality of the subset  $|X \cap Y|$  and  $X$ .

Also, the measure of non-specificity  $C_2$  is defined as:

$$\begin{aligned} C_2(m) &= I(m) \\ &= \sum_{X \in F(m)} m(X) \cdot \log_2(|X|), \end{aligned} \quad (9)$$

where  $|X|$  refers to the cardinality of a subset  $X$ .

2) *Conflict Between the BoEs*: The second class of evaluation criteria relates to conflicting information, which is usually represented by  $m_{\oplus}(\emptyset)$  and distance measures. Two conflict measures are used in this paper: Shafer's weight of conflict ( $C_3$ ), and the interval distance between evidences ( $C_4$ ):

$$\begin{aligned} C_3(m) &= Conf(m) \\ &= \frac{1}{M} \sum_{i=1}^M [-\log_2(1 - K(m, m_i))], \end{aligned} \quad (10)$$

where  $K$  is the conflict between two BoEs  $m(\cdot)$  and  $m_i(\cdot)$  calculated by Eq.(4) and  $M$  is the number of sensor readings.

The criterion  $C_4$  is based on the interval distance  $d_{BI}$ , see [35], [36], that is:

$$\begin{aligned} C_4(m) &= d_{BI}^E(m) = \frac{1}{M} \sum_{i=1}^M d_{BI}(m, m_i) \\ &= \frac{1}{M} \sum_{i=1}^M \left[ \sqrt{n_c \cdot \sum_{X \in F(m)} [d^I(BI(X), BI_i(X))]^2} \right] \end{aligned} \quad (11)$$

where,  $n_c = 1/2^{(p-1)}$ ,  $M$  is the number of sensor readings, and  $p$  is the number of FEs in  $\Theta$ ,  $BI(X) \triangleq [Bel(X), Pl(X)]$ , and  $BI_i(X) \triangleq [Bel_i(X), Pl_i(X)]$ , and

$$\begin{aligned} d^I([a, b], [a', b']) &= \left[ \left[ \frac{a+b}{2} - \frac{a'+b'}{2} \right]^2 \right. \\ &\quad \left. + \frac{1}{3} \left[ \frac{b-a}{2} - \frac{b'-a'}{2} \right]^2 \right]^{\frac{1}{2}} \end{aligned}$$

## IV. NEW COMBINATION APPROACH BASED ON BF-TOPSIS

### A. Construction of Scoring Matrix

At first, there exists  $A_i$ ,  $i = 1, \dots, M$  sensors, and each sensor gives a corresponding reading  $m_i$ ,  $i = 1, \dots, \theta_p$  according to eq. (12)

$$\begin{matrix} & X_1 & X_2 & \dots & X_{2^{|\Theta|}} \\ \begin{matrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{matrix} & \begin{bmatrix} m_1(X_1) & m_1(X_2) & \dots & m_1(X_{2^{|\Theta|}}) \\ m_2(X_1) & m_2(X_2) & \dots & m_2(X_{2^{|\Theta|}}) \\ \dots & \dots & \dots & \dots \\ m_M(X_1) & m_M(X_2) & \dots & m_M(X_{2^{|\Theta|}}) \end{bmatrix} \end{matrix} \quad (12)$$

where  $X \in 2^\Theta$ .

Then, one calculates the evaluation of all sensors from the perspective of imprecision ( $C_1$  and  $C_2$ ) and conflict ( $C_3$  and  $C_4$ ), and then constructs the following scoring matrix  $S$  defined by eq. (13)

$$\begin{matrix} & A_1 & A_2 & \dots & A_i & \dots & A_M \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{matrix} & \begin{bmatrix} S_1(A_1) & S_1(A_2) & \dots & S_1(A_i) & \dots & S_1(A_M) \\ S_2(A_1) & S_2(A_2) & \dots & S_2(A_i) & \dots & S_2(A_M) \\ S_3(A_1) & S_3(A_2) & \dots & S_3(A_i) & \dots & S_3(A_M) \\ S_4(A_1) & S_4(A_2) & \dots & S_4(A_i) & \dots & S_4(A_M) \end{bmatrix} \end{matrix} \quad (13)$$

Without loss of generality, we just use the general mathematical symbol  $C_j$ ,  $j = 1, \dots, N$  (in this paper,  $N = 4$ ) to represent one of the mentioned four criteria for convenience in the following sections. That is:  $j = 1$ ,  $C_1 \triangleq St(\cdot)$ ;  $j = 2$ ,  $C_2 \triangleq I(\cdot)$ ;  $j = 3$ ,  $C_3 \triangleq Conf(\cdot)$ ;  $j = 4$ ,  $C_4 \triangleq d_{BI}^E(\cdot)$ .

### B. Construction of BBAs for Multi-Criteria Decision Making (MCDM) Problems

In traditional mono-criterion problems, the weights of sensors  $A_i$ ,  $i = 1, \dots, M$  are calculated according to a single criterion. However, in MCDM problems, the direct weights associated with different criteria can be very different. Therefore, efficient fusion techniques must be developed in order to provide the global evaluating solution to solve the MCDM problem. For this aim, original BF-TOPSIS is used to Estimate the Ranking Vector (ERV) from all evidences that support or refute each sensor thanks to BBAs. First, the FoD is the set of sensors, that is  $A \triangleq \{A_1, A_2, \dots, A_M\}$ . The construction of BBAs is based on the method of construction presented in [29]

$$m_j(A_i) \triangleq Bel_j(A_i), \quad (14)$$

$$m_j(\bar{A}_i) \triangleq Bel_j(\bar{A}_i) = 1 - Pl_j(A_i), \quad (15)$$

$$m_j(A_i \cup \bar{A}_i) \triangleq Pl_j(A_i) - Bel_j(A_i), \quad (16)$$

where  $m_j(A_i)$  means the support belief in favor of  $A_i$  according to criterion  $C_j$ ,  $m_j(\bar{A}_i)$  means the support belief against  $A_i$  according to  $C_j$  and  $m_j(A_i \cup \bar{A}_i)$  means the uncertainty degree whether support or against  $A_i$  based on  $C_j$ .

$Bel_j(A_i)$ ,  $Pl_j(A_i)$  and  $Bel_j(\bar{A}_i)$  in Eqs.(15) and (16) are defined as follows:

$$\begin{aligned} Sup_j(A_i) &\triangleq \sum_{k \in \{1, \dots, M\} | S_j(A_k) \leq S_j(A_i)} |S_j(A_i) - S_j(A_k)|, \end{aligned} \quad (17)$$

$$\begin{aligned} Bel_j(A_i) &\triangleq \frac{Sup_j(A_i)}{A_{\max}^{j-Sup}} \\ &= \frac{Sup_j(A_i)}{\max(Sup_j(A_1), Sup_j(A_2), \dots, Sup_j(A_M))}, \end{aligned} \quad (18)$$

$$Inf_j(A_i) \triangleq - \sum_{k \in \{1, \dots, M\} | S_j(A_k) > S_j(A_i)} |S_j(A_i) - S_j(A_k)| \quad (19)$$

$$\begin{aligned} Bel_j(\bar{A}_i) &\triangleq \frac{Inf_j(A_i)}{A_{\min}^{j-Inf}} \\ &= \frac{Inf_j(A_i)}{\min(Inf_j(A_1), Inf_j(A_2), \dots, Inf_j(A_M))}. \end{aligned} \quad (20)$$

and

$$Pl_j(A_i) \triangleq 1 - Bel_j(\bar{A}_i). \quad (21)$$

### C. Calculation of Criteria Weights

In original BF-TOPSIS [29], the weights of criteria are often chosen subjectively which limits the applications of this new multi-criteria strategy in practice. In this paper, we automatically determine the importance of each criterion from the scoring matrix without manual intervention.

1) *Normalized Scoring Matrix According to the Max-Min Scaling*: Here, we first transform all values of  $S_j(A_j)$  in S (Eq.(13)) into the same measurement scale based on Max-Min scaling.

*Definition 1*: The normalized scoring matrix is defined by

$$S'_j(A_i) = \frac{S_j(A_i) - A_{\min}^{j-S}}{A_{\max}^{j-S} - A_{\min}^{j-S}}, \quad (22)$$

where  $i \in \{1, \dots, M\}$ ,  $j \in \{1, \dots, N\}$  and  $A_{\max}^{j-S}$  and  $A_{\min}^{j-S}$  refer to the maximum value and minimum value in  $[S_j(A_1), S_j(A_2), \dots, S_j(A_M)]$ .

The aim of this Max-Min scaling is to transform the original scoring linearly so that all the values of elements in scoring matrix are within the interval  $[0, 1]$ .

2) *Pairwise Comparison Matrix for All Criteria*: Based on Eq.(22), we can calculate the pairwise comparison results for all criteria according to the following definition:

*Definition 2*: The pairwise comparison matrix is defined by

$$pc_{h'h} = \frac{\sum_{i=1}^M \exp(S'_{h'}(A_i))}{\sum_{i=1}^M \exp(S'_h(A_i))}, \quad (23)$$

where  $h' \in \{1, \dots, N\}$ , and  $h \in \{1, \dots, N\}$ . In this paper, the value of parameter N is 4, which represents the number of the involved criteria.

Obviously, Eq.(23) is consistent because all pairs of criteria satisfy  $pc_{h'l} \times pc_{lh} = pc_{h'h}$ .

*Proof*:

$$\begin{aligned} pc_{h'l} \times pc_{lh} &= \frac{\sum_{i=1}^M \exp(S'_{h'}(A_i))}{\sum_{i=1}^M \exp(S'_l(A_i))} \times \frac{\sum_{i=1}^M \exp(S'_l(A_i))}{\sum_{i=1}^M \exp(S'_h(A_i))} \\ &= \frac{\sum_{i=1}^M \exp(S'_{h'}(A_i))}{\sum_{i=1}^M \exp(S'_h(A_i))} = pc_{h'h}. \end{aligned}$$

*End Proof*.

3) *Weights Determination for All Criteria*: Next, we can calculate the weight of each criterion  $v(C_j)$  based on the following expressions:

$$v(C_j) = \frac{\sum_{h=1}^N pc'_{jh}}{M}, \quad (24)$$

$$pc'_{jh} = \frac{pc_{jh}}{\sum_{l=1}^N pc_{lh}}, \quad (25)$$

where  $j = 1, \dots, N$  and  $N = 4$  in this paper;  $M$  is the number of sensor readings.

### D. Steps of BF-TOPSIS Algorithm

- *Step 1*: From the scoring matrix  $S$ , compute BBAs  $m_j(A_i)$ ,  $m_j(\bar{A}_i)$  and  $m_j(A_i \cup \bar{A}_i)$  using Eqs. (15)–(16), and then construct  $vec_j(A_i) \triangleq [m_j(A_i), m_j(\bar{A}_i), m_j(A_i \cup \bar{A}_i)]$  for each sensor  $A_i$  according to the criterion  $C_j$ ;
- *Step 2*: For each sensor  $A_i$ , also construct the best ideal BBA  $vec_j^{best}(A_i) \triangleq [m_j^{best}(A_i), 0, 0] \triangleq [1, 0, 0]$  and the worst ideal BBA  $vec_j^{worst}(A_i) \triangleq [0, m_j^{worst}(\bar{A}_i), 0] \triangleq [0, 1, 0]$ ; Then compute the belief interval distance [35], [36]  $d_{BI}(vec_j(A_i), vec_j^{best}(A_i))$  and  $d_{BI}(vec_j(A_i), vec_j^{worst}(A_i))$ ;
- *Step 3*: Compute the weighted average of distance  $d_{BI}(vec_j(A_i), vec_j^{best}(A_i))$  and distance  $d_{BI}(vec_j(A_i), vec_j^{worst}(A_i))$  with the relative importance weighting factor  $v(C_j)$  of each criterion  $C_j$ , that is

$$dis^{best}(A_i) \triangleq \sum_{j=1}^N v(C_j) \cdot d_{BI}(vec_j(A_i), vec_j^{best}(A_i)), \quad (26)$$

$$dis^{worst}(A_i) \triangleq \sum_{j=1}^N v(C_j) \cdot d_{BI}(vec_j(A_i), vec_j^{worst}(A_i)), \quad (27)$$

where  $vec_j(A_i) \triangleq [m_j(A_i), m_j(\bar{A}_i), m_j(A_i \cup \bar{A}_i)]$ ,  $vec_j^{best}(A_i) \triangleq [1, 0, 0]$  and  $vec_j^{worst}(A_i) \triangleq [0, 1, 0]$ .

- *Step 4*: The final weights  $\omega$  of the sensor  $A_i$  with respect to ideal best solution  $A^{best}$  is then defined by

$$\omega(A_i) \triangleq \frac{dis^{worst}(A_i)}{dis^{worst}(A_i) + dis^{best}(A_i)} \quad (28)$$

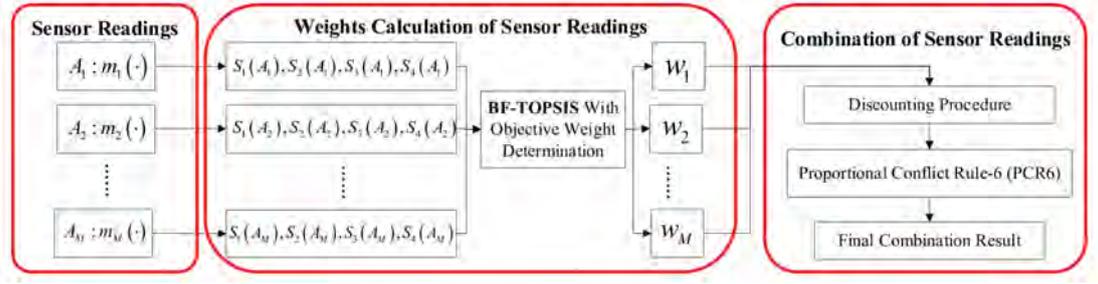


Figure 1. The framework of our proposed multi-criteria discounting combination method.

### E. Discounting sensor readings and PCR6 combination

- All the weights of the involved sensors are constructing a weight vector, which is denoted as  $Weight = \{\omega(A_1), \omega(A_2), \dots, \omega(A_M)\}$  based on Eq.(28) and then discount all involved BBAs with Eq.(7).
- PCR6 Combination Rule: According to the sensor reading of  $A_i, i \in \{1, \dots, M\}$  and PCR6 combination rule, we can globally combine all the involved sensor readings with their corresponding weights:  $m_{fusion} = PCR6(m_1^{\omega(A_1)}(\cdot), m_1^{\omega(A_2)}(\cdot), \dots, m_1^{\omega(A_M)}(\cdot))$ .

### F. A Proposed Combination Method and Fusion Process

Because the unreliable sensors often provide conflicting and imprecision information which may lead to counter-intuitive results in traditional combination methods, the novel multi-criteria discounting combination method is used. Here, we would like to emphasize that the involved criteria mentioned in this paper can be modified to any criteria according to the actual demand in the process of the objective weight estimation.

The process of our multi-criteria discounting combination method is illustrated in Fig.1 and in order to help the readers to reproduce this new combination method proposed in this paper, pseudo code is given in Algorithm 1.

## V. NUMERICAL SIMULATIONS AND DISCUSSIONS

In this part, we make comprehensive comparisons between the proposed combination procedure and other classical combination rules. Also, several tests and comparative analysis are illustrated in details. Independent random runs of Monte Carlo simulations are generated to observe the appealing behaviors of the proposed approach. All simulations results were obtained with MATLAB R2018a running with a hardware of Intel Core i7-5600U CPU at 2.60GHz and with 8G RAM.

### A. Target Recognition Context

In this case, six sensors give their corresponding readings which consider the class of the same target in Table I. The common FoD of these sensors is  $\Theta = \{\theta_1 \triangleq \text{target 1}, \theta_2 \triangleq \text{target 2}, \theta_3 \triangleq \text{target 3}\}$ . Among the given BBAs in Table I, we can notice that four of these sensors (1, 2, 4 and 5) give the maximum belief to  $\theta_1$ . On the contrary, sensor 3 assigns most

---

### Algorithm 1: The Proposed Multi-Criteria Discounting Combination Method

---

**Input :** Sensor Readings:  
 $A_1 : m_1(\cdot), A_2 : m_2(\cdot), \dots, A_M : m_M(\cdot)$ .

**Output:** The Fused Final BBA  $m_{fusion}(\cdot)$ .

```

1 for  $j = 1, \dots, N$  do
2   for  $i = 1, \dots, M$  do
3      $S_j(A_i) = C_j(m_i(\cdot))$ ;
4   end
5 end
6 for  $j = 1, \dots, N$  do
7   for  $i = 1, \dots, M$  do
8      $S'_j(A_i) = \frac{S_j(A_i) - A_{\min}^{j-S}}{A_{\max}^{j-S} - A_{\min}^{j-S}}$ ;
9   end
10 end
11 for  $j = 1, \dots, N$  do
12   for  $i = 1, \dots, M$  do
13      $Sup_j(A_i) \triangleq$ 
14        $\sum_{k \in \{1, \dots, M\} | S_j(A_k) \leq S_j(A_i)} |S_j(A_i) - S_j(A_k)|$ ;
15      $Inf_j(A_i) \triangleq$ 
16        $-\sum_{k \in \{1, \dots, M\} | S_j(A_k) > S_j(A_i)} |S_j(A_i) - S_j(A_k)|$ ;
17   end
18 end
19 for  $j = 1, \dots, N$  do
20   for  $i = 1, \dots, M$  do
21      $m_j(A_i) \triangleq Bel_j(A_i)$ ;
22      $m_j(\bar{A}_i) \triangleq Bel_j(\bar{A}_i) = 1 - Pl_j(A_i)$ ;
23      $m_j(A_i \cup \bar{A}_i) \triangleq Pl_j(A_i) - Bel_j(A_i)$ ;
24   end
25 end
26 for  $j = 1, \dots, N$  do
27   for  $i = 1, \dots, M$  do
28      $\omega(A_i) = \frac{dis^{worst}(A_i)}{dis^{worst}(A_i) + dis^{best}(A_i)}$ ;
29   end
30 Discounting Step: Using Eq.(7) Based on Weight;
31 Fusion Step:  $m_{Fusion} = PCR6(m_1(\cdot), \dots, m_M(\cdot))$ ;
32 return The Fused BBA  $m_{Fusion}(\cdot)$ 

```

---

Table I  
SENSOR'S CORRESPONDING BBAs  $m_i(\cdot)$ .

	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	$m_6$
$\theta_1$	0.75	0.4	0	0.35	0.5	0.05
$\theta_2$	0.1	0.2	0.9	0.15	0.1	0.1
$\theta_3$	0.05	0.1	0.1	0.25	0	0
$\{\theta_1, \theta_2\}$	0	0.3	0	0.2	0	0.3
$\{\theta_1, \theta_3\}$	0	0	0	0	0	0.2
$\{\theta_2, \theta_3\}$	0	0	0	0	0.15	0.1
$\{\theta_1, \theta_2, \theta_3\}$	0.1	0	0	0.05	0.25	0.25

of its belief to  $\theta_2$ . Accordingly, sensor 3 is highly conflicting with the mentioned four sensors (1, 2, 4 and 5).

According to Eq.(13), we can evaluate each sensor reading  $m_i$  by calculating  $S_j(m_i)$ , for all  $i = 1, \dots, 6$  and  $j = 1, \dots, 4$ . All results of  $S_j(m_i)$  are listed in the scoring matrix (Table II).

Table II  
SCORING MATRIX  $S_j(m_i)$ .

$S_j(m_i)$	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	$m_6$
$St(\cdot)$	0.6771	0.9591	0.4690	1.1508	0.6959	0.4721
$I(\cdot)$	0.1585	0.3000	0	0.2792	0.5462	0.9962
$Conf(\cdot)$	0.8813	0.6374	1.2641	0.7562	0.5317	0.3225
$d_{BI}^E I(\cdot)$	0.3785	0.2586	0.6115	0.2682	0.2622	0.3127

As we can see in Table II, the sorted rankings with the involved criteria are quite different:

$$\begin{aligned}
 St(\cdot) &: m_3 \succ m_6 \succ m_1 \succ m_5 \succ m_2 \succ m_4, \\
 I(\cdot) &: m_3 \succ m_1 \succ m_4 \succ m_2 \succ m_5 \succ m_6, \\
 Conf(\cdot) &: m_6 \succ m_5 \succ m_2 \succ m_4 \succ m_1 \succ m_3, \\
 d_{BI}^E I(\cdot) &: m_2 \succ m_5 \succ m_4 \succ m_6 \succ m_1 \succ m_3.
 \end{aligned}$$

This phenomenon indicates that it is not appropriate to evaluate the involved BBAs based on single criterion and a robust multi-criteria strategy is necessary. For this, we first calculate the positive and negative evidence supports of all sensor readings which are illustrated in Table III and Table IV.

Table III  
EVIDENTIAL SUPPORTS  $Sup_j(m_i)$ .

$Sup_j(m_i)$	$St(\cdot)$	$I(\cdot)$	$Conf(\cdot)$	$d_{BI}^E I(\cdot)$
$m_1$	0.7746	1.4877	0.3828	0.2330
$m_2$	0.1917	0.9425	0.9893	0.5404
$m_3$	1.6101	2.2802	0	0
$m_4$	0	1.0047	0.6330	0.4980
$m_5$	0.7180	0.4500	1.4123	0.5221
$m_6$	1.5944	0	2.4581	0.3646

In Table III  $Sup_j(m_i)$  is called the positive support of  $m_i$  according to criterion  $C_j$ . Because the four involved criteria we mentioned before are such that the smaller the value, the better the BoE. Thus, if the value of the BoE  $m_i$  in Table II is small according to  $C_j$  (such as  $m_3$  in  $St(\cdot)$ ), the support value  $Sup_1(m_3)$  will be the largest according to  $St(\cdot)$  in Table III.

Table IV  
EVIDENTIAL SUPPORTS  $Inf_j(m_i)$ .

$Inf_j(m_i)$	$St(\cdot)$	$I(\cdot)$	$Conf(\cdot)$	$d_{BI}^E I(\cdot)$
$m_1$	-0.4131	-0.1585	-1.2774	-0.4123
$m_2$	-1.5223	-0.4623	-0.4206	0
$m_3$	0	0	-3.1914	-1.5773
$m_4$	-2.4897	-0.4000	-0.7769	-0.0157
$m_5$	-0.4696	-1.4472	-0.2092	0.0036
$m_6$	-0.0031	-3.6972	0	-0.1492

According to Eq.(15)-Eq.(16), the specific supporting BBAs  $m_s(\cdot)$  of all involved sensors are given in Fig. 2.

In Fig. 2 (a) and (b), we can find that  $m_3$  (orange plot) receives the fully support belief (because  $m_s(m_3) = 1$ ) from the perspective of the imprecision criteria ( $St(\cdot)$  and  $I(\cdot)$ ). This is because  $m_3$  is the only Bayesian BBAs. On the contrary, since  $m_3$  is highly conflicting with other BBAs, the support against  $m_3$  in Fig. 2 (c) and (d) is largest (because  $m_s(\bar{m}_3) = 1$ ) according to the conflict measures ( $Conf(\cdot)$  and  $d_{BI}^E I(\cdot)$ ). The supporting degrees of  $m_3$  under different criteria are totally inconsistent, which directly indicates that it is difficult to evaluate the reliability of sensor readings comprehensively depending on a single criterion. In our proposed discounting combination method, the reliability factors are given by multi-criteria based on BF-TOPSIS.

According to Eqs. (22)–(25), the corresponding weights of all involved criteria are:

$$\begin{aligned}
 v(St) &= 0.2581, \\
 v(I) &= 0.1519, \\
 v(Conf) &= 0.2581, \\
 v(d_{BI}^E I) &= 0.1519.
 \end{aligned}$$

Then, all weights of the involved six sensor readings are given in Table V based on Eq. (28).

Table V  
WEIGHTS OF ALL SENSOR READINGS  $\omega(m_i)$ .

	$dis^{best}(m_i)$	$dis^{worst}(m_i)$	$\omega(m_i)$
$m_1$	2.8450	3.8938	0.5778
$m_2$	2.6323	4.0353	<b>0.6838</b>
$m_3$	2.8284	2.8284	0.5000
$m_4$	3.1726	3.3888	0.5165
$m_5$	2.2802	4.2988	0.6534
$m_6$	1.8371	3.9732	0.6052

As we can see in Table V,  $m_2$  receives the highest degree of reliability and the highly conflicting BBA  $m_3$  gets the lowest reliability degree. It is worth mentioning that the four weighting values of the involved criteria are automatically calculated based on the input sensor readings. To make a direct comparison between manual adjustment of criteria weights (subjective) and objective weighting used in BF-TOPSIS, we first set the weights of  $St(\cdot)$  and  $I(\cdot)$  to 1 and set the weights of  $Conf(\cdot)$  and  $d_{BI}^E I(\cdot)$  to 0 and we can observe that in this case the weight of  $m_3$  increases to 1 in Fig.3 (a) and

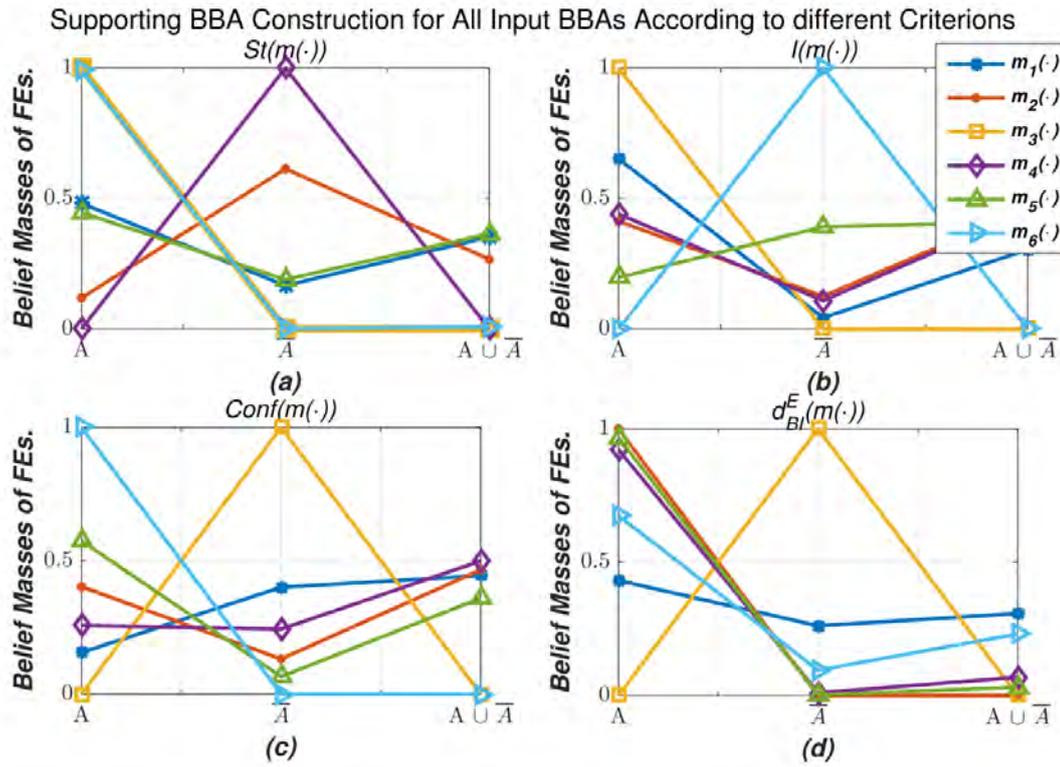


Figure 2. Construction of BBAs based on four distinct criteria. (a) FEs in supporting BBA. (b) FEs in supporting BBA. (c) FEs in supporting BBA. (d) FEs in supporting BBA.

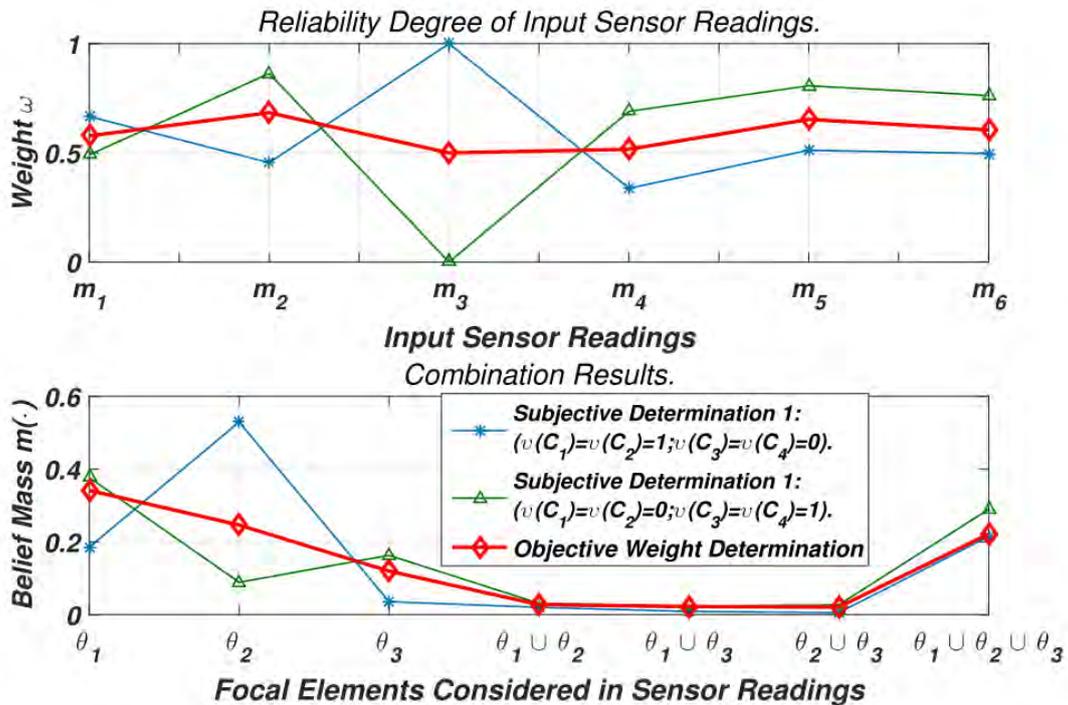


Figure 3. Comparison of subjective and objective weights of criteria for BF-TOPSIS.

accordingly, the mass of  $\theta_2$  becomes the largest in Fig.3 (b). When we modify subjectively the corresponding weights of different criteria, the corresponding weights of all BoEs and the masses of all FEs will also change. Thus, the more criteria are considered, the more weights need to be set, which makes subjective determination of criteria weighting factors more difficult in BF-TOPSIS. In this new method, we do not set the weights of all involved criteria in advance and we can see in Fig.3 (a) that the weight of  $m_3$  is calculated appropriately (red plot) and the final decision is  $\theta_1$ . Based on the obtained weights calculated by this new method, we can use PCR6 rule to fuse all discounted sensor readings. Fig. 3 (b) illustrates the corresponding belief mass of the considered FE.

### B. Fusion of High Conflicting Sensor Readings

Here we show that our multi-criteria discounting combination method can fuse high conflicting sensor readings and give reasonable belief mass distribution. We herein consider two independent BBAs  $m_1$  and  $m_2$  which are defined over the identical FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . And we also assume that these two BBAs are in highly conflict which are given as follows:

$$\begin{cases} m_1(\{\theta_1\}) = 0.99, m_1(\{\theta_3\}) = 0.01, \\ m_2(\{\theta_2\}) = 0.9, m_2(\{\theta_3\}) = 0.1. \end{cases} \quad (29)$$

In order to be able to see whether our approach can handle highly conflicting problems and how it differs from other classical approaches, we summarize the combination results in Table VI.

We can note that the Dempster's rule directly leads to counter-intuitive results because  $\theta_3$  receives the total mass of belief by using this classical rule. Martin's approach and Jiang's approach assign most values of belief to FE  $\{\theta_1, \theta_2, \theta_3\}$  and after probabilistic transformation ( $BetP(\cdot)$ ), these two methods can give the final decision:  $\theta_1$ . Frikha's approaches in [26], [27] and our approach can directly draw the conclusion that  $\theta_1$  is the final decision based on the principle of maximum probability. In addition, we also give a comparison of the computational time of each mentioned approach. Because the multi-criteria methods need to process more fusion steps, it takes more time than the classical method (Dempster's rule, Martin's method and Jiang's method). However, in this paper, those methods with lower computational complexity (such as Dempster's rule) are not the optimal choice because the criterion for judging the pros and cons of the fusion algorithms here is to give reasonable and correct decision results in face of highly conflicting fusion problems.

### C. Combination of Conflict and Imperfect BBAs

Assume that three sensors 1, 2 and 3 providing three BBAs  $m_1, m_2$  and  $m_3$  defined over the same FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  in Table VII.

As we can see, the most supported element is  $\theta_1$  by sensor 1. Contrary to sensor 1, it is  $\theta_2$  in sensors 2 and 3.

From Table VIII, we can find that the weight of sensor 1 is the lowest ( $\omega(m_1) = 0.3109$  and  $\omega(m_2) = 0.6834$ ,  $\omega(m_3) = 0.6578$ ) which means that  $m_1(\cdot)$  is the most imprecision and

also in conflict with  $m_2$  and  $m_3$ . This result is in some degree consistent with Frinkha's method [26], which also agrees that sensor  $m_1$  is much more less important than  $m_2$  and  $m_3$ . So, based on the results of the obtained weights in Table VIII, we can say that sensor 1 should not play an important rule in the combination result which means that  $m_1$  needs to be discounted.

From the given three BBAs, we can find that  $m_1$  gives the most mass of belief to FE  $\theta_1$ . At the same time,  $m_2$  and  $m_3$  considerably support  $\theta_2$ . However,  $m_2$  and  $m_3$  are not in conflict with  $m_1$  because these two BBAs also assign appropriate masses to  $\theta_1$ . When we combine these three BBAs with Dempster's rule, Jiang's approach or Martin's approach, we can see that the final combination results of these methods are almost the same in Table IX (the final decision is  $\theta_1$ ). This is because that only one conflict criterion is applied for weight calculation in Martin's method or Jiang's method. As we can see in Table VIII, conflict measure could not capture the difference between  $m_1$  and  $m_2, m_3$  which leads to almost same weights of these three BBAs. However, in our proposed method based on BF-TOPSIS, multi-criteria (imprecision and conflict measure) are both cooperated into the weight calculation which can give more comprehensive evaluations. In Table VIII,  $m_1$  receive the lowest weight compared to  $m_2$  and  $m_3$  which means that  $m_2$  and  $m_3$  have the most important effect on the final combination result. In Table IX, we can see that our proposed method supports  $\theta_2$  as the final decision and this conclusion is consistent with the other multi-criteria approaches (PROMETHEE II and AHP).

### D. Random Runs Generated by Monte Carlo Simulations

In this part, we make 50 Monte Carlo simulations and in each simulation, we generate 25 random BBAs over the identical FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The first six sensors are in favour of  $\{\theta_1\}$ , the next eight ones are in favour of  $\{\theta_3\}$ , whereas the last eleven sensors are again in favour of  $\{\theta_1\}$ . A mass function focused on  $\{\theta_1\}$  has four focal elements:  $\{\theta_1\}, \{\theta_2\}, \{\theta_3\}$  and  $\{\theta_1, \theta_2\}$  with  $m(\{\theta_1\}) = 0.45 + x$ ;  $m(\{\theta_2\}) = 0.15 - y$ ;  $m(\{\theta_3\}) = 0.15 - x$ ; and  $m(\{\theta_1, \theta_2\}) = 0.25 + y$ . The values  $x$  and  $y$  are randomly generated in the intervals of [0.01, 0.15] and [0.01, 0.10] according the uniform distribution, respectively. Besides, the average values of the weights  $\omega$  of the involved sensors and the pignistic probability  $BetP(\theta)$  are generated in the 50 Monte Carlo simulations.

Due to the fact that the weight of one BoE is independent of the other BoEs according to the class of imperfection criteria, in this case, the evolution of the weight of each sensor when we sequentially add BBAs is definitely affected by conflict criteria.

Figures 4–6 indicate the average of reliability degrees for each group of BoEs (first six BoEs, next eight BoEs, the last 11 BoEs). In each step, one BoE is added, and the discounting factors are calculated for all BoEs. The discounting factors are obtained by averaging the 50 Monte-Carlo simulations. In Fig.4, we can observe that the weights of sensors (1-6) decrease as soon as the next eight BoEs are added. This

Table VI  
COMBINATION OF HIGHLY CONFLICTING BBAs.

	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_1, \theta_2\}$	$\{\theta_1, \theta_3\}$	$\{\theta_2, \theta_3\}$	$\{\theta_1, \theta_2, \theta_3\}$	Decision	Computational time
Dempster's rule	0	0	1.0000	0	0	0	0	$\theta_3$	35.972ms
Martin's method [25]	0.239	0.217	0.027	0	0	0	0.517	$\theta_1$	49.421ms
Jiang's method [37]	0.049	0.044	0.005	0	0	0	0.902	$\theta_1$	49.728ms
Frikhas's approach [26]	0.984	0	0.016	0	0	0	0	$\theta_1$	81.139ms
Frikhas's approach [27]	0.9262	0.0738	0	0	0	0	0	$\theta_1$	73.233ms
<b>Our proposed method</b>	<b>0.5365</b>	<b>0.2089</b>	<b>0.0186</b>	0	0	0	<b>0.2360</b>	$\theta_1$	72.23ms

Table VII  
SENSOR'S CORRESPONDING BBA'S  $m_i(\cdot)$ .

	$m_1$	$m_2$	$m_3$
$\theta_1$	0.38	0.3	0.28
$\theta_2$	0.15	0.4	0.42
$\theta_3$	0.15	0	0
$\{\theta_1, \theta_2\}$	0.15	0.3	0.3
$\{\theta_1, \theta_3\}$	0.03	0	0
$\{\theta_2, \theta_3\}$	0.03	0	0
$\{\theta_1, \theta_2, \theta_3\}$	0.03	0	0

Table VIII  
RELIABILITY DEGREES.

$\omega$	$\{m_1\}$	$\{m_2\}$	$\{m_3\}$
Martin's method [25]	0.961	0.989	0.988
Jiang's method [37]	0.822	0.854	0.848
Frikhas's approach [26]	0.643	1.0000	0.989
Frikhas's approach [27]	0.8019	0.9428	0.9407
Our proposed method	0.3109	0.6834	0.6578

Table IX  
COMBINATION OF  $m_1, m_2$  AND  $m_3$ .

$m(\cdot)$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_1, \theta_2\}$	$\{\theta_1, \theta_3\}$	$\{\theta_2, \theta_3\}$	$\{\theta_1, \theta_2, \theta_3\}$	Decision	Computational time
Dempster's rule	0.5028	0.4582	0	0.0400	0	0	0	$\theta_1$	48.471ms
Martin's method [25]	0.491	0.462	0	0.047	0	0	0	$\theta_1$	56.251ms
Jiang's method [37]	0.452	0.438	0.005	0.092	0.002	0.002	0.009	$\theta_1$	56.274ms
Frikhas's approach [26]	0.418	0.5000	0	0.082	0	0	0	$\theta_2$	87.380ms
Frikhas's approach [27]	0.3938	0.4923	0.0039	0.1090	0	0	0	$\theta_2$	75.563ms
<b>Our proposed method</b>	<b>0.2285</b>	<b>0.3090</b>	<b>0.0081</b>	<b>0.2060</b>	<b>0.0012</b>	<b>0.0012</b>	<b>0.2460</b>	$\theta_2$	73.832ms

phenomenon is mainly due to that these eight sensors (sensor 7-14) are in conflict with the initial sensor readings. In Fig.5, at the beginning stage, there are no BoEs which are in conflict with the first six sensors, the weights of sensors 7-14 are relatively low. After that, their corresponding reliability degrees tend to increase because of the increment of those sensors which support  $\theta_3$ . Finally, these weights of the 7-14 sensors decrease because the remaining BoEs (Sensors 17-25) support  $\theta_1$ . It can be seen that the proposed method in this paper can give the corresponding weights of each sensor reading effectively when conflicts occur. This feature is extremely important in complex dynamic fusion problems.

We also show the evolution of  $BetP(\theta_i)$  for all  $i = 1, 2, 3$  using our proposed approach and other mentioned methods (especially the two multi-criteria methods) in Figures 7–9. As shown in Fig. 7, at the beginning, our method supports:  $\theta_1$ , then, with the increment of the number of conflicting BoEs,

our approach reacts slowly. However, when receiving many conflicting BoEs which support:  $\theta_3$ , the value of  $BetP(\theta_1)$  can decrease rapidly and then the mass of belief of  $\theta_3$  increases when the second group is added in Fig. 9. Besides, we can notice that the related reaction of our method when the second groups of BoEs are added in Fig. 8. Compared to our proposed method, the value of  $BetP(\theta_3)$  obtained by Dempster's rule in Fig. 9 is always zero which proves that Dempster's rule cannot be applied to fuse efficiently highly conflicting BoEs; The changing trends of the mentioned multi-criteria methods (Frikha's two methods in [26] and [27]) can prove that they can deal with the conflicting fusion problem to a certain extent, but it is not sensitive enough compared with our proposed method in this paper. For example, in Fig. 9, with the increment of the number of BoEs supporting  $\theta_3$ ,  $BetP(\theta_3)$  obtained by our method is much larger than the values of  $BetP(\theta_3)$  derived from these two methods [26] and [27];

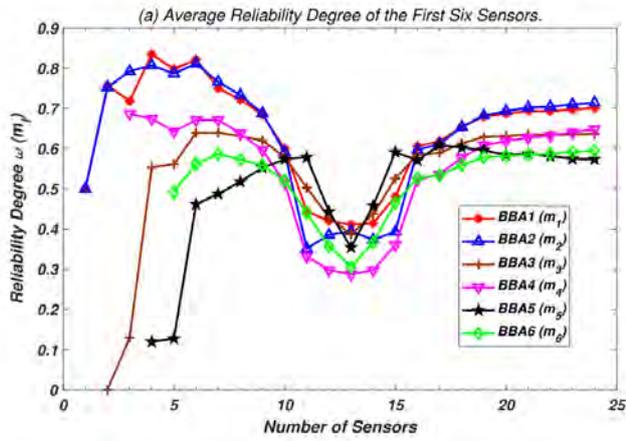


Figure 4. Evolution of the average reliability degrees (sensor 1–sensor 6).

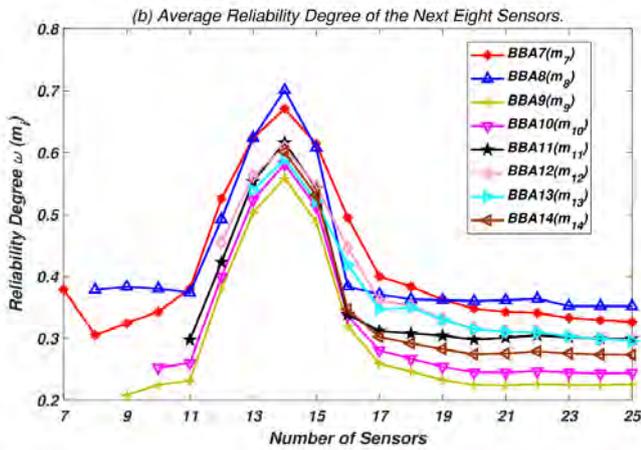


Figure 5. Evolution of the average reliability degrees (sensor 7–sensor 14).

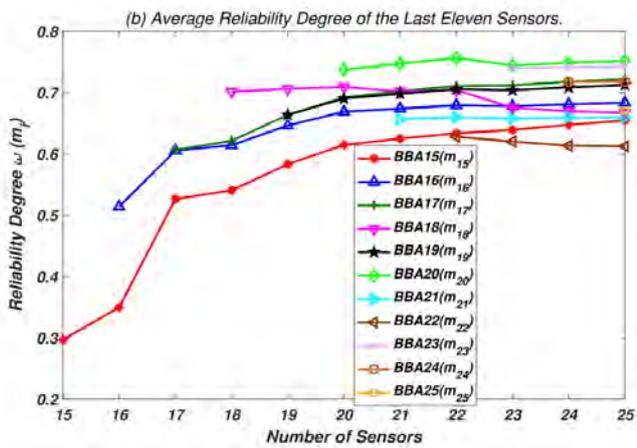


Figure 6. Evolution of the average reliability degrees (sensor 15–sensor 25).

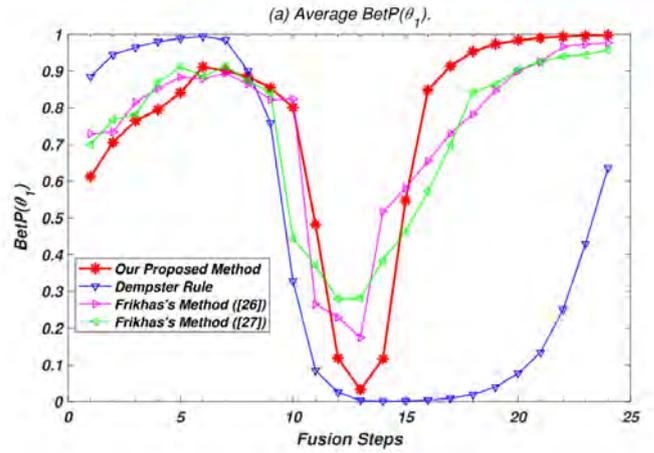


Figure 7.  $BetP(\theta_1)$  using the different methods.

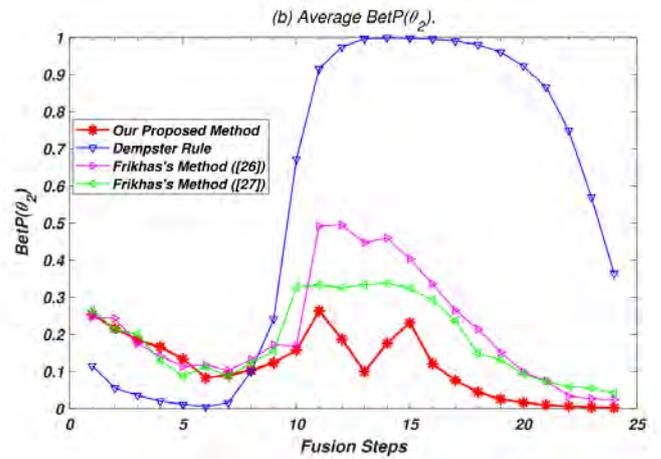


Figure 8.  $BetP(\theta_2)$  using the different methods.

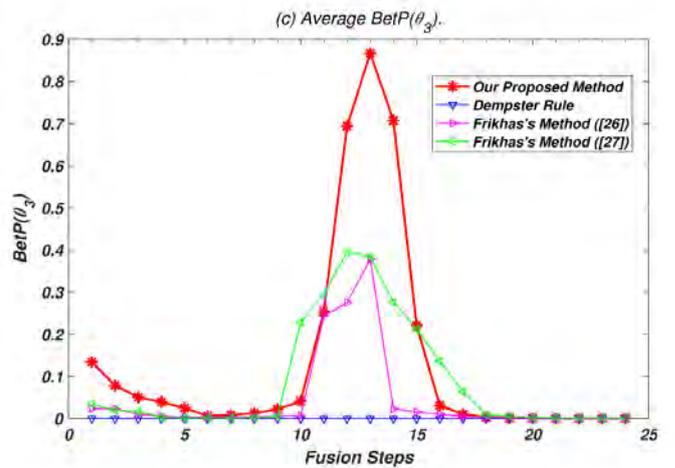


Figure 9.  $BetP(\theta_3)$  using the different methods.

As shown in Fig. 7, the convergence speed of  $BetP(\theta_1)$  in our method is also significantly faster (when the BBAs increase from 15th to 25th fusion step). It can be seen from the Monte Carlo simulations that the proposed method can adapt to the dynamic changes of the environment and give the correct decisions.

## VI. CONCLUSION

In this work, a novel multi-criteria discounting combination rule has been proposed and presented. The BF-TOPSIS approach with an original objective criteria weighting method has been used to evaluate the reliability of the involved sensor readings. The procedure of assessing the reliability degree is based on two categories of criteria. The first class is the degree of imprecision (contradiction and imprecision) and the second class is the conflict degree between sensor readings (conflict and interval distance). After discounting the original sensor reading, all involved BBAs are combined with PCR6 fusion rule for decision-making support.

In order to prove the efficiency of our proposed approach, several simulations have been provided to illustrate the applicability and efficiency of our method. Also, meaningful comparisons were made with other classical approaches. Our results and the analysis of the performance obtained show that our approach is effective in dealing with conflict issues because of multi-criteria strategy adopted. Consequently, this approach can help to reduce counter-intuitive behaviors and biased readings.

In future work, we will focus on the impact of each criterion in the combination results and more criteria will be taken into account. Also, more investigations will be done to explore the difference of performances between global fusion with PCR6 in this paper and sequential fusion with PCR6. We plan to analyze the performance of this new approach for real-world sceneries using real data sets.

## REFERENCES

- [1] Z. Zhang, W. Zhang, H.-C. Chao, C.-F. Lai, *Toward belief function-based cooperative sensing for interference resistant industrial wireless sensor networks*, IEEE Trans. on Ind. Informat., Vol. 12(6), pp. 2115–2126, Dec. 2016.
- [2] D. Alshamaa, F. Mourad-Chehade, P. Honeine, *Tracking of mobile sensors using belief functions in indoor wireless networks*, IEEE Sensors J., Vol. 18(1), pp. 310–319, Jan. 2017.
- [3] E. Daniel, *Optimum wavelet-based homomorphic medical image fusion using hybrid genetic-grey wolf optimization algorithm*, IEEE Sensors J., Vol. 18(16), pp. 6804–6811, Aug. 2018.
- [4] Q. Jiang, X. Jin, J. Hou, S.-J. Lee, S. Yao, *Multi-sensor image fusion based on interval type-2 fuzzy sets and regional features in non-sampled shearlet transform domain*, IEEE Sensors J., Vol. 18(6), pp. 2494–2505, March 2018.
- [5] A. Dallil, M. Oussalah, A. Ouldali, *Sensor fusion and target tracking using evidential data association*, IEEE Sensors J., Vol. 13(1), pp. 285–293, Jan. 2013.
- [6] R. Rucco et al., *Type and location of wearable sensors for monitoring falls during static and dynamic tasks in healthy elderly: A review*, Sensors, Vol. 18(5), p. 1613, May 2018.
- [7] P.S. Rossi, D. Ciunzo, T. Ekman, H. Dong, *Energy detection for MIMO decision fusion in underwater sensor networks*, IEEE Sensors J., Vol. 15(3), pp. 1630–1640, Mar. 2015.
- [8] D. Peralta, I. Triguero, S. Garc a, Y. Saeys, J.M. Benitez, F. Herrera, *Distributed incremental fingerprint identification with reduced database penetration rate using a hierarchical classification based on feature fusion and selection*, Knowledge Based Systems, Vol. 126, pp. 91–103, June 2017.
- [9] D. Peralta, I. Triguero, S. Garc a, F. Herrera, J.M. Benitez, *DPD-DF: A dual phase distributed scheme with double fingerprint fusion for fast and accurate identification in large databases*, Information Fusion, Vol. 32, pp. 40–51, Nov. 2016.
- [10] Z.-G. Liu, J. Dezert, Q. Pan, G. Mercier, *Combination of sources of evidence with different discounting factors based on a new dissimilarity measure*, Decision Support Systems, Vol. 52(1), pp. 133–141, 2011.
- [11] F.A.N. Palmieri, D. Ciunzo, *Objective priors from maximum entropy in data classification*, Information Fusion, Vol. 14(2), pp. 186–198, April 2013.
- [12] F.A.N. Palmieri, D. Ciunzo, *Data fusion with entropic priors*, in Proc. of 20th Italian Workshop on Neural Nets (WIRN10), Vietri sul Mare, Salerno, Italy, May 27–29, 2010.
- [13] D. Ciunzo, P.S. Rossi, *Decision fusion with unknown sensor detection probability*, IEEE Signal Processing Letters, Vol. 21(2), pp. 208–212, Feb. 2014.
- [14] J.-Y. Wu, C.-W. Wu, T.-Y. Wang, T.-S. Lee, *Channel-aware decision fusion with unknown local sensor detection probability*, IEEE Trans. on Signal Process., Vol. 58(3), pp. 1457–1463, March 2010.
- [15] D. Ciunzo, A.D. Maio, P.S. Rossi, *A systematic framework for composite hypothesis testing of independent Bernoulli trials*, IEEE Signal Process. Lett., Vol. 22(9), pp. 1249–1253, Sept. 2015.
- [16] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Comput. Intell., Vol. 4(3), pp. 244–264, Sept. 1988.
- [17] S. Populaire, P. Ginetet, J. Blanc, T. Denoux, *Fusion of expert knowledge with data using belief functions: A case study in wastewater treatment*, in Proc. of 5th Int. Conf. on Information Fusion, Vol. 2, pp. 1613–1618, July 2002.
- [18] E. Lefevre, O. Colot, P. Vannoorenberghe, *Reply to the comments of R. Haenni on the paper “Belief functions combination and conflict management”*, Information Fusion, Vol. 4(1), pp. 63–65, March 2003.
- [19] P. Smets, R. Kennes, *The transferable belief model*, Artif. Intell., Vol. 66(2), pp. 191–234, April 1994.
- [20] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, NJ, USA, 1976.
- [21] J. Klein, O. Colot, *Automatic discounting rate computation using a dissent criterion*, in Proc. of Workshop on Theory of Belief Functions, pp. 1–6, March 2010.
- [22] J. Schubert, *Conflict management in Dempster-Shafer theory using the degree of falsity*, Int. J. Approx. Reasoning, Vol. 52(3), pp. 449–460, March 2011.
- [23] Y. Yang, D. Han, C. Han, *Discounted combination of unreliable evidence using degree of disagreement*, Int. J. Approx. Reasoning, Vol. 54(8), pp. 1197–1216, Oct. 2013.
- [24] L.G. de Oliveira Silva, A.T. de Almeida-Filho, *A multicriteria approach for analysis of conflicts in evidence theory*, Information Sciences, Vols. 346–347, pp. 275–285, June 2016.
- [25] A. Martin, A.-L. Jousselme, C. Osswald, *Conflict measure for the discounting operation on belief functions*, in Proc. of 11th Int. Conf. on Information Fusion (Fusion 2008), pp. 1–8, June–July, 2008.
- [26] A. Frikha, *On the use of a multi-criteria approach for reliability estimation in belief function theory*, Information Fusion, Vol. 18, pp. 20–32, July 2014.
- [27] A. Frikha, H. Moalla, *Analytic hierarchy process for multi-sensor data fusion based on belief function theory*, Eur. J. Oper. Res., Vol. 241(1), pp. 133–147, 2015.
- [28] A. Sarabi-Jamab, B.N. Araabi, *How to decide when the sources of evidence are unreliable: A multi-criteria discounting approach in the Dempster-Shafer theory*, Inf. Sci., Vols. 448–449, pp. 233–248, June 2018.
- [29] J. Dezert, D. Han, H. Yin, *A new belief function based approach for multi-criteria decision-making support*, in Proc. of 19th Int. Conf. on Information Fusion (Fusion 2016), pp. 782–789, July 2016, .
- [30] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, in Advances and Applications of DSMT for Information Fusion, Vol. 2, pp. 69–88, F. Smarandache, and J. Dezert, Eds. Rehoboth, NM, USA: American Research Press Rehoboth, 2006.

- [31] F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion*, Collected Works, Vol. 4. Rehoboth, NM, USA: American Research Press, 2015. <https://www.onera.fr/staff/jean-dezert/references>
- [32] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. 16th of Int. Conf. on Information Fusion, July 2013, pp. 1119–1126.
- [33] J. Vejnarová, G.J. Klir, *Measure of strife in Dempster-Shafer theory*, Int. J. General Syst., Vol. 22(1), pp. 25–42, Feb. 1993.
- [34] D. Dubois, H. Prade, *A note on measures of specificity for fuzzy sets*, Int. J. Gen. Syst., Vol. 10(4), pp. 279–283, 1985.
- [35] D. Han, J. Dezert, Y. Yang, *New distance measures of evidence based on belief intervals*, in Proc. of Belief Conf., Springer, Oxford, U.K., Sept. 2014.
- [36] D. Han, J. Dezert, Y. Yang, *Belief interval-based distance measures in the theory of belief functions*, IEEE Trans. on Syst., Man, Cybern. Syst., Vol. 48(6), pp. 833–850, June 2018.
- [37] W. Jiang, A. Zhang, Q. Yang, *A new method to determine evidence discounting coefficient*, in Proc. of Int. Conf. Intell. Comput., pp. 882–887, Sept. 2008.

# A New Image Registration Algorithm Based on Evidential Reasoning

Zhe Zhang<sup>a</sup>, Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>

<sup>a</sup>MOE KLINNS Lab, Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>c</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China.

Emails: zhangzsmg@gmail.com, deqhan@xjtu.edu.cn, jean.dezert@onera.fr, jiafeiyi@mail.xjtu.edu.cn

Originally published as: Z. Zhang, D. Han, J. Dezert, Y. Yang, *A New Image Registration Algorithm Based on Evidential Reasoning*, Sensors 2019, 19, 1091, 2019, and reprinted with permission.

**Abstract**—Image registration is a crucial and fundamental problem in image processing and computer vision, which aims to align two or more images of the same scene acquired from different views or at different times. In image registration, since different keypoints (e.g., corners) or similarity measures might lead to different registration results, the selection of keypoint detection algorithms or similarity measures would bring uncertainty. These different keypoint detectors or similarity measures have their own pros and cons and can be jointly used to expect a better registration result. In this paper, the uncertainty caused by the selection of keypoint detector or similarity measure is addressed using the theory of belief functions, and image information at different levels are jointly used to achieve a more accurate image registration. Experimental results and related analyses show that our proposed algorithm can achieve more precise image registration results compared to several prevailing algorithms.

**Keywords:** image registration, evidential reasoning, belief functions, uncertainty.

## I. INTRODUCTION

Image registration is a fundamental problem encountered in image processing, e.g., image fusion [1] and image change detection [2]. It refers to the alignment of two or more images of the same scene taken at different time, from different sensors, or from different viewpoints. Image registration plays an increasingly important role in applications of surveillance [3], remote-sensing [4] and medical imaging [5].

For a collection of images to be registered, one is chosen as the reference image and the others are selected as sensed images. Image registration align each sensed image to the reference image by finding the correspondence between all pixels in the image pair and estimating the spatial transformation from the sensed image to the reference image. In this paper, we just consider the image registration between two images, i.e., there is only one sensed image together with a given reference image.

Current image registration techniques that based on image domain can be generally divided into two categories [6]: the sparse methods and dense methods. There are also some methods based on transform domain, like Fourier-Mellin transformation method [7]. The transform domain based methods are often used for image registration with similarity transformation

model. In this paper, we focus on the image domain based methods.

The sparse methods [8] extracts and matches salient features from the reference image and sensed image and then estimates the spatial transformation between the two images based on these matched features. Line features (e.g., edges) and point features (corners, line intersections and gravities of regions) all can be used for image registration. Corner features are the mostly used features and can be manually selected or automatically detected by Harris [9], FAST (Features from Accelerated Segment Test) [10], SIFT (Scale-Invariant Feature Transform) [11], SURF (Speeded-Up Robust Features) [12], DAISY [13], ORB (Oriented FAST and Rotated BRIEF) [14], KAZE [15], etc.

In contrast to the sparse methods, the dense methods [16] do not detect features from the image pair but search the optimal spatial transformation directly that can best match all the pixels in the image pair. Similarity (resp. dissimilarity) measures are defined to quantify the independency (resp. dependency) between the pair of images. Various similarity and dissimilarity measures have been proposed [17] such as RMSE (Root-Mean-Squared Error), PSNR (Peak Signal to Noise Ratio), Spearman's Rho [18], NCC (Normalized Cross-correlation Coefficient) and MI (Mutual Information). It should be noted that dense methods based on RMSE or PSNR cannot handle the cases with illumination variation since these two similarity/dissimilarity measures are very sensitive to illumination changes.

Both the sparse methods and dense methods involve uncertainty problems. For the sparse methods, keypoints obtained from different keypoint detectors describe different corner features of the image. Therefore, image registrations based on different keypoint detectors would obtain different spatial transformations. For the dense methods, different similarity (dissimilarity) measures quantify the difference between the pair of images from different aspects so that image registrations based on different similarity (dissimilarity) measures would obtain different spatial transformations. These different spatial transformations obtained have their own pros and cons, and the selection of the spatial transformation (the selection of

the feature detector or similarity measure indeed) would bring uncertainty.

To deal with the uncertainty caused by the particular selection of feature detector or similarity (dissimilarity) measure, one feasible way is to combine these registration transformations obtained from different feature detection methods or similarity measures to obtain a better registration result. The belief functions introduced in Dempster–Shafer Theory (DST) [19] of evidence offer a powerful theoretical tool for uncertainty modeling and reasoning; therefore, we propose a fusion based image registration method using belief functions. In this paper, the spatial transformations obtained from different feature detection algorithms or similarity measures compose the frame of discernment (FOD) and their uncertainties are modeled using belief functions. In uncertainty modeling, image information at different levels, i.e., image’s intensities, edges and phase angles, are jointly used to evaluate the beliefs about image transformations. Then, these uncertainties are further handled through the evidence combination of the above multiple information. The final registration result is obtained according to the combined evidence.

This paper is an extension of our previous work in [20] where the basic idea is briefly presented. The main added values with respect to [20] are as follows. First, the transformation model between the reference image and sensed image is more comprehensive. We use similarity transformation model in [20] but use projective transformation model in this paper, which is more general since all similarity transformations are examples of projective transformations. Second, the keypoints used in the sparse approach in [20] are manually selected. To reduce the subjective influence to the registration result, in this paper, the keypoints are generated from detection algorithms. Accordingly, feature matching and mismatching removal are added after the keypoint detection. Third, when modeling uncertainties, one more information source, i.e., image’s phase angle information, is considered in this work. Fourth, more experiments and analyses are provided for performance evaluation and analysis.

The rest of this paper is organized as follows. The basics of image registration are introduced in Section II. The basics of evidence theory are introduced in Section III. The proposed image registration method is introduced in Section IV with emphasis of uncertainty modeling and handling. Evaluation method is introduced in Section V. Experiment results of the proposed method and other registration methods are presented and compared in Section VI. Concluding remarks are given in Section VII.

## II. BASICS OF IMAGE REGISTRATION

For two (or more) images of the same scene taken at different time, from different sensors, or from different viewpoints, one is chosen as the reference image ( $R$ ) and the other one is chosen as the sensed image ( $S$ ). In this paper, we focus on the projective transformation model between the reference image and sensed image, which is a commonly used model in image registration [16]. Denote pixel coordinates in the

reference image  $R$  as  $(v, w)$  and their mapping counterparts in the sensed image  $S$  as  $(g, h)$ . The projective transformation from  $R$  to  $S$  can be expressed based on the homogeneous coordinates (Homogeneous coordinates can easily express the translation transformation as matrix multiplications while Cartesian coordinates cannot) as

$$[g \ h \ 1] = [v \ w \ 1]T = [v \ w \ 1] \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \quad (1)$$

The similarity transformation and affine transformation are important specializations of the projective transformation, as illustrated in Table I.

Table I  
PROJECTIVE TRANSFORMATION AND ITS TWO SPECIALIZATIONS.

Similarity	Affine	Projective
$\begin{bmatrix} s \cos \theta & s \sin \theta & 0 \\ -s \sin \theta & s \cos \theta & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix}$	$\begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix}$	$\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix}$
		

The purpose of image registration is to estimate the transformation  $T$  to align the sensed image  $S$  with the reference image  $R$  by

$$[v' \ w' \ 1] = [g \ h \ 1]T^{-1}, \quad (2)$$

where  $(v', w')$  and  $(g, h)$  denote pixel coordinates in registered sensed image  $S'$  and sensed image  $S$ , respectively. Current image registration techniques can be divided into two categories [6] in general, including the sparse method and dense method. Basics of these two methods are introduced below.

### A. Sparse Image Registration and Its Uncertainty

The feature detection and feature matching are two critical steps in the sparse methods. The flow chart of the sparse approach is illustrated in Figure 1, where each functional block is detailed in the sequel.

1) *Feature Detection*: Corner features are the mostly used features in image registration due to their invariance to imaging geometry [6]. Some early keypoint detectors, like Harris and FAST, are very sensitive to image scale changes so that have poor performance when the sensed images have different scales with the reference image. The most well-known SIFT detector shows good robustness to illumination, orientation and scale changes. Most scale invariant detectors, like SIFT, SURF, ORB and BRISK, detect and describe features at different scale levels by building or approximating the Gaussian scale space of the image. In a different way, KAZE detects features in a nonlinear scale space built using

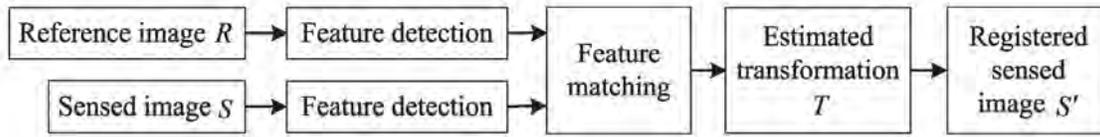


Figure 1. Flow chart of sparse approach.

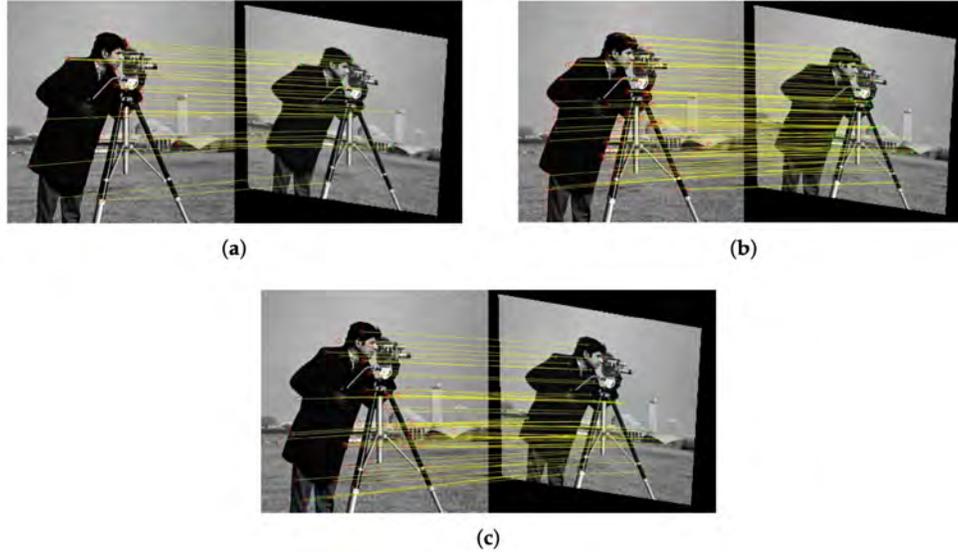


Figure 2. Different keypoint pairs detected by different keypoint detectors. (a) BRISK, (b) KAZE, (c) SURF.

efficient additive operator splitting techniques and variable conductance diffusion.

2) *Feature Matching*: To align the sensed image and the reference image, the detected keypoints in the two images are matched first by comparing their local feature characterized by descriptors. Generally, if the two keypoints' descriptors are similar, the two keypoints are likely to be a matched pair. Given a keypoint  $t$  in the reference image, there might be a set of candidates in the sensed image having similar descriptor with  $t$ . Among these candidates,  $t$ 's real counterpart should have the closest distance with  $t$ , and at the same time its distance should be much closer than other candidates' distances.

The accuracy of the keypoints' matching affects the accuracy of the transformation's estimation. The mismatched keypoint pairs should be further removed before estimating the transformation. RANSAC (RANdom SAMple Consensus) [21] and MSAC (M-estimator SAMple and Consensus) [22] are often used to deal with this problem. A recent RANRESAC (RANdom RESAMple Consensus) [23] algorithm has been proposed to remove mismatched keypoint pairs for noisy image registration. Besides the accuracy of the keypoints' matching, the distribution of matched pairs over the image space is another key factor to obtain a high-quality estimation of transformation.

3) *Transformation Estimation*: With all the matched keypoint pairs, the transformation matrix  $T$  can be estimated using Eq. (1). Since  $T$  has eight degrees of freedom, four point correspondences (with no three collinear) are needed to obtain the unique solution of  $T$  according to Cramer's rule.

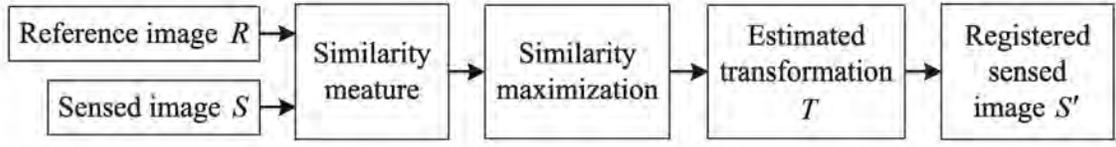
Normally, the amount of the matched keypoint pairs is more than four and  $T$  can be estimated using the least squares (LS) fitting technique [6] by searching the minimum sum of the Euclidean distances between all the matched keypoints:

$$\hat{T} = \arg \min \sum_i d(\text{cor}_i^R, \text{cor}_i^{S'}), \quad (3)$$

where  $\text{cor}_i^R = (v_i, w_i)$  represents the coordinate of the  $i$ -th matched keypoint in the reference image and  $\text{cor}_i^{S'} = (v'_i, w'_i)$  represents the coordinate of the  $i$ -th matched keypoint in the registered sensed image transformed from the sensed image using Eq. (2).

4) *Uncertainty Encountered in Sparse Approach*: Since different keypoint detection algorithms detect different kinds of corner features, the detected keypoints are usually different, as shown in Figure 2.

Image registrations based on different matched keypoint pairs would in general yield different spatial transformations to align two images. Different transformations obtained have their own pros and cons. Therefore, the selection of keypoint detection algorithms would bring uncertainty problem to the registration results.



### B. Dense Image Registration and Its Uncertainty

The dense image registration estimates the optimal transformation  $T$  by searching the largest similarity (or the smallest dissimilarity) between the reference image  $R$  and the registered sensed image  $S' = T(S)$ :

$$\hat{T} = \arg \min Sim(R, T(S)), \quad (4)$$

where  $Sim$  is a chosen similarity measure.

The flow chart of the dense approach is illustrated in Figure 3, where each functional block is detailed in the sequel.

1) *Similarity Measure*: Various similarity (or dissimilarity) measures have been proposed. Here we briefly introduce the commonly used MI, NCC and PSNR measures.

#### 1) MI (Mutual Information) measure:

The MI measure between images  $A$  and  $B$  is defined by

$$MI(A, B) = \sum_{a=0}^{255} \sum_{b=0}^{255} p_{AB}(a, b) \log \frac{p_{AB}(a, b)}{p_A(a)p_B(b)}, \quad (5)$$

where  $p_{AB}$  is the joint probability distribution function (PDF) of images  $A$  and  $B$ , and  $p_A$  and  $p_B$  are the marginal PDFs of  $A$  and  $B$ , respectively.  $MI(A, B)$  is larger when  $A$  and  $B$  are more similar.

#### 2) NCC (Normalized Cross-Correlation) measure:

For given images  $A$  and  $B$  with size of  $M \times N$ , NCC measure between them is

$$NCC(A, B) = \sum_{x=1}^M \sum_{y=1}^N \frac{(A(x, y) - \mu_A)(B(x, y) - \mu_B)}{\sigma_A \sigma_B}, \quad (6)$$

where  $A(x, y)$  and  $B(x, y)$  are the pixels' intensities in images  $A$  and  $B$  at  $(x, y)$ , respectively;  $\mu_A$  and  $\mu_B$  are the mean intensities of  $A$  and  $B$ , respectively;  $\sigma_A$  and  $\sigma_B$  are the standard deviation intensities of  $A$  and  $B$ , respectively.  $NCC(A, B)$  is larger when  $A$  and  $B$  are more similar.

#### 3) PSNR (Peak Signal-to-Noise Ratio) measure:

PSNR measure between images  $A$  and  $B$  is defined by

$$PSNR(A, B) = 10 \times \log_{10} \left( \frac{255^2}{MSE(A, B)} \right), \quad (7)$$

where  $MSE(A, B) = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [A(x, y) - B(x, y)]^2$ . PSNR( $A, B$ ) is larger when  $A$  and  $B$  are more similar. Since PSNR measure is very sensitive to illumination changes, it cannot be used for image registration when there are illumination variations between image pairs.

2) *Transformation Estimation*: The estimation for transformation  $T$ , i.e., Eq. (4), is always a non-convex problem and is not so easy to obtain the global maximum [24]. Therefore, advanced optimization methods [25], or intelligent optimization approaches (like genetic, or particle swarm algorithms, etc.) are often used to estimate the optimal transformation  $T$ .

3) *Uncertainty Encountered in Dense Approach*: Since different similarity (dissimilarity) measures compare two images from different aspects, their calculated similarities (dissimilarities) between the reference image and registered sensed image are different. Image registration based on different measures would obtain different spatial transformations to align two images and they have their own pros and cons. Therefore, the selection of similarity (dissimilarity) measure would bring uncertainty problem to the registration results.

To deal with the uncertainty caused by the selection of feature detection algorithms or similarity measures, one feasible way is to combine the registration transformations ( $T_1, T_2, \dots, T_Q$ ) obtained from different feature detection algorithms (or different similarity measures) to expect a better registration result. We propose an evidential reasoning [19] based image registration algorithm to generate a combined transformation from  $T_1, T_2, \dots, T_Q$  thanks to the ability of belief functions for uncertainty modeling and reasoning. Basics of the theory of belief functions are recalled first below.

### III. BASICS OF DEMPSTER-SHAFFER EVIDENCE THEORY

Dempster-Shafer evidence theory (DST) [19] is a theoretical framework for uncertainty modeling and reasoning. In DST, elements in the frame of discernment (FOD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_Q\}$  are mutually exclusive and exhaustive. The power set of  $\Theta$ , i.e.,  $2^\Theta$ , is the set of all subsets of  $\Theta$ . For example, if  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ , then

$$2^\Theta = \{\{\emptyset\}, \{\theta_1\}, \{\theta_2\}, \{\theta_3\}, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \{\theta_2, \theta_3\}, \{\theta_1, \theta_2, \theta_3\}\}.$$

The basic belief assignment (BBA, also called mass function) is defined by a function  $m: 2^\Theta \mapsto [0, 1]$ , satisfying

$$m(\emptyset) = 0, \text{ and } \sum_{A \subseteq \Theta} m(A) = 1, \quad (8)$$

where  $m(A)$  depicts the evidence support to the proposition  $A$ .  $A$  is called a focal element when  $m(A) > 0$ . If there is only one element in  $A$ , like  $\{\theta_1\}$  and  $\{\theta_2\}$ ,  $A$  is called the singleton element; if there are more than one element in  $A$ , e.g.,  $\{\theta_1, \theta_2\}$  and  $\{\theta_1, \theta_2, \theta_3\}$ ,  $A$  is called the compound

element. The belief assigned to a compound element represents the degree of ambiguity for the multiple elements.

The plausibility function ( $Pl$ ) and belief function ( $Bel$ ) are defined as follows:

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B), \quad (9)$$

$$Bel(A) = \sum_{B \subseteq A} m(B). \quad (10)$$

Dempster's combination rule [19] for combining two distinct pieces of evidence is defined as

$$(m_1 \oplus m_2)(A) = \begin{cases} 0, & \text{if } A = \emptyset, \\ \frac{1}{1-K} \sum_{B \cap C = A} m_1(B)m_2(C) & \text{if } A \neq \emptyset. \end{cases} \quad (11)$$

Here,  $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$  denotes the total conflict or contradictory mass assignments.

An alternative fusion rule PCR6 [26] for the combination of two sources is defined as

$$m_{12}^{PCR6}(A) = m_{12}^{Conj}(A) + \sum_{A \cap Y = \emptyset} \left[ \frac{m_1(A)^2 m_2(Y)}{m_1(A) + m_2(Y)} + \frac{m_2(A)^2 m_1(Y)}{m_2(A) + m_1(Y)} \right], \quad (12)$$

where  $m_{12}^{Conj}(A)$  is the conjunctive rule defined as

$$m_{12}^{Conj}(A) = \sum_{B \cap C = A} m_1(B)m_2(C). \quad (13)$$

General PCR6 formula for the combination of more than two sources is given in [26].

For a probabilistic decision-making, Smets defined the pignistic probability transformation [27] to obtain the probability measure  $BetP$  from a BBA

$$BetP(\theta_i) = \sum_{A \subseteq \Theta | \theta_i \in A} \frac{m(A)}{|A|}. \quad (14)$$

where  $|A|$  is the cardinality of  $A$ . The decision can be made by choosing the element in FOD whose  $BetP$  value is the highest one and higher than a preset threshold. Other types of probability transformation methods can be found in [26,28].

#### IV. IMAGE REGISTRATION BASED ON EVIDENTIAL REASONING

To deal with the uncertainty caused by the choice of keypoint detectors in the sparse approach or the choice of similarity measure in the dense approach, we propose an image registration method based on evidential reasoning. Suppose that the spatial transformation between the reference image and sensed image is projective. Our purpose is to estimate the transformation matrix to align two images. Unlike the prevailing methods estimating the transformation matrix from single method of keypoint detection or similarity (dissimilarity) measure, we estimate the transformation matrix by jointly

utilizing different keypoint detection methods or similarity measures.

To use belief functions for image registration, one should define the frame of discernment (FOD) first. The FOD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_Q\}$ , where  $Q$  is the amount of transformations obtained from different single feature detection algorithms or different single similarity measures. We first model the beliefs for every proposition  $A \subseteq \Theta$  using BBAs.  $A$  can be single transformation in FOD or a set of transformations in FOD. One BBA depicts the support to each proposition  $A$  from one evidence source. The BBA allocations from different evidence sources describes the uncertainty of the transformations in FOD. Next, the BBAs are combined to generate the combined BBA  $m_c$  depicting the fused support to each proposition  $A$ . Then, the combined transformation  $T_c$  is generated from the combined BBA  $m_c$ . Finally, the registered sensed image  $S'_c$  is transformed from the sensed image using Eq. (2). During this process, the resampling [29] is needed to determine the intensity of each pixel in  $S'_c$ . Figure 4 illustrates the flow chart of this new proposed method. It should be noted that the classical interpretation of BFT assumes that the final estimation should be in the FOD. In this work, we relax this assumption and the final transformation is a combination result of those in the FOD.

#### A. Uncertainty Modeling

If the similarity between the reference image  $R$  and registered sensed image  $S'_i$  is large, the corresponding transformation  $T_i$  is quite accurate and should be allocated large support ( $S'_i$  is transformed from sensed image  $S$  by  $T^{-1}$ ). Here we use NCC (other similarity or dissimilarity measures, e.g., MI, are also appropriate to quantify the similarity here) to measure the similarity between  $R$  and  $S'_i$ :

$$NCC_i = \sum_{x=1}^M \sum_{y=1}^N \frac{(R(x,y) - \mu_R)(S'_i(x,y) - \mu_{S'_i})}{\sigma_R \sigma_{S'_i}}, \quad (15)$$

where  $\mu_R$  and  $\mu_{S'_i}$  are the mean intensities of  $R$  and  $S'_i$ , respectively;  $\sigma_R$  and  $\sigma_{S'_i}$  are the standard deviation intensities of  $R$  and  $S'_i$ , respectively.

Since multi-source information can help to reduce the uncertainty through evidence combination, we use different levels of image information to quantify the similarity between  $R$  and  $S'_i$ . The similarity can be calculated from the gray images, edge feature images or reconstructed images using phase angle as shown in Figure 5. Their corresponding  $NCC_i$  are denoted as  $NCC_i(G)$ ,  $NCC_i(E)$  and  $NCC_i(P)$ , respectively. The edge detection method used in Figure 5-b is the Canny detector [30]. More details of the image reconstruction from phase angle information can be found in [29].

The value range of  $NCC_i(\cdot)$  is  $[-1, 1]$ . According to our experiments, most values of  $NCC_i(\cdot)$  are larger than 0. Before allocating BBAs, we first enlarge the differences of  $NCC_i(\cdot)$  within  $[0, 1]$  using function  $y = e^{x-1}$ , as illustrated in Figure 6. Each level of image information (gray images ( $G$ ), edge

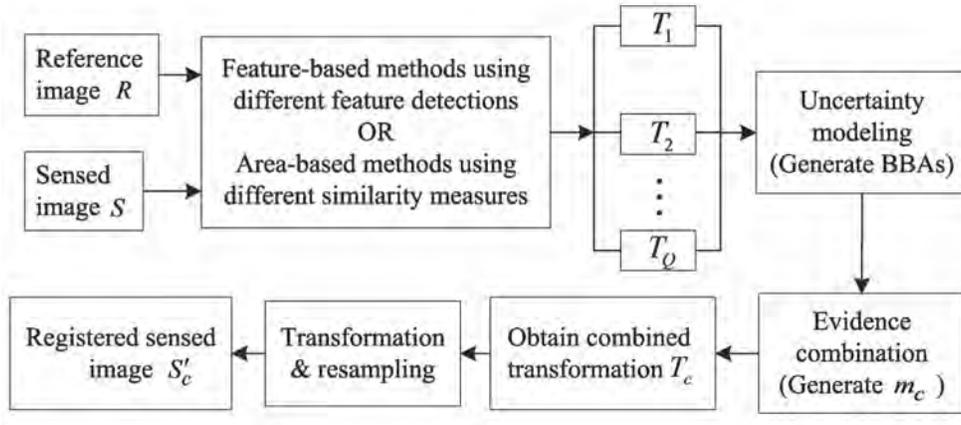


Figure 4. Flow chart of the proposed image registration.

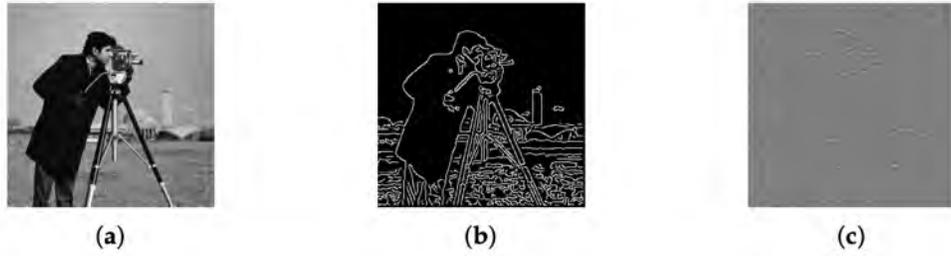
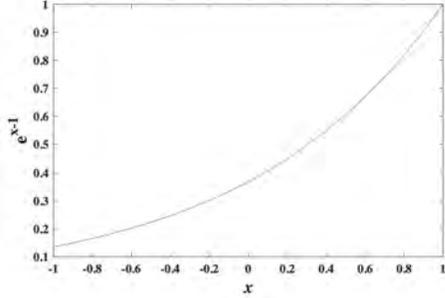


Figure 5. Image information at different levels. (a) Gray image. (b) Edge feature image. (c) Reconstructed image using phase angle.


 Figure 6. The curve of function  $e^{x-1}$ .

feature images ( $E$ ) and reconstructed images using phase angle ( $P$ ) can be viewed as one evidence source and their corresponding  $y = e^{\text{NCC}_i(\cdot)-1}$  can be used to assign beliefs for transformation  $T_i$ :

$$\begin{cases} m_G(T_i) = \frac{e^{\text{NCC}_i(G)-1}}{\sum_{j=1}^Q e^{\text{NCC}_j(G)-1}} \\ m_E(T_i) = \frac{e^{\text{NCC}_i(E)-1}}{\sum_{j=1}^Q e^{\text{NCC}_j(E)-1}} \\ m_P(T_i) = \frac{e^{\text{NCC}_i(P)-1}}{\sum_{j=1}^Q e^{\text{NCC}_j(P)-1}} \end{cases} \quad (16)$$

### B. Fusion-Based Registration

After obtaining BBAs  $m_G$ ,  $m_E$  and  $m_P$ , we generate the combined BBA  $m_c$  using a combination rule denoted

symbolically with  $\oplus$ :

$$m_c(\cdot) = [m_G \oplus m_E \oplus m_P](\cdot). \quad (17)$$

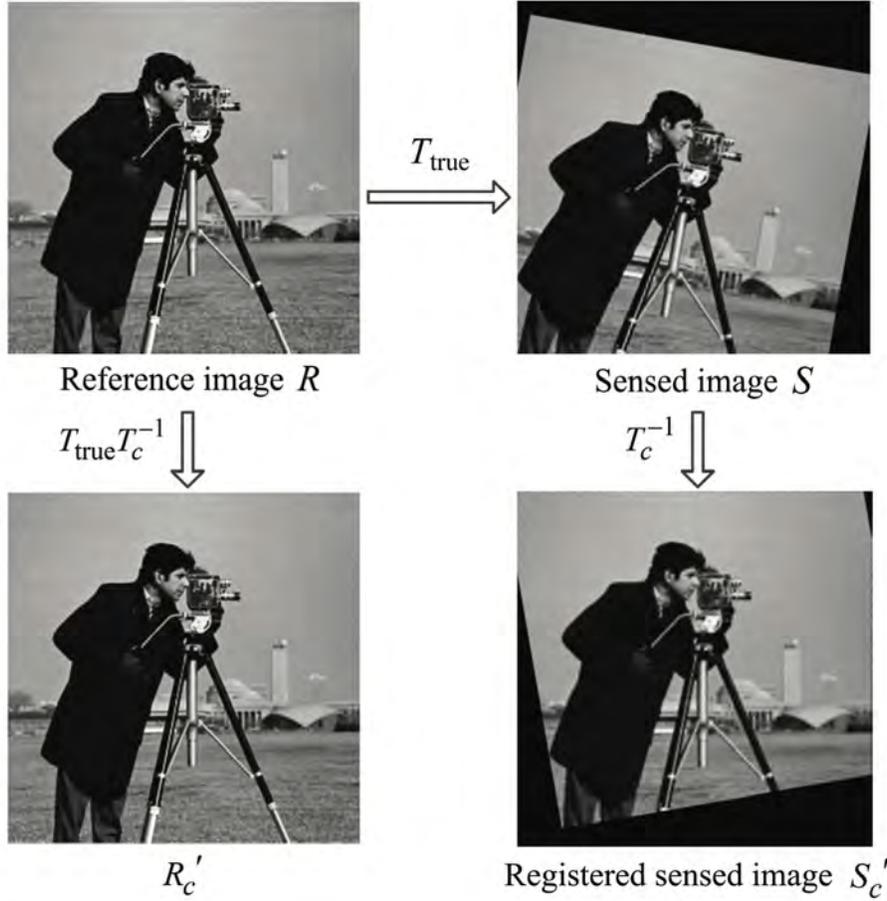
$m_c(T_i)$  describes the combined evidence support to  $T_i$  (a  $3 \times 3$  matrix with 6 unknown parameters). The combined transformation  $T_c$  is computed by

$$T_c^{-1} = \sum_{i=1}^Q m_c(T_i) T_i^{-1}. \quad (18)$$

Finally, the registered sensed image  $S'_c$  can be obtained using Eq. (2) following the resampling.

### V. EVALUATION OF IMAGE REGISTRATION

Since the purpose of image registration is to align the reference image  $R$  and sensed image  $S$  to a single coordinate frame, one popular evaluation method for the registration result is to quantify the difference (usually quantified by Root-Mean-Squared Error (RMSE)) between  $R$  and the registered sensed image  $S'_c$  [31,32]. However, since  $S'_c$  is transformed from the sensed image  $S$ , which may have less information than  $R$  ( $S$  may be part of  $R$  or have lower resolution than  $R$  since  $R$  and  $S$  can be taken from different views or taken by different cameras), the difference between  $R$  and  $S'_c$  could be large even when the estimated transformation  $T_c$  equals to the true transformation  $T_{\text{true}}$  from the reference image  $R$  to the sensed image  $S$ , as shown in Figure 7. Therefore, this kind of evaluation method is not accurate enough.


 Figure 7. Relationship among  $R$ ,  $S$ ,  $R'_c$ , and  $S'_c$ .

Another popular evaluation method is to quantify the difference between the reference image  $R$  and image  $R'_c$ , which is transformed from  $R$  by the transformation matrix  $T_{\text{true}} T_c^{-1}$  [16,33], as shown in Figure 7. The mapping relationship between pixel at  $(v, w)$  in image  $R$  and pixel at  $(v', w')$  in image  $R'_c$  satisfies

$$[v' \ w' \ 1] = [v \ w \ 1] T_{\text{true}} T_c^{-1}, \quad (19)$$

when the registration is absolutely accurate,  $T_c = T_{\text{true}}$  and  $R'_c = R$ .

In this paper, we evaluate the registration performance by quantifying the difference between the reference image  $R$  and image  $R'_c$  using AAID (average absolute intensity difference) [16]:

$$\text{AAID}(R, R'_c) = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N |R(x, y) - R'_c(x, y)|. \quad (20)$$

AAID( $R, R'_c$ ) is smaller when the registration result is better.

## VI. EXPERIMENTS

To verify the performance of our new proposed image registration method, we provide experiments on noise-free images and noisy images, respectively. Image registration

under the noisy condition is difficult since the noise pixels bring difficulties for keypoints' detection and matching and reduce the accuracy for similarity measure. For the sparse method, experiment results based on BRISK [34], KAZE [15] and SURF [12] feature detection algorithms are provided for comparison. For the dense method, experiment results based on MI, PSNR and NCC similarity measures are provided for comparison. For the noisy image registration, the experiment result of RANRESAC (a recently proposed method for noisy image registration) [23] is also provided for comparison.

### A. Sparse Image Registration Results

We first do experiments on actual data to illustrate the effectiveness of the proposed method. The reference image and sensed image are taken from different cameras with different views, as shown in Figure 8. BRISK, KAZE and SURF feature detections are used for generating transformations  $T_1$ ,  $T_2$  and  $T_3$ , respectively. When deriving combined BBAs in Eq. (17), an alternative fusion rule PCR6, which is more robust than Dempster's rule [26], is also used for comparison.

The registered results of the proposed method are illustrated in Figure 9. From Figure 9, the proposed method can successfully align the sensed image to the reference image illustrating that the proposed method is effective for actual data.

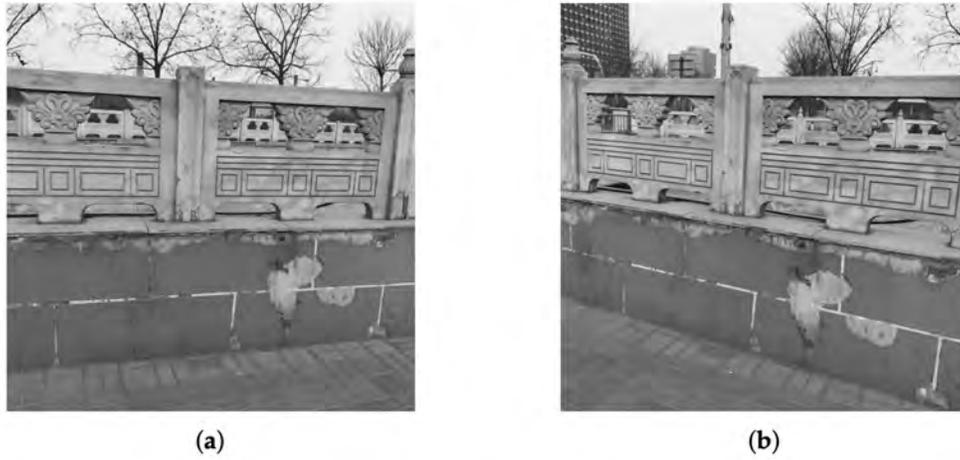


Figure 8. Fence image pair. (a) Reference image. (b) Sensed image.

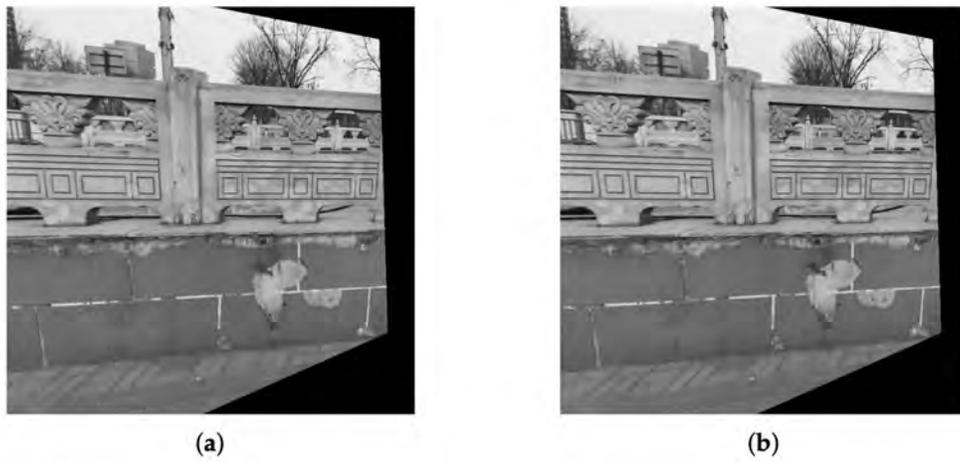


Figure 9. Registered results of the proposed methods for Fence image. (a) Dempster's rule. (b) PCR6.

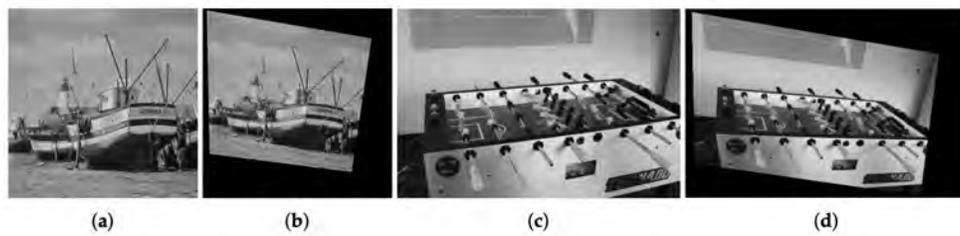


Figure 10. Boats image pair and Foosball image pair. (a) Boats reference image. (b) Boats sensed image. (c) Foosball reference image. (d) Foosball sensed image.

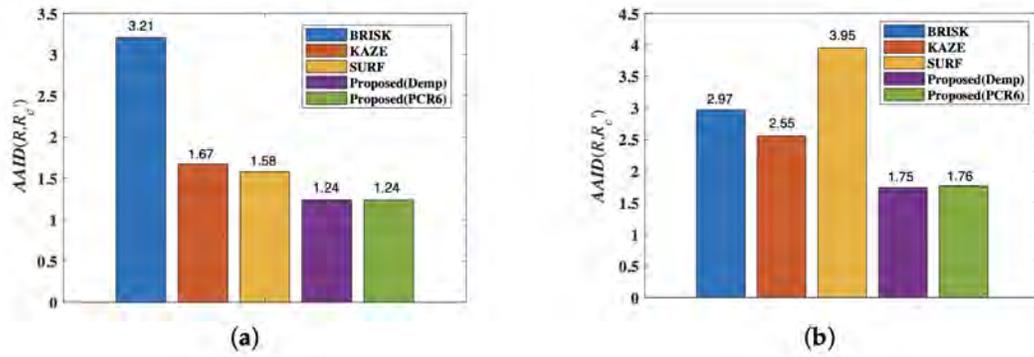


Figure 11. AAID evaluations of registration results for Boats image pair and Foosball image pair. (a) Boats. (b) Foosball.

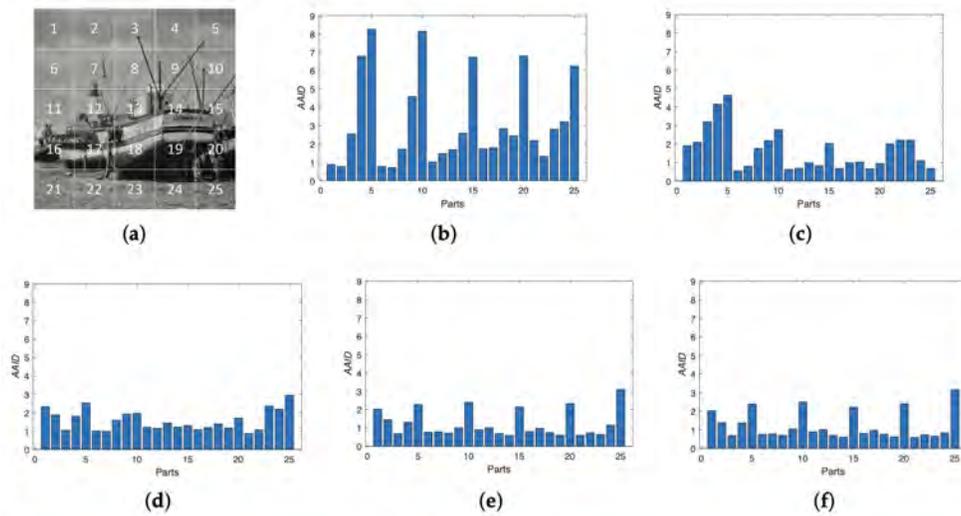


Figure 12. Spatial partition of the AAID evaluation for Boats image. (a) Partition method. (b) BRISK. (c) KAZE. (d) SURF. (e) Proposed (Demp). (f) Proposed (PCR6).

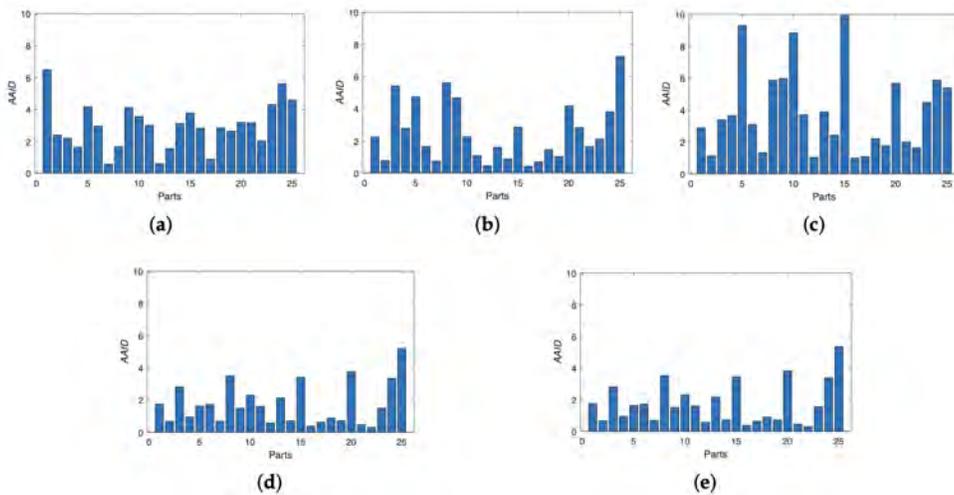


Figure 13. Spatial partition of the AAID evaluation for Foosball image. (a) BRISK. (b) KAZE. (c) SURF. (d) Proposed (Demp). (e) Proposed (PCR6).

To quantify the accuracy of the registration results, the actual transformation between the reference image and sensed image is needed and we do experiments on simulated images. We first do experiments on Boats image (The reference image can be found at <https://imagej.nih.gov/ij/images/boats.gif>) and Foosball image (sample image from the MATLAB), as shown in Figure 10.

The AAID evaluations of these registration results for Boats image and Foosball image are compared in Figure 11, where Demp represents the Dempster's combination rule. According to Figure 11, the proposed fusion-based method achieves much better registration result (smaller AAID) than algorithms based on BRISK, KAZE or SURF feature detections, respectively.

Furthermore, we also analyzed the spatial partition of the AAID evaluation for each result by evenly dividing the reference image into  $5 \times 5$  parts (as shown in Figure 12a) and calculating the AAID between the reference image and the registration result in each part. The AAID spatial partition results for Boats image and Foosball image are illustrated in Figures 12 and 13, respectively. For Boats image, the AAID of BRISK and KAZE results varies significantly for different parts while the SURF result is relatively uniform; the proposed methods have low and similar AAID in most parts while the rightmost parts (parts 5, 10, 15, 20 and 25) have significant larger AAID. For Foosball image, the AAID spatial partition of all these results are uneven.

Then, we consider the noisy image registration and do experiments on West Concord image pair (sample image from the MATLAB) with zero-mean Gaussian noise (variance is 0.01), as shown in Figure 14. The AAID evaluations for these registration results are compared in Figure 15, where the proposed fusion-based methods achieve better performance (smaller AAID) than RANRESAC and methods based on BRISK, KAZE and SURF feature detections, respectively. The spatial partition of the AAID evaluation for each result is illustrated in Figure 16, where the KAZE result is the most uneven one.

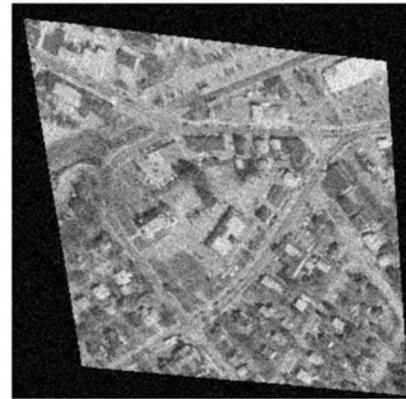
### B. Dense Image Registration Results

Since the optimization of dense registration is intractable when the solution space has high dimensions, we simplify the transformation model to rigid transformation here. The solution space for rigid model only has three dimensions: one for rotation and two for translations in horizontal and vertical directions, respectively. We first provide experiments on Concord image and Hestain image (sample images from the MATLAB) as shown in Figure 17, where the sensed image is transformed from the reference image through the rotation ( $\theta = 10^\circ$  in anti-clockwise) and translation ( $(t_v, t_h) = (-10, 5)$ ) successfully.

In the proposed dense approach, MI, PSNR and NCC similarity measures are used for generating transformations  $T_1$ ,  $T_2$  and  $T_3$ , respectively. The AAID evaluations of these registration results for the Concord image and Hestain image are compared in Figure 18, where the proposed fusion-based



(a)



(b)

Figure 14. West Concord image pair. (a) Reference image. (b) Sensed image.

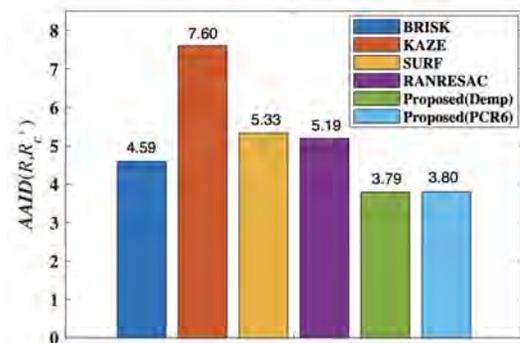


Figure 15. AAID evaluations of registration results for West Concord image pair.

methods achieve much better registration results (smaller AAID) than algorithms based on MI, PSNR or NCC similarities, respectively. The AAID spatial partition results for Concord image and Hestain image are illustrated in Figures 19 and 20, respectively. For these two images, the AAID results of the proposed methods are smaller in the downside parts compared with those in upside parts.

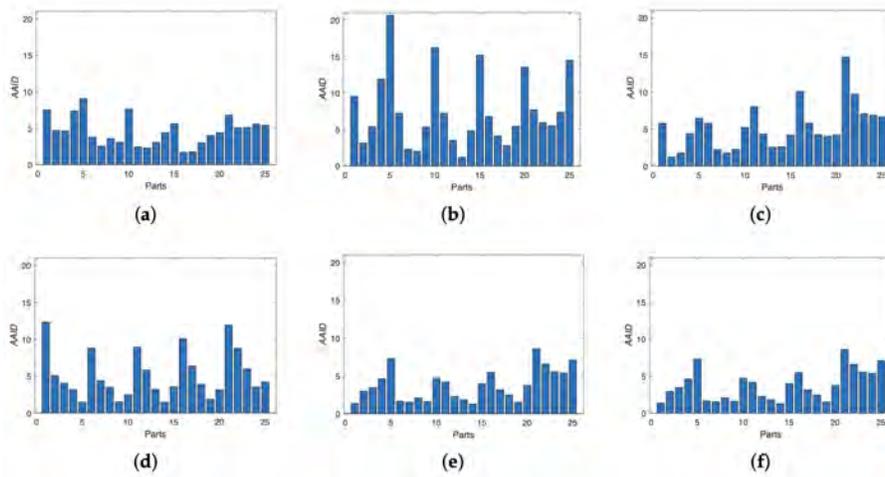


Figure 16. Spatial partition of the AAID evaluation for West Concord image. (a) BRISK. (b) KAZE. (c) SURF. (d) RANRESAC. (e) Proposed (Demp). (f) Proposed (PCR6).

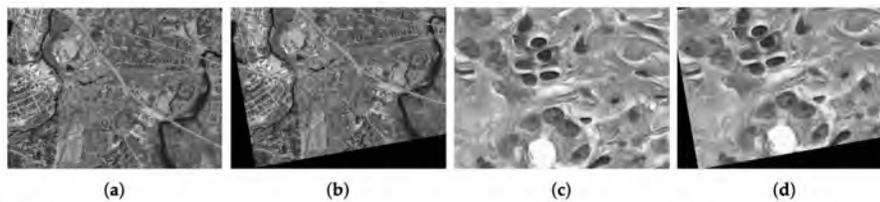


Figure 17. Concord image pair and Hestain image pair. (a) Concord reference image. (b) Concord sensed image. (c) Hestain reference image. (d) Hestain sensed image.

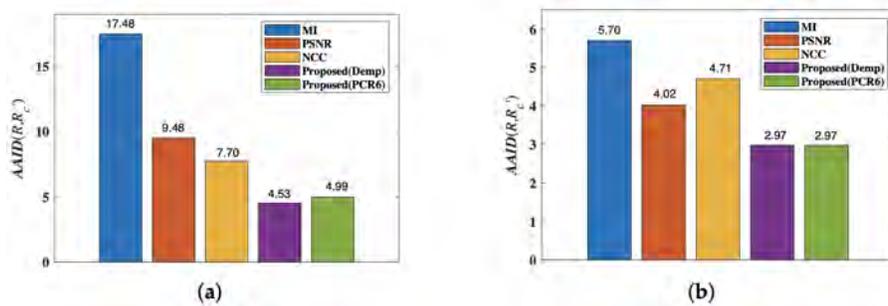


Figure 18. AAID evaluations of registration results for Concord image pair and Hestain image pair. (a) Concord. (b) Hestain.

Then, we consider the noisy image condition and implement experiments on Lifting Body image pair (sample image from the MATLAB) with zero-mean Gaussian noise (variance is 0.01), as shown in Figure 21. The sensed image is transformed from the reference image through the rotation ( $\theta = -10^\circ$ ) and translation ( $(t_v, t_h) = (-10, 5)$ ), successively.

The spatial partition of the AAID evaluation for each result is illustrated in Figure 22, and the AAID evaluations for these registration results are compared in Figure 23. From these two figures, the proposed fusion-based methods achieve better performance and the rightmost parts (parts 5, 10, 15, 20 and 25) have larger AAID than other parts.

According to all the experiments, the proposed fusion-based methods achieve better registration results than those prevailing ones (BRISK, KAZE, SURF, MI, PSNR and NCC). For noisy image registration, the proposed methods also obtain better performance than RANRESAC. This indicates that the theory of belief function can well deal with the uncertainty brought by the selection of keypoint detection algorithms or similarity measures, and the jointly use of the different keypoint detections or similarity measures is effective. Furthermore, from the above provided experiments one sees that the choice of combination rule does not affect the registration performance that much.

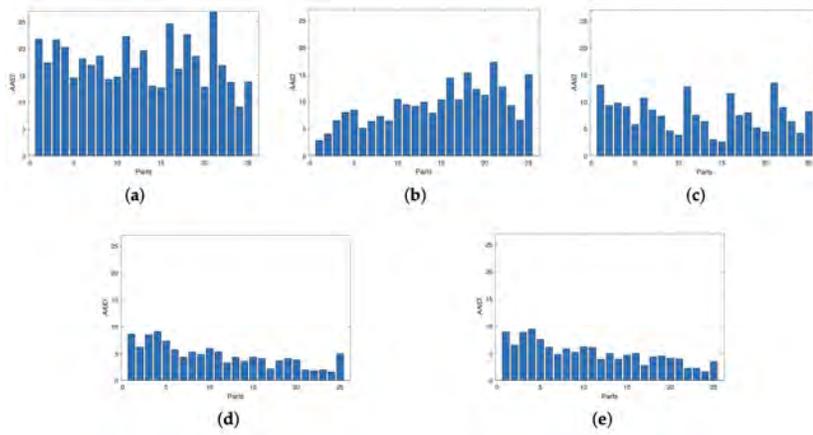


Figure 19. Spatial partition of the AAID evaluation for Concord image. (a) MI. (b) PSNR. (c) NCC. (d) Proposed (Demp). (e) Proposed (PCR6).

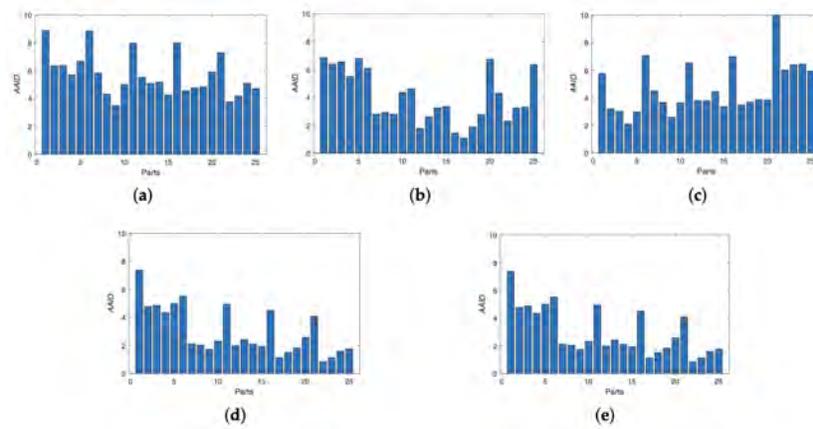


Figure 20. Spatial partition of the AAID evaluation for Hestain image. (a) MI. (b) PSNR. (c) NCC. (d) Proposed (Demp). (e) Proposed (PCR6).

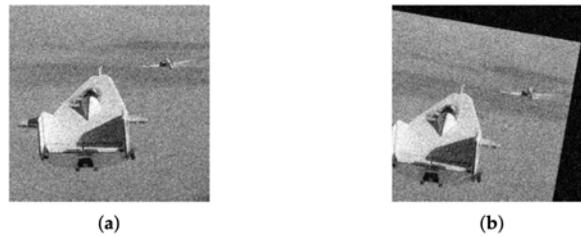


Figure 21. Lifting Body image pair. (a) Reference image. (b) Sensed image.

C. Computational Cost

The computational cost is an important criterion to evaluate an algorithm. We counted the computational costs of the above sparse algorithms and dense algorithms for Cameraman image (Figure 5-a) on a Windows 10 Enterprise system equipped with Intel Core i7-7700HQ CPU at 2.80 GHz and 16.00 GB RAM. The platform is MATLAB R2018a. The average execution time comparisons for the sparse algorithms and dense algorithms are provided in Tables II and III, respectively. Each average execution time is calculated from 100 runs of experiments.

Table II  
AVERAGE EXECUTION TIME (IN s) COMPARISON FOR SPARSE ALGORITHMS.

Method	Noise-Free Images	Noisy Images
BRISK	0.2847	0.2832
KAZE	0.1348	0.1304
SURF	0.0431	0.0437
RANRESAC	—	6.2648
Proposed (Demp)	0.3934	0.3933
Proposed (PCR6)	0.3938	0.3989

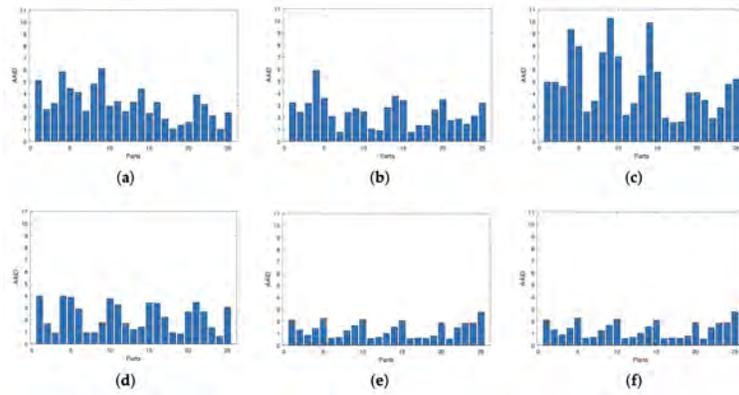


Figure 22. Spatial partition of the AAID evaluation for Lifting Body image. (a) MI. (b) PSNR. (c) NCC. (d) RANRESAC. (e) Proposed (Demp). (f) Proposed (PCR6).

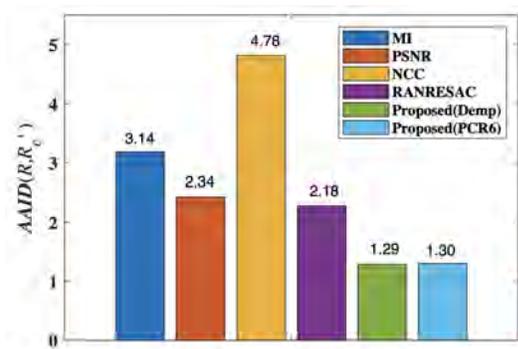


Figure 23. AAID evaluations of registration results for Lifting Body image pair.

Table III  
AVERAGE EXECUTION TIME (IN  $s$ ) COMPARISON FOR DENSE ALGORITHMS.

Method	Noise-Free Images	Noisy Images
MI	16.7622	16.4789
PSNR	12.8583	13.4214
NCC	14.7187	15.1050
Proposed (Demp)	17.0734	16.7945
Proposed (PCR6)	17.0812	16.8729

From Tables II and III, the dense algorithms need more execution time than the sparse algorithms. Furthermore, since the proposed fusion-based method combines the registration transformations generated from the three sparse methods (BRISK, KAZE and SURF) or the three dense methods (PSNR, MI and NCC) and these three methods can be executed in parallel, and the execution time of the proposed fusion-based method is longer than the most time-consuming one among the three methods.

#### D. Discussion of BBA Generation

The BBA generated in Eq. (16) is Bayesian BBA, where all its focal elements are singletons. People in the community of belief function theory may prefer to use the compound focal elements, which usually seems better than only using single-

tons in Bayesian BBAs. We have also designed experiments of generating non-Bayesian BBAs for image registration using FCOWA-ER (Fuzzy-Cautious Ordered Weighted Averaging with Evidential Reasoning) [35] method. In detail, when multiple image information (image's intensities, edges and phase angle) are simultaneously considered, image registration can be viewed as a multi-criteria decision making problem. FCOWA-ER (Fuzzy-Cautious Ordered Weighted Averaging with Evidential Reasoning) [35] is a decision making approach under multi-criteria with uncertainty and it generates non-Bayesian BBAs using  $\alpha$ -cut method (The  $\alpha$ -cut method used in FCOWA-ER boils down to the Dubois and Prade allocation [36] in this case) when modeling uncertainties. According to the experimental results, non-Bayesian BBAs obtain similar registration results with Bayesian BBAs. Since Bayesian BBAs are easier to generate than non-Bayesian BBAs, we recommend Bayesian BBAs for image registration and do not provide the non-Bayesian BBA based method in this work.

## VII. CONCLUSION

In this paper, we proposed a new image registration algorithm based on evidential reasoning. The uncertainty encountered in image registration is taken into account and modeled by belief functions. Image information at different levels are jointly used to achieve a more effective registration. Experimental results show that the proposed algorithm can improve the precision of image registration.

The generation of BBA is crucial in evidential reasoning and most methods are proposed based on applications. In this paper, we generate BBAs from three different image information, i.e., intensity, edge and phase angle. In future work, other image information, such as texture feature and gradient feature, will also be considered and jointly used in image registration. Furthermore, we will attempt to apply the proposed method to color image registration. Different color channels of the color image provide different image information and can be jointly used in image registration. We will also focus on the comparison with the state-of-the-art approaches based on convolutional neural networks (CNN).

## REFERENCES

- [1] B. Meher, S. Agrawal, R. Panda, A. Abraham, *A survey on region based image fusion methods*, Information Fusion, Vol. 48, pp. 119–132, 2019.
- [2] V.A. Krylov, G. Moser, S.B. Serpico, J. Zerubia, *False discovery rate approach to unsupervised image change detection*, IEEE Trans. on Image Processing, Vol. 25, pp. 4704–4718, 2016.
- [3] A. Torabi, G. Massé, G. Bilodeau, *An iterative integrated framework for thermal-visible image registration, sensor fusion, and people tracking for video surveillance applications*, Computer Vision and Image Understanding, Vol. 116, pp. 210–221, 2012.
- [4] G. Zhang, Q. Wu, T. Wang, R. Zhao, M. Deng, B. Jiang, X. Li, H. Wang, Y. Zhu, F. Li, *Block Adjustment without GCPs for Chinese Spaceborne SAR GF-3 Imagery*, Sensors, Vol. 18, 4023, 2018.
- [5] G. Saygili, M. Staring, E.A. Hendriks, *Confidence estimation for medical image registration based on stereo confidences*, IEEE Trans. Med. Imaging, Vol. 35, pp. 539–549, 2016.
- [6] B. Zitová, J. Flusser, *Image registration methods: A survey*, Image Vis. Comput., Vol. 21, pp. 977–1000, 2003.
- [7] X. Guo, Z. Xu, Y. Lu, Y. Pang, *An Application of Fourier-Mellin Transform in Image Registration*, in Proc. of the Int. Conf. on Computer and Information Technology, Shanghai, China, 21–23 September 2005.
- [8] E. Ask, O. Enqvist, L. Svärm, F. Kahl, G. Lippolis, *Tractable and reliable registration of 2D point sets*, in Proc. of the 13th European Conf. on Computer Vision, Zürich, Switzerland, pp. 393–406, 6–12 Sept. 2014.
- [9] C. Harris, M. Stephens, *A combined corner and edge detector*, in Proc. of the 4th Alvey Vision Conference, pp. 147–151, Manchester, UK, 31 August–2 September 1988.
- [10] E. Rosten, T. Drummond, *Machine learning for high-speed corner detection*, in Proc. of the 9th European Conf. on Computer Vision, pp. 430–443, Graz, Austria, 7–13 May 2006.
- [11] D.G. Lowe, *Object recognition from local scale-invariant features*, in Proc. of the Int. Conf. on Computer Vision, pp. 1150–1157, Corfu, Greece, 20–25 September 1999.
- [12] H. Bay, T. Tuytelaars, L.V. Gool, *Surf: Speeded up robust features*, Comput. Vis. Image Underst., Vol. 110, pp. 346–359, 2018.
- [13] E. Tola, V. Lepetit, P. Fua, *Daisy: An efficient dense descriptor applied to wide baseline stereo*, IEEE Trans. on Pattern Anal. Mach. Intell., Vol. 32, pp. 815–830, 2010.
- [14] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, *ORB: An efficient alternative to SIFT or SURF*, in Proc. of the IEEE Int. Conf. on Computer Vision, Barcelona, Spain, 6–13 November 2011.
- [15] P.F. Alcantarilla, A. Bartoli, A.J. Davison, *KAZE Features*, in Proc. of the European Conf. on Computer Vision, pp. 214–227, Florence, Italy, 7–13 October 2012.
- [16] A.A. Goshtasby, *Image Registration: Principles, Tools and Methods*, Springer: Berlin, Germany, 2012.
- [17] S. Santini, R. Jain, *Similarity measures*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 21, pp. 871–883, 1999.
- [18] C. Spearman, *The proof and measurement of association between two things*, Int. J. of Epidemiol., Vol. 39, pp. 1137–1150, 2010.
- [19] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press: Princeton, NJ, USA, 1976.
- [20] D. Han, J. Dezert, S. Li, C. Han, Y. Yang, *Image registration based on evidential reasoning*, in Proc. of the 16th Int. Conf. on Information Fusion, pp. 1143–1150, Istanbul, Turkey, 9–12 July 2013.
- [21] M.A. Fischler, R.C. Bolles, *Readings in Computer Vision: Issues, Problem, Principles, and Paradigms*, Morgan Kaufmann: Burlington, VT, USA, 1987.
- [22] P.H.S Torr, D.W. Murray, *The development and comparison of robust methods for estimating the fundamental matrix*, Int. J. Comput. Vis., Vol. 24, pp. 271–300, 1997.
- [23] A. Nakazawa, *Noise stable image registration using random resample consensus*, in Proc. of the 23rd Int. Conf. on Pattern Recognition, pp. 853–858, Cancun, Mexico, 4–8 December 2016.
- [24] P. Jain, P. Kar, *Nonconvex Optimization for Machine Learning*, Found. Trends Mach. Learn., Vol. 10, pp. 142–336, 2017.
- [25] D.T. Pham, *Intelligent Optimisation Techniques*, Springer: Berlin, Germany, 2000.
- [26] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion: Collected Works*, Vol. 4, American Research Press: Rehoboth, NM, USA, 2015.
- [27] P. Smets, *The transferable belief model*, Artif. Intell., Vol. 66, pp. 191–234, 1994.
- [28] D. Han, J. Dezert, Z. Duan, *Evaluation of probability transformations of belief functions for decision making*, IEEE Trans. on Syst. Man and Cybern: Systems, Vol. 46, pp. 93–108, 2016.
- [29] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, (3rd edition), Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2008.
- [30] J. Canny, *A computational approach to edge detection*, IEEE Trans. on Pattern Anal. Mach. Intell., Vol. 8, pp. 679–698, 1986.
- [31] D.P.L. Ferreira, E. Ribeiro, C.A.Z. Barcelos, *A variational approach to non-rigid image registration with Bregman divergences and multiple features*, Pattern Recognit., Vol. 77, pp. 237–247, 2018.
- [32] X. Xia, G. Dang, Y. Yao, J. Liang, *Image registration model and algorithm for multi-focus images*, Pattern Recognit. Lett., Vol. 86, pp. 26–30, 2017.
- [33] J.P.W. Pluim, B. Likar, F.A. Gerritsen, *Biomedical Image Registration*, Springer: Berlin, Germany, 2006.
- [34] S. Leutenegger, M. Chli, R.Y. Siegwart, *BRISK: Binary Robust invariant scalable keypoints*, in Proc. of the IEEE Int. Conf. on Computer Vision, Barcelona, Spain, 6–13 November 2011.
- [35] D. Han, J. Dezert, J.-M. Tacnet, C. Han, *A fuzzy-cautious OWA approach with evidential reasoning*, in Proc. of the Int. Conf. on Information Fusion (Fusion 2012), pp. 278–285, Singapore, 9–12 July 2012.
- [36] D. Dubois, H. Prade, *Representation and Combination of Uncertainty with Belief Functions and Possibility Measures*, Comput. Intell., Vol. 1, pp. 244–264, 1988.

# Two applications of Inter-Criteria Analysis With Belief Functions

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Stefka Fidanova<sup>b</sup>, Deqiang Han<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, stefka@parallel.bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, A. Tchamova, S. Fidanova, D. Han, *Two Applications of Inter-Criteria Analysis with Belief Functions*, in a special issue of Cybernetics and Information Technologies Journal, Bugarian Academy of Sciences Editor, Vol. 20, No. 5, pp. 38–59, 2020, and reprinted with permission.

**Abstract**—In this paper we present two applications of a new Belief Function-based Inter-Criteria Analysis (BF-ICrA) approach for the assessment of the degree of redundancy of criteria involved in multi-criteria decision-making (MCDM) problems. This BF-ICrA method allows to simplify the original MCDM problem by suppressing redundant criteria (if any) and thus diminish the complexity of MCDM problem. This approach is appealing for solving large MCDM problems whose solution requires the fusion of many belief functions. We show how this approach can be used in two distinct fields of applications: The GPS Surveying Problem, and the car selection problem.

**Keywords:** Inter-Criteria Analysis, ICrA-BF, MultiCriteria Decision Making, MCDM, belief functions.

## I. INTRODUCTION

In a Multi-Criteria Decision-Making (MCDM) problem we consider a set of alternatives (or objects)  $\mathbf{A} \triangleq \{A_1, A_2, \dots, A_M\}$  ( $M > 2$ ), and a set of criteria  $\mathbf{C} \triangleq \{C_1, C_2, \dots, C_N\}$  ( $N \geq 1$ ). We search for the best alternative  $A^*$  given the available information expressed by a  $M \times N$  score matrix (also called benefit or payoff matrix)  $\mathbf{S} \triangleq [S_{ij} = C_j(A_i)]$ , and (eventually) the importance factor  $w_j \in [0, 1]$  of each criterion  $C_j$  with  $\sum_{j=1}^N w_j = 1$ . The set of normalized weighting factors is denoted by  $\mathbf{w} \triangleq \{w_1, w_2, \dots, w_N\}$ . Depending on the context of the MCDM problem, the score  $S_{ij}$  of each alternative  $A_i$  with respect to each criteria  $C_j$  can be interpreted either as a cost (i.e. an expense), or as a reward (i.e. a benefit). By convention and without loss of generality<sup>1</sup> we will always interpret the score as a reward having monotonically increasing preference. Thus, the best alternative  $A_j^*$  for a given criteria  $C_j$  will be the one providing the highest reward/benefit.

The MCDM problem is not easy to solve because the scores are usually expressed in different (physical) units and different scales. This necessitates a choice of score/data normalization yielding rank reversal problems [1], [2]. Usually there is no same best alternative choice  $A^*$  for all criteria, so a compromise must be established to provide a reasonable

and acceptable solution of the MCDM problem for decision-making support.

Many MCDM methods exist, see references in [3]. Most popular methods are AHP<sup>2</sup> [4], ELECTRE<sup>3</sup> [5], TOPSIS<sup>4</sup> [6], [7]. In 2016 and 2017, we did develop BF-TOPSIS methods [3], [8] based on Belief Functions (BF) to improve the original TOPSIS approach to avoid data normalization and to deal also with imprecise score values as well. It appears however that the complexity of these new BF-TOPSIS methods can become a bottleneck for their use in large MCDM problems because of the fusion step of basic belief assignments required for the implementation of the BF-TOPSIS. That is why a simplification of the MCDM problem (if possible) is very welcome in order to save computational time and resources. This is the motivation of the present work.

For this aim we propose a new Inter-Criteria Analysis (ICrA) based on belief functions for identifying and estimating the possible degree of agreement (i.e. redundancy) between some criteria driven from the data (score values). This permits to remove all redundant criteria of the original MCDM problem and thus solving a simplified (almost) equivalent MCDM problem faster and at lower computational cost. ICrA has been developed originally by Atanassov et al. [9]–[11] based on Intuitionistic Fuzzy Sets [12], and it has been applied in different fields like medicine [13]–[15], optimization [16]–[20], workforce planning [21], competitiveness analysis [22], radar detection [23], ranking [24]–[27], etc. In this paper we improve ICrA approach thanks to belief functions introduced by Shafer in [28] from original Dempster's works [29]. We will refer it as BF-ICrA method in the sequel.

After a short presentation of basics of belief functions in section II, we present Atanassov's ICrA method in section III and discuss its limitations. In Section IV we present the new BF-ICrA approach based on a new construction of Basic Belief Assignment (BBA) matrix from the score matrix and a new establishment of Inter-Criteria belief matrix. In section

<sup>2</sup>Analytic Hierarchy Process

<sup>3</sup>ELimination Et Choix Taduisant la REalité

<sup>4</sup>Technique for Order Preference by Similarity to Ideal Solution

<sup>1</sup>because it suffices to multiply the scores values by  $-1$  to reverse the preference ordering.

V a method of simplification of MCDM using BF-ICrA is proposed. Two distinct applications of BF-ICrA are presented in VI with concluding remarks in Section VII.

## II. BASICS OF THE THEORY OF BELIEF FUNCTIONS

To follow classical notations of the theory of belief functions, also called Dempster-Shafer Theory (DST) [28], we assume that the answer (i.e. the solution, or the decision to take) of the problem under concern belongs to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are exclusive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A BBA (or mass function) associated with a given source of evidence is defined [28] as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are usually interpreted respectively as lower and upper bounds of unknown (possibly subjective) probability measure [29]. They are defined by<sup>5</sup>

$$\text{Bel}(A) \triangleq \sum_{B \subseteq A, B \in 2^\Theta} m(B), \quad \text{and} \quad \text{Pl}(A) \triangleq 1 - \text{Bel}(\bar{A}). \quad (1)$$

If  $m(A) > 0$ ,  $A$  is called a focal element of  $m(\cdot)$ . When all focal elements are singletons then  $m(\cdot)$  is called a *Bayesian BBA* [28] and its corresponding  $\text{Bel}(\cdot)$  function is homogeneous to a probability measure. The vacuous BBA, or VBBA for short, representing a totally ignorant source is defined as  $m_v(\Theta) = 1$ . The main challenge of the decision-maker consists to combine efficiently the possible multiple BBAs  $m_s(\cdot)$  given by  $s > 1$  distinct sources of evidence to obtain a global (combined) BBA, and to make a final decision from it. Historically the combination of BBAs is accomplished by Dempster's rule proposed by Shafer in DST. Because Dempster's rule presents several serious problems (insensitivity to the level of conflict between sources in some cases, inconsistency with bounds of conditional probabilities when used for belief conditioning, dictatorial behavior, counter-intuitive results), many fusion rules have been proposed in the literature as alternative to Dempster's rule, see [30], Vol. 2 for a detailed list of fusion rules. We will not detail here all the possible combination rules but just mention that the Proportional Conflict Redistribution rule no. 6 (PCR6) proposed by Martin and Osswald in [30] (Vol. 3) is one of the most serious alternative rule for BBA combination available so far.

## III. ATANASSOV'S INTER-CRITERIA ANALYSIS (ICrA)

Atanassov's Inter-Criteria Analysis (ICrA) approach is based on a  $M \times N$  score matrix<sup>6</sup>  $\mathbf{S} \triangleq [S_{ij} = C_j(A_i), i = 1, \dots, M, j = 1, \dots, N]$ , and intuitionistic fuzzy pairs [12] including two membership functions  $\mu(\cdot)$  and  $\nu(\cdot)$ . Mathematically, an intuitionistic fuzzy set (IFS)  $A$  is denoted by  $A \triangleq \{(x, \mu_A(x), \nu_A(x)) | x \in E\}$ , where  $E$  is the set of possible values of  $x$ ,  $\mu_A(x) \in [0, 1]$  defines the membership of

$x$  to the set  $A$ , and  $\nu_A(x) \in [0, 1]$  defines the non-membership of  $x$  to the set  $A$ , with the restriction  $0 \leq \mu_A(x) + \nu_A(x) \leq 1$ . The ICrA method consists to build an  $N \times N$  Inter-Criteria (IC) matrix from the score matrix  $\mathbf{S}$ . The elements of the IC matrix consist of all intuitionistic fuzzy pairs  $(\mu_{jj'}, \nu_{jj'})$  whose components express respectively the degree of agreement and the degree of disagreement between criteria  $C_j$  and  $C_{j'}$  for  $j, j' \in \{1, 2, \dots, N\}$ . For a given column  $j$  (i.e. criterion  $C_j$ ), it is always possible to compare with  $>$ ,  $<$  and  $=$  operators all the scores  $S_{ij}$  for  $i = 1, 2, \dots, M$  because the scores of each column are expressed in same unit. The construction of IC matrix can be used to search relations between the criteria because the method compares homogeneous data relatively to a same column. In [32] Atanassov prescribes to normalize the score matrix before applying ICrA as follows

$$S_{ij}^{\text{norm}} = (S_{ij} - S_j^{\text{min}}) / (S_j^{\text{max}} - S_j^{\text{min}}), \quad (2)$$

if one wants to apply it in the dual manner for the search of InterObjects analysis (IOBa).

Because we focus on ICrA only, we don't need to apply a score matrix normalization because each column of the score matrix represents the values of a same criterion for different alternatives, and the criterion values are expressed with the same unit (e.g.  $m$ ,  $m^2$ ,  $\text{sec}$ ,  $Kg$ , or  $\text{€}$ , etc).

### A. Construction of Inter-Criteria matrix

The construction of the  $N \times N$  IC matrix, denoted<sup>7</sup>  $\mathbf{K}$ , is based on the pairwise comparisons between every two criteria along all evaluated alternatives. Let  $K_{jj'}^\mu$  be the number of cases in which the inequalities  $S_{ij} > S_{i'j}$  and  $S_{ij'} > S_{i'j'}$  hold simultaneously, and let  $K_{jj'}^\nu$  be the number of cases in which the inequalities  $S_{ij} > S_{i'j}$  and  $S_{ij'} < S_{i'j'}$  hold simultaneously. Because the total number of comparisons between the alternatives is  $M(M-1)/2$  then one always has necessarily

$$0 \leq K_{jj'}^\mu + K_{jj'}^\nu \leq \frac{M(M-1)}{2}, \quad (3)$$

or equivalently after the division by  $\frac{M(M-1)}{2} > 0$

$$0 \leq \frac{2K_{jj'}^\mu}{M(M-1)} + \frac{2K_{jj'}^\nu}{M(M-1)} \leq 1. \quad (4)$$

This inequality permits to define the elements of  $N \times N$  IC matrix  $\mathbf{K} = [K_{jj'}]$  as intuitionistic fuzzy (IF) pairs  $K_{jj'} = (\mu_{jj'}, \nu_{jj'})$  where

$$\mu_{jj'} \triangleq \frac{2K_{jj'}^\mu}{M(M-1)}, \quad \text{and} \quad \nu_{jj'} \triangleq \frac{2K_{jj'}^\nu}{M(M-1)}. \quad (5)$$

$\mu_{jj'}$  measures the degree of agreement between criteria  $C_j$  and  $C_{j'}$ , and  $\nu_{jj'}$  measures their degree of disagreement. By construction the IC matrix  $\mathbf{K}$  is always a symmetric matrix.

<sup>5</sup>where the symbol  $\triangleq$  means *equal by definition*.

<sup>6</sup>called index matrix by Atanassov in [31].

<sup>7</sup>We use  $\mathbf{K}$  because it corresponds to the first letter of word Kriterium, meaning criteria in German. The letter  $C$  is being already in use.

The computation of  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  can be done explicitly thanks to Atanassov's formulas [32]

$$K_{jj'}^\mu = \sum_{i=1}^{M-1} \sum_{i'=i+1}^M [\text{sgn}(S_{ij} - S_{i'j})\text{sgn}(S_{ij'} - S_{i'j'}) + \text{sgn}(S_{i'j} - S_{ij})\text{sgn}(S_{i'j'} - S_{ij'})], \quad (6)$$

and

$$K_{jj'}^\nu = \sum_{i=1}^{M-1} \sum_{i'=i+1}^M [\text{sgn}(S_{ij} - S_{i'j})\text{sgn}(S_{i'j'} - S_{ij'}) + \text{sgn}(S_{i'j} - S_{ij})\text{sgn}(S_{ij'} - S_{i'j'})], \quad (7)$$

where the signum function  $\text{sgn}(\cdot)$  used by Atanassov is defined as follows

$$\text{sgn}(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases} \quad (8)$$

Actually the values of  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  depend on the choice of  $\text{sgn}(x)$  function<sup>8</sup>. That is why in [21], [33], the authors propose different algorithms implemented under Java in an ICRA software yielding different  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  values for making the analysis and to reduce the dimension (complexity) of the original MCDM problem.

### B. Inter-criteria analysis

Once the Inter-Criteria matrix  $\mathbf{K} = [K_{jj'}]$  of intuitionistic fuzzy pairs is calculated one needs to analyze it to decide which criteria  $C_j$  and  $C_{j'}$  are in strong agreement (or positive consonance) reflecting the correlation between  $C_j$  and  $C_{j'}$ , in strong disagreement (or negative consonance) reflecting non correlation between  $C_j$  and  $C_{j'}$ , or in dissonance reflecting the uncertainty situation where nothing can be said about the non correlation or the correlation between  $C_j$  and  $C_{j'}$ . If one wants to identify the set of criteria  $C_{j'}$  for  $j' \neq j$  that are strongly correlated with  $C_j$  then we can sort  $\mu_{jj'}$  values in descending order to identify those in strong positive consonance with  $C_j$ . In [25], [26], the authors propose a qualitative scale to refine the levels of consonance and dissonance and for helping the decision making procedure. A dual approach based on  $\nu_{jj'}$  values can be made to determine the set of criteria that are not correlated with  $C_j$ . An other approach [10], [27] proposes to define two thresholds  $\alpha, \beta \in [0; 1]$  for the positive and negative consonance respectively against which the components  $\mu_{jj'}$  and  $\nu_{jj'}$  of  $K_{jj'} = (\mu_{jj'}, \nu_{jj'})$  will be compared. The correlations between the criteria  $C_j$  and  $C_{j'}$  are called "positive consonance", "negative consonance" or "dissonance" depending on their  $\mu_{jj'}$  and  $\nu_{jj'}$  values with respect to chosen thresholds  $\alpha$  and  $\beta$ , see [22] for details. More precisely,  $C_j$  and  $C_{j'}$  are in

- $(\alpha, \beta)$  positive consonance (i.e. correlated):  
If  $\mu_{jj'} > \alpha$  and  $\nu_{jj'} < \beta$ .
- $(\alpha, \beta)$  negative consonance (i.e. no correlated):

<sup>8</sup>for instance if we use  $\text{sgn}(x) = 1$  if  $x \geq 0$  and  $\text{sgn}(x) = 0$  if  $x < 0$ , we will obtain, in general, other  $K_{jj'}^\mu$  and  $K_{jj'}^\nu$  values.

If  $\mu_{jj'} < \beta$  and  $\nu_{jj'} > \alpha$ .

- $(\alpha, \beta)$  dissonance (i.e. full uncertainty): Otherwise.

At the beginning of ICRA development it was not very clear how these intuitionistic fuzzy (IF) pairs  $(\mu_{jj'}, \nu_{jj'})$  had to be used and that is why Atanassova [34], [35] proposed to handle both components of the IF pair. For this, she interpreted pairs  $(\mu_{jj'}, \nu_{jj'})$  as points located in the elementary  $TFU$  triangle, where the point  $T$  of coordinate  $(1, 0)$  represents the maximal positive consonance (i.e. the true consonance), the point  $F$  with coordinate  $(0, 1)$  represents the maximal negative consonance (i.e. the falsity), and the point  $U$  with coordinates  $(0, 0)$  represents the maximal dissonance (i.e. the uncertainty). From this interpretation it becomes easy to identify the top of consonant IF pairs  $(\mu_{jj'}, \nu_{jj'})$  that fall in bottom right corner of  $(TFU)$  triangle limited by vertical line from  $x$ -axis  $x = \alpha$ , and horizontal line from  $y$ -axis  $y = \beta$ . The set of consonant IF pairs are then ranked according to their Euclidean distance  $d_{C_j C_{j'}}^T$  with respect to  $T$  point of coordinate  $(1, 0)$  defined by

$$d_{C_j C_{j'}}^T = d((1, 0), (\mu_{jj'}, \nu_{jj'})) = \sqrt{(1 - \mu_{jj'})^2 + \nu_{jj'}^2}. \quad (9)$$

In the MCDM context only the criteria that are negatively consonant (or uncorrelated) must be kept for solving MCDM and saving computational resources because they have no (or only very low) dependency with each other, so that each uncorrelated criterion provides useful information. The set of criteria that are positively consonant (if any), called the consonant set, indicates somehow a redundancy of information between the criteria belonging to it in term of decisional behavior. Therefore all these positively consonant criteria must be represented by only one representative criterion that will be kept in the MCDM analysis to simplify MCDM problem. Also all the criteria that are deemed strongly dissonant (if any) could be taken out of the original MCDM problem because they only introduce uncertainty in the decision-making.

### C. General comments on ICRA

Although appealing at the first glance, the classical ICRA approach induces the following comments:

- 1) The IF values  $\mu_{jj'}$  and  $\nu_{jj'}$  can be easily interpreted in the belief function framework. Indeed, the belief and plausibility of (positive) consonance between criteria  $C_j$  and  $C_{j'}$  can be directly linked to the values  $\mu_{jj'}$  and  $\nu_{jj'}$  by taking  $Bel_{jj'}(\theta) = \mu_{jj'}$  and  $Pl_{jj'}(\theta) = 1 - \nu_{jj'}$ . Moreover  $U_{jj'}(\theta) = Pl_{jj'}(\theta) - Bel_{jj'}(\theta) = 1 - \nu_{jj'} - \mu_{jj'}$  represents the dissonance (the uncertainty about the correlation) of the criteria  $C_j$  and  $C_{j'}$ . Here the proposition  $\theta$  means: the criteria  $C_j$  and  $C_{j'}$  are totally positively consonant (i.e. totally correlated) and the frame of discernment is defined as  $\Theta \triangleq \{\theta, \bar{\theta}\}$ , where  $\bar{\theta}$  means: the criteria  $C_j$  and  $C_{j'}$  are totally negatively

consonant (uncorrelated). From this, one can define any BBA  $m_{jj'}(\theta)$ ,  $m_{jj'}(\bar{\theta})$  and  $m_{jj'}(\theta \cup \bar{\theta})$  of  $2^\Theta$  by

$$m_{jj'}(\theta) = \mu_{jj'}, \quad (10)$$

$$m_{jj'}(\bar{\theta}) = \nu_{jj'}, \quad (11)$$

$$m_{jj'}(\theta \cup \bar{\theta}) = 1 - \mu_{jj'} - \nu_{jj'}. \quad (12)$$

- 2) The construction of  $\mu_{jj'}$  and  $\nu_{jj'}$  proposed in the classical ICRA is disputable because it is only based on counting the valid “>” or “<” inequalities but it does not exploit how bigger and how smaller the scores values are in each comparison done in the construction of the Inter-Criteria Matrix  $\mathbf{K}$ . Therefore the construction of  $\mu_{jj'}$  and  $\nu_{jj'}$  is actually only a very crude method to estimate IF pairs.
- 3) The construction of the Inter-Criteria Matrix  $\mathbf{K}$  is in fact not unique as reported in [33]. This will yield different results in general.
- 4) The exploitation of the ICRA method depends on the choice of  $\alpha$  and  $\beta$  thresholds that will impact the final result.
- 5) The classical ICRA method cannot deal directly with imprecise or missing score values.

#### IV. A NEW ICRA METHOD BASED ON BELIEF FUNCTIONS

We present in this section a new ICRA method, called BF-ICRA for short, based on belief functions that circumvents most of the aforementioned drawbacks of classical ICRA. Here we show how to get more precisely the Inter-Criteria Belief Matrix and how to exploit it for MCDM simplification.

##### A. Construction of BBA matrix from the score matrix

From any non-zero score matrix  $\mathbf{S} = [S_{ij}]$ , we can construct the  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  as follows

$$m_{ij}(A_i) = Bel_{ij}(A_i), \quad (13)$$

$$m_{ij}(\bar{A}_i) = Bel_{ij}(\bar{A}_i) = 1 - Pl_{ij}(A_i), \quad (14)$$

$$m_{ij}(A_i \cup \bar{A}_i) = Pl_{ij}(A_i) - Bel_{ij}(A_i). \quad (15)$$

Assuming  $A_{\max}^j \neq 0$  and  $A_{\min}^j \neq 0$ , we take<sup>9</sup>

$$Bel_{ij}(A_i) \triangleq Sup_j(A_i)/A_{\max}^j, \quad (16)$$

$$Bel_{ij}(\bar{A}_i) \triangleq Inf_j(A_i)/A_{\min}^j, \quad (17)$$

where  $A_{\max}^j \triangleq \max_i Sup_j(A_i)$  and  $A_{\min}^j \triangleq \min_i Inf_j(A_i)$  and with

$$Sup_j(A_i) \triangleq \sum_{k \in \{1, \dots, M\} | S_{kj} \leq S_{ij}} |S_{ij} - S_{kj}|, \quad (18)$$

$$Inf_j(A_i) \triangleq - \sum_{k \in \{1, \dots, M\} | S_{kj} \geq S_{ij}} |S_{ij} - S_{kj}|. \quad (19)$$

The entire justification of these formulas can be found in our previous works [3]. For another criterion  $C_{j'}$  and the  $j'$ -th column of the score matrix we will obtain another set of BBA values  $m_{ij'}(\cdot)$ . Applying this method for each column

of the score matrix we are able to compute the BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  whose each component is in fact a triplet  $(m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i))$  of BBA values in  $[0, 1]$  such that  $m_{ij}(A_i) + m_{ij}(\bar{A}_i) + m_{ij}(A_i \cup \bar{A}_i) = 1$  for all  $i = 1, \dots, M$  and  $j = 1, \dots, N$ .

##### B. Construction of Inter-Criteria Matrix from BBA matrix

The next step of BF-ICRA approach is the construction of the  $N \times N$  Inter-Criteria Matrix  $\mathbf{K} = [K_{jj'}]$  from  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  where elements  $K_{jj'}$  corresponds to the BBA  $(m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  about positive consonance  $\theta$ , negative consonance  $\bar{\theta}$  and uncertainty between criteria  $C_j$  and  $C_{j'}$  respectively. The principle of construction of the triplet  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  is based on two steps that will be detailed in the sequel:

- Step 1: For each alternative  $A_i$ , we first compute the BBA  $(m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  for any two criteria  $j, j' \in \{1, 2, \dots, N\}$ .
- Step 2: The BBA  $(m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  is then obtained by the combinations of the  $M$  BBA  $m_{jj'}^i(\cdot)$ .

We present the details of each step of BF-ICRA method.

##### Step 1: Construction of BBA $m_{jj'}^i(\cdot)$

The mass of belief  $m_{jj'}^i(\theta)$  represents the degree of agreement between the BBA  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$  for the alternative  $A_i$ , and  $m_{jj'}^i(\bar{\theta})$  represents the degree of disagreement between  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$ . The mass  $m_{jj'}^i(\theta \cup \bar{\theta})$  is the degree of uncertainty about the agreement (or disagreement) between  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$  for the alternative  $A_i$ . The calculation of  $m_{jj'}^i(\theta)$  could be envisaged in several manners.

The first manner would consist to consider the degree of conflict [28]  $k_{jj'}^i \triangleq \sum_{X, Y \in \Theta | X \cap Y = \emptyset} m_{ij}(X) m_{ij'}(Y)$  and consider the Bayesian BBA  $m_{jj'}^i(\theta) = 1 - k_{jj'}^i$ ,  $m_{jj'}^i(\bar{\theta}) = k_{jj'}^i$  and  $m_{jj'}^i(\theta \cup \bar{\theta}) = 0$ . Instead of using Shafer's conflict, the second manner would consist to use a normalized distance  $d_{jj'}^i = d(m_{ij}, m_{ij'})$  to measure the closeness between  $m_{ij}(\cdot)$  and  $m_{ij'}(\cdot)$ , and then consider the Bayesian BBA modeling defined by  $m_{jj'}^i(\theta) = 1 - d_{jj'}^i$ ,  $m_{jj'}^i(\bar{\theta}) = d_{jj'}^i$  and  $m_{jj'}^i(\theta \cup \bar{\theta}) = 0$ . These two manners however are not very satisfying because they always set to zero the degree of uncertainty between the agreement and disagreement of the BBA, and the second manner depends also on the choice of the distance metric. So, we propose a more appealing third manner of the BBA modeling of  $m_{jj'}^i(\theta)$ ,  $m_{jj'}^i(\bar{\theta})$ , and  $m_{jj'}^i(\theta \cup \bar{\theta})$ . For this, we consider two sources of evidences (SoE) indexed by  $j$  and  $j'$  providing the BBA  $m_{ij}$  and  $m_{ij'}$  defined on the simple FoD  $\{A_i, \bar{A}_i\}$  and denoted  $m_{ij} = [m_{ij}(A_i), m_{ij}(\bar{A}_i), m_{ij}(A_i \cup \bar{A}_i)]$  and  $m_{ij'} = [m_{ij'}(A_i), m_{ij'}(\bar{A}_i), m_{ij'}(A_i \cup \bar{A}_i)]$ . We also denote  $\Theta = \{\theta, \bar{\theta}\}$  the FoD about the relative state of the two SoE, where  $\theta$  means that the two SoE agree,  $\bar{\theta}$  means that they disagree and  $\theta \cup \bar{\theta}$  means that we don't know. Then the BBA modeling is based on the important remarks

- Two SoE are in total agreement if both commit their maximum belief mass to the element  $A_i$  or to element

<sup>9</sup>If  $A_{\max}^j = 0$  then  $Bel_{ij}(A_i) = 0$ , and if  $A_{\min}^j = 0$  then  $Pl_{ij}(A_i) = 1$ .

$\bar{A}_i$ . So they perfectly agree if  $m_{ij}(A_i) = m_{ij'}(A_i) = 1$ , or if  $m_{ij}(\bar{A}_i) = m_{ij'}(\bar{A}_i) = 1$ . Therefore the pure degree of agreement<sup>10</sup> between two sources is modeled by

$$m_{jj'}^i(\theta) = m_{ij}(A_i)m_{ij'}(A_i) + m_{ij}(\bar{A}_i)m_{ij'}(\bar{A}_i). \quad (20)$$

- Two SoE are in total disagreement if each one commits its maximum mass of belief to one element and the other to its opposite, that is if one has  $m_{ij}(A_i) = 1$  and  $m_{ij'}(\bar{A}_i) = 1$ , or if  $m_{ij}(\bar{A}_i) = 1$  and  $m_{ij'}(A_i) = 1$ . Hence the pure degree of disagreement<sup>11</sup> between two sources is modeled by

$$m_{jj'}^i(\bar{\theta}) = m_{ij}(A_i)m_{ij'}(\bar{A}_i) + m_{ij}(\bar{A}_i)m_{ij'}(A_i). \quad (21)$$

- All possible remaining products between components of  $m_{ij}$  and  $m_{ij'}$  reflect the part of uncertainty we have about the SoE (i.e. we don't know if they agree or disagree). Hence the degree of uncertainty between the two sources is modeled by

$$\begin{aligned} m_{jj'}^i(\theta \cup \bar{\theta}) &= m_{ij}(A_i)m_{ij'}(A_i \cup \bar{A}_i) \\ &\quad + m_{ij}(\bar{A}_i)m_{ij'}(A_i \cup \bar{A}_i) \\ &\quad + m_{ij}(A_i \cup \bar{A}_i)m_{ij'}(A_i) \\ &\quad + m_{ij}(A_i \cup \bar{A}_i)m_{ij'}(\bar{A}_i) \\ &\quad + m_{ij}(A_i \cup \bar{A}_i)m_{ij'}(A_i \cup \bar{A}_i). \end{aligned} \quad (22)$$

By construction  $m_{jj'}^i(\cdot) = m_{ij'}^i(\cdot)$ , hence this BBA modeling permits to build a set of  $M$  symmetrical Inter-Criteria Belief Matrices (ICBM)  $\mathbf{K}^i = [K_{jj'}^i]$  of dimension  $N \times N$  relative to each alternative  $A_i$  whose components  $K_{jj'}^i$  correspond to the triplet of BBA values  $m_{jj'}^i = (m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}))$  modeling the belief of agreement and of disagreement between  $C_j$  and  $C_{j'}$  based on  $A_i$ . One has also<sup>12</sup>  $m_{jj'}^i(\theta), m_{jj'}^i(\bar{\theta}), m_{jj'}^i(\theta \cup \bar{\theta}) \in [0, 1]$  and  $m_{jj'}^i(\theta) + m_{jj'}^i(\bar{\theta}) + m_{jj'}^i(\theta \cup \bar{\theta}) = 1$ . This BBA construction can be easily extended for modeling the agreement, disagreement and uncertainty of  $n > 2$  criteria  $C_{j_1}, \dots, C_{j_n}$  altogether if needed by taking

$$\begin{aligned} m_{j_1 \dots j_n}^i(\theta) &= \prod_{k=1}^n m_{ij_k}(A_i) + \prod_{k=1}^n m_{ij_k}(\bar{A}_i), \\ m_{j_1 \dots j_n}^i(\bar{\theta}) &= \sum_{\substack{X_{j_1}, \dots, X_{j_n} \in \{A_i, \bar{A}_i\} \\ X_{j_1} \cap \dots \cap X_{j_n} = \emptyset}} \prod_{k=1}^n m_{ij_k}(X_{j_k}), \\ m_{j_1 \dots j_n}^i(\theta \cup \bar{\theta}) &= 1 - m_{j_1 \dots j_n}^i(\theta) - m_{j_1 \dots j_n}^i(\bar{\theta}). \end{aligned}$$

## Step 2: Construction of BBA $m_{jj'}(\cdot)$ (fusion step)

Once all the BBAs  $m_{jj'}^i(\cdot)$  ( $i = 1, \dots, M$ ) are calculated one combines them to get the component  $K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))$  of the Inter-Criteria Belief matrix (ICBM)  $\mathbf{K} = [K_{jj'}]$ . This fusion step can be done in many ways depending on the combination rule chosen by the user. If the number of alternatives  $M$  is not too large we recommend

<sup>10</sup>or positive consonance according Atanassov's terminology.

<sup>11</sup>or negative consonance according Atanassov's terminology.

<sup>12</sup>because  $(m_{ij}(A_i) + m_{ij}(\bar{A}_i) + m_{ij}(A_i \cup \bar{A}_i))(m_{ij'}(A_i) + m_{ij'}(\bar{A}_i) + m_{ij'}(A_i \cup \bar{A}_i)) = 1 \cdot 1 = 1$ .

to combine the BBAs  $m_{jj'}^i(\cdot)$  with PCR6 fusion rule [30] (Vol. 3) because of known deficiencies of Dempster's rule. If  $M$  is too large to prevent PCR6 working on computer, we can just use the simple averaging rule of combination in these high dimensional MCDM problems. Simple Matlab™ code for PCR6 rule can be found in [42] for convenience.

The computational complexity of BF-ICrA is of course higher than the complexity of ICrA because it makes a more precise evaluation of local and global inter-criteria belief matrices with respect to Intuitionistic Fuzzy inter-criteria matrices of ICrA. The overall reduction of the computational burden of the original MCDM problem thanks to BF-ICrA depends highly on the problem under concern, the complexity and cost to evaluate each criteria involved in it, as well as the number of redundant criteria identified by BF-ICrA method.

## V. SIMPLIFICATION OF ORIGINAL MCDM THANKS TO BF-ICrA

Once the global Inter-Criteria Belief Matrix  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$  is calculated, we need to identify and cluster the criteria that are in strong agreement, in strong disagreement, and those on which we are uncertain. For identifying the criteria that are in very strong agreement, we evaluate the distance of each component of  $K_{jj'}$  with the BBA representing the best agreement state and characterized by the specific BBA<sup>13</sup>  $m_T(\theta) = 1$ . From a similar approach we can also identify, if we want, the criteria that are in very strong disagreement using the distance of  $m_{jj'}(\cdot)$  with respect to the BBA representing the best disagreement state characterized by the specific BBA  $m_F(\bar{\theta}) = 1$ . As alternative of Jousselme's distance [37], we use the  $d_{BI}(\cdot, \cdot)$  distance based on belief interval [36] because it is a good method for measuring the distance  $d(m_1, m_2)$  between the two BBAs<sup>14</sup>  $m_1(\cdot)$  and  $m_2(\cdot)$  over the same FoD. It is defined by

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))}, \quad (23)$$

where the *Belief-Intervals* are defined by  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)]$  and computed from  $m_1(\cdot)$  and  $m_2(\cdot)$  thanks to formula (1).  $d_W(BI_1(X), BI_2(X))$  is Wassertein's distance between intervals calculated by

$$\begin{aligned} d_W([a_1, b_1], [a_2, b_2]) &= \\ &= \sqrt{\left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2}, \end{aligned} \quad (24)$$

and  $N_c = 1/2^{|\Theta|-1}$  is a factor to get  $d_{BI}(m_1, m_2) \in [0, 1]$ .

Because all criteria that are in strong agreement somehow contain redundant (correlated) information and behave similarly from decision-making standpoint, we propose to simplify

<sup>13</sup>We use the index  $T$  in the notation  $m_T(\cdot)$  to refer that the agreement is true, and  $F$  in  $m_F(\cdot)$  to specify that the agreement is false.

<sup>14</sup>Here  $m_1(\cdot) = m_{jj'}(\cdot)$ , and  $m_2(\cdot) = m_T(\cdot)$  or  $m_2(\cdot) = m_F(\cdot)$ .

the original MCDM problem by keeping in the MCDM only criteria that are non redundant. The remaining criteria can be eventually weighted by their degree of importance reflecting the number of different criteria that are in agreement through this BF-ICrA approach.

For instance, if one has a seven criteria MCDM problem and if criteria  $C_1$ ,  $C_2$  and  $C_3$  are in strong agreement we will only select one remaining criterion among  $\{C_1, C_2, C_3\}$  and we give it a weight of  $w_1 + w_2 + w_3$ . Moreover if  $C_4$  and  $C_5$  are in strong agreement also we will only select one remaining criterion among  $\{C_4, C_5\}$  and we give it a weight of  $w_4 + w_5$ , and we will use the weight  $w_6$  for  $C_6$ , and  $w_7$  for  $C_7$ . Hence the original MCDM problem will reduce to a four simplified MCDM problem that can be solved using BF-TOPSIS method already presented in details in [3] and in [8], or with AHP [4] if one prefers, or with any other chosen method that the system-designer may prefer.

The strategy for selecting the most representative criterion among a set of redundant criteria is not unique and depends mainly on the cost necessary (i.e. human efforts, data mining, computational resources, etc) for getting the values of the score matrix of the problem under concern. The least costly criteria may be a good option of selection.

In [38], we provided simple detailed examples for BF-ICrA where we selected the representative criterion as being the one with smallest index. So in the aforementioned example the simplified MCDM problem will reduce to a  $M \times 4$  MCDM problem involving only four criteria  $C_1$ ,  $C_4$ ,  $C_6$  and  $C_7$ .

The BF-ICrA method proposed in this work allows also, in principle, to make a refined analysis (if necessary) based on IC matrices  $K_{jj}^i$ , about the origin of disagreement between criteria with respect to each alternative  $A_i$  in order to identify the potential inconsistencies in original MCDM problem. This aspect is not developed in this paper and has been left for future investigations. It is worth mentioning that the analysis of the number of redundant criteria versus time improvements that could be proposed as an effective measure of performance of this approach depends highly of the application under consideration and the difficulty (and cost) to get the value of each criteria.

## VI. TWO APPLICATIONS OF BF-ICrA

In this section we present two applications of the BF-ICrA approach. The first one is for Global Positioning System (GPS) Surveying Problems (GSP) presented in [39], and the second one is for the car selection problem.

### A. Application of BF-ICrA for the GPS surveying problem

GPS surveying is an NP-hard problem. For designing Global Positioning System surveying network, a given set of earth points must be observed consecutively. The survey cost is the sum of the distances to go from one point to another one. This kind of problems is hard to be solved with traditional numerical methods. Here we use BF-ICrA to analyze an Ant Colony Optimization (ACO) algorithm developed to provide

near-optimal solutions for Global Positioning System surveying problem.

GPS satellites continuously transmit radio signals to the Earth while orbiting it. A receiver, with unknown position on Earth, has to detect and convert the signals received from all of the satellites into useful measurements. These measurements would allow a user to compute a three-dimensional coordinate position: location of the receiver. Any GPS observation is proven to have biases, hence, in order to survey an appropriate combination of measurement processing strategies must be used to minimize their effect on the positioning results. Differencing data collected simultaneously from two or more GPS receivers to several GPS satellites allows to eliminate or significantly reduce most of the biases. The GPS network can be defined as set of stations  $(a_1, a_2, \dots, a_n)$ , which are co-ordinated by placing receivers  $(X1, X2, \dots)$  on them to determine sessions  $(a_1a_2, a_1a_3, a_2a_3, \dots)$  among them. The problem is to search for the best order in which these sessions can be organized to give the best schedule. Thus, the schedule can be defined as a sequence of sessions to be observed consecutively. The solution is represented by linear graph with weighted edges. The nodes represent the stations and the edges represent the moving cost. The objective function of the problem is the cost of the solution which is the sum of the costs (time) to move from one point to another one,  $C(V) = \sum C(a_i, a_j)$ , where  $a_i a_j$  is a session in solution  $V$ . For example if the number of points (stations) is 4, a possible solution is  $V = (a_1, a_3, a_2, a_4)$  and it can be represented by linear graph  $a_1 \rightarrow a_3 \rightarrow a_2 \rightarrow a_4$ . The moving costs are as follows:  $C(a_1, a_3), C(a_3, a_2), C(a_2, a_4)$ . Thus the cost of the solution is  $C(V) = C(a_1, a_3) + C(a_3, a_2) + C(a_2, a_4)$ . In practice, determining how each GPS receiver should be moved between stations to be surveyed in an efficient manner taking into account some important factors such as time, cost etc. The problem is to search for the best order, with respect to the time, in which these sessions can be observed to give the cheapest schedule or to minimize  $C(V)$ . The initial data is a cost matrix, which represents the cost (time, or distance) of moving a receiver from one point to another. Solving such problems - GSPs - to optimality requires a very high computational time. Therefore, meta-heuristic methods are used to provide near-optimal solutions for large networks within acceptable amount of computational effort. In this paper, we consider the Max-Min Ant System (MMAS) meta-heuristic [40] and we present it briefly in the next subsection.

Real ants foraging for food lay down quantities of pheromone (chemical cues) marking the path that they follow. An isolated ant moves essentially at random but an ant encountering a previously laid pheromone will detect it and decide to follow it with high probability and thereby reinforce it with a further quantity of pheromone. The repetition of the above mechanism represents the auto-catalytic behavior of real ant colony where the more the ants follow a trail, the more attractive that trail becomes.

The ACO algorithm uses a colony of artificial ants that behave as cooperative agents in a mathematics space were

they are allowed to search and reinforce pathways (solutions) in order to find the optimal ones. The problem is represented by graph and the ants walk on the graph to construct solutions. The solution is represented by path in the graph. After initialization of the pheromone trails, ants construct feasible solutions, starting from random nodes, then the pheromone trails are updated. At each step ants compute a set of feasible moves and select the best one (according to some probabilistic rules) to carry out the rest of the tour. The transition probability  $p_{ij}$ , to chose the node  $j$  when the current node is  $i$ , is based on the heuristic information  $\eta_{ij}$  and pheromone trail level  $\tau_{ij}$  of the move, where  $i, j = 1, \dots, n$ .

$$p_{ij} = \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum_{k \in U_{\text{unused}}} \tau_{ik}^\alpha \eta_{ik}^\beta}. \quad (25)$$

The higher value of the pheromone and the heuristic information, the more profitable is to select this move and resume the search. In the beginning, the initial pheromone level is set to a small positive constant value  $\tau_0$  and then ants update this value after completing the construction stage. ACO algorithms adopt different criteria to update the pheromone level.

In our implementation we use MAX-MIN Ant System (MMAS) [40], [41], which is ones of the best ant approaches. In MMAS the main is using fixed upper bound  $\tau_{max}$  and lower bound  $\tau_{min}$  of the pheromone trails. Thus accumulation of big amount of pheromone by part of the possible movements and repetition of same solutions is partially prevented. The main features of MMAS are:

The aim of using only one solution is to make solution elements, which frequently occur in the best found solutions, get large reinforcement. Pheromone trail update is given by:

$$\tau_{ij} \leftarrow \rho \tau_{ij} + \Delta \tau_{ij}, \quad (26)$$

where

$$\Delta \tau_{ij} = \begin{cases} 1/C(V_{best}) & \text{if } (i, j) \in \text{best solution,} \\ 0 & \text{otherwise,} \end{cases}$$

and  $V_{best}$  is the iteration best solution and  $i, j = 1, \dots, n$ .

To avoid stagnation of the search, the range of possible pheromone value on each movement is limited to an interval  $[\tau_{min}, \tau_{max}]$ .  $\tau_{max}$  is an asymptotic maximum of  $\tau_{ij}$  and  $\tau_{max} = 1/(1 - \rho)C(V^*)$ , while  $\tau_{min} = 0.087\tau_{max}$ . Where  $V^*$  is the optimal solution, but it is unknown, therefore we use  $V_{best}$  instead of  $V^*$ .

When all ants have completed their solutions, the pheromone level is updated by applying the global update rule. Only the pheromone corresponding to the best found solution is increased by the similar to the MMAS way. The global update rule is intended to provide a greater amount of pheromone on the paths of the best solution. It is a kind of intensification of the search around the best found solution. We use heuristic information equals to one over the cost of the session.

Here, we analyze the experimental results obtained using MMAS algorithm. For this, we use real data from Malta and

Seychelles GPS networks composed of 38 sessions and 71 sessions respectively denoted GSP1 and GSP2. We use also 6 larger test problems range from 100 to 443 sessions denoted GSP3, ..., GSP8. The results are obtained by performing 30 independent runs, for every experiment. The details of our MMAS implementation are given in [43]. So in our GSP example we consider 8 GSP criteria  $C_i = GSP_i, i = 1, \dots, 8$  and six average costs as results  $A_1, \dots, A_6$ , where  $A_1$  is the cost average for the first 5 runs,  $A_2$  the cost average for the first 10 runs,  $A_3$  for the first 15 runs, ... and finally  $C_6$  for all the 30 runs. Table I shows the values of averaged costs obtained for this problem. It corresponds to the transpose of the score matrix  $\mathbf{S}$ .

Hence in this problem  $M = 6$  and  $N = 8$ , and  $\mathbf{S} = [S_{ij}]$  is a  $6 \times 8$  score matrix. Based on classical ICRA approach, one gets the following IC matrices<sup>15</sup>  $\mathbf{K}^\mu$  and  $\mathbf{K}^\nu$  matrices given in (27) and (28).

The element  $K_{jj'}^\mu$  of matrix  $\mathbf{K}^\mu$  expresses the *degree of agreement* between criteria  $C_j = GSP_j$  and  $C_{j'} = GSP_{j'}$ , whereas the element  $K_{jj'}^\nu$  of matrix  $\mathbf{K}^\nu$  expresses the *degree of disagreement* between  $C_j = GSP_j$  and  $C_{j'} = GSP_{j'}$ .

Based on these results, one sees that ACO algorithm performs similarly for  $GSP_2, GSP_4, GSP_5$  and  $GSP_8$  because they are all in high agreement. Indeed  $\mu_{jj'}$  values for  $j, j' \in \{2, 4, 5, 8\}$  are quite high (greater than 70%). They are GPS networks with different numbers of sessions, but may have a similar structure, therefore, the value of agreement is high. For other networks, we can conclude that they have very different structure.

What is worth noting is that there appears also a strong agreement of GSP1 with GSP8 because  $\mu_{18} = 0.87$ . But because GSP8 is also in strong agreement with GSP2, GSP4, GSP5 and with GSP1 it is logically expected that GSP1 should be also in agreement with GSP2, GSP4, GSP5, which is unfortunately not the case based on this classical ICRA. This example points out some inconsistency of ICRA result because of the too crude method of estimation of the degree of agreement and disagreement between criteria based on IFS.

Now if we consider the same example with the same score matrix  $\mathbf{S}$  (built from Table I), we obtain the following IC Belief matrices<sup>16</sup>  $\mathbf{K}(\theta)$  and  $\mathbf{K}(\bar{\theta})$  given by (29) and (30). From ICBM  $\mathbf{K}(\theta)$  and  $\mathbf{K}(\bar{\theta})$  we compute the matrix  $\mathbf{D}(\theta)$  of distance of  $m_{jj'}(\cdot)$  to the full agreement state with BBA  $m_F(\theta) = 1$  based on  $d_{BI}(\cdot)$  distance. We get the following distances to full agreement  $\mathbf{D}(\theta)$  given in (31).

The element  $D_{jj'}$  represents the agreement distance between  $C_j$  and  $C_{j'}$ , the lower the better. From the values of elements of  $\mathbf{D}(\theta)$  matrix one sees clearly that ACO performs similarly for GSP2, GSP4 and GSP5 because distances  $D_{24}$ ,

<sup>15</sup>For presentation convenience and due to typesetting column width, we decompose et present the IC matrix  $\mathbf{K} = [K_{jj'} = (K_{jj'}^\mu, K_{jj'}^\nu)]$  into two distinct matrices  $\mathbf{K}^\mu = [K_{jj'}^\mu]$  and  $\mathbf{K}^\nu = [K_{jj'}^\nu]$ .

<sup>16</sup>For presentation convenience, the ICBM  $\mathbf{K} = [K_{jj'} = (m_{jj'}(\theta), m_{jj'}(\bar{\theta}), m_{jj'}(\theta \cup \bar{\theta}))]$  is decomposed into three matrices  $\mathbf{K}(\theta) = [K_{jj'}^\theta = m_{jj'}(\theta)]$ ,  $\mathbf{K}(\bar{\theta}) = [K_{jj'}^{\bar{\theta}} = m_{jj'}(\bar{\theta})]$  and  $\mathbf{K}(\theta \cup \bar{\theta}) = [K_{jj'}^{\theta \cup \bar{\theta}} = 1 - m_{jj'}(\theta) - m_{jj'}(\bar{\theta})]$ .

	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$A_6$
$C_1 = GSP1$	899.00	898.00	898.33	898.50	899.40	899.50
$C_2 = GSP2$	916.40	915.60	922.47	924.80	924.72	922.07
$C_3 = GSP3$	41336.40	41052.40	40991.93	40935.90	40832.20	40910.60
$C_4 = GSP4$	3244.80	3303.30	3327.00	3344.55	3345.60	3341.93
$C_5 = GSP5$	1656.20	1660.80	1663.93	1664.95	1666.96	1665.90
$C_6 = GSP6$	1673.60	1683.50	1690.73	1688.75	1690.24	1692.67
$C_7 = GSP7$	3420.00	3430.70	3433.13	3426.85	3429.44	3428.57
$C_8 = GSP8$	3758.20	3755.70	3758.73	3760.50	3760.80	3765.80

Table I  
TRANSPOSE OF THE SCORE MATRIX  $\mathbf{S} = [S_{ij}]$  OF GSP PROBLEM.

$$\mathbf{K}^\mu = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 1 & 0.60 & 0.27 & 0.67 & 0.73 & 0.67 & 0.33 & 0.87 \\ 0.60 & 1 & 0.27 & 0.80 & 0.73 & 0.53 & 0.47 & 0.73 \\ 0.27 & 0.27 & 1 & 0.07 & 0 & 0.20 & 0.40 & 0.13 \\ 0.67 & 0.80 & 0.07 & 1 & 0.93 & 0.73 & 0.53 & 0.80 \\ 0.73 & 0.73 & 0 & 0.93 & 1 & 0.80 & 0.60 & 0.87 \\ 0.67 & 0.53 & 0.20 & 0.73 & 0.80 & 1 & 0.67 & 0.80 \\ 0.33 & 0.47 & 0.40 & 0.53 & 0.60 & 0.67 & 1 & 0.47 \\ 0.87 & 0.73 & 0.13 & 0.80 & 0.87 & 0.80 & 0.47 & 1 \end{bmatrix} \end{matrix} \quad (27)$$

$$\mathbf{K}^\nu = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 0 & 0.40 & 0.73 & 0.33 & 0.27 & 0.33 & 0.67 & 0.13 \\ 0.40 & 0 & 0.73 & 0.20 & 0.27 & 0.47 & 0.53 & 0.27 \\ 0.73 & 0.73 & 0 & 0.93 & 1 & 0.80 & 0.60 & 0.87 \\ 0.33 & 0.20 & 0.93 & 0 & 0.07 & 0.27 & 0.47 & 0.20 \\ 0.27 & 0.27 & 1 & 0.07 & 0 & 0.20 & 0.40 & 0.13 \\ 0.33 & 0.47 & 0.80 & 0.27 & 0.20 & 0 & 0.33 & 0.20 \\ 0.67 & 0.53 & 0.60 & 0.47 & 0.40 & 0.33 & 0 & 0.53 \\ 0.13 & 0.27 & 0.87 & 0.20 & 0.13 & 0.20 & 0.53 & 0 \end{bmatrix} \end{matrix} \quad (28)$$

$$\mathbf{K}(\theta) = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 0.9098 & 0.6732 & 0.1791 & 0.5968 & 0.6106 & 0.5620 & 0.1659 & 0.7789 \\ 0.6732 & 0.9546 & 0.0364 & 0.8983 & 0.8783 & 0.8341 & 0.5532 & 0.7016 \\ 0.1791 & 0.0364 & 0.8722 & 0.0172 & 0.0154 & 0.0178 & 0.0366 & 0.1137 \\ 0.5968 & 0.8983 & 0.0172 & 0.9552 & 0.9146 & 0.9163 & 0.7395 & 0.6092 \\ 0.6106 & 0.8783 & 0.0154 & 0.9146 & 0.8917 & 0.8778 & 0.6922 & 0.6315 \\ 0.5620 & 0.8341 & 0.0178 & 0.9163 & 0.8778 & 0.9060 & 0.7630 & 0.6441 \\ 0.1659 & 0.5532 & 0.0366 & 0.7395 & 0.6922 & 0.7630 & 0.8587 & 0.2484 \\ 0.7789 & 0.7016 & 0.1137 & 0.6092 & 0.6315 & 0.6441 & 0.2484 & 0.8508 \end{bmatrix} \end{matrix} \quad (29)$$

$$\mathbf{K}(\bar{\theta}) = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 0.0207 & 0.1941 & 0.5385 & 0.2578 & 0.1757 & 0.2117 & 0.5335 & 0.0399 \\ 0.1941 & 0.0166 & 0.8323 & 0.0486 & 0.0298 & 0.0513 & 0.1808 & 0.0682 \\ 0.5385 & 0.8323 & 0.0117 & 0.9002 & 0.8754 & 0.8548 & 0.7062 & 0.5486 \\ 0.2578 & 0.0486 & 0.9002 & 0.0187 & 0.0216 & 0.0204 & 0.0606 & 0.1193 \\ 0.1757 & 0.0298 & 0.8754 & 0.0216 & 0.0170 & 0.0201 & 0.0558 & 0.0832 \\ 0.2117 & 0.0513 & 0.8548 & 0.0204 & 0.0201 & 0.0154 & 0.0390 & 0.0726 \\ 0.5335 & 0.1808 & 0.7062 & 0.0606 & 0.0558 & 0.0390 & 0.0110 & 0.3495 \\ 0.0399 & 0.0682 & 0.5486 & 0.1193 & 0.0832 & 0.0726 & 0.3495 & 0.0100 \end{bmatrix} \end{matrix} \quad (30)$$

$$\mathbf{D}(\theta) = [D_{jj'} = d_{BI}(m_{jj'}, m_T)] = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{matrix} & \begin{bmatrix} 0.0590 & 0.2633 & 0.6845 & 0.3331 & 0.2892 & 0.3314 & 0.6893 & 0.1406 \\ 0.2633 & 0.0321 & 0.8987 & \mathbf{0.0767} & \mathbf{0.0803} & 0.1135 & 0.3230 & 0.1950 \\ 0.6845 & 0.8987 & 0.0774 & 0.9418 & 0.9306 & 0.9192 & 0.8381 & 0.7241 \\ 0.3331 & \mathbf{0.0767} & 0.9418 & 0.0326 & \mathbf{0.0566} & 0.0552 & 0.1706 & 0.2668 \\ 0.2892 & \mathbf{0.0803} & 0.9306 & \mathbf{0.0566} & 0.0679 & 0.0770 & 0.1958 & 0.2404 \\ 0.3314 & 0.1135 & 0.9192 & 0.0552 & 0.0770 & 0.0592 & 0.1494 & 0.2293 \\ 0.6893 & 0.3230 & 0.8381 & 0.1706 & 0.1958 & 0.1494 & 0.0849 & 0.5626 \\ 0.1406 & 0.1950 & 0.7241 & 0.2668 & 0.2404 & 0.2293 & 0.5626 & 0.0892 \end{bmatrix} \end{matrix} \quad (31)$$

$D_{25}$ , and  $D_{45}$  are very small. Also we see that GSP6 is also in good agreement with GSP4 and GSP5 but is relatively less in agreement with GSP2 because  $D_{26} = 0.1135$ .

As we see there is no inconsistency in this new BF-ICrA method with respect to what provides classical ICrA because with BF-ICrA we have a much better and precise estimation of degrees of agreement and disagreement between criteria for making the analysis thanks to a proper belief functions modeling.

### B. Application of BF-ICrA for the car selection problem

Let's consider another concrete problem related to car selection. Suppose one has a limited budget of 12000€ and one wants to buy a new car based on multiple criteria. A set of potential cars under 12K€ that present interest with respect to some criteria is obtained initially from a search on the web. How to apply BF-ICrA to simplify the selection process, and how to make the final choice of the car to buy?

Here we consider a set of ten small urban cars  $\{A_1, A_2, \dots, A_{10}\}$  as follows:

- $A_1 =$  DACIA SANDERO SCe 75;
- $A_2 =$  RENAULT CLIO TCe 75;
- $A_3 =$  SUZUKI CELERIO 1.0 VVT Avantage;
- $A_4 =$  FORD KA+ Ka+ 1.2 70 ch S&S Essential;
- $A_5 =$  MITSUBISHI SPACE STAR 1.0 MIVEC 71;
- $A_6 =$  KIA PICANTO 1.0 essence MPi 67 ch BVM5;
- $A_7 =$  HYUNDAI I10 1.0 66 BVM5 Initia;
- $A_8 =$  CITROEN C1 VTi 72 S&S Live;
- $A_9 =$  TOYOTA AYGO 1.0 VVT-i x;
- $A_{10} =$  PEUGEOT 108 VTi 72ch S&S BVM5 Like;

We consider the following seventy criteria related to price, dimensions, engine and consumption of the car for making the choice of the best car to buy:

- $C_1$  is the price (€);
- $C_2$  is the length (mm);
- $C_3$  is the height (mm);
- $C_4$  is the width without mirror (mm);
- $C_5$  is the wheelbase (mm);
- $C_6$  is the max loading volume (L);
- $C_7$  is the tank capacity (L);
- $C_8$  is the unloaded weight (Kg);
- $C_9$  is the cylinder volume( $\text{cm}^3$ );
- $C_{10}$  is the acceleration 0-100 Km/h (s);
- $C_{11}$  is the max speed (Km/h);
- $C_{12}$  is the power (Kw);
- $C_{13}$  is the horse power (hp);
- $C_{14}$  is the mixed consumption (L/100Km);
- $C_{15}$  is the extra-urban consumption (L/100Km);
- $C_{16}$  is the urban consumption (L/100Km);
- $C_{17}$  is the CO2 emission level (g/Km)

The score matrix  $\mathbf{S} = [S_{ij}]$  is built from information extracted from car-makers technical characteristics available on the world wide web<sup>17</sup> site. For the chosen cars, the corresponding original score matrix is given by (32).

<sup>17</sup><https://automobile.choisir.com/comparateur/voitures-neuves>

For criteria  $C_1, C_4, C_8$ , and  $C_{14}$  to  $C_{17}$  we consider that smaller is better. For other criteria larger is better. To make the preference order homogeneous in the score matrix, we multiply values of columns  $C_1, C_4, C_8$ , and  $C_{14}$  to  $C_{17}$  by -1 so that our MCDM problem is described by a modified score matrix with homogeneous preference order ("larger is better") for each column before applying the BF-ICrA method.

After applying BF-ICrA method (with PCR6 fusion rule in step 2) we obtain the following IC Belief matrices  $\mathbf{K}(\bar{\theta}) = [m_{jj'}(\bar{\theta})]$ ,  $\mathbf{K}(\bar{\theta}) = [m_{jj'}(\bar{\theta})]$  and  $\mathbf{K}(\theta \cup \bar{\theta})$ <sup>18</sup> given by (33) and (34).

From ICBM  $\mathbf{K}(\theta)$  and  $\mathbf{K}(\bar{\theta})$  we compute the matrix  $\mathbf{D}(\theta) = [D_{jj'} = d_{BI}(m_{jj'}, m_T)]$  of distance of the BBA  $m_{jj'}(\cdot)$  with respect to the full agreement state having BBA  $m_F(\theta) = 1$  based on  $d_{BI}(\cdot)$  distance. We get the distances to full agreement given in (35).

The element  $D_{jj'}$  represents the agreement distance between  $C_j$  and  $C_{j'}$ , the lower the better.

From the analysis of elements of  $D_{jj'}$  one sees clearly that criteria  $C_{14}, C_{15}, C_{16}$  and  $C_{17}$  are in very strong agreement and will behave very similarly for the preference ordering which is not very surprising because they are all related with energy consumption. Hence only one criteria among of these four criteria be used to simplify the MCDM car selection problem. We decide to keep only criteria  $C_{16}$  (urban consumption) in simplified MCD because urban displacements will be the main use of the car. One sees clearly that  $C_2, C_5$  and  $C_7$  are also in very strong agreement and so they will behave very similarly for the preference ordering. One decides to keep only the criterion  $C_7$  (tank capacity) which we consider more important than criteria  $C_2$  and  $C_5$  because it is linked to autonomy of the car. From BF-ICrA, one sees that tank capacity is linked with the dimensions of the car (mainly its length and wheelbase), which makes perfectly sense. Also we can note that criteria  $C_{12}$  and  $C_{13}$  are not too far since their distance is only 0.1403 and we can simplify a bit more the MCDM problem by taking only criterion  $C_{12}$  (the power) instead of keeping  $C_{12}$  and  $C_{13}$ .

Thanks to BF-ICrA, we can simplify the original MCDM car selection problem by removing redundant criteria and keeping only those which bring useful information. So our simplified MCDM car selection problem is characterized by the  $10 \times 11$  score matrix given in (36).

From this reduced score matrix, we can apply classical MCDM techniques to find the final preference order for making final decision and selectioning the car to buy. For this, one needs to define the importance  $imp(C_j)$  of each criteria  $C_j$  involved in the score matrix above. For simplicity, the importance of each criteria  $C_j$  is expressed as a value in  $\{1, 2, 3, 4, 5\}$ , where 1 means *the least important*, and 5 means *the most important*. In this car selection example we take  $imp(C_1) = imp(C_{16}) = 5$ ,  $imp(C_6) = imp(C_7) = 4$ ,  $imp(C_{10}) = imp(C_{11}) = imp(C_{12}) = 3$ ,  $imp(C_8) =$

<sup>18</sup>The ICBM  $\mathbf{K}(\theta \cup \bar{\theta})$  is obtained from  $\mathbf{K}(\theta)$  and  $\mathbf{K}(\bar{\theta})$  by taking  $\mathbf{K}(\theta \cup \bar{\theta}) = [K_{jj'}^{\theta \cup \bar{\theta}} = 1 - m_{jj'}(\theta) - m_{jj'}(\bar{\theta})]$ .

$$\mathbf{S} \triangleq \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \\ A_8 \\ A_9 \\ A_{10} \end{matrix} & \begin{bmatrix} 7990 & 4069 & 1523 & 1733 & 2589 & 1200 & 50 & 969 & 998 & 14.2 & 158 & 55 & 75 & 5.2 & 4.5 & 6.5 & 117 \\ 10990 & 4063 & 1448 & 1732 & 2589 & 1146 & 45 & 1138 & 898 & 12.3 & 178 & 56 & 75 & 5 & 4.2 & 6.3 & 113 \\ 9790 & 3600 & 1530 & 1600 & 2425 & 1053 & 35 & 815 & 998 & 13.9 & 155 & 50 & 68 & 3.9 & 3.6 & 4.5 & 89 \\ 10350 & 3941 & 1524 & 1774 & 2490 & 1029 & 42 & 1063 & 1198 & 14.6 & 164 & 51 & 70 & 5.1 & 4.4 & 6.3 & 117 \\ 10990 & 3795 & 1505 & 1665 & 2450 & 910 & 35 & 865 & 999 & 16.7 & 172 & 52 & 71 & 4.6 & 4.1 & 5.3 & 105 \\ 11000 & 3595 & 1485 & 1595 & 2400 & 1010 & 35 & 860 & 998 & 14.3 & 161 & 49 & 67 & 4.4 & 3.7 & 5.6 & 106 \\ 11050 & 3665 & 1500 & 1660 & 2385 & 1046 & 40 & 1008 & 998 & 14.7 & 156 & 49 & 66 & 5.1 & 4.3 & 6.5 & 117 \\ 11550 & 3466 & 1465 & 1615 & 2340 & 780 & 35 & 840 & 998 & 14 & 160 & 53 & 72 & 3.7 & 3.4 & 4.3 & 85 \\ 11590 & 3465 & 1460 & 1615 & 2340 & 812 & 35 & 915 & 998 & 13.8 & 160 & 51 & 69 & 4.1 & 3.6 & 4.9 & 93 \\ 11950 & 3475 & 1460 & 1615 & 2340 & 780 & 35 & 840 & 998 & 12.6 & 160 & 53 & 72 & 3.7 & 3.4 & 4.3 & 85 \end{bmatrix} \end{matrix} \quad (32)$$

$$\mathbf{K}(\theta) = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \\ C_{10} \\ C_{11} \\ C_{12} \\ C_{13} \\ C_{14} \\ C_{15} \\ C_{16} \\ C_{17} \end{matrix} & \begin{bmatrix} 0.7610 & 0.6456 & 0.6005 & 0.0722 & 0.6689 & 0.6518 & 0.6988 & 0.1152 & 0.3728 & 0.3624 & 0.1885 & 0.3273 & 0.3528 & 0.0836 & 0.0593 & 0.0981 & 0.1024 \\ 0.6456 & 0.8905 & 0.4994 & 0.0281 & 0.8716 & 0.7718 & 0.8635 & 0.0579 & 0.1022 & 0.2123 & 0.5406 & 0.6069 & 0.6062 & 0.0572 & 0.0411 & 0.0746 & 0.0760 \\ 0.6005 & 0.4994 & 0.8352 & 0.2913 & 0.4792 & 0.5196 & 0.4764 & 0.4315 & 0.5934 & 0.5300 & 0.1100 & 0.1676 & 0.2137 & 0.2570 & 0.1837 & 0.3016 & 0.2678 \\ 0.0722 & 0.0281 & 0.2913 & 0.8523 & 0.0403 & 0.0899 & 0.0402 & 0.7553 & 0.0688 & 0.1874 & 0.0941 & 0.1048 & 0.1066 & 0.7690 & 0.7849 & 0.7521 & 0.7520 \\ 0.6689 & 0.8716 & 0.4792 & 0.0403 & 0.8730 & 0.7741 & 0.8602 & 0.0684 & 0.0588 & 0.1916 & 0.5275 & 0.6167 & 0.6093 & 0.0889 & 0.0650 & 0.1098 & 0.1148 \\ 0.6518 & 0.7718 & 0.5196 & 0.0899 & 0.7741 & 0.8063 & 0.7533 & 0.0863 & 0.1126 & 0.2059 & 0.3050 & 0.3777 & 0.4096 & 0.0528 & 0.0493 & 0.0528 & 0.0572 \\ 0.6988 & 0.8635 & 0.4764 & 0.0402 & 0.8602 & 0.7533 & 0.9492 & 0.0455 & 0.1989 & 0.1959 & 0.5019 & 0.6660 & 0.6169 & 0.0964 & 0.0783 & 0.1028 & 0.1371 \\ 0.1152 & 0.0579 & 0.4315 & 0.7553 & 0.0684 & 0.0863 & 0.0455 & 0.8060 & 0.2495 & 0.3144 & 0.0877 & 0.1262 & 0.1472 & 0.7398 & 0.7042 & 0.7632 & 0.7424 \\ 0.3728 & 0.1022 & 0.5934 & 0.0688 & 0.0588 & 0.1126 & 0.1989 & 0.2495 & 0.8901 & 0.6005 & 0.0252 & 0.0187 & 0.0211 & 0.1409 & 0.1061 & 0.1684 & 0.1092 \\ 0.3624 & 0.2123 & 0.5300 & 0.1874 & 0.1916 & 0.2059 & 0.1959 & 0.3144 & 0.6005 & 0.7484 & 0.1268 & 0.0447 & 0.0500 & 0.1628 & 0.1087 & 0.2057 & 0.1720 \\ 0.1885 & 0.5406 & 0.1100 & 0.0941 & 0.5275 & 0.3050 & 0.5019 & 0.0877 & 0.0252 & 0.1268 & 0.7809 & 0.5442 & 0.4851 & 0.2404 & 0.2154 & 0.2709 & 0.2783 \\ 0.3273 & 0.6069 & 0.1676 & 0.1048 & 0.6167 & 0.3777 & 0.6660 & 0.1262 & 0.0187 & 0.0447 & 0.5442 & 0.7845 & 0.7665 & 0.2940 & 0.2388 & 0.3387 & 0.3693 \\ 0.3528 & 0.6062 & 0.2137 & 0.1066 & 0.6093 & 0.4096 & 0.6169 & 0.1472 & 0.0211 & 0.0500 & 0.4851 & 0.7665 & 0.7765 & 0.2708 & 0.2234 & 0.3128 & 0.3303 \\ 0.0836 & 0.0572 & 0.2570 & 0.7690 & 0.0889 & 0.0528 & 0.0964 & 0.7398 & 0.1409 & 0.1628 & 0.2404 & 0.2940 & 0.2708 & 0.8921 & 0.8678 & 0.8986 & 0.9066 \\ 0.0593 & 0.0411 & 0.1837 & 0.7849 & 0.0650 & 0.0493 & 0.0783 & 0.7042 & 0.1061 & 0.1087 & 0.2154 & 0.2388 & 0.2234 & 0.8678 & 0.8628 & 0.8634 & 0.8735 \\ 0.0981 & 0.0746 & 0.3016 & 0.7521 & 0.1098 & 0.0528 & 0.1028 & 0.7632 & 0.1684 & 0.2057 & 0.2709 & 0.3387 & 0.3128 & 0.8986 & 0.8634 & 0.9141 & 0.9168 \\ 0.1024 & 0.0760 & 0.2678 & 0.7520 & 0.1148 & 0.0572 & 0.1371 & 0.7424 & 0.1092 & 0.1720 & 0.2783 & 0.3693 & 0.3303 & 0.9066 & 0.8735 & 0.9168 & 0.9263 \end{bmatrix} \end{matrix} \quad (33)$$

$$\mathbf{K}(\bar{\theta}) = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \\ C_{10} \\ C_{11} \\ C_{12} \\ C_{13} \\ C_{14} \\ C_{15} \\ C_{16} \\ C_{17} \end{matrix} & \begin{bmatrix} 0.0073 & 0.0559 & 0.0413 & 0.5165 & 0.0452 & 0.0334 & 0.0503 & 0.3839 & 0.1473 & 0.0725 & 0.2356 & 0.1333 & 0.1205 & 0.5677 & 0.6001 & 0.5615 & 0.5636 \\ 0.0559 & 0.0232 & 0.2781 & 0.8331 & 0.0262 & 0.0469 & 0.0527 & 0.7364 & 0.2368 & 0.2329 & 0.1259 & 0.1136 & 0.1168 & 0.8256 & 0.8285 & 0.8187 & 0.8251 \\ 0.0413 & 0.2781 & 0.0199 & 0.4166 & 0.2692 & 0.1566 & 0.3470 & 0.2421 & 0.0132 & 0.0353 & 0.5231 & 0.4722 & 0.4097 & 0.5242 & 0.5658 & 0.5041 & 0.5527 \\ 0.5165 & 0.8331 & 0.4166 & 0.0164 & 0.7815 & 0.6067 & 0.8499 & 0.0324 & 0.3318 & 0.2199 & 0.5259 & 0.5691 & 0.5472 & 0.0623 & 0.0396 & 0.0880 & 0.0972 \\ 0.0452 & 0.0262 & 0.2692 & 0.7815 & 0.0222 & 0.0407 & 0.0434 & 0.6918 & 0.3119 & 0.2453 & 0.1233 & 0.0963 & 0.1065 & 0.7609 & 0.7695 & 0.7530 & 0.7560 \\ 0.0334 & 0.0469 & 0.1566 & 0.6067 & 0.0407 & 0.0153 & 0.0724 & 0.5777 & 0.3117 & 0.2007 & 0.2177 & 0.1845 & 0.1727 & 0.7491 & 0.7309 & 0.7676 & 0.7671 \\ 0.0503 & 0.0527 & 0.3470 & 0.8499 & 0.0434 & 0.0724 & 0.0175 & 0.8074 & 0.1993 & 0.2931 & 0.1976 & 0.1074 & 0.1416 & 0.8036 & 0.8032 & 0.8152 & 0.7871 \\ 0.3839 & 0.7364 & 0.2421 & 0.0324 & 0.6918 & 0.5777 & 0.8074 & 0.0166 & 0.0819 & 0.1206 & 0.5366 & 0.4927 & 0.4454 & 0.0566 & 0.0559 & 0.0569 & 0.0753 \\ 0.1473 & 0.2368 & 0.0132 & 0.3318 & 0.3119 & 0.3117 & 0.1993 & 0.0819 & 0.0004 & 0.0085 & 0.4128 & 0.4864 & 0.4643 & 0.2182 & 0.2972 & 0.1824 & 0.2371 \\ 0.0725 & 0.2329 & 0.0333 & 0.2199 & 0.2453 & 0.2007 & 0.2931 & 0.1206 & 0.0085 & 0.0045 & 0.2490 & 0.4253 & 0.4127 & 0.2898 & 0.3474 & 0.2535 & 0.3006 \\ 0.2356 & 0.1259 & 0.5231 & 0.5259 & 0.1233 & 0.2177 & 0.1976 & 0.5366 & 0.4128 & 0.2490 & 0.0096 & 0.0790 & 0.1087 & 0.4037 & 0.3761 & 0.4053 & 0.4043 \\ 0.1333 & 0.1136 & 0.4722 & 0.5691 & 0.0963 & 0.1845 & 0.1074 & 0.4927 & 0.4864 & 0.4253 & 0.0790 & 0.0157 & 0.0179 & 0.3934 & 0.4018 & 0.3820 & 0.3633 \\ 0.1205 & 0.1168 & 0.4097 & 0.5472 & 0.1065 & 0.1727 & 0.1416 & 0.4454 & 0.4643 & 0.4127 & 0.1087 & 0.0179 & 0.0163 & 0.4209 & 0.4233 & 0.4092 & 0.4013 \\ 0.5677 & 0.8256 & 0.5242 & 0.0623 & 0.7609 & 0.7491 & 0.8036 & 0.0566 & 0.2182 & 0.2898 & 0.4037 & 0.3934 & 0.4209 & 0.0247 & 0.0244 & 0.0282 & 0.0271 \\ 0.6001 & 0.8285 & 0.5658 & 0.0396 & 0.7695 & 0.7309 & 0.8032 & 0.0559 & 0.2972 & 0.3474 & 0.3761 & 0.4018 & 0.4233 & 0.0244 & 0.0181 & 0.0339 & 0.0332 \\ 0.5615 & 0.8187 & 0.5041 & 0.0880 & 0.7530 & 0.7676 & 0.8152 & 0.0569 & 0.1824 & 0.2535 & 0.4053 & 0.3820 & 0.4092 & 0.0282 & 0.0339 & 0.0257 & 0.0272 \\ 0.5636 & 0.8251 & 0.5527 & 0.0972 & 0.7560 & 0.7671 & 0.7871 & 0.0753 & 0.2371 & 0.3006 & 0.4043 & 0.3633 & 0.4013 & 0.0271 & 0.0332 & 0.0272 & 0.0240 \end{bmatrix} \end{matrix} \quad (34)$$

$$\mathbf{D}(\theta) = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} \\ \begin{matrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \\ C_9 \\ C_{10} \\ C_{11} \\ C_{12} \\ C_{13} \\ C_{14} \\ C_{15} \\ C_{16} \\ C_{17} \end{matrix} & \begin{bmatrix} 0.1401 & 0.2225 & 0.2434 & 0.7318 & 0.2054 & 0.2114 & 0.1901 & 0.6506 & 0.4113 & 0.3907 & 0.5493 & 0.4320 & 0.4128 & 0.7489 & 0.7766 & 0.7383 & 0.7369 \\ 0.2225 & 0.0709 & 0.3946 & 0.9034 & **0.0827** & 0.1471 & **0.0977** & 0.8414 & 0.5985 & 0.5349 & 0.3081 & 0.2659 & 0.2675 & 0.8848 & 0.8945 & 0.8726 & 0.8750 \\ 0.2434 & 0.3946 & 0.1014 & 0.5689 & 0.4016 & 0.3319 & 0.4383 & 0.4161 & 0.2387 & 0.2821 & 0.7145 & 0.6605 & 0.6078 & 0.6368 & 0.6948 & 0.6039 & 0.6445 \\ 0.7318 & 0.9034 & 0.5689 & 0.0904 & 0.8721 & 0.7634 & 0.9054 & 0.1515 & 0.6548 & 0.5438 & 0.7242 & 0.7382 & 0.7272 & 0.1545 & 0.1370 & 0.1742 & 0.1780 \\ 0.2054 & **0.0827** & 0.4016 & 0.8721 & 0.0805 & 0.1436 & 0.0958 & 0.8146 & 0.6524 & 0.5514 & 0.3145 & 0.4537 & 0.2618 & 0.8372 & 0.8536 & 0.8225 & 0.8214 \\ 0.2114 & 0.1471 & 0.3319 & 0.7634 & 0.1436 & 0.1165 & 0.1673 & 0.7520 & 0.6222 & 0.5261 & 0.4767 & 0.4227 & 0.4001 & 0.8501 & 0.8432 & 0.8589 & 0.8565 \\ 0.1901 & **0.0977** & 0.4383 & 0.9054 & 0.0958 & 0.1673 & 0.0355 & 0.8820 & 0.5295 & 0.5681 & 0.3585 & 0.2302 & 0.2715 & 0.8541 & 0.8632 & 0.8565 & 0.8253 \\ 0.6506 & 0.8414 & 0.4161 & 0.1515 & 0.8146 & 0.7520 & 0.8820 & 0.1171 & 0.4588 & 0.4349 & 0.7325 & 0.6920 & 0.6597 & 0.1689 & 0.1890 & 0.1558 & 0.1947 \\ 0.4113 & 0.5985 & 0.2387 & 0.6548 & 0.6524 & 0.6222 & 0.5295 & 0.4588 & 0.0636 & 0.2331 & 0.7125 & 0.7476 & 0.7367 & 0.5695 & 0.6200 & 0.5405 & 0.5947 \\ 0.3907 & 0.5349 & 0.2821 & 0.5438 & 0.5514 & 0.5261 & 0.5681 & 0.4349 & 0.2331 & 0.1466 & 0.5893 & 0.7070 & 0.6988 & 0.5852 & 0.6389 & 0.5466 & 0.5845 \\ 0.5493 & 0.3081 & 0.7145 & 0.7242 & 0.3145 & 0.4767 & 0.3585 & 0.7325 & 0.7125 & 0.5893 & 0.1294 & 0.2887 & 0.3331 & 0.5907 & 0.5922 & 0.5748 & 0.5704 \\ 0.4320 & 0.2659 & 0.6605 & 0.7382 & 0.2537 & 0.4227 & 0.2302 & 0.6920 & 0.7476 & 0.7070 & 0.2887 & 0.1292 & **0.1403** & 0.5571 & 0.5907 & 0.5278 & 0.5030 \\ 0.4128 & 0.2675 & 0.6078 & 0.7272 & 0.2618 & 0.4001 & 0.2715 & 0.6597 & 0.7367 & 0.6988 & 0.3331 & **0.1403** & 0.1340 & 0.5819 & 0.6086 & 0.5541 & 0.5411 \\ 0.7489 & 0.8848 & 0.6368 & 0.1545 & 0.8372 & 0.8501 & 0.8541 & 0.1689 & 0.5695 & 0.5852 & 0.5907 & 0.5571 & 0.5819 & 0.0705 & **0.0842** & **0.0882** & **0.0832** \\ 0.7766 & 0.8945 & 0.6948 & 0.1370 & 0.8536 & 0.8432 & 0.8632 & 0.1890 & 0.6200 & 0.6389 & 0.5922 & 0.5907 & 0.6086 & **0.0842** & 0.0849 & **0.0902** & **0.0642** \\ 0.7383 & 0.8726 & 0.6039 & 0.1742 & 0.8225 & 0.8589 & 0.8565 & 0.1558 & 0.5405 & 0.5466 & 0.5748 & 0.5278 & 0.5541 & **0.0682** & **0.0902** & 0.0584 & **0.0575** \\ 0.7369 & 0.8750 & 0.6445 & 0.1780 & 0.8214 & 0.8565 & 0.8253 & 0.1746 & 0.5947 & 0.5845 & 0.5704 & 0.5030 & 0.5411 & **0.0682** & **0.0842** & **0.0575** & 0.0509 \end{bmatrix} \end{matrix} \quad (35)$$

$$\mathbf{S} \triangleq \begin{matrix} & C_1 & C_3 & C_4 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{16} \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \\ A_8 \\ A_9 \\ A_{10} \end{matrix} & \begin{bmatrix} 7990 & 1523 & 1733 & 1200 & 50 & 969 & 998 & 14.2 & 158 & 55 & 6.5 \\ 10990 & 1448 & 1732 & 1146 & 45 & 1138 & 898 & 12.3 & 178 & 56 & 6.3 \\ 9790 & 1530 & 1600 & 1053 & 35 & 815 & 998 & 13.9 & 155 & 50 & 4.5 \\ 10350 & 1524 & 1774 & 1029 & 42 & 1063 & 1198 & 14.6 & 164 & 51 & 6.3 \\ 10990 & 1505 & 1665 & 910 & 35 & 865 & 999 & 16.7 & 172 & 52 & 5.3 \\ 11000 & 1485 & 1595 & 1010 & 35 & 860 & 998 & 14.3 & 161 & 49 & 5.6 \\ 11050 & 1500 & 1660 & 1046 & 40 & 1008 & 998 & 14.7 & 156 & 49 & 6.5 \\ 11550 & 1465 & 1615 & 780 & 35 & 840 & 998 & 14 & 160 & 53 & 4.3 \\ 11590 & 1460 & 1615 & 812 & 35 & 915 & 998 & 13.8 & 160 & 51 & 4.9 \\ 11950 & 1460 & 1615 & 780 & 35 & 840 & 998 & 12.6 & 160 & 53 & 4.3 \end{bmatrix} \end{matrix} \quad (36)$$

$imp(C_9) = 2$  and  $imp(C_3) = imp(C_4) = 1$ , which means that the price of the car and its urban consumption are the most important criteria for us, and its height and its width are the least important ones. From these importance values and after normalization, we get the following vector of relative weights of criteria

$$\mathbf{w} = \left[ \frac{5}{33} \quad \frac{1}{33} \quad \frac{1}{33} \quad \frac{4}{33} \quad \frac{4}{33} \quad \frac{2}{33} \quad \frac{2}{33} \quad \frac{3}{33} \quad \frac{3}{33} \quad \frac{3}{33} \quad \frac{5}{33} \right]$$

When using different BF-TOPSIS methods [3], [8], we will obtain the following preference orders

- with BF-TOPSIS1 method:  $A_2 \succ A_1 \succ A_4 \succ A_7 \succ A_5 \succ A_6 \succ A_{10} \succ A_9 \succ A_8 \succ A_3$
- with BF-TOPSIS2 method:  $A_2 \succ A_1 \succ A_4 \succ A_7 \succ A_5 \succ A_6 \succ A_{10} \succ A_9 \succ A_8 \succ A_3$
- with BF-TOPSIS3 method:  $A_2 \succ A_1 \succ A_4 \succ A_7 \succ A_5 \succ A_{10} \succ A_9 \succ A_6 \succ A_8 \succ A_3$
- with BF-TOPSIS4 method:  $A_2 \succ A_1 \succ A_4 \succ A_7 \succ A_5 \succ A_{10} \succ A_9 \succ A_6 \succ A_8 \succ A_3$

When using classical AHP method [4], we obtain the following preference order<sup>19</sup>.

$$A_2 \succ A_1 \succ A_4 \succ A_7 \succ A_5 \succ A_6 \succ A_9 \succ A_8 \succ A_3 \succ A_{10}$$

From the results of the BF-TOPSIS methods and AHP (with double normalization of score matrix), one sees that  $A_2$  car (RENAULT CLIO TCe 75) will be the best car to buy, and the car  $A_1$  (DACIA SANDERO SCe 75) will be the second best car to buy, whereas  $A_3$  (SUZUKI CELERIO 1.0 VVT Avantage) will be the worst one according to BF-TOPSIS or  $A_{10}$  according to AHP. Because the AHP and BF-TOPSIS methods are based on very different principles it is not surprising that preference order can change in the results of the methods, but what is most important from decision-making standpoint is the stability of the order of first best solutions. In this example, the car  $A_2$  is always the best car selection to make with BF-TOPSIS or with AHP method based on the chosen criteria involved in this MCDM problem and their importance weights.

## VII. CONCLUSION

In this paper we have presented a new method called BF-ICrA which helps to simplify (when it is possible) Multi-Criteria Decision-Making problems based on inter-criteria analysis and belief functions. This method is in the spirit of Atanassov's method but proposes a better construction of Inter-Criteria Matrix that fully exploits all information of the score matrix, and the closeness measure of agreement between criteria based on belief interval distance. In fact, BF-ICrA proposes a more precise and refined method for estimating the degree of agreement and disagreement between criteria which use the whole information available in the

<sup>19</sup>Here we did apply a two steps normalization of the score matrix. At first we normalize  $\mathbf{S}$  according to (2) and in a second step each column is renormalized by dividing each element of the column by the sum of its elements. If we apply only first normalization step we obtain with AHP the preference order  $A_2 \succ A_4 \succ A_1 \succ A_7 \succ A_5 \succ A_6 \succ A_9 \succ A_8 \succ A_{10} \succ A_3$ .

data. This BF-ICrA approach could, in theory, also deal with imprecise or missing score values using the technique presented in [8]. We have shown two concrete applications of BF-ICrA method. The first one related with the GPS surveying problem has been addressed in order to overcome the potential inconsistencies of the results generated by the classical ICrA method. Instances containing from 38 to 443 sessions have been solved using MMAS algorithm and we did compare the performance of ACO algorithms applied to eight GPS networks. Our results shows that ACO can provide fast near-optimal solution for observing GPS networks, and could help to improve the services based on GPS networks. From this new Inter-Criteria Analysis we are able to identify some relations and dependences between the considered eight GSPs and MMAS algorithm performance. In our second application, we have shown how a typical (no so simple) multi-criteria car selection problem can be addressed and solved by this BF-ICrA method coupled with BF-TOPSIS methods. This shows the usefulness and potential of this new technique to solve MCDM problems.

## ACKNOWLEDGEMENTS

This paper is partially supported by the National Scientific Fund of Bulgaria under grant DFNI DN12/5 and by Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program and co-financed by the European Union through the European structural and Investment funds.

## REFERENCES

- [1] J. Barzilai, B. Golany, *AHP rank reversal, normalization and aggregation rules*, INFOR Vol. 32, No. 2, pp. 57–63, 1994.
- [2] D. Pavlicic, *Normalization affects the results of MADM methods*, Yugoslav J. of Operations Research, Vol. 11, No. 2, pp. 251–265, 2001.
- [3] J. Dezert, D.Han, H. Yin, *A New Belief Function Based Approach for Multi-Criteria Decision-Making Support*, Proc. of Fusion 2016 Conf.
- [4] T. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, 1980.
- [5] X. Wang, E. Triantaphyllou, *Ranking irregularities when evaluating alternatives by using some ELECTRE methods*, Omega, Vol. 36(1), 2008.
- [6] C.L. Hwang, K. Yoon, *Multiple Attribute Decision Making*, in Lecture Notes in Economics and Math. Syst. 186, Springer-Verlag, Berlin, 1981.
- [7] Y.J. Lai, T.Y. Liu, C.L. Hwang, C.L., *TOPSIS for MODM*, European Journal of Operational Research, Vol. 76, No. 3, pp. 486–500, 1994.
- [8] J. Dezert, D. Han, J.-M. Tacnet, *Multi-Criteria Decision-Making with Imprecise Scores and BF-TOPSIS*, Proc. of Fusion 2017 Conf.
- [9] K. Atanassov, D. Mavrov, V. Atanassova, *Intercriteria Decision Making. A New Approach for Multicriteria Decision Making, Based on Index Matrices and Intuitionistic Fuzzy Sets*, Issues in IFS and Generalized Nets, Vol. 11, pp. 1–8, 2014.
- [10] K. Atanassov, V. Atanassova, G. Gluhchev, *InterCriteria Analysis: Ideas and problems*, Notes on IFS, Vol. 21, No. 1, pp. 81–88, 2015.
- [11] K. Atanassov et al., *An approach to a constructive simplification of multiagent multicriteria decision making problems via intercriteria analysis*, C.R. de l'Acad. Bulgare des Sci., Vol. 70, No. 8, 2017.
- [12] K. Atanassov, *On Intuitionistic Fuzzy Sets Theory*, Springer, 2012.
- [13] S. Todorova et al., *Blood Plasma Thermograms Dataset Analysis by Means of InterCriteria and Correlation Analyses for the Case of Colorectal Cancer*, Int. J. of BIO Automation, Vol. 20(1), pp 115–124, 2016.
- [14] S. Krumova, et al., *Intercriteria analysis of calorimetric data of blood serum proteome*, in Bioch. et Biophys. Acta, Gen. Subjects, 1861, 2017.
- [15] B. Zaharieva et al. *InterCriteria decision making approach for Behtereve's disease analysis*, in Int. J. of Bioautom, Vol. 22(2), 2018.
- [16] T. Pencheva et al., *InterCriteria Analysis of Genetic Algorithm Parameters in Parameter Identification*, Notes on IFS, Vol. 21(2), 2015.

- [17] S. Sotirov, et al., Application of the Intuitionistic Fuzzy InterCriteria Analysis Method to a Neural Network Preprocessing Procedure, Proc of 9th EUSFLAT, pp. 1559–1564, Gijon, Spain, June 30–July 7th, 2015.
- [18] O. Roeva et al., InterCriteria analysis of a model parameters identification using genetic algorithm, Proc. of the Federated Conf. on Computer Science and Information Systems 5, pp. 501–506, 2015.
- [19] M. Angelova, O. Roeva, T. Pencheva, InterCriteria analysis of crossover and mutation rates relations in simple genetic algorithm, Proc. of Conf on Computer Sci. and Inf. Syst, Vol. 5, pp. 419–424, 2015.
- [20] O. Roeva, S. Fidanova, M. Paprzycki, InterCriteria analysis of ACO and GA hybrid algorithms, Stud. Comput. Intell. 610, pp. 107–126, 2016.
- [21] O. Roeva et al., InterCriteria Analysis of ACO Performance for Workforce Planning Problem, Studies in Comp. Intell. 795, Springer 2019.
- [22] V. Atanassova et al., Discussion on the threshold values in the InterCriteria decision making approach. Notes on Intuitionistic Fuzzy Sets 20(2), pp. 94–99, 2014.
- [23] L. Doukowska, V. Atanassova, InterCriteria Analysis approach in radar detection threshold analysis, Notes on IFS, Vol. 21(4), 2015.
- [24] L. Doukowska et al., InterCriteria Analysis Applied to EU Micro, Small, Medium and Large Enterprises, Proc. 5th Int. Symp. on BMSD, 2015.
- [25] V. Bureva et al., Application of the InterCriteria decision making method to Bulgarian universities ranking, Int. Workshop on IFSs, Bulgaria, 2015.
- [26] V. Bureva, E. Sotirova, H. Panayotov, The InterCriteria decision making method to Bulgarian university ranking system, Annual of Informatics Section, Vol. 8, pp. 54–70, 2015–2016.
- [27] M. Krawczak et al., Application of the InterCriteria Decision Making Method to Universities Ranking, Adv. in Intell. Syst. and Comp., Vol 401, pp. 365–372, Springer, 2016.
- [28] G. Shafer, A Mathematical Theory of Evidence, Princeton Press, 1976.
- [29] A. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. of Math. Stat., Vol. 38, pp. 325–339, 1967.
- [30] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, American Research Press, Vol. 1–4, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>
- [31] K. Atanassov, Index Matrices: Towards an Augmented Matrix Calculus, Springer, Cham, 2014.
- [32] K. Atanassov et al., InterCriteria analysis over normalized data, Proc. of 8th IEEE Int. Conf. on Intelligent Syst., 2016.
- [33] N. Ikonov, P. Vassilev, O. Roeva, ICRAData software for InterCriteria analysis, Int. J. BioAutomation, Vol. 22, No. 2, 2018.
- [34] V. Atanassova, Interpretation in the Intuitionistic Fuzzy Triangle of the Results Obtained by the InterCriteria Analysis, Proc. of 16th World Congr. of IFSA, Atlantis Press, 2015.
- [35] V. Atanassova et al., Traversing and ranking of elements of an intuitionistic fuzzy set in the intuitionistic fuzzy interpretation triangle, Adv. in Intell. Syst. and Comp., Vol. 401, pp. 161–174, 2016.
- [36] D. Han, J. Dezert, Y. Yang, New Distance Measures of Evidence based on Belief Intervals, Proc. of Belief 2014, Oxford, UK, Sept. 2014.
- [37] A.-L. Jousselme, D. Grenier, É. Bossé, A new distance between two bodies of evidence, Information Fusion, Vol. 2, No. 2, pp. 91–101, 2001.
- [38] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, Simplification of Multi-Criteria Decision-Making Using Inter-Criteria Analysis and Belief Functions, Proc. of Fusion 2019 Int. Conf, Ottawa, Canada, July 2-5, 2019.
- [39] S. Fidanova, J. Dezert, A. Tchamova, Inter-Criteria Analysis Based on Belief Functions for GPS Surveying Problems, Proc. of Int. Symposium on INnovations in Intelligent SysTems and Applications (INISTA), Sofia, Bulgaria, July 3–5, 2019.
- [40] T. Stutzle, H.H. Hoos, MAX-MIN Ant System, In: Dorigo, M., Stutzle, T., Di Caro, G. (eds.) Future Generation Computer Systems, Vol. 16, pp. 889–914, 2000.
- [41] M. Dorigo, L.M. Gambardella, Ant colony system: a cooperative learning approach to the traveling salesman problem, IEEE Trans. Evol. Comput. 1, pp. 5–66, 1997.
- [42] F. Smarandache, J. Dezert, J.-M. Tacnet, Fusion of sources of evidence with different importances and reliabilities, Proc. of Fusion Conf 2010.
- [43] S. Fidanova, V. Atanassova, O. Roeva, Ant Colony Optimization Application to GPS Surveying Problems: InterCriteria Analysis. In: Atanassov K. et al. (eds) Uncertainty and Imprecision in Decision Making and Decision Support: Cross-Fertilization, New Models and Applications. IWIFSGN 2016. Advances in Intelligent Systems and Computing, Vol 559. Springer, Cham, 2018.

# DSmT-Based Fusion Strategy for Human Activity Recognition in Body Sensor Networks

Yilin Dong<sup>a</sup>, Xinde Li<sup>a, b</sup>, Jean Dezert<sup>c</sup>,  
 Mohammad Omar Khyam<sup>d</sup>, Md. Noor-A-Rahim<sup>e</sup>, Shuzhi Sam Ge<sup>f</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing, China.

<sup>b</sup>The School of Cyber Science and Engineering, Southeast University, Nanjing, China.

<sup>c</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>d</sup>School of Electrical Engineering, Central Queensland University, Melbourne, Australia.

<sup>e</sup>School of Computer Science and IT, University College Cork, Ireland.

<sup>f</sup>Social Robotics Laboratory, ECE Dept., Interactive Digital Media Institute, National Univ. of Singapore, Singapore.

Emails: yldong@shmtu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr,  
 m.khyam@cqu.edu.au, m.rahim@cs.ucc.ie, samge@nus.edu.sg

Originally published as: Y. Dong, X. Li, J. Dezert, M.O. Khyam, Md. Noor-A-Rahim, S. Ge, *Dezert-Smarandache Theory-Based Fusion for Human Activity Recognition in Body Sensor Networks*, IEEE Trans. on Industrial Informatics, Vol. 16(11), pp. 7138–7149, November 2020, and reprinted with permission.

**Abstract**—Multi-sensor fusion strategies have been widely applied in Human Activity Recognition (HAR) in Body Sensor Networks (BSNs). However, the sensory data collected by BSNs systems are often uncertain or even incomplete. Thus, designing a robust and intelligent sensor fusion strategy is necessary for high-quality activity recognition. In this paper, Dezert-Smarandache Theory (DSmT) is used to develop a novel sensor fusion strategy for HAR in BSNs, which can effectively improve the accuracy of recognition. Specifically, in the training stage, the Kernel Density Estimation (KDE) based models are first built and then precisely selected for each specific activity according to the proposed discriminative functions. After that, a structure of Basic Belief Assignment (BBA) can be constructed, using the relationship between the test data of unknown class and the selected KDE models of all considered types of activities. In order to deal with the conflict between the obtained BBAs, Proportional Conflict Redistribution-6 (PCR6) is applied to fuse the acquired BBAs. Moreover, the missing data of the involved sensors are addressed as ignorance in the framework of the DSmT without manual interpolation or intervention. Experimental studies on two real-world activity recognition datasets (The OPPORTUNITY dataset; Daily and Sports Activity Dataset (DSAD)) were conducted, and the results showed the superiority of our proposed method over some state-of-the-art approaches proposed in the literature.

**Keywords:** HAR, Multi-sensor fusion, Belief function theory, KDE, DSmT.

## I. INTRODUCTION

Human Activity Recognition (HAR) has spawned intense researches in the past decades and continues to be an active research area [1], [2], [3], [4]. These HAR systems have enabled several practical applications, such as health monitoring [5], physical activity [6] and gesture detection. Recently, multi-sensor fusion for activity recognition is playing an increasing role in HAR field and many strategies have been proposed (see [7] for more references). Generally speaking, multi-sensor fusion strategies can be mainly categorized into three level categories depending on the abstraction level used for data

processing: data-fusion level [8], feature-fusion level [9] and decision-level fusion [10]. Among all these three fusion levels, decision-level fusion output is a unique decision obtained from local decision of multiple (homogeneous or heterogeneous) sensors. The fusion in this level has many advantages: communication bandwidth saving, allowing the combination of the heterogeneous sensors. In this paper, the main topic thus focus on decision-level fusion area. Two most common used approaches for this level of fusion are majority voting [11] and naive bayes [12]. However, complex sensory data, especially when these data are uncertain or even incomplete, make these two methods unsuitable for HAR. Two classical scenarios are described as follows:

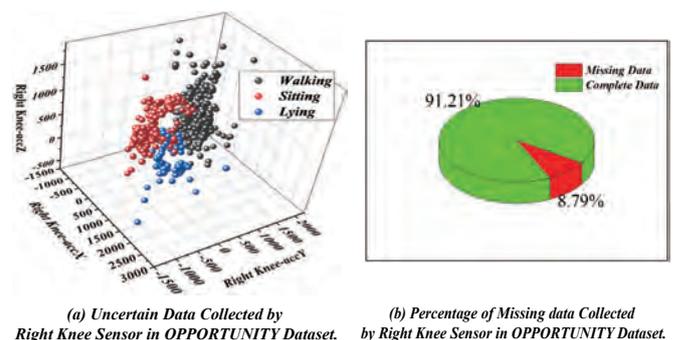


Figure 1. Uncertain and incomplete sensory data in OPPORTUNITY dataset.

- 1) **Uncertain sensory data in HAR problem.** In order to intuitively discuss the uncertainty of sensory data, one of the involved sensor in UCI OPPORTUNITY dataset [13], [14] was randomly selected and parts of the original data of three activities derived from the chosen sensor were drawn in Fig.1(a). As we can see from Fig.1(a), some objects that are very close can sometimes truly originate from different classes. Such objects are

really difficult to classify correctly into a particular class using the given information. In this case, we call this data uncertain when it can belong to different specific classes with probability mass assignments to estimate;

- 2) **Incomplete sensory data in HAR problem.** Missing data frequently occur during the measurement of wearable-based activity recognition. As we can see in Fig.1(b), sensory data with incomplete pattern occupy an important proportion which cannot be easily neglected in OPPORTUNITY dataset. The traditional ways to cope with these feature vectors, which include missing data, are to interpolate or delete the whole vector. However, interpolation or deletion is not the wise choice which may bring noise and information loss to the recognition system.

The aforementioned discussions motivate our study, where HAR in Body Sensor Networks (BSNs) is implemented based on belief function theory [15]. Belief function allows to model uncertainty and to fuse Basic Belief Assignments (BBAs) built from sensors' measurements. Within this theory, information fusion relies on the use of a combination rule allowing the pieces of evidences (drawn from sensor readings) expressed in a common frame of discernment to be combined. Among all available combination rules, Dempster's rule proposed by Shafer in Dempster-Shafer theory [15] is the most well-known rule still used in many applications even if it remains very controversial. Recently, Chen et al. [16] proposed a new method based on Dempster-Shafer theory to improve human action recognition by using the fusion of depth camera and inertial sensors. Although the recognition results mentioned in [16] is good, two key issues are ignored by authors: 1) In Dempster-Shafer theory, there exists an assumption that hypotheses considered should be exclusive. However, in HAR, activities to be identified often fail to satisfy the characteristics of mutual exclusion. For example, the intersection between "Walking" and "Running" can be defined as "Standing" or intermediate transition state "Walking to Running" [17]; 2) Dempster's rule cannot solve high conflict issues and even very low conflict issues in specific cases, which have been widely discussed in [18], [19].

To solve those mentioned drawbacks in Dempster-Shafer theory, Dezert and Smarandache proposed Dezert-Smarandache Theory (DSMT) [18] to solve multi-sensor fusion problems, with more reasonable assumptions and better combination rules, which is more appropriate to handle HAR problems. In this paper, a new use of DSMT is proposed to solve HAR issues thanks to a novel decision-level fusion strategy based on DSMT. Such DSMT-based HAR can be used for online activity recognition system because of its higher recognition accuracy and lower recognition delay, which can meet the required response speed in real-time recognition systems (less than 200ms) [2]. Specifically, the main contributions of this work are summarized as follows:

- A novel DSMT-based fusion strategy for HAR in BSNs is proposed;

- Kernel Density Estimation (KDE) models are constructed based on the sensor readings, and those selected KDE models of all considered classes are applied to calculate BBAs in DSMT;
- The missing data in original sensor readings are also modeled by vacuous BBA (i.e. the total ignorance source of evidence) in DSMT without any manual interpolation;
- The efficiency of our fusion system with two activities recognition open datasets is demonstrated.

This paper is organized as follows: Section II provides an inventory of the basic concepts of DSMT. Section III provides a description of the new proposed fusion method. Section IV includes the experimental results and discussions. The final section V contains a brief conclusion.

## II. BASICS OF DSMT

In DSMT framework, the BBAs are defined on the so-called *hyper-power set* (or Dedekind's lattice) denoted  $D^\Theta \triangleq (\Theta, \cup, \cap)$  whose cardinalities follows Dedekind's numbers sequence, see [18], Vol.1 for details and examples. A (generalized) BBA, called a mass function,  $m(\cdot)$  is defined by the mapping:  $D^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{A \in D^\Theta} m(A) = 1$ .

To palliate the drawbacks of Dempster's rule, Martin et.al [20] proposed a very interesting combination rule: PCR6. Due to its good performance, it is widely applied in recent applications. We recall that the PCR6 formula for the combination of two BBAs coincides with PCR5 formula originally developed by Smarandache and Dezert in [18]. The combination of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  by the PCR5 rule is given as follows: for  $m_{PCR5}(\emptyset) = 0$  and  $\forall A \in D^\Theta$

$$m_{PCR6}(A) = m_{PCR5}(A) = m_{12}(A) + \sum_{B \in D^\Theta \setminus \{A\} | A \cap B = \emptyset} \left[ \frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right], \quad (1)$$

where  $m_{12}(A) = \sum_{B, C \in D^\Theta | B \cap C = A} m_1(B) m_2(C)$ .

The combinations of more than two BBAs altogether with PCR5 and with PCR6 fusion rule in general provide different results. The choice of PCR6 with respect to PCR5 was justified at first by Martin and Osswald in [20] from a specific application, and then theoretically by Smarandache and Dezert in [21]. The general formula of PCR6 for combining more than two BBAs was given in details in [20] with examples.

## III. DSMT-BASED FUSION STRATEGY FOR HAR IN BSNs

### A. The Flow Chart of Our Proposed Method

Before entering in the detailed presentation of our DSMT-based fusion strategy, we briefly introduce it through the flowchart of Fig.2 for convenience. Specifically, in the training stage, multiple KDE models are derived from the raw sensor readings so as to build the model pool. Then, the representative model is selected for a particular activity based on our proposed discriminative functions. After that, when the test sample comes, the corresponding BBA is calculated through each activity representative model. Finally, these BBAs are

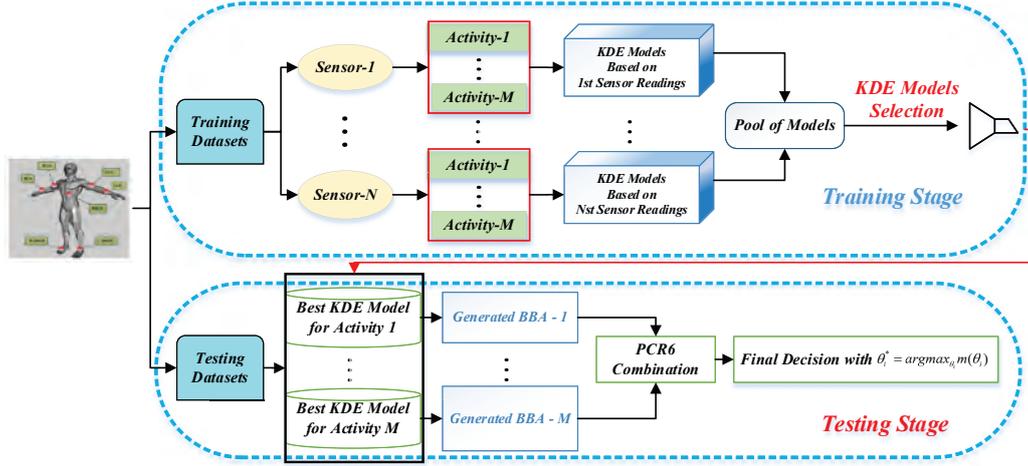


Figure 2. DSMT-Based Fusion Strategy for HAR in BSNs.

combined with PCR6 rule, from which we make the final decisions.

### B. Mathematical Definitions of Daily Activities in DSMT

The goal of our work is to recognize human daily activities thanks to DSMT-based framework. Thus, the basic mathematical definitions of the interested activities need to be given. We assume that the finite frame of discernment considered in our activity recognition problem is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The corresponding hyper-power set of  $\Theta$  is denoted  $D^\Theta$ . Singletons in  $D^\Theta$  are used to represent the simple daily activity such as  $\theta_1 \triangleq \text{Standing}$ ,  $\theta_2 \triangleq \text{Sitting}$ ,  $\theta_3 \triangleq \text{Lying}$  and so on. Disjunctive focal elements in  $D^\Theta$  represent the coarse-grained activities. For example,  $\theta_1 \cup \theta_2 \cup \theta_3 \triangleq \text{Static Activity}$ . Also, if  $\theta_4 \triangleq \text{Walking}$ ,  $\theta_5 \triangleq \text{Running}$ , then  $\theta_4 \cup \theta_5$  is regarded as *Dynamic Activity*. Following the definition line of disjunctive focal elements,  $\theta_1 \cup \theta_2 \dots \cup \theta_n$  represents the whole unknown activity. Besides, the conjunctive focal elements in  $D^\Theta$  can be used to stand for the transition activity like  $\theta_1 \cap \theta_2 \triangleq \text{Standing to Sitting}$  or  $\theta_1 \cap \theta_2 \triangleq \text{Sitting to Standing}$  because  $\theta_1 \cap \theta_2 = \theta_2 \cap \theta_1$  and  $\theta_2 \cap \theta_3 \triangleq \text{Sitting to Lying or Lying to Sitting}$ . In this paper, we only consider a restricted hyper-power set, which is denoted as  $D_{restricted}^\Theta = \{\theta_1, \theta_2, \dots, \theta_n, \theta_1 \cup \theta_2 \dots \cup \theta_n\}$ . In  $D_{restricted}^\Theta$ , only two types of focal elements exist: one is the singleton, which represents the simple activity and another is  $\theta_1 \cup \theta_2 \dots \cup \theta_n$ , which represents the unknown activity. More complicated situations involving less restricted hyper-power sets will be discussed in our future work.

### C. Training Model Stage

In the training stage, the KDE model is employed to fit the sensor readings. The most suitable KDE model to distinguish a certain activity is then selected to be regarded as the specific activity representative model. Among the process of this training stage, two main steps are involved:

1) *Construction of KDE Models*: We assume that there are  $M$  kinds of activities that need to be classified and the original dataset collected from the wearable sensors are denoted as  $\mathbf{x}_{ij}$ ,  $i = 1, \dots, M$  and  $j = 1, \dots, N$ . Here,  $M$  represents the types of activities to be classified and  $N$  is the number of sensors. Thus, based on the Eq.(2), the KDE model of the specific activity is derived from the sensor readings by

$$f_{ij}(\mathbf{x}_{ij}) = \frac{1}{Q} \cdot \sum_{q=1}^Q K_h(x - x_q^{ij}) = \frac{1}{Qh} \cdot \sum_{q=1}^Q K\left(\frac{x - x_q^{ij}}{h}\right). \quad (2)$$

where  $f(\mathbf{x}_{ij})$  is the KDE model of  $\mathbf{x}_{ij}$  which represents the model of the  $j$  sensor for the  $i$  activity;  $K(\cdot)$  is the kernel function which can be 'normal', 'epanechnikov', 'box' and 'triangle';  $h$  is the smoothing parameter (the bandwidth) of the KDE model. In this paper, the value of  $h$  is the adaptive bandwidth selected by the method presented in [22]; The parameter  $Q$  is the dimension of  $\mathbf{x}_{ij}$ .

2) *Selection of the Best Discriminative KDE Model for the Specific Activity*: As we can see from Eq.(2), each activity can have  $N$  KDE models and we need to select the most discriminative KDE model in order to reduce the computational complexity and the interference model. Once the unique KDE model for each activity is selected, one can easily determine a specific sensor to identify activity because there is one-to-one correspondence between the KDE models and the wearable sensors. We propose two novel discriminant evaluation functions as follows:

For the specific activity  $\theta_s$ ,  $s \in \{1, \dots, M\}$ , the value of Sum of Statistical Difference (SSD) of the  $j$ ,  $j = 1, \dots, N$  the KDE model is calculated as follows:

$$\begin{aligned} SSD_{\theta_s}(j) &= [\Psi(f_{\theta_s j}) - \Psi(f_{1j})] + \dots + [\Psi(f_{\theta_s j}) - \Psi(f_{Mj})] \\ &= (M - 1) \cdot \Psi(f_{\theta_s j}) - \sum_{i=1, i \neq s}^M \Psi(f_{\theta_i j}). \end{aligned} \quad (3)$$

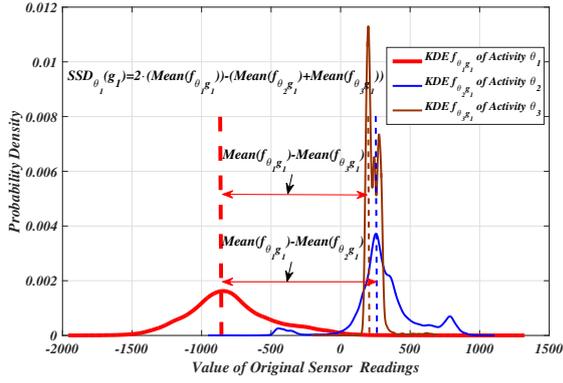
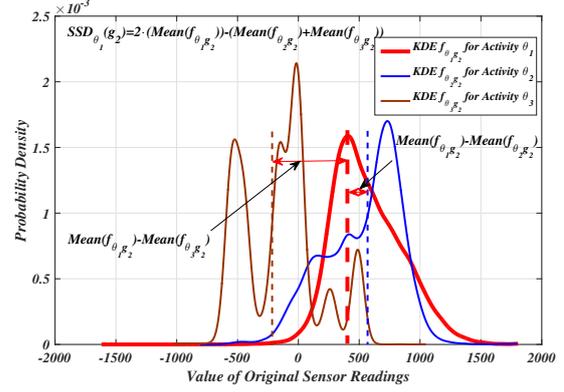

 (a) KDE models of  $g_1$  sensor for three activities  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ .

 (b) KDE models of  $g_2$  sensor for three activities  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ .

Figure 3. Selection of the KDE Model for the Specific Activity Based on the Principle of SSD Function.

In Eq.(3),  $\theta_s$  is one of the specific activity among the  $M$  considered activities;  $j$  is the sensor readings of the  $j$  sensor;  $\Psi(\cdot)$  calculates the statistical characteristic value of the derived distribution of the KDE model  $f_{\theta_s,j}$ . In this paper,  $\Psi(\cdot) = Mean(\cdot)$ , that is the average value of sensor readings.

The principle of selecting KDE model based on SSD is quite simple: for the specific activity  $\theta_s$ , if the SSD value of the  $j$ ,  $j = 1, \dots, N$  sensor is large, it means that this  $j$  KDE model of  $\theta_s$  has a better discriminative ability. Here, a simple illustrative example was extracted from the OPPORTUNITY dataset experiment in Section IV to show the principle of SSD. As we can see in Fig.3, for the specific activity  $\theta_1$ , the value of  $SSD_{\theta_1}(g_1) = (Mean(f_{\theta_1,g_1}) - Mean(f_{\theta_2,g_1})) + (Mean(f_{\theta_1,g_1}) - Mean(f_{\theta_3,g_1}))$  (Fig.3(a)) is larger than  $SSD_{\theta_1}(g_2) = (Mean(f_{\theta_1,g_2}) - Mean(f_{\theta_2,g_2})) + (Mean(f_{\theta_1,g_2}) - Mean(f_{\theta_3,g_2}))$  (Fig.3(b)). Here,  $g_1$  and  $g_2$  represent the  $g_1$  sensor and the  $g_2$  sensor. It can be clearly seen in Fig.3 that KDE model ( $f_{\theta_1,g_1}$ ) has the higher discriminative ability than KDE model ( $f_{\theta_1,g_2}$ ) for activity  $\theta_1$ .

In order to measure the distances between probability density functions of each pair of KDEs models, another well-known choice for such measurement is Kullback-Leibler (KL) divergence defined by, see [23]:

$$Div_{KL}(f_{p_1}||f_{p_2}) = \sum_i f_{p_1}(i) \log \frac{f_{p_1}(i)}{f_{p_2}(i)}. \quad (4)$$

Here  $f_{p_1}$  and  $f_{p_2}$  are two discrete probability density functions. Similar to  $Div_{KL}$ , another well-known divergence is Jensen-Shannon (JS) divergence defined by:

$$Div_{JS}(f_{p_1}||f_{p_2}) = \frac{1}{2} [Div_{KL}(f_{p_1}||f_{p_2}) + Div_{KL}(f_{p_2}||f_{p_1})]. \quad (5)$$

Based on Eq.(4) and Eq.(5), another discriminative evaluation function is given to measure the discriminative ability between different KDE models, which is named as Sum of Divergence Difference (SDD): For the specific activity

$\theta_s, s \in \{1, \dots, M\}$ , the value of SDD of the  $j$ ,  $j = 1, \dots, N$  KDE model is calculated as follows:

$$SDD_{\theta_s}(j) = \sum_{i,i \neq s}^{M-1} \Upsilon(f_{\theta_s,j}, f_{\theta_i,j}). \quad (6)$$

In Eq.(6),  $\theta_s$  is the specific class of daily activity;  $\Upsilon(\cdot)$  represents the divergence function. In this paper,  $\Upsilon(\cdot)$  is defined as  $KL$  (Eq.(4)) or  $JS$  (Eq.(5)). It is worth noting that in order to make the statements more clear in the following sections, we will directly use the  $Mean(\cdot)$  to represent that SSD criterion is applied for selecting KDE models in the process of activity recognition. Similarly,  $Div_{KL}(\cdot)$  or  $Div_{JS}(\cdot)$  mean that SDD is applied and  $Div_{KL}(f_{p_1}||f_{p_2})$  or  $Div_{JS}(f_{p_1}||f_{p_2})$  is used in SDD criterion to measure the difference between two distributions. For each activity  $\theta_1, \theta_2, \dots, \theta_M$ , the best discriminative  $M$  KDE models  $f_{\theta_i}, i = 1, \dots, M$  can be selected and denoted as follows:

$$\begin{bmatrix} f_{\theta_1,g_1} & f_{\theta_2,g_1} & \dots & f_{\theta_M,g_1} \\ f_{\theta_1,g_2} & f_{\theta_2,g_2} & \dots & f_{\theta_M,g_2} \\ \vdots & \vdots & \ddots & \vdots \\ f_{\theta_1,g_M} & f_{\theta_2,g_M} & \dots & f_{\theta_M,g_M} \end{bmatrix} \quad (7)$$

and  $g_1, g_2, \dots, g_M \in [1, N]$ . Each of  $g_i, i \in \{1, \dots, M\}$  represents the selected wearable sensor number.

#### D. Testing Stage

When the test sample becomes available, the corresponding BBA is calculated through each KDE model of each activity. Finally, we combine all related BBAs with PCR6 rule and we make the final decisions from the combined BBAs.

1) *BBAs Calculation*: In this paper, the considered frame of discernment is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_M\}$ . Each focal element in  $\Theta$  represents one kind of activity and here we just consider a simplified  $D_{restricted}^\Theta = \{\theta_1, \theta_2, \dots, \theta_M, \theta_1 \cup \theta_2 \cup \dots \cup \theta_M\}$ . We consider a testing vector  $\mathbf{x}$  with unknown class and we want to identify the label of  $\mathbf{x}$  corresponding to the activity it

belongs to. Next, we use the following equations to calculate the BBAs ( $m_1(\cdot), m_2(\cdot), \dots, m_M(\cdot)$ ):

$$\begin{aligned} m_1(\theta_1) &= f_{\theta_1 g_1}(x(g_1)), \dots, m_1(\theta_M) = f_{\theta_M g_1}(x(g_1)); \\ m_2(\theta_1) &= f_{\theta_1 g_2}(x(g_2)), \dots, m_2(\theta_M) = f_{\theta_M g_2}(x(g_2)); \\ &\vdots \\ m_M(\theta_1) &= f_{\theta_1 g_M}(x(g_M)), \dots, m_M(\theta_M) = f_{\theta_M g_M}(x(g_M)). \end{aligned}$$

It is worth noting that when the value of one feature is missing, we directly assign "1" to  $m(\theta_1 \cup \theta_2 \cup \dots \cup \theta_M)$  which means in this case, we cannot obtain the valuable decision information. Besides, in order to make sure that the derived BBAs satisfy the normalization condition, the following normalization applies:

- If  $m_i(\theta_1) + \dots + m_i(\theta_M) \leq 1$ , then  $m_i(\theta_1 \cup \dots \cup \theta_M) = 1 - (m_i(\theta_1) + \dots + m_i(\theta_M))$ ;
- If  $m_i(\theta_1) + \dots + m_i(\theta_M) > 1$ , then  $m_i(\theta_k) = \frac{m_i(\theta_k)}{\sum_{k=j, \dots, M} m_i(\theta_j)}$  for  $k = 1, \dots, M$ , and  $m_i(\theta_1 \cup \dots \cup \theta_M) = 0$ .

2) *Global Fusion with PCR6 and Decision Making*: After obtaining the  $M$  BBAs, the PCR6 fusion rule is used to fuse all these BBAs which is denoted symbolically by

$$m_{fusion} = PCR6(m_1, m_2, \dots, m_M). \quad (8)$$

Then the final decision of the predicted class of  $\mathbf{x}$  can be made as  $\theta_i^* = \operatorname{argmax}_{\theta_i} m_{fusion}(\theta_i)$ , where  $\theta_i$  is a focal element of the  $D_{restricted}^{\ominus}$  based on the max of belief mass.

The DSMT-Based Activity Recognition technique is described in Algorithm 1 for convenience.

#### IV. PERFORMANCE EVALUATION

##### A. Datasets

The performance of the proposed DSMT-Based HAR was evaluated on the following two open HAR datasets. The first one is **UCI OPPORTUNITY dataset** [13], [14]. The details of this dataset can be found in OPPORTUNITY UCI dataset<sup>1</sup>. Three basic activities were classified: Walking, Sitting and Lying; The other one is **UCI DSAD**<sup>2</sup>. The details of the DSAD can be found in [24]. In this dataset, five common daily activities including Sitting, Standing, Lying, Walking and Running were classified to prove the effectiveness of our proposed method.

##### B. Measures of Performance

As measures of the performance of our activity recognition system, the classical Accuracy, Precision, Recall, and F1-score [7] have been used. They are defined by

$$Accuracy = \frac{1}{n} \sum_{k=1}^n \frac{TP_k + TN_k}{TP_k + TN_k + FP_k + FN_k}, \quad (9)$$

<sup>1</sup><http://archive.ics.uci.edu/ml/datasets/OPPORTUNITY+Activity+Recognition>.

<sup>2</sup><http://archive.ics.uci.edu/ml/datasets/Daily+and+Sports+Activities>.

#### Algorithm 1: DSMT-Based HAR

---

**Input:** Sequential original data  $\mathbf{x}_{ij}, i = 1, \dots, M, j = 1, \dots, N, K = 'Normal'$ .

**Output:** The Predicted Class of Unknown data  $\mathbf{x}^*$ .

1 **Initialize:** Cross Validation ( $\mathbf{x}_{ij}$ )  $\rightarrow$   $\mathbf{x}_{training}, \mathbf{x}_{testing}$ ;

2 **Training Stage:**

3 **for**  $i = 1, \dots, M$  **do**

4     **for**  $j = 1, \dots, N$  **do**

5          $f_{ij}(\mathbf{x}_{ij}) = \frac{1}{Qh} \cdot \sum_{q=1}^Q K(\frac{x-x_{ij}^q}{h})$ ;

6     **end**

7 **end**

8 **for**  $i = 1, \dots, M$  **do**

9     **for**  $j = 1, \dots, N$  **do**

10          $SSD_{\theta_s}(j) = (M-1) \cdot \Psi(f_{\theta_s j}) - \sum_{i=1, i \neq s}^M \Psi(f_{ij})$ ;

11         **or**

12          $SDD_{\theta_s}(j) = \sum_{i, i \neq s}^{M-1} \Upsilon(f_{\theta_s j}, f_{\theta_i j})$ ;

13     **end**

14      $g_i = \max(SSD_{\theta_i})$  or  $g_i = \max(SDD_{\theta_i})$ ;

15 **end**

16  $f_{matrix} = f_{\theta_1 g_1}, \dots, f_{\theta_M g_1}; f_{\theta_1 g_2}, \dots, f_{\theta_M g_2}; \dots; f_{\theta_1 g_M}, \dots, f_{\theta_M g_M}$ ;

17 **Testing Stage:**

18  $D_{restricted}^{\ominus} = \{\theta_1, \theta_2, \dots, \theta_M, \theta_1 \cup \theta_2 \cup \dots \cup \theta_M\}$ ;

19 **for**  $i = 1, \dots, M$  **do**

20      $m_i(\theta_1) = f_{\theta_1 g_i}(x^*(g_i)), \dots, m_i(\theta_M) = f_{\theta_M g_i}(x^*(g_i))$ ;

21 **end**

22 **if**  $m_i(\theta_1) + \dots + m_i(\theta_M) \leq 1$  **then**

23      $m_i(\theta_1 \cup \theta_2 \cup \dots \cup \theta_M) = 1 - (m_i(\theta_1) + \dots + m_i(\theta_M))$ ;

24 **end**

25 **else if**  $m_i(\theta_1) + \dots + m_i(\theta_M) > 1$  **then**

26     Normalization of BBAs  $m_i(\theta_1), \dots, m_i(\theta_M)$ ;

27 **end**

28 **Fusion Step:**  $m_{fusion} = PCR6(m_1(\cdot), \dots, m_M(\cdot))$ ;

29 **Decision Step:** Take as decision the maximum of belief mass of focal elements  $\theta_i^* = \operatorname{argmax}_{\theta_i} m_{fusion}(\theta_i)$ ;

30 **final** ;

31 **return** Predicted Class of  $\mathbf{x}^*$ ;

---

$$Precision = \frac{1}{n} \sum_{k=1}^n \frac{TP_k}{TP_k + FP_k}, \quad (10)$$

$$Recall = \frac{1}{n} \sum_{k=1}^n \frac{TP_k}{TP_k + FN_k}, \quad (11)$$

$$F1 - Score = \frac{1}{n} \sum_{k=1}^n (2 \cdot \frac{precision_k \cdot recall_k}{precision_k + recall_k}), \quad (12)$$

where  $k$  denotes class index and  $n$  is the number of classes. True Positives ( $TP_k$ ): the number of correctly recognized class examples; True Negatives ( $TN_k$ ): the number of correctly recognized examples that do not belong to the class; False Positives ( $FP_k$ ): examples that were either incorrectly assigned to the class; False Negatives ( $FN_k$ ): not recognized as class examples.

##### C. Results on UCI OPPORTUNITY dataset

1) *Effectiveness of the Selection of  $\Psi(\cdot)$  and  $\Upsilon(\cdot)$  in Eq.(3) and Eq.(6)*: The selections of  $\Psi(\cdot)$  in SSD and  $\Upsilon(\cdot)$  in SDD were quite crucial to the representative KDE models for all involved activities. Thus, the relevant comparisons about the

Table I  
THE SELECTED SENSORS IN OPPORTUNITY DATASET BASED ON  $Mean(\cdot)$ ,  $Div_{KL}(\cdot)$ ,  $Div_{JS}(\cdot)$ .

2*Subject (lr)2-4 (lr)5-7 (lr)8-10	$SSD : \Psi(\cdot) = Mean(\cdot)$			$SDD : \Upsilon_1(\cdot) = Div_{KL}(\cdot)$			$SDD : \Upsilon_2(\cdot) = Div_{JS}(\cdot)$		
	Walking	Sitting	Lying	Walking	Sitting	Lying	Walking	Sitting	Lying
Subject 1	LLA-accX	RLA-accX	Back-magX	LLA-magX	RKN-accZ	Back-magZ	LWR-accY	RKN-accZ	LShoe-accZ
Subject 2	LLA-accX	RLA-accX	LShoe-accZ	LLA-magX	HIP-accY	Back-magZ	RKN-accY	RKN-accZ	Back-magX
Subject 3	RH-accY	LLA-magX	RShoe-accZ	Back-magX	Back-magZ	Back-accZ	RKN-accY	Back-magZ	RShoe-accY
Subject 4	LWR-accY	RH-accY	Back-magX	Back-magZ	LUA-accY	Back-accZ	LUA-accY	LUA-accY	Back-accX

\*According to [25], each triaxial (x,y,z) sensor unit has 3-degree of freedom. And in this Table, all the meanings of the involved sensors are: Left Lower Arm (LLA);Right Lower Arm (RLA);Right Knee (RKN);Left Wrist (LWR);Left Shone (LShone);Hips (HIP);Right Hand (RH);Right Shoe (RShoe);Left Upper Arm (LUA);Accelerator x axis (accX);Magnetic Z-axis (magZ). More details about OPPORTUNITY Dataset can be referred to [25].

recognition rates were given in Fig.4 when  $\Psi(\cdot)$  and  $\Upsilon(\cdot)$  were set to (1)  $\Psi(\cdot) = Mean(\cdot)$ , (2)  $\Upsilon_1(\cdot) = Div_{KL}(\cdot)$ , (3)  $\Upsilon_2(\cdot) = Div_{JS}(\cdot)$ , respectively. As we can see in Fig.4, our proposed method based on these three discriminative functions<sup>3</sup> distinguished three mentioned activities in Opportunity dataset (four subjects) very well, which indirectly proved the effectiveness of  $Mean(\cdot)$ ,  $Div_{KL}$ ,  $Div_{JS}$  in measuring the difference between the distributions of activities. Besides, all the three generated models had the highest recognition accuracy on Subject 1. However, the sensors selected by each function were quite different, and the corresponding involved sensors were listed in Table I. It can be found that the sensitivity of sensors to different daily activities varied, and was influenced by their locations of deployment. Sensors located on the arm such as left lower arm, right hand, left wrist were more likely to identify “Walking” but sensors located on the Back or shoes had higher recognition rates of ”Lying” than other sensors. This directly indicates that it is not feasible or wise to rely on a single sensor deployed in a single location to identify various kinds of activities [26]. This is also our motivation to use multi-sensor fusion strategy based on DSMT to solve activity recognition problems.

2) *Recognition Rate versus Training Percentage*: In this experiment, we did modify the percentage of training set and investigated the relationship between the training percentage and the classification accuracy of our proposed method on OPPORTUNITY dataset. It is worth mentioning that the discriminative function chosen here was SSD (Eq.(3)) and  $\Psi(\cdot) = Mean(\cdot)$ . Since our experiments were conducted based on ten-fold cross validation method, it is convenient for us to test the relationship between recognition rate and training percentage. According to the principle of ten-fold cross validation, the original datasets were first randomly divided into ten equal parts. And then, in the first experiment, we first treated 10% data as training dataset and the remaining 90% data were used as testing dataset; And in the second experiment, 20% datasets were used for training and the remaining 80% for testing, and so on, until the last experiment which we used 90% datasets for training and the last 10% datasets for

<sup>3</sup>As we introduced in Definition 1 and Definition 2,  $\Psi(\cdot)$  means that SSD (Eq.(3)) is used to choose the best KDE models and  $\Upsilon_1(\cdot)$ ,  $\Upsilon_2(\cdot)$  means that SDD (Eq.(6)) is applied in our activity recognition model.

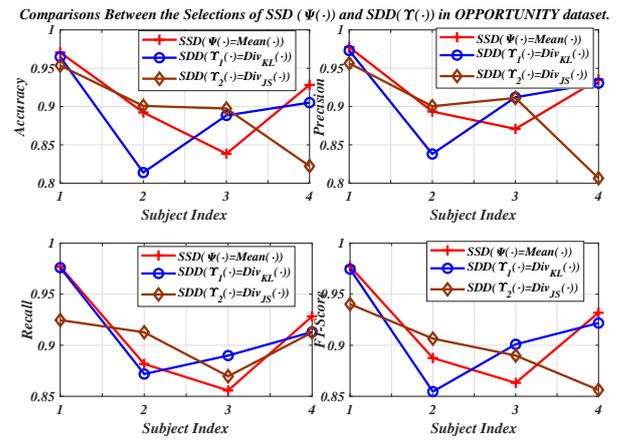


Figure 4. Effectiveness of the selection of  $SSD(\Psi(\cdot))$  and  $SDD(\Upsilon(\cdot))$  in OPPORTUNITY dataset.

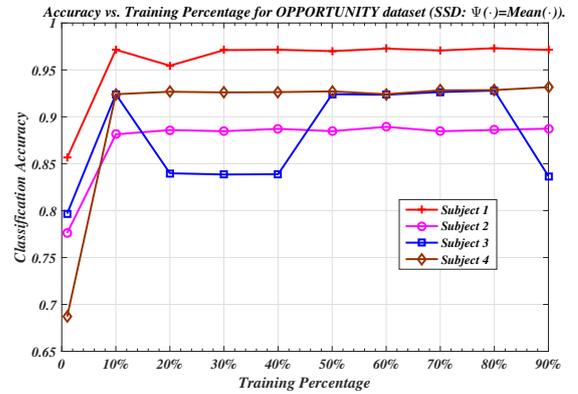


Figure 5. Classification accuracy vs. training percentage for the OPPORTUNITY dataset.

testing. Besides, in order to further observe the performance of the proposed method, we divided the original data into 100 equal parts on the basis of one hundred cross-validation. And then one of the equal parts was randomly selected as the training datasets (1%) and the remaining (99%) were regarded as testing datasets. The average accuracy rates of all these ten experiments was shown in Fig.5, which showed that even if

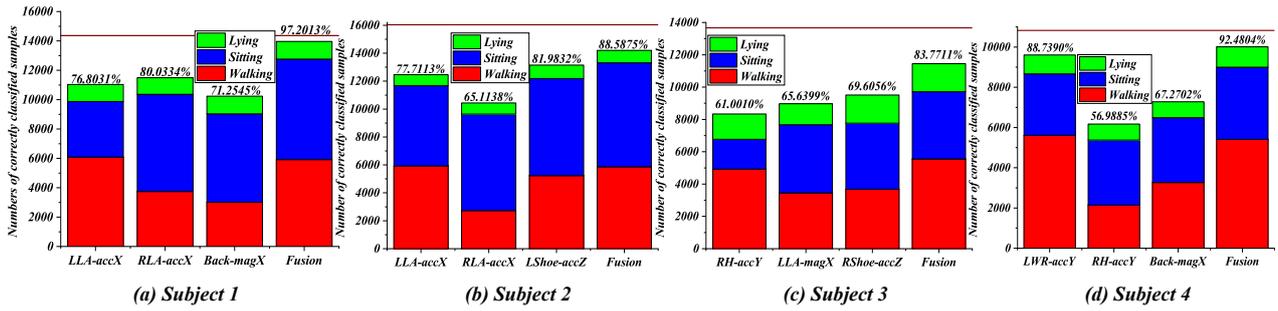


Figure 6. Comparisons Between Base Classifiers and Fused Classifiers in OPPORTUNITY dataset.

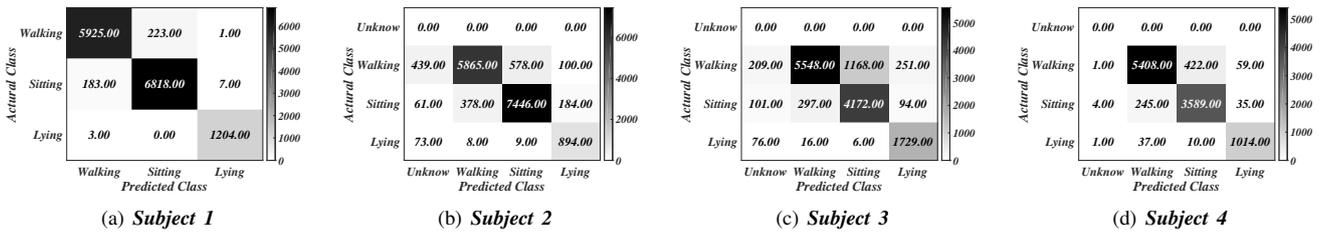


Figure 7. Confusion matrices of Four Subjects in OPPORTUNITY dataset.

there were few training samples, the model proposed in this paper still gave higher recognition accuracy.

3) *Comparison Between Base Classifiers and Fused Classifiers in OPPORTUNITY dataset*: In order to deeply analyze the relationship between base classifiers and fused classifier in our proposed model, the detailed comparisons were given in Fig.6. Based on the results presented in Fig.4, the discriminative function chosen here was SSD (Eq.(3)) and  $\Psi(\cdot) = \text{Mean}(\cdot)$ . In Fig.6, the x-axis represents the KDE model corresponding to the selected sensor, the y-axis represents the number of correctly classified test samples, the value above each histogram represents the classification recognition rate corresponding to each KDE model, and the solid line at the top of the histogram represents the total number of test samples. As we can see from Fig.6: (1) the recognition accuracy of the fused model was significantly improved compared with that of the base classifier; (2) the performance of based classifiers were obviously different. Among these mentioned base classifiers, RH-accY in subject 4 had the lowest rate: 56.9885% and LWR-accY also in subject 4 had the highest rate: 88.7390%. The main reason for the performance difference of the based classifiers is that we looked for the relative *best* KDE model for the specific activity based on our proposed SSD or SDD, not the absolute *best* KDE model for all activities. More concretely, in subject 1, the specific KDE model corresponding to LLA-accX had the best classification only for Walking; the specific KDE model corresponding to RLA-accX had the best classification only for Sitting and the specific KDE model corresponding to Back-magX had the best classification only for Lying. In this way, we could effectively guarantee the degree of diversity among

base classifiers, which is really important for ensemble fusion [11].

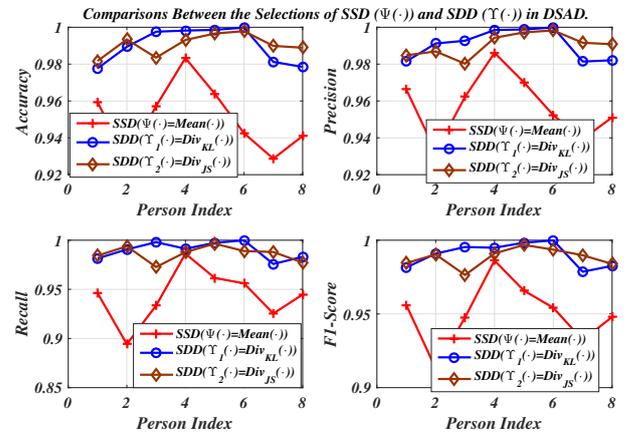
4) *Comparisons with State-of-the-art Approaches Based on Monte-Carlo Simulation*: In this part, we further gave the confusion matrix (Fig.7) of the four subjects in OPPORTUNITY dataset based on our proposed method. It is worth noting that in the confusion matrix of subject 2-4, there existed a special label “UNKNOWN” which was quite different from the three mentioned activities: Walking, Sitting and Lying. This “UNKNOWN” label occurred in our DSMT-Based method because of the missing value in original sensor readings. When the current sensor reading was NULL or missing value, the maximum belief mass ( $1'$ ) was assigned to the focal element ( $\Theta$ ) which meant at current time, we really did not know the actual class. Modeling missing or NULL information is the feature of our proposed method in this paper, which is quite different from the traditional supplementation of NULL or missing information by interpolation. In this way, our proposed method can reduce the risk of misjudgment without guaranteeing any changes to the original data. Besides, we repeated 50 experiments and recorded the recognition rates of all four subjects in Table II. Among the mentioned classical approaches, the performance of k-Nearest Neighbours and Nearest Centroid Classifier were heavily affected by the number of 'k'-closest samples and the centroid of each class. These two principles of classification were difficult to work very well when there existed uncertain data in HAR problem. Linear discriminative analysis and quadratic discriminant analysis based on the assumption that the features are normally distributed are obviously unsuitable in HAR problems. Extreme learning machine has been successfully applied for the task of HAR.

Table II  
 COMPARISON WITH STATE-OF-THE-ART RESULTS ON UCI OPPORTUNITY DATASET.

2*Reported Methods	Accuracy				1*Average Computational Cost
(lr)2-5	Subject1	Subject2	Subject3	Subject4	
Extreme Learning Machine [27]	0.7056±0.1123	0.7126±0.0687	0.6587±0.0295	0.7154±0.1414	13.6175 ms
Linear Discriminant Analysis [28]	0.7859±0.0246	0.8147±0.0274	0.7346±0.0318	0.7913±0.0419	11.0537 ms
Nearest Centroid Classifier [14]	0.8305±0.0312	0.8718±0.0289	0.7647±0.0185	0.8185±0.0152	10.3426 ms
K-Nearest Neighbours ( $k = 5$ ) [14]	0.8995±0.0015	0.8516±0.0101	0.8383±0.0291	0.8516±0.0091	11.6340 ms
Quadratic Discriminant Analysis [14]	0.9143±0.0076	0.8517±0.0078	0.8562±0.0218	0.8216±0.0214	13.5754 ms
Naive Bayes [12]	0.8742±0.0015	0.8401±0.0053	0.8210±0.0315	0.8517±0.0091	15.7027 ms
Ensemble-Extreme Learning Machine(Majority Voting) [11]	0.9142±0.0098	0.8843± 0.0144	<b>0.8714±0.0156</b>	0.8830±0.0144	29.5384 ms
<b>New Method (HAR DSMT-based)</b>	<b>0.9714±0.0014</b>	<b>0.8869±0.0026</b>	0.8439±0.0199	<b>0.9262±0.0025</b>	-
Computational Testing Time For Each Individual Sample	8.6545 ms	14.2733 ms	7.5581 ms	7.6887 ms	9.5436 ms

 Table III  
 THE SELECTED SENSORS IN DSAD BASED ON  $Div_{JS}(\cdot)$ .

2*Subject	$SDD : \Upsilon_2(\cdot) = Div_{JS}(\cdot)$				
(lr)2-6	Sitting	Standing	Lying	Walking	Running
Person 1	$RA_{zgyro}$	$LA_{zmag}$	$LA_{zacc}$	$LA_{zmag}$	$RA_{xacc}$
Person 2	$RL_{zacc}$	$RA_{ymag}$	$RL_{yacc}$	$LA_{xmag}$	$T_{xgyro}$
Person 3	$T_{yacc}$	$T_{xmag}$	$RA_{yacc}$	$RA_{xmag}$	$LA_{ymag}$
Person 4	$LL_{zacc}$	$RL_{xacc}$	$RA_{yacc}$	$RA_{xmag}$	$LA_{xmag}$
Person 5	$LL_{xmag}$	$LL_{zmag}$	$RL_{yacc}$	$LA_{xmag}$	$LA_{zmag}$
Person 6	$RL_{xmag}$	$RL_{xacc}$	$T_{ygyro}$	$RA_{xacc}$	$T_{zmag}$
Person 7	$RL_{yacc}$	$LL_{xacc}$	$T_{zmag}$	$RA_{zmag}$	$T_{xacc}$
Person 8	$RL_{zacc}$	$LA_{xacc}$	$LA_{zacc}$	$T_{xmag}$	$LA_{ymag}$


 Figure 8. Effectiveness of the selection of  $SSD(\Psi(\cdot))$  and  $SDD(\Upsilon(\cdot))$  in DSAD.

And for extreme learning machine, sigmoid activation function was utilized and the number of hidden nodes was set to 100. However, due to the randomness of the algorithm, the results of extreme learning machine were unstable and had a wide variability. As we can observe in Table II, our method gave the highest activity recognition accuracy in subject-1, subject-2 and subject-4, and Ensemble-Extreme Learning Machine (Majority Voting) gave the highest recognition accuracy in subject 3. In addition to the comparison of classification accuracy, we also showed the testing time for each individual sample of our proposed method in Table II. Our method was running in MATLAB R2018b with a hardware of Intel Quad Core i5-4670 CPU at 3.4GHz and 16G RAM. As shown in Table II, our proposed method was significantly more efficient than other general listed methods. The low recognition delay of our method was mainly because in the testing phase, only the data of selected sensors in the testing sample participates in the BBA calculation. The low-recognition delay also showed its potential for the application in online activity recognition systems, because such real-time activity recognition often requires the predictions are updated 1-5 times/s [2].

#### D. Results on UCI DSAD

1) *Effectiveness of the Selection of  $\Psi(\cdot)$  and  $\Upsilon(\cdot)$  in Eq.(3) and Eq.(6):* Similar to the discussions in OPPORTUNITY dataset, we also gave the performance comparisons between the selections of  $\Psi(\cdot)$  and  $\Upsilon(\cdot)$  in DSAD. First, the comparisons of recognition accuracy with different evaluation criterion was shown in Fig.8 when  $\Psi(\cdot)$  and  $\Upsilon(\cdot)$  were set to (1)  $\Psi(\cdot) = Mean(\cdot)$ , (2)  $\Upsilon_1(\cdot) = Div_{KL}(\cdot)$ , (3)  $\Upsilon_2(\cdot) = Div_{JS}(\cdot)$ , respectively. Different from the phenomenon in Fig.4, our proposed method based on  $Div_{KL}(\cdot)$  and  $Div_{JS}(\cdot)$  could give higher recognition accuracy in DSAD. Due to the robust performance of our proposed method based on  $\Upsilon(\cdot) = Div_{JS}(\cdot)$  in DSAD, in the following experiments, the discriminative function  $Div_{JS}$  was applied in Eq.(6). Besides, the sensors selected by  $Div_{JS}$  were also listed in Table III. It can be found that the sensitivity of sensors to different daily activities varied, and was influenced by their locations of deployment and the types of sensors. In Table III,  $T$ : Torso;  $RA$ : Right Arm;  $LA$ : Left Arm;  $RL$ : Right Leg;  $LL$ : Left Leg;  $x, y, zacc$ :  $x, y, z$  accelerometers;  $x, y, zmag$ :

$x, y, z$  magnetometers;  $x, y, zgyro$  :  $x, y, z$  gyroscopes.

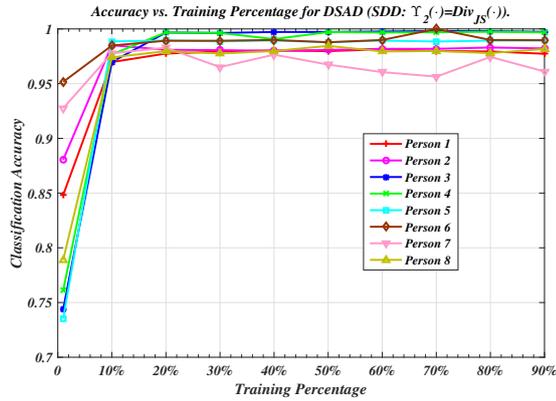


Figure 9. Classification accuracy vs. training percentage for DSAD.

2) *Recognition Rate versus Training Percentage*: In this part, we also varied the percentage of training set and investigated the relationship between the training percentage and the classification accuracy of our proposed method on DSAD. Similar to the experiments in OPPORTUNITY dataset, here we also conducted ten independent experiments. The average accuracy rates of all ten experiments can be seen in Fig.9. From these results, we could also draw the same conclusion as from the proposed method, i.e. classification accuracies for DSAD could reach a high level, without a large amount of training samples.

3) *Comparison Between Base Classifiers and Fused Classifiers in DSAD*: Similar to the experiments in OPPORTUNITY dataset, we also analyzed the relationship between base classifier and fused classifier in DSAD, which was shown in Fig.10. As we can see from Fig.10: (1) when the classification difference between base classifiers were quite obvious, the final performance of fused model could be substantially improved. For example, in person 4, the range of classification accuracy of all base classifiers was [RA-yacc: 74.0080%, LL-zacc:93.6386%] and the final rate of fused model was 98.6185%; (2) On the contrary, when the performances between base classifiers were close, the performance of final fused model was not substantially improved. For example, in person 7, all five base classifiers had similar recognition rates: 86.6024%, 91.9036%, 87.2459%, 91.9036%, 91.9036% and the performance of the final fused model was 92.4498%. These two groups of phenomena further verified the rationality of the modeling strategy proposed in this paper: base KDE model was only selected for the specific activity, which did guaranty the diversities between base models.

4) *Comparison with State-of-the-art Approaches Based on Monte-Carlo Simulation*: In this part, we further gave the confusion matrix (Fig.11) of the eight persons in DSAD based on our proposed method. As we can see in Fig.11, our method had a higher recognition rate in identifying the activities of all mentioned persons. Besides, we further repeated 50 experiments and compared DSMT-based method

Table IV  
COMPARISON WITH STATE-OF-THE-ART RESULTS ON UCI DSAD.

Reported Methods	Accuracy	Computational Cost
Artificial Neural Networks [24]	0.743	23.2442 ms
Bayesian Decision Making [24]	0.758	27.4170 ms
K-Nearest Neighbours [24]	0.860	20.2664 ms
Support Vector Machines [24]	0.876	25.9724 ms
differential Recurrent Neural Networks [29]	0.8956	50.9993 ms
pFTA-Learn + K-Nearest Neighbors [30]	0.9018	19.4653 ms
<b>New Method (HAR DSMT-based)</b>	<b>0.9515</b>	<b>17.0964 ms</b>

with the other traditional method in references in Table IV. All parameters involved in the mentioned state-of-the-art models were consistent with those mentioned in the literature, which were not listed in detail here. For k-Nearest Neighbours, the performance of this method changed for different values of  $k$ . A value of  $k = 5$  gave the best results, therefore the accuracy of the k-Nearest Neighbours algorithm was provided for  $k = 5$  in Table IV. For support vector machine, following the one-versus-the-rest method, each type of activity was assumed as the first class and the remaining 4 activity types were grouped into the second class. The overall accuracy rate of support vector machine was calculated as 87.6%. Besides, we also conducted performance comparison between our technique and differential Recurrent Neural Networks (the related source codes for dRNN could be downloaded from [29]). As shown in Table IV, our proposed method with DSMT-based fusion strategy could achieve even higher accuracy than traditional approaches. Although SVM and dRNN were powerful models for classification and they were not able to properly combine the characteristics of multiple sensors; conversely, DSMT-based approach was especially designed to effectively fuse these information from multi-sensor readings, which proved to be very effective for HAR in BSNs. Besides, we also showed the testing time for each individual sample of our proposed method in Table IV. Results showed that DSMT-based HAR takes shorter time than other classical methods.

## V. CONCLUSION

In this paper, we addressed the challenge of HAR problem in BSNs from the perspective of multi-sensor fusion strategy and exploited the unique DSMT-Based fusion strategy. In this novel fusion strategy, there were two points worth mentioning: 1) unlike traditional fusion strategy, not all sensor readings were used for modeling and fusing, only the selected representative sensors were finally fused; 2) BBA of each test sample was constructed according to KDE models. Besides, the vacuous BBA was directly given when test sample had incomplete pattern. Extensive performance evaluations on two wearable sensor-based HAR datasets (OPPORTUNITY dataset and DSAD) demonstrated that the proposed approach outperformed start-of-the-art methods in accuracy. In our future work, we will explore the performance of the proposed method in complex activity recognition. In this work, our proposed

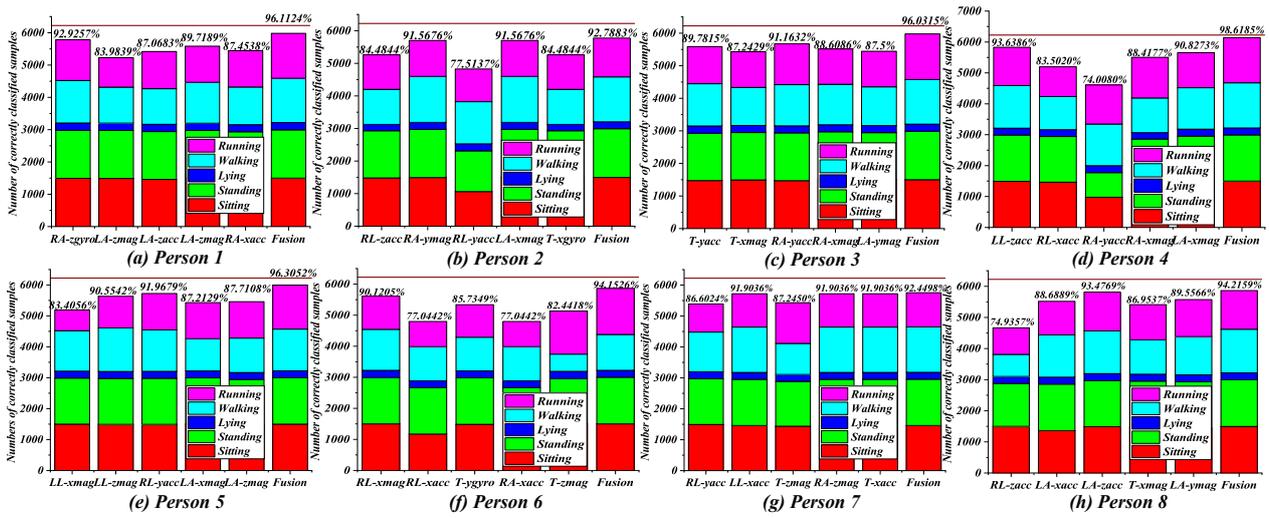


Figure 10. Comparisons Between Base Classifiers and Fused Classifiers in DSAD.

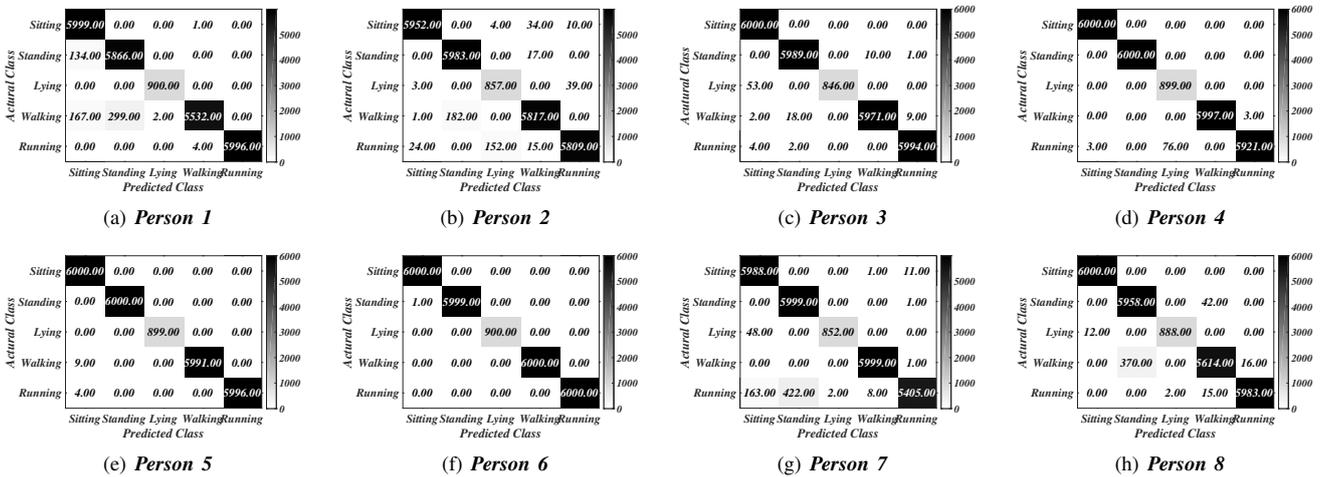


Figure 11. Confusion matrices of 8 Persons in DSAD.

DSmT-based model was currently trained and tested offline. In our future research works, we will investigate and test how such new model can be applied to an online activity recognition system in real-time.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61573097 and Grant 91748106, in part by the Key Laboratory of Integrated Automation of Process Industry under Grant PALN201704, in part by the Advanced Research Project of the 13th Five-Year Plan under Grant 31511040301, in part by Science and Technology on Information System Engineering Laboratory, No:05201905, in part by the Fundamental Research Funds for the Central Universities under Grant 3208008401, in part by the Qing Lan Project and Six Major Top-Talent Plan, and in part by the Priority Academic Program Development of

Jiangsu Higher Education Institutions (*Corresponding author: Xinde Li.*).

REFERENCES

- [1] Z. Chen, L. Zhang, Z. Cao, *Distilling the knowledge from handcrafted features for human activity recognition*, IEEE Trans. on Industrial Informatics, Vol. 14, No. 10, pp. 4334–4342, 2018.
- [2] I. Andrey, *Real-time human activity recognition from accelerometer data using convolutional neural networks*, Applied Soft Computing, Vol. 62, pp. 915–922, 2018.
- [3] Z. Chen, Q. Zhu, Y. Soh, L. Zhang, *Robust human activity recognition using smartphone sensors via CT-PCA and online SVM*, IEEE Trans. on Industrial Informatics, Vol. 13, No. 6, pp. 3070–3080, 2017.
- [4] J.C. Davila, A.M. Cretu, M. Zaremba, *Wearable sensor data classification for human activity recognition based on an iterative learning framework*, Sensors, Vol. 17, No. 6, pp. 1287–1304, 2017.
- [5] T.G. Lee, S.H. Lee, *Design of wearable bio-patch system platform in human healthcare environment*, Indian Journal of Science and Technology, Vol. 8, No. 17, 2015.

- [6] S.A. Khowaja, B.N. Yahya, S.L. Lee, *Hierarchical classification method based on selective learning of slacked hierarchy for activity recognition systems*, Expert Systems with Applications, Vol. 88, pp. 165–177, 2017.
- [7] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, *Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges*, Information Fusion, Vol. 35, pp. 68–80, 2017.
- [8] D. Schuldhuis, H. Leutheuser, B.M. Eskofier, *Towards big data for activity recognition: A novel database fusion strategy*, in Proc. of 9th Int. Conf. on Body Area Networks (BodyNets 2014), pp. 97–103, London, Great Britain, 29th Sep.–1st Oct., 2014.
- [9] C. Chen, R. Jafari, N. Kehtarnavaz, *A survey of depth and inertial sensor fusion for human action recognition*, Multimedia Tools and Applications, Vol. 76, No. 3, pp. 1–21, 2017.
- [10] A. Bulling, U. Blanke, B. Schiele, *A tutorial on human activity recognition using body-worn inertial sensors*, ACM Computing Surveys (CSUR), Vol. 46, No. 3, pp. 1–33, 2014.
- [11] Z. Chen, C. Jiang, L. Xie, *A novel ensemble ELM for human activity recognition using smartphone sensors*, IEEE Trans. on Industrial Informatics, Vol. 15, No. 5, pp. 2691–2699, 2019.
- [12] S. Saeedi, N. El-Sheimy, *Activity recognition using fusion of low-cost sensors on a smartphone for mobile navigation application*, Micromachines, Vol. 6, No. 8, pp. 1100–1134, 2015.
- [13] R. Daniel, C. Alberto, R. Mirco, H. Thomas, *Collecting complex activity data sets in highly rich networked sensor environments*, in Proc. of 2010 7th Int. Conf. on Networked Sensing Systems (INSS), Kassel, Germany, 15–18 June 2010.
- [14] R. Chavarriaga, H. Sagha, A. Calatroni, *The opportunity challenge: A benchmark database for on-body sensor-based activity recognition*, Pattern Recognition Letters, Vol. 34, pp. 2033–2042, 2013.
- [15] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [16] C. Chen, R. Jafari, N. Kehtarnavaz, *Improving human action recognition using fusion of depth camera and inertial sensors*, IEEE Trans. on Human-Machine Systems, Vol. 45, No. 1, pp. 51–61, 2015.
- [17] J.L. Reyes-Ortiz, L. Oneto, A. Sama, X. Parra, D. Anguita, *Transition-aware human activity recognition using smartphones*, Neurocomputing, Vol. 171, No. C, pp. 754–767, 2016.
- [18] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, Vol. 1-4, American Research Press, 2004–2015. <https://www.onera.fr/staff/jean-dezert/references>
- [19] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Int. Conf. on Information Fusion (Fusion 2012), pp. 655–660, Singapore, July 2012.
- [20] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, in [18], Vol. 2, pp. 69–88, 2006.
- [21] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of Int. Conf. on Information Fusion (Fusion 2013), pp. 1119–1126, Istanbul, Turkey, July 9–12, 2013.
- [22] T.M. Davies, C.R. Flynn, M.L. Hazelton, *On the utility of asymptotic bandwidth selectors for spatially adaptive kernel density estimation*, Statistics and Probability Letters, Vol. 138, pp. 75–81, 2018.
- [23] K. Stephens, A.G. Bors, *Modelling of interactions for the recognition of activities in groups of people*, Digital Signal Processing, Vol. 79, pp. 34–46, 2018.
- [24] K. Altun, B. Barshan, O. Tunnl, *Comparative study on classifying human activities with miniature inertial and magnetic sensors*, Pattern Recognition, Vol. 43, pp. 3605–3620, 2010.
- [25] H. Sagha, S.T. Digumarti, J.d.R. Millán, R. Chavarriaga, A. Calatroni, D. Roggen, G. Tröster, *Benchmarking classification techniques using the opportunity human activity dataset*, IEEE International Conference on Systems, Man, and Cybernetics, pp. 233–240, Anchorage, AK, USA, 09-12 October, 2011.
- [26] A.M. Khan, Y.-K. Lee, S.Y. Lee, T.-S. Kim, *A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer*, IEEE Trans. on Information Technology in Biomedicine, Vol. 14, No. 5, pp. 1166–1172, 2010.
- [27] J. Cao, W. Li, C. Ma, Z. Tao, *Optimizing multi-sensor deployment via ensemble pruning for wearable activity recognition*, Information Fusion, Vol. 41, pp. 68–79, 2018.
- [28] R. Chavarriaga, H. Sagha, J.d.R. Millán, *Ensemble creation and reconfiguration for activity recognition: An information theoretic approach*, IEEE International Conference on Systems, Man, and Cybernetics, pp. 233–240, Anchorage, AK, USA, 09-12 October, 2011.
- [29] V. Veeriah, N. Zhuang, G.-J. Qi, *Differential recurrent neural networks for action recognition*, in Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 4041–4049, Santiago, Chile, Dec. 7–13, 2015.
- [30] J. Ye, G. Qi, N. Zhuang, H. Hu, K.A. Hua, *Learning compact features for human activity recognition via probabilistic first-take-all*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 42, No. 1, pp. 126–139, 2020.



# A Study on Human Learning Ability During Classification of Motion and Colour Visual Cues and Their Combination

Albena Tchamova<sup>a</sup>, Jean Dezert<sup>b</sup>, Nadejda Bocheva<sup>c</sup>,  
Pavlina Konstantinova<sup>d</sup>, Bilyana Genova<sup>c</sup>, Miroslava Stefanova<sup>c</sup>

<sup>a</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>b</sup>The French Aerospace Lab, F-91761 Palaiseau, France.

<sup>c</sup>Institute of Neurobiology, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>d</sup>European Polytechnical University, Pernik, Bulgaria.

Emails: tchamova@bas.bg, jean.dezert@onera.fr, nbbocheva@hotmail.com,  
pavlina.konstantinova@gmail.com, b.genova@abv.bg, mirad\_st@abv.bg

Originally published as: A. Tchamova, J. Dezert, N. Bocheva, P. Konstantinova, B. Genova, M. Stefanova, *A Study on Human Learning Ability During Classification of Motion and Colour Visual Cues and Their Combination*, Cybernetics and Information Technologies Journal (CITJ), issue 1/2021, March 2021, and reprinted with permission.

**Abstract**—The paper presents a study on the human learning process during the classification of stimuli, defined by motion and color visual cues and their combination. Because the classification dimension and the features that define each category are uncertain, we modeled the learning curves using Bayesian inference and more precisely the Normalized Conjunctive Consensus rule, and also on the base of the more efficient probabilistic Proportional Conflict Redistribution rule no.5 (pPCR5) defined within Dezert-Smarandache Theory (DSMT) of plausible and paradoxical reasoning. Our goal is to study how these rules succeed to model consistently both: human individual and group behaviour during the learning of the associations between the stimuli and the responses in categorization tasks varying by the amount of relevant stimulus information. The effect of age on this process is also evaluated.

**Keywords:** Vision, Human Perception, Classification, Color cue, Motion cue, Cues Combination, DSMT, probabilistic Proportional Redistribution rule no.5 (pPCR5), Normalized Conjunctive Consensus (NCC) rule.

## I. INTRODUCTION

In everyday activities, humans often have to classify objects and events in different categories. The process of classification requires the acquisition of the common characteristics of the members of a category. Depending on the category structure, three different ways are assumed to be employed in classification [1]: rule-based, incremental learning, or memorization of all exemplars. Rule-based classification is supposed to involve sequential hypothesis testing to uncover the rule of categorization. The incremental learning is supposed to be related to finding the category boundaries in cases when the stimulus categories are defined by more than one feature and no simple rule describes the category membership. It involves forming associations between a set of features and the responses. The third way to find the category assignments is by memorizing the associations the responses for each combination of stimulus features.

When the stimuli for categorization are multidimensional and not all features are relevant for their classification, an important question is how humans find out the proper stimulus characteristics for category membership. To answer this question, [2] tested whether a normative strategy based on probabilistic inference could describe the process of category learning. Their modeling data imply that the decision making based on Bayesian inference is computationally too demanding and that humans use suboptimal strategies in the process of categorization.

In the present study, we used multidimensional visual stimuli that were divided into categories by rules of different complexity. The change in the rule of classifications changes in the amount of irrelevant information. We will compare human cue combination performance in arbitrary (unstructured) classification task with modeled combination performance, based on particular fusion rules. In the presented study we will apply and compare the performances of the following fusion rules: the Normalized Conjunctive Consensus (NCC), and the probabilistic Proportional Conflict Redistribution rule no.5 (pPCR5) defined within DSMT to model the human process of cue integration. We will focus on how the human age influences the process of classification as the experimental evidence implies that various brain structures and processes are involved in the different categorization tasks [3] and they do change differently with ageing [4].

This paper is organized as follows. In section II we present briefly the principles of the used fusion rules, applied to model the human cue integration in a classification task. Section III is devoted to the experimental strategy, methods, procedures, stimulus, and subjects participating in the experiments. The results obtained are described and analysed in Section IV. In section V fusion rules performance is presented and in section VI the trends are illustrated. Conclusions are made in Section VII.

## II. FUSION RULES FOR MODELLING VISUAL CUE COMBINATION

Various fusion rules exist in the literature to deal with uncertain data. They are based on different mathematical models and on different methods for transferring the conflicting mass onto the meaningful hypotheses about the problem under consideration. In this paper, we use the Normalized Conjunctive Consensus (NCC) rule and the Probabilistic Proportional Conflict Redistribution rule no.5 (pPCR5), defined within Dezert-Smarandache Theory (DSMT) of Plausible and Paradoxical Reasoning [5]. Both these rules are described in detail in [6].

## III. EXPERIMENTS

Three experiments were performed.

### A. Stimuli

The stimuli were dynamic patterns that differed by the motion direction, the spatial distribution, shape, and the color of the moving elements. The moving elements were either spheres or cubes. Two conditions were simulated – in one of them the elements were positioned on a plane, in the other they were randomly positioned in depth. The simulated motion could have 4 different directions: to the left, to the right, forward, or backward. As a result, eight different moving patterns were generated: a movement to the left among a cloud of elements, movement to the right among a cloud of elements, movement forward among a cloud of elements, movement backward among a cloud of elements, horizontal translation to the left, horizontal translation to the right, movement forward towards a plane, and movement backward from a plane. The moving elements – spheres or cubes, could have one of 4 colors: red, blue, green, or yellow. Of all possible combinations of movements, shape, and color of elements (64 in total: 8 movements × 4 colors × 2 shapes of moving elements) we randomly selected 16 combinations. The characteristics of the chosen stimuli are given in Table I.

Table I  
CHARACTERISTICS OF THE STIMULI USED IN THE STUDY.

Number of stimuli	Disposition of elements	Motion direction	Color	Shape of elements
1	cloud	forward	green	cube
2	cloud	backward	yellow	sphere
3	cloud	backward	green	cube
4	cloud	backward	red	cube
5	cloud	right	green	sphere
6	cloud	right	yellow	cube
7	cloud	left	blue	cube
8	cloud	left	green	cube
9	wall	forward	red	sphere
10	wall	forward	yellow	cube
11	wall	forward	blue	sphere
12	wall	backward	blue	sphere
13	wall	right	red	sphere
14	wall	right	blue	sphere
15	wall	left	yellow	sphere
16	wall	left	red	cube

### B. Experimental conditions

Three experiments were performed. They differed by the classification rule used to separate the stimuli into two categories. In Experiment 1 the stimuli were divided arbitrarily by the movement type that resulted from the disposition of the elements and the direction of motion, whereas the shape and the color of the elements were irrelevant. In Experiment 2, the stimuli were divided randomly into two categories based on their color, whereas the elements' spatial disposition, motion direction, and the shape of elements were irrelevant. In Experiment 3, the stimuli were randomly divided into two groups based on the combination of the motion direction, elements' disposition, and color. As in Experiments 1 and 2, the shape of the elements was irrelevant. Table II presents the separation of the elements in two categories for the three experiments.

Table II  
CHARACTERISTICS OF THE STIMULI USED FOR DIVIDING THE STIMULI IN CATEGORIES IN EXPERIMENTS 1-3.

Exper. 1	Exper. 2	Experiment 3			
Cat.1	Cat.2	Cat.1	Cat.2	Cat.1	Cat.2
cloud, right	cloud, left	red	blue	cloud, right, green	cloud, left, blue
cloud, backward	cloud, forward	yellow	green	wall, left, red	cloud, left, green
wall, right	wall, left			wall, left, yellow	cloud, right, yellow
wall, backward	wall, forward			wall, right, red	cloud, forward, green
				wall, right, blue	cloud, backward, red
				wall, forward, red	cloud, backward, green
				wall, forward, blue	cloud, backward, yellow
				wall, forward, yellow	wall, backward, blue

As is evident, the classification of the stimuli in Experiment 3 could be done either by trying to find the combination of the stimulus characteristics, or as a rule-with-exception as all cloud stimuli except 1 are in Category 2, and all wall stimuli except one are in Category 1.

### C. Experimental Procedure

Before each experiment, the calibration of an eye-movement recording device was performed. In addition to the standard calibration, each experimental session started with a sequential presentation of a dot at different positions (center, left, left corner, up, right corner, right, down) 10 degrees from the screen center. The dot changed position after 1.5 sec.

Each experiment started with the sequential presentation of all stimuli. At the end of the stimulus sequence, all stimuli

were presented again in a different order and the Subject has to describe the stimulus characteristics â shape and color of elements, the direction of motion, and disposition of elements. This preliminary session aims to acquaint the subjects with the stimulus set and the stimulus characteristics. The experimental session consisted of 128 stimuli – 8 repetitions of each stimulus in random order.

Before the presentation of each stimulus, a warning signal is given. The stimulus duration was 1 sec. During stimulus presentation, a fixation dot with a diameter of 0.5 deg. of arc was shown in the middle of the screen during stimulus presentation. Five hundred milliseconds after stimulus disappearance, two figures â a triangle and a star (with a size of approximately  $4 \times 4$  deg. of arc) appeared at 10 degrees to the left or right from the fixation point. On every trial, the position of these figures was randomly selected. The Subjectâs task was to select to which stimuli corresponds the star, and to which â the triangle. They were required to keep fixation at the fixation point during stimulus presentation and to make a saccade to the selected figure. They had to press the left mouse button if the selected figure is to the left of the fixation point, and the right mouse button â if the selected figure was to the right. In the case of correct choice, a high tone was played, whereas in case of incorrect choice a low tone was played. The subjects were told that at the beginning they could only guess, but during the progression of the experiment, by trial-and-error, they would be able to find the proper classification of the stimuli in categories.

#### D. Method

The order of the experiments was contra-balanced between the subjects. The number of the experimental sessions depended on the Subjectâs performance – if the number of correct responses was low, the participants started a new session after a short break. However, even if the performance of a subject was still not good, no subject participated in more than three experimental sessions. The experiments were separated by 3 to 7 days to avoid inference from previously learned categorization.

#### E. Apparatus

The stimuli were presented on a black background with a custom program written in Python with OpenGL. They were presented on the computer screen operated in refresh rate 60 Hz and resolution  $1280 \times 1024$  pixels, 21â Dell Trinitron with Nvidia Quadro 900XGL graphic board. The stimulus observation was binocular from a distance of 57 cm.

The eye movements of the participants were recorded by Jazz novo eye tracking system (Ober Consulting Sp. Z o.o.) [7].

#### F. Subjects

17 young subjects, aged 18–38 years (median = 22 years) and 17 elderly subjects, aged 63–75 years (median = 67 years) took part in the study.

#### IV. PERFORMANCE EVALUATION OF AGE-RELATED OBSERVERS GROUPS

The experimental goal of our study is directed to characterize human decision making in a classification task influenced by:

- motion information only,
- color information only,
- combined motion and color information.

As the stimuli were randomly assigned to different categories based on their visual characteristics, the test Subjects have to find the correct association between the stimuli and the outcome by trial-and-error. While in the classification of objects or events in categories a generalization of their characteristics is needed, in arbitrary categorization a specific representation of the stimuli is required. As the stimuli in all experiments were the same, one possibility is that irrespective of the categorization rule, the participants will represent them in working memory by all cues. In this case, their performance will be similar in all conditions and the memory load will be equivalent. Also, if unstructured categorization is based on procedural memory [8], the number of features used to categorize the stimuli would be irrelevant. However, the experiments in our study could be also characterized as rule-based with rules of varying complexity that change the amount of irrelevant information. A more efficient way to classify the stimuli is to represent them in memory only by the cues determined by the categorization rule ignoring the irrelevant stimulus characteristics.

An example of the performance from the experiments of an occasional test person is shown in figure 1. It represents the proportion of correct responses in blocks of 16 trials. This information is processed and analyzed to get conclusions about the characteristics used for classification in different categories.

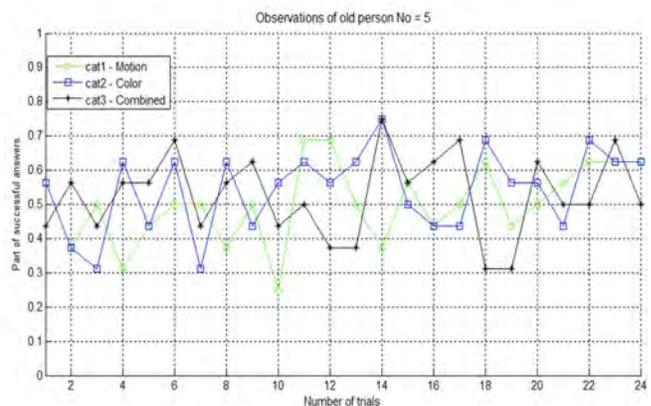


Figure 1. Observations from an occasional test person.

The question is whether the people rely and base their responses on a single source of information, or on a combined one, and also which type of information utilized is more informative in the decision process and corresponds best to the rule used for separating the stimuli in categories. The

participants belong to two age groups: Young and Old. Hence, also the influence of human age on the assessment of the decision will be evaluated. The evaluation is made on the base of experimental learning curves, obtained for all different experimental categories and for each subject in all age-contingent groups. The learning curve represents the change in the correct responses with some measure of the experience gained i.e. the number of trials.

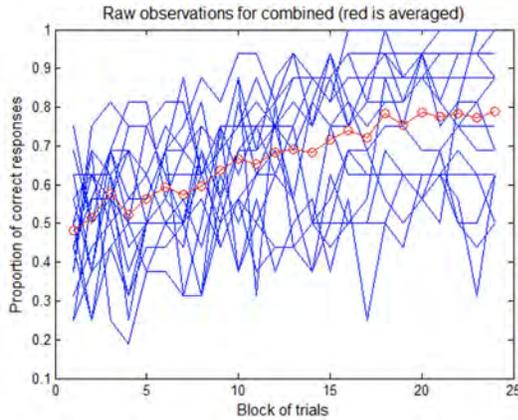


Figure 2. Raw observations (blue), averaged over subjects (red).

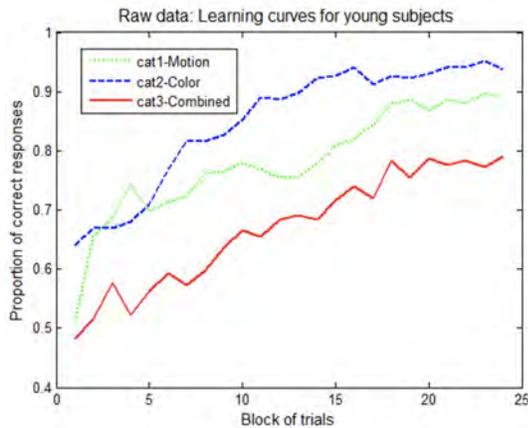


Figure 3. Experimental learning curves for the 3 categorization rules.

Figure 2 presents the learning curves of all young subjects in the case when the stimuli were divided arbitrarily into two groups based on the combination of the color and motion of the stimulus elements. The figure clearly shows the large individual differences in task performance. It also demonstrates that with the increase of blocks, the performance of the group improves. Figure 3 represents the averaged learning curves for all subjects in the young group. It implies that the performance gradually increases and the rate of increase is different for the different categorization rules.

*A. Evaluation of the perception in Young observers group*

The comparison of the performance in the motion and color conditions show that in the Young group only 6 out

of 17 observers have the best performance for the motion condition, 9 observers effectively utilized the color information showing the best performance in this case, and only 1 observer shows best performance in the combined condition (for two other observers the performance in the combined condition is equivalent to that in a single-cue case). For one observer the learning performance is equivalent to both single-cue conditions. The cumulative curve representing the distribution of the average correct responses (on the base of 17 subjects in the group) of the young subjects is shown in Fig. 4.

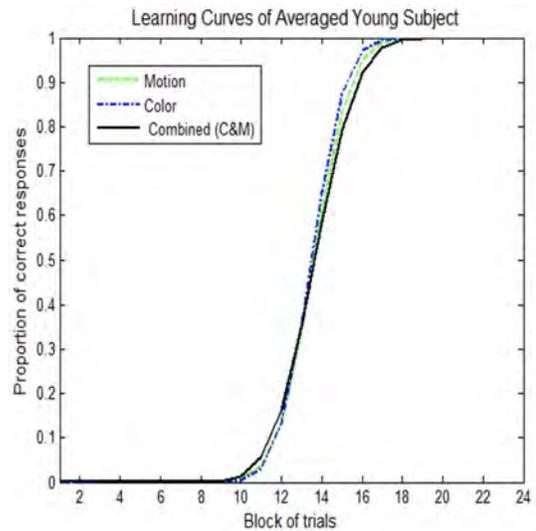


Figure 4. Learning curves of the averaged young subject.

*B. Evaluation of the perception in Old observers group*

The comparison of the performance in the motion and color conditions shows that in the Old group 9 out of 17 observers have the best performance for the motion condition, 6 observers show better performance using the color information. For 2 out of 17 observers the learning curves for both motion and color information could be considered as equivalent. The performance of averaged Old test person on the base of 17 subjects in the group is shown in Fig. 5. Here again, the best performance of the single-cue category is confirmed though the difference between the three conditions is not significant.

It can be summarized that the participants in each group learn best the association between the stimuli and the response when the categorization rule is based on a single cue. This implies that the memory representation of the stimulus characteristics is determined by the categorization rule and the participants are able to ignore the stimulus features irrelevant to the category membership.

It is interesting to compare young and old test persons for the same conditions. We put together the learning curves of young and old subjects in figure 6 for the categories *Motion* and *Color*, and category *Combined* in figure 7.

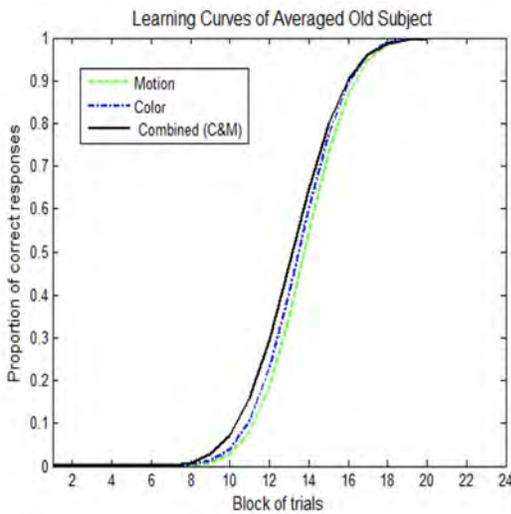


Figure 5. Learning curves of an averaged old subject.

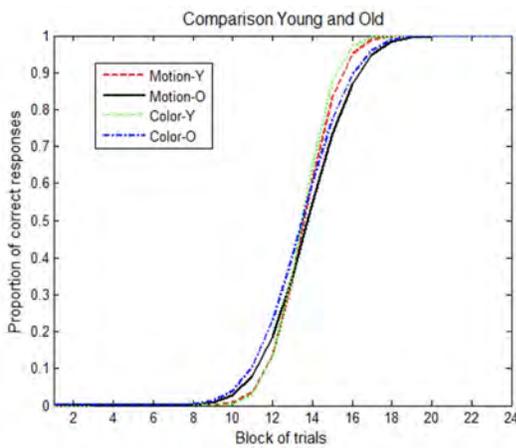


Figure 6. Learning curves of Young and Old for *Color* and *Motion* categories.

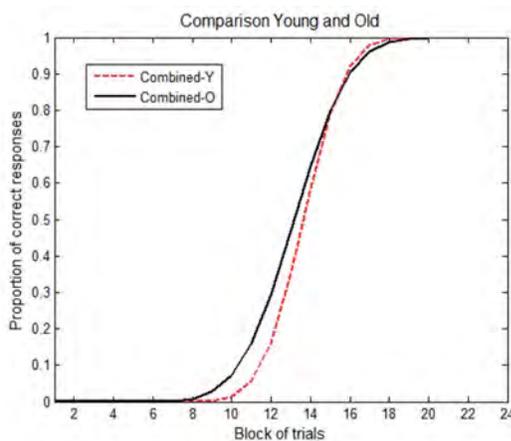


Figure 7. Learning curves of Young and Old for *Combined* category.

#### V. pPCR5 AND NCC RULES PERFORMANCE FOR PREDICTING HUMAN'S WAY OF MOTION AND COLOR COMBINATION IN DECISION MAKING

The main question here is which fusion rule - pPCR5 or NCC used to combine available motion and color information

predicts more adequately human cue integration in deciding the stimulus category?

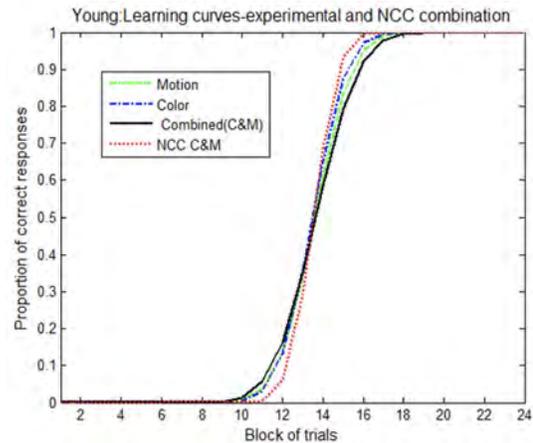


Figure 8. Learning curves for experimental categories and mathematically obtained NCC result for the averaged *young* subject.

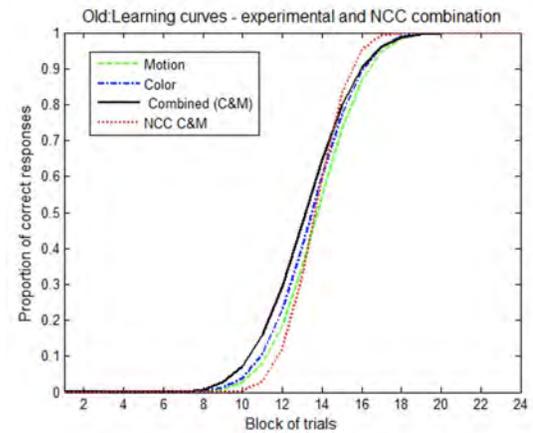


Figure 9. Learning curves for experimental categories and mathematically obtained NCC result for the averaged *old* subject.

In order to answer this question, we need to make a comparison between experimentally obtained and predicted (via pPCR5 and NCC rules) learning curves for combined categories (motion and color), for the two age groups.

In Figures 8 and 9, the results of mathematical modeling by NCC is shown, and on figures 10 and 11 results of applying pPCR5 are presented.

In Figures 12 and 13 the comparison of the empirical and both mathematical methods are given.

The results of the mathematical modeling based on both rules predict better performance than observed in the experimental data. This conclusion concerns the averaged learning curves for the two groups. However, due to the large individual differences in each group, the average learning curve might not be representative of group performance. In section VI we present a different approach to describe the learning in the two groups and apply the same mathematical modeling to it.

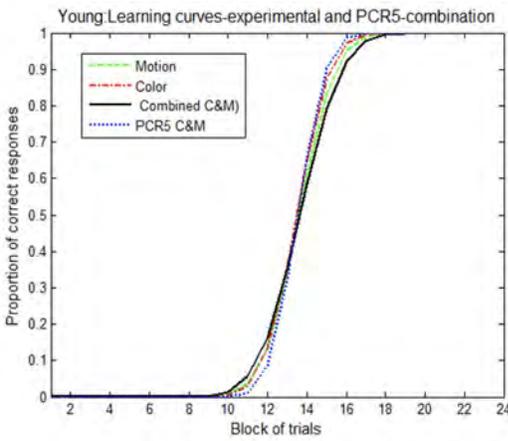


Figure 10. Learning curves for experimental categories and mathematically obtained PCR5 for result for the averaged *young* subject

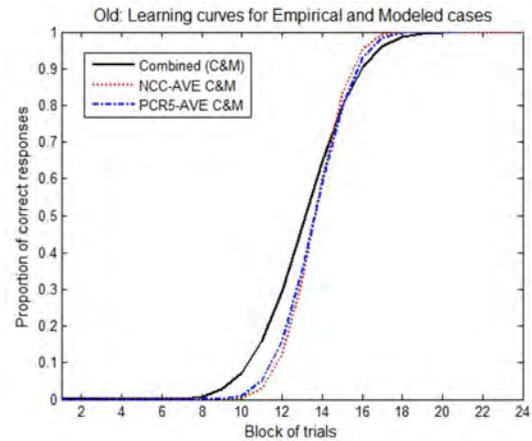


Figure 13. Learning curves for empirical and mathematically modelled cases for *old*.

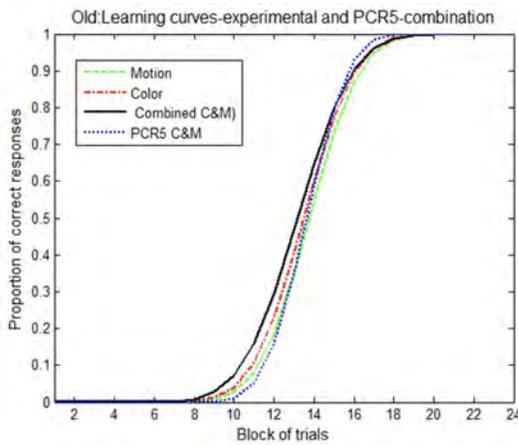


Figure 11. Learning curves for experimental categories and mathematically obtained PCR5 result for the averaged *old* subject.

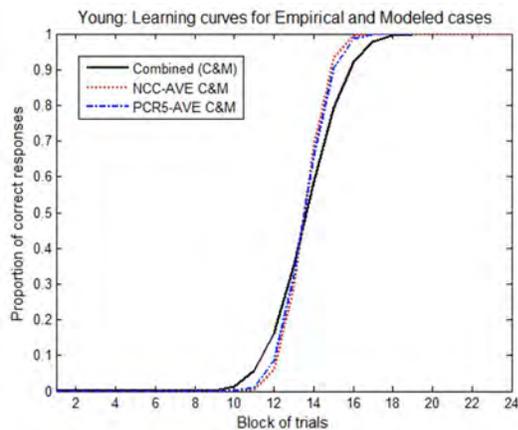


Figure 12. Learning curves for empirical and mathematically modelled cases for *young*.

Another comparison between the methods is provided on the base of the *goodness-of-fit test* [9], that is an important

application of chi-squared criteria:

$$\chi^2 = \sum_{j=1}^J \frac{(O_j - E_j)^2}{E_j}$$

where  $\chi^2$  is an index of the agreement between an *observed(O)/experimental* and *expected(E)/predicted* via particular fusion rule sample values of the learning curve. For our case,  $J = 24$  represents the number of independent observations. The critical value of the test for  $\nu = J - 1 = 23$  degrees of freedom at the assumed probability  $p = 0.1$  is  $\chi^2 = 32.0$  [9].

The respective results are given in Table III - for the young group, in Table IV - for the old persons' group.

In general, the results show that both fusion rules - NCC and pPCR5 succeed to predict adequately human performance for the two age groups. Only for one subject from table IV, the NCC modeling is not adequate - because its NCC error is bigger than the defined critical value  $\chi^2 = 32.0$ . Thus, contrary to the case of the average learning curves where the NCC and pPCR5 predict better performance than obtained experimentally, both rules can describe well the individual learning curves in both age groups.

The results for young and old test-persons are presented in Tables III and IV respectively<sup>1</sup>.

#### VI. COMMON TRENDS OF AGE-RELATED OBSERVER GROUPS

The goal here is to find the common trend, concerning the performance of the two groups. For this purpose, we consider each group as a set of different sources of evidence, associated with each person in the group. That way the young group consists of 17 (young subjects) sources of evidence, which should be combined all together via pPCR5 and NCC fusion rules.

<sup>1</sup>\*\*\*\* means missing information from the test person.

Table III  
CHI-SQUARED VALUES FOR YOUNG SUBJECTS.

Subject	(Motion and Color) pPCR5	(Motion and Color) NCC
1	25.4366	0.4469
2	0.0747	0.5870
3	0.8255	6.4822
4	0.5054	1.2087
5	28.2272	1.0270
6	1.2497	3.6587
7	23.9470	1.4417
8	0.3935	0.3370
9	8.5851	0.3969
10	0.5281	3.1269
11	16.0195	3.7373
12	14.5225	9.4381
13	1.0158	0.0756
14	11.4884	0.3826
15	22.0764	0.3815
16	17.7547	1.1310
17	28.0757	0.2587

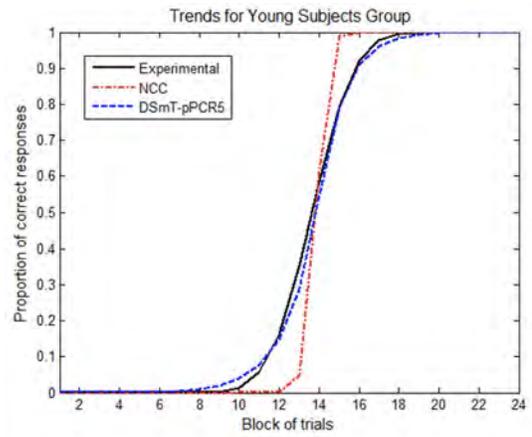


Figure 14. Trends for Young Subjects Group.

Table IV  
CHI-SQUARED VALUES FOR OLD SUBJECTS.

Subject	(Motion and Color) pPCR5	(Motion and Color) NCC
1	0.3587	1.7369
2	*****	*****
3	0.7524	0.9697
4	13.1982	14.2576
5	6.2634	11.0246
6	8.0005	11.0022
7	0.9172	3.3941
8	0.0441	0.0923
9	10.0382	9.8545
10	0.0471	0.2213
11	11.3722	15.2725
12	2.3003	<b>38.7541</b>
13	2.4354	31.1429
14	24.7161	9.5761
15	*****	*****
16	0.8704	9.2390
17	6.5457	21.1361

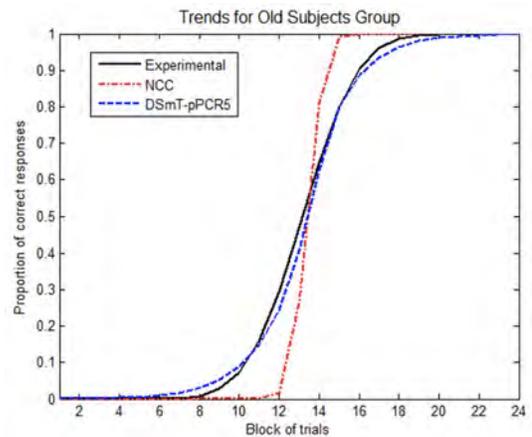


Figure 15. Trends for Old Subjects Group.

The combined individual behaviours in a particular group are estimated, revealing its intrinsic behaviour as a whole, reducing uncertainties associated with individual performances. All the tested subjects in age groups are considered as independent and equally reliable sources of information because each subject provides his/her learning curve, associated with the motion and color condition and should be taken into account with equal weights to derive these trends. Our goal is to find out which combinational rule (pPCR5 or NCC) is able to model correctly and adequately such human age-contingent group trends in the process of decision making. The results obtained for experimental and estimated (via the fusion rules) trends, concerning the cues combination groups' performance, are presented in Figures 14 and 15. It can be seen that the learning curves obtained by the pPCR5 fusion rule in the two figures – for young and old test persons, are more close to the experimentally defined target curve.

pPCR5 fusion rule predicts more correctly the human model of decision making, than the NCC rule, utilizing all the

available information (Motion and Color), even in case of conflict. NCC based trends are very sensitive to the sources (different subjects' learning curves) with the bigger means, neglecting that way part of the available information and acting as an amplifier of the information by reducing the variances

## VII. DISCUSSION

This paper presented a study on the human classification of stimuli defined by motion and color visual cues and their combination. The influence of human age on this process was evaluated. The results obtained show age-related differences in the performance of the subjects in estimating the human classification based on both single- and multi-cue classification rules.

Our experimentally obtained data for young observers suggest a smaller effect of the motion information, while for the older observers the color information has less effect. Hence, the learning performance differs depending on the categorization rule and the age of the participants. All age-related groups have difficulties to divide the stimuli in groups based on combined (motion and color) information. This

finding implies that this condition presents a greater memory load to all observers than the single cue conditions.

In the classification task of multi-cue stimuli, there is uncertainty about which dimension and which feature along this dimension determine the correct response. Hence, the observers have to determine not only the classification rule that specifies the categorization dimension, but also which exemplars that differ by this dimension fall in one or the other category. In contrast to the previous studies [2, 10, 11] testing Bayesian inference in classification studies and learning, we do not model the explicit performance of each subject or group based on the available cues. Instead, our approach reminds the analysis of cue combination in perception studies where the proportion of correct responses is related to stimulus strength. Thus, in our analyses, the experience gained during the task performance is regarded similarly to stimulus characteristics in detection or discrimination perceptual tasks. We performed a comparison between experimentally obtained and predicted (via pPCR5 and NCC rules) learning curves for combined condition (motion and color), for the two age groups and applied the goodness-of-fit test, one important application of chi-squared criteria, to evaluate the correspondence of the experimental and the model data. The results suggest that the predictions of the models outperform human performance for both age groups. This finding differs from our previous results [6] on cue combination in evaluating the heading direction from texture and motion cues. However, it coincides with the conclusion in [2] that human subjects perform suboptimally in categorization tasks.

Both the NCC and the pPCR5 rules predict well the individual learning curves with a slight advantage of the pPCR5 rule as it fits well all the learning curves. This finding provides evidence for the relevance of our approach for analysis of the learning curves.

We evaluated the common trend in the performance of the two age groups by considering each group member as an independent source of information. The obtained trends are better described by the pPCR5 rule than by the NCC rule. The best fit of the group behavior by the PCR5 rule is due to its properties to utilize all the available information even in a case of conflict between the individual learning curves. It is an appropriate characteristic of the group data as it preserves the idiosyncrasies in the performance of each individual and hence, represents effectively the process of decision making in classification tasks for different age groups.

#### VIII. ACKNOWLEDGMENT

The reported work is a part of and was supported by the project DN02/3/2016 “Modelling of voluntary saccadic eye movements during decision making” funded by the Bulgarian Science Fund.

#### REFERENCES

- [1] F.G. Ashby, V.V. Valentin, *Multiple systems of perceptual category learning: theory and cognitive tests*, In Handbook of categorization in cognitive science, 2nd Ed. (H. Cohen & C. Lefebvre Editors), pp. 157–188, Elsevier, New York, 2017.
- [2] R.C. Wilson, Y. Niv, *Inferring relevance in a changing world*, Frontiers in human neuroscience, Vol. 5, 189, 2012.
- [3] D. Shohamy, C.E. Myers, J. Kalanithi, M.A. Gluck, *Basal ganglia and dopamine contributions to probabilistic category learning*, Neuroscience and biobehavioral reviews, 32(2), pp. 219–236, 2008.
- [4] P.B. Kalra, J. Gabrieli, A.S. Finn, *Evidence of stable individual differences in implicit learning*, Cognition, 190, pp. 199–211, 2019.
- [5] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, Vols. 1–4, American Research Press, 2004–2015.
- [6] A. Tchamova, J. Dezert, P. Konstantinova, N. Bocheva, B. Genova, M. Stefanova, *Human Heading Perception Based on Form and Motion Combination*, 2018 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA 2018), 3–5 July 2018.
- [7] <http://www.ober-consulting.com/9/lang/1>
- [8] C.A. Seger, C.M. Cincotta, *The roles of the caudate nucleus in human classification learning*, J. of Neurosci., Vol. 25, pp. 2941–2951, 2005.
- [9] J. Matre, G. Gilbreath, *Statistics for Business and Economics*, 3rd Edition, 1987.
- [10] A.J. Yu, P. Dayan, *Uncertainty, neuromodulation, and attention*, Neuron 46, pp. 681–692, 2005
- [11] S.J. Gershman, J.D. Cohen, Y. Niv, *Learning to selectively attend*, in 32nd Annual Conference of the Cognitive Science Society, Portland, 2010.

# Improvement of Proportional Conflict Redistribution Fusion Rules for Levee Characterization

Théo Dezert<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>IFSTTAR, GERS, GeoEND, F-44344 Bouguenais, France.

<sup>b</sup>The French Aerospace Lab, ONERA/DTIS, Palaiseau, France.

Emails: theo.dezert@univ-eiffel.fr, jean.dezert@onera.fr

Originally published as: T. Dezert, J. Dezert, *Improvement of Proportional Conflict Redistribution Fusion Rules for Levee Characterization*, in Proc. of ESREL 2021 Int. Conference, Angers, France, September 19–23, 2021, and reprinted with permission.

**Abstract**—Levee security assessment is a complex expert assessment process based on several heterogeneous data. In our previous research works, we applied information fusion techniques to characterize flood protection levees. We used the proportional conflict redistribution rule no.6 (PCR6) proposed in DSMT (Dezert Smarandache Theory) framework to combine data from geotechnical and geophysical investigation methods. However, in some cases, this rule can generate non satisfactory results. Indeed, the uncertainty between several hypotheses (lithological materials) is overestimated after the fusion process, which is detrimental to decision making in the end. This result occurs because the PCR6 rule does not preserve the neutrality of the vacuous belief assignment, which can be judged as being a counter-intuitive behavior. To overcome this problem we present an improved rule that preserves the neutrality of vacuous belief assignments in the fusion process. Hence, the redistribution of the partial conflict masses using this new rule does not overestimate the masses associated with partial uncertainties. To illustrate the use of this new fusion rule in a levee characterization problematic, we simulate data acquisition. Two geophysical investigation campaigns (electrical resistivity tomography and multi-channel analysis of surface waves methods) and a geotechnical acquisition campaign (core drillings with particle size analysis) are numerically simulated on an earthen structure. The objective is to compare and discuss the fusion results obtained using this new rule with respect to the methodology based on the original PCR6 rule as well as to demonstrate the enhancement of the levee characterization.

**Keywords:** belief functions, levee, cross-disciplinary approach, natural hazards, fusion rules, risk management, proportional conflict redistribution rule.

## I. INTRODUCTION

This work is part of a problematic of levee characterization for flood protection. Indeed, these hydraulic works are mostly old and heterogeneous and their rupture can lead to disastrous consequences such as human, economic and environmental losses. Since many different materials and construction methods exist, each flood protection embankment is unique, and the nature of its structure goes hand in hand with its environment [1]. The structures are more or less subject to breakage in weak areas under specific loads. Reducing the risk of levee rupture requires an improvement of their diagnosis and therefore to enhance their characterization. First, it relies on technical surveys able to determine if specific pathologies that could lead to failure mechanisms are present in the levee structure.

Methodologies for the evaluation of these structures usually include geotechnical and geophysical investigation methods. Geophysical methods are mainly non-intrusive and provide physical information on large volumes of subsoil but with potential significant uncertainties. Geotechnical investigation methods, on the other hand, are intrusive and provide more punctual information spatially, but also more precise. These two sets of methods are complementary. Information fusion is a helpful technique to combine geotechnical and geophysical data in a complex processing for the levee security assessment based on several heterogeneous data. The processing of the data from geophysical and geotechnical investigation methods and their fusion, taking into account their imperfections and associated spatial distributions, is an essential issue for the evaluation of earthen levees. A cross-disciplinary fusion approach for the characterization of lithological materials within the structures has recently been proposed in the mathematical framework of belief functions [2].

In this paper, we present a flawed behavior of PCR6 combination rule attributed to the non neutrality of the vacuous BBA (Basic Belief Assignment), and we propose an improvement to this rule (PCR6<sup>+</sup>) in order to ensure the neutrality property of the vacuous BBA. This improvement helps in reducing the level of uncertainty in fusion results by discarding ignorant sources for each partial conflict. To demonstrate the pertinence and advantages of PCR6<sup>+</sup> over PCR6, we compare the obtained results for i) a simple numerical example and for ii) the fusion of simulated geophysical and geotechnical data on an earthen levee.

## II. BELIEF FUNCTIONS

Based on preliminary works done in [3], [4], Shafer has introduced the belief functions (BF) in [5] to model epistemic uncertainty, to reason about uncertainty and to combine uncertain information. The theory of belief functions is also known as Dempster-Shafer Theory (DST) in the literature. We assume that the answer<sup>1</sup> of the problem under concern belongs to a known (or given) finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are exhaustive and exclusive<sup>2</sup>. The set of all subsets of  $\Theta$

<sup>1</sup>i.e. the solution, or the decision to take.

<sup>2</sup>This is so-called Shafer's model of FoD [6].

(including empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . In this section we recall the main definitions related with BF and introduce briefly the conjunctive and Dempster-Shafer rules of combinations.

#### A. Main Definitions

A (normal) basic belief assignment (BBA) associated with a given source of evidence is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The real number  $m(A)$  is called the mass of  $A$  committed by the source of evidence. The subset  $A \in 2^\Theta$  is called a focal element (FE) of the BBA  $m(\cdot)$  if and only if  $m(A) > 0$ . The set of all the focal elements of a BBA  $m(\cdot)$  is denoted  $\mathcal{F}_\Theta(m) = \{X \in 2^\Theta | m(X) > 0\}$ . The set  $\mathcal{F}_\Theta(m)$  has at least one focal element, and at most  $2^{|\Theta|} - 1$  focal elements because one has always  $m(\emptyset) = 0$  by the definition of a normal BBA - see [5]. Belief and plausibility functions are respectively defined from  $m(\cdot)$  by

$$Bel(A) = \sum_{X \in 2^\Theta | X \subseteq A} m(X), \quad (1)$$

$$Pl(A) = \sum_{X \in 2^\Theta | A \cap X \neq \emptyset} m(X), \quad (2)$$

where  $\bar{A}$  represents the complement of  $A$  in  $\Theta$ , that is  $\bar{A} \triangleq \Theta - \{A\} = \{X | X \in \Theta \text{ and } X \notin A\}$ . The symbol  $\triangleq$  means *equal by definition* and the minus symbol denotes the set difference operator - see [7], [8].

$Bel(A)$  and  $Pl(A)$  are usually interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$ . The width  $Pl(A) - Bel(A)$  of the belief interval  $[Bel(A), Pl(A)]$  is usually called the *uncertainty on  $A$*  but it represents in fact the imprecision on the probability of  $A$  granted by the source of evidence. When all the focal elements of a BBA  $m(\cdot)$  are singletons this BBA is called a *Bayesian BBA* and its corresponding  $Bel(\cdot)$  function is equal to  $Pl(\cdot)$  and they are homogeneous to a (subjective) probability measure  $P(\cdot)$ . The vacuous BBA (VBBA for short) representing a totally ignorant source is defined as  $m_v(\Theta) = 1$ .

#### B. Conjunctive Combination Rule

We consider  $S \geq 2$  distinct reliable sources of evidence characterized by their BBA  $m_s(\cdot)$  ( $s = 1, \dots, S$ ) defined on the same frame of discernment  $\Theta^3$ . Their conjunctive fusion, denoted  $\text{Conj}(m_1, m_2, \dots, m_S)$ , corresponds to a (non proper) BBA defined for all  $A \in 2^\Theta$  by

$$m_{1,2,\dots,S}^{\text{Conj}}(A) = \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = A}} \prod_{i=1}^S m_i(X_{j_i}), \quad (3)$$

where  $\mathbf{X}_j \triangleq (X_{j_1}, X_{j_2}, \dots, X_{j_S})$  is a possible  $S$ -uple of focal elements, where  $j_1 \in \{1, \dots, \mathcal{F}_1\}$ ,  $j_2 \in \{1, \dots, \mathcal{F}_2\}$ ,  $\dots$ , and  $j_S \in \{1, \dots, \mathcal{F}_S\}$ . The element  $X_{j_i}$  is the focal

<sup>3</sup>For notation simplicity, we omit  $\Theta$  lower index in the notations of sets of focal elements  $\mathcal{F}_\Theta(m_1), \dots, \mathcal{F}_\Theta(m_S)$ , and their cardinalities are simply written as  $\mathcal{F}_1, \dots$ , and  $\mathcal{F}_S$ .

element of the BBA  $m_i(\cdot)$  that makes the  $i$ -th component of the  $j$ -th  $S$ -uple  $\mathbf{X}_j$ . The set  $\mathcal{F}(m_1, \dots, m_S)$  is the set of all possible  $S$ -uples. The cardinality of the set  $\mathcal{F}(m_1, \dots, m_S)$  is noted  $\mathcal{F}$  for convenience. The total conflicting mass, denoted  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)$ , is given by

$$m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) = \sum_{\substack{\mathbf{X}_j \in \mathcal{F}(m_1, \dots, m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = \emptyset}} \prod_{i=1}^S m_i(X_{j_i}). \quad (4)$$

This fusion rule is commutative and associative, and the vacuous BBA  $m_v$  has a neutral impact, that is  $\text{Conj}(m_1, m_2, \dots, m_S, m_v) = \text{Conj}(m_1, m_2, \dots, m_S)$ . Its main drawback is that it does not generate a proper BBA because  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset) > 0$  in general. Because the empty set  $\emptyset$  is the absorbing element for the conjunctive operation, this rule generates  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)$  that quickly tends to one after only few steps of a sequential fusion processing of the sources which is not very useful for decision-making support. The main interest of this rule is its ability to identify the partial conflicts, and to provide a measure of the total level of conflict  $m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)$  between the sources which can be used to manage (select or discard) the sources in the fusion process if one prefers, see [2] for instance.

#### C. Dempster-Shafer Combination Rule

Dempster-Shafer (DS) rule of combination is the emblematic rule of combination proposed by Shafer in his Mathematical Theory of Evidence (see [5]) which is based on Dempster's early works (see [3], [4]). DS rule is nothing but the normalized version of the conjunctive rule. Hence, DS combination of  $S > 1$  BBAs  $m_s(\cdot)$  ( $s = 1, \dots, S$ ) defined on the same frame of discernment  $\Theta$ , denoted as  $\text{DS}(m_1, m_2, \dots, m_S)$ , is a proper BBA defined by  $m_{1,2,\dots,S}^{\text{DS}}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{1,2,\dots,S}^{\text{DS}}(A) = \frac{m_{1,2,\dots,S}^{\text{Conj}}(A)}{1 - m_{1,2,\dots,S}^{\text{Conj}}(\emptyset)}. \quad (5)$$

DS fusion rule is commutative and associative, and the vacuous BBA  $m_v$  has also a neutral impact for this rule, but its justification and behavior have been disputed over the years from many counter-examples involving high or low conflicting sources (from both theoretical and practical standpoints) as reported in [9], [10], [11]. In our applications that are related with risk assessment and safety, we do not prefer to use DS rule because of its very serious problems. Actually, many alternative rules of combination exist<sup>4</sup>, and among them we focus on the new interesting rule based on the proportional conflict redistribution no. 6 (PCR6) principle (see [6], Vol. 3 for details) which is presented in the next section.

### III. PCR6 COMBINATION RULE

#### A. PCR6 General Formula

The PCR6 rule of combination has been proposed in [12], [13] as an interesting alternative of original PCR rule of

<sup>4</sup>see [6], Vol. 2 for a detailed list of many fusion rules.

combination no. 5 (PCR5) proposed in [14], [15]. The PCR6 rule coincides with the PCR5 rule when one combines only two sources (i.e. two BBAs defined on the same FoD). The difference between PCR5 and PCR6 rules lies in the way the proportional conflict redistribution is done as soon as three (or more) sources are involved in the fusion. For notation convenience, we define

$$\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) \triangleq \prod_{i=1}^S m_i(X_{j_i}). \quad (6)$$

If  $X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S} = \emptyset$ , then we use the more concise notation  $\pi_j(\emptyset)$  instead of  $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S})$ , and  $\pi_j(\emptyset)$  is called a *conflicting mass product*.

The PCR6 fusion of  $S > 2$  BBAs is obtained by  $m_{1,2,\dots,S}^{\text{PCR6}}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$\begin{aligned} m_{1,2,\dots,S}^{\text{PCR6}}(A) &= m_{1,2,\dots,S}^{\text{Conj}}(A) \\ &+ \sum_{j \in \{1,\dots,S\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \left[ \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \right. \\ &\quad \left. \cdot \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right], \quad (7) \end{aligned}$$

where  $\wedge$  is the logical conjunction<sup>5</sup>.

We use this general PCR6 formula because it is more easy to implement and to improve than the original formula given in [12] and in [13]. The PCR6 rule is quasi-associative and it offers a more refined conflict redistribution than DS rule but it is more complex, and it does not preserve the neutrality of the vacuous BBA. PCR6 is simpler to implement than PCR5. When  $S > 2$ , PCR6 is better than PCR5 for decision-making as shown in [12]. Matlab<sup>TM</sup> codes of PCR5 and PCR6 fusion rules can be found in [6], [16], and also from the BFAS<sup>6</sup> repository. The PCR5 formula can be obtained from the PCR6 formula by just replacing the two summation operators on  $i \in \{1, \dots, S\} | X_{j_i} = A$  appearing in (7) by the two product operators on  $i \in \{1, \dots, S\} | X_{j_i} = A$ , that is

$$\sum_{i \in \{1,\dots,S\} | X_{j_i} = A} \rightarrow \prod_{i \in \{1,\dots,S\} | X_{j_i} = A}.$$

### B. Drawback of PCR6 Rule

The PCR6 (resp. PCR5) rule of combination is not associative which means that the fusion of the BBAs must be done using general formula (7) if one has more than two BBAs to combine, which is not very convenient. Therefore, the sequential PCR6 (resp. PCR5) combination of  $S > 2$  BBAs are not in general equal to the global PCR6 (resp. PCR5) fusion of the  $S$  BBAs altogether because the order of the combination of the sources does matter in the sequential combination. Moreover, the PCR6 rule (resp. PCR5) can become computationally intractable for combining a large number of sources and for working with large FoD. This is a

well-known limitation of this rule, but this is the price to pay to get better results than with DS rule. Aside the complexity of this rule, it is worth to mention that the neutral impact property of the vacuous BBA  $m_v$  is lost in general when considering the PCR6 (or PCR5) combination of  $S > 2$  BBAs altogether because of the proportional conflict redistribution principles used in PCR6 (resp. PCR5) rule. The non neutral impact of the vacuous BBA is clearly a drawback because it is naturally expected that the vacuous BBA must not impact the fusion result in the fusion process because the vacuous BBA brings no useful information to exploit. Also a BBA that is close to the vacuous BBA should not have a strong impact on the fusion result because it brings only a very little valuable information. This can be seen as a flaw of the behavior of PCR6 (resp. PCR5) rule of combination. To emphasize clearly this flaw, we give in the example 1 a case where the mass committed to some partial uncertainties can increase more than necessary with PCR6 rule if we have a BBA which is close (or equal) to the vacuous BBA, which is detrimental for the quality of the fusion result and for decision-making (because the result is more uncertain than it should be, and consequently the decision is more difficult to make).

**Example 1:** consider  $\Theta = \{A, B, C, D, E\}$  and the three BBAs listed in Table I.

TABLE I  
THE THREE BBAs TO COMBINE.

Focal Elements	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$
$B$	0.05	0.05	0
$A \cup B$	0.65	0.05	0
$C \cup D$	0.05	0.50	0
$A \cup B \cup C \cup D$	0.15	0.05	0
$E$	0.10	0.35	0.01
$\Theta$	0	0	0.99

Here  $m_3(\cdot)$  is not equal to the vacuous BBA but it is very close to the vacuous BBA because  $m_3(\Theta)$  is close to one. The results<sup>7</sup> of the fusion PCR6( $m_1, m_2$ ), and the fusion PCR6( $m_1, m_2, m_3$ ) are given in Table II.

TABLE II  
 $m_{1,2}^{\text{PCR6}}(\cdot)$  AND  $m_{1,2,3}^{\text{PCR6}}(\cdot)$  RESULTS.

Focal Elements	$m_{1,2}^{\text{PCR6}}(\cdot)$	$m_{1,2,3}^{\text{PCR6}}(\cdot)$
$B$	0.054877	0.048939
$A \cup B$	0.406987	0.247656
$C \cup D$	0.312886	0.204005
$A \cup B \cup C \cup D$	0.024917	0.013439
$E$	0.200333	0.101731
$\Theta$	0	0.384230

<sup>5</sup>i.e.  $x \wedge y$  means that conditions  $x$  and  $y$  are both true.

<sup>6</sup>Belief Functions and Applications Society, see <https://www.bfasociety.org/>.

<sup>7</sup>The numerical values have been rounded to their sixth digit.

One sees that combining the BBAs  $m_1, m_2$  with the BBA  $m_3$  (where  $m_3$  is close to vacuous BBA, and therefore  $m_3$  is almost non-informative) generates a big increase of the belief of the uncertainty in the resulting BBA. This behavior is clearly counter-intuitive because if the source is almost vacuous, only a small degradation of the uncertainty is expected and in the limit case when  $m_3$  is the vacuous BBA no impact of  $m_3$  on the fusion result should occur. Because of this flawed behavior, we propose in the next section an improvement of PCR6 rule (called PCR6<sup>+</sup> fusion rule) in order to preserve the neutrality of the vacuous BBA.

#### IV. IMPROVEMENT OF PCR6 RULE

The very simple and basic idea to improve the PCR6 conflict redistribution principle is to discard the elements that contain the other elements implied in the conflict mass product  $\pi_j(\emptyset)$  calculation. Indeed, the elements discarded are regarded as non informative and not useful for making the conflict redistribution. To illustrate clearly this point, let's consider again Example 1 and the conflicting product

$$\pi_{16}(\emptyset) = m_1(A \cup B)m_2(C \cup D)m_3(\Theta).$$

With PCR6, the redistribution of  $\pi_{16}(\emptyset)$  follows

$$\begin{aligned} \frac{x_{16}(A \cup B)}{m_1(A \cup B)} &= \frac{x_{16}(C \cup D)}{m_2(C \cup D)} = \frac{x_{16}(\Theta)}{m_3(\Theta)} \\ &= \frac{\pi_{16}(\emptyset)}{m_1(A \cup B) + m_2(C \cup D) + m_3(\Theta)}, \end{aligned}$$

which is not very efficient because  $\Theta$  is not the source of conflict in this case since  $A \cup B \subseteq \Theta$  and  $C \cup D \subseteq \Theta$ . The conflict exists only because  $(A \cup B) \cap (C \cup D) = \emptyset$ . In the improved version of PCR6 rule, denoted PCR6<sup>+</sup>, the conflicting product  $\pi_{16}(\emptyset)$  will be redistributed only to  $A \cup B$  and to  $C \cup D$  but not to  $\Theta$ . With PCR6<sup>+</sup> rule we will make the new (simpler) redistribution of  $\pi_{16}(\emptyset)$  according to

$$\frac{x_{16}(A \cup B)}{m_1(A \cup B)} = \frac{x_{16}(C \cup D)}{m_2(C \cup D)} = \frac{\pi_{16}(\emptyset)}{m_1(A \cup B) + m_2(C \cup D)}.$$

##### A. PCR6<sup>+</sup> general formula

The general expression of PCR6<sup>+</sup> (and also PCR5<sup>+</sup>) is presented in details, with many examples and Matlab<sup>TM</sup> codes in [17]. Here, due to space limitation, we just recall its expression for convenience. Actually, PCR6<sup>+</sup> fusion rule is the proper modification of PCR6 formula (7) taking into account the selection of focal elements on which the proportional redistribution must apply thanks to the value of their keeping-

index. More precisely, the PCR6<sup>+</sup> fusion of  $S > 2$  BBAs is obtained by  $m_{1,2,\dots,S}^{\text{PCR6}^+}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$\begin{aligned} m_{1,2,\dots,S}^{\text{PCR6}^+}(A) &= m_{1,2,\dots,S}^{\text{Conj}}(A) \\ &+ \sum_{j \in \{1,\dots,S\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \\ &\left[ \left( \kappa_j(A) \sum_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \cdot \right. \\ &\left. \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \kappa_j(X) \sum_{i \in \{1,\dots,S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right], \end{aligned} \quad (8)$$

where  $\kappa_j(A)$  and  $\kappa_j(X)$  are respectively the keeping indexes of elements  $A$  and  $X$  involved in the conflicting product  $\pi_j(\emptyset)$ , that are calculated by the formula

$$\kappa_j(X_{j_i}) \triangleq 1 - \prod_{\substack{X_{l'}, X_l \in \mathcal{X}_j | X_{l'} \neq X_l \\ |X_{j_i}| \leq |X_l| \\ |X_{l'}| \leq |X_l|}} \delta_j(X_{l'}, X_l). \quad (9)$$

$\mathcal{X}_j = \{X_1, \dots, X_{s_j}, s_j \leq S\}$  is the set of all distinct components of the  $S$ -uple  $\mathbf{X}_j$  related with the conflicting product  $\pi_j(\emptyset)$ . The term  $\delta_j(X_{l'}, X_l)$  is the binary *containing indicator* of  $X_l$  with respect to  $X_{l'} \in \mathcal{X}_j$  that characterizes if  $X_l$  contains (includes)  $X_{l'}$  in wide sense, or not. More precisely,  $\delta_j(X_{l'}, X_l)$  is defined by

$$\delta_j(X_{l'}, X_l) \triangleq \begin{cases} 1 & \text{if } X_{l'} \subseteq X_l, \\ 0 & \text{if } X_{l'} \not\subseteq X_l. \end{cases} \quad (10)$$

The value  $\kappa_j(X_{j_i}) = 1$  stipulates that the focal element  $X_{j_i} \in \mathbf{X}_j$  must receive some proportional redistribution from the conflicting mass  $\pi_j(\emptyset)$ , and  $\kappa_j(X_{j_i}) = 0$  indicates that  $X_{j_i} \in \mathbf{X}_j$  will not be involved in the proportional redistribution of  $\pi_j(\emptyset)$ .

Note that  $\kappa_j(\Theta) = 0$  if  $\Theta \in \mathcal{X}_j$  because  $\Theta$  always includes all other focal elements of  $\mathcal{X}_j$  and  $\Theta$  has the highest cardinality. For a given FoD and a given number of BBAs to combine, it is always possible to calculate off-line the values of the keeping-indexes of focal elements for all combinations leading to conflicting products  $\pi_j(\emptyset) > 0$ . We can verify that formula (8) is consistent with PCR6 formula (7) when all keeping indexes are equal to one. The fusion rule (8) is commutative and non associative, and the vacuous BBA  $m_v$  has a neutral impact in PCR6<sup>+</sup> rule - see proof in [17].

##### B. Example 1 revisited with PCR6<sup>+</sup>

Consider the example 1 with the three BBAs given in table I. If we combine the BBAs  $m_1$  and  $m_2$ , we have  $\text{PCR6}^+(m_1, m_2) = \text{PCR6}(m_1, m_2)$  because these rules coincide when combining two BBAs. If we make the PCR6<sup>+</sup> fusion of the three BBAs altogether we obtain different results which is normal, because for  $S > 2$  one has in general  $\text{PCR6}^+(m_1, \dots, m_S) \neq \text{PCR6}(m_1, \dots, m_S)$ . For this example we get results shown in Table III.

We can verify that the result obtained by PCR6<sup>+</sup> fusion rule is more judicious than with PCR6 rule because the fusion of the almost vacuous BBA  $m_3(\cdot)$  has a very little impact in the fusion result as we intuitively expect. This is because the PCR6<sup>+</sup> combination rule discards the ignorant (or almost ignorant) information. With  $m_{1,2,3}^{\text{PCR6}^+}(\cdot)$ , the largest mass is allocated to  $A \cup B$  as with<sup>8</sup>  $m_{1,2}^{\text{PCR6}^+}(\cdot)$ , and contrariwise to  $m_{1,2,3}^{\text{PCR6}}(\cdot)$  when using the PCR6 fusion rule - see results in Table II.

TABLE III  
 $m_{1,2}^{\text{PCR6}^+}(\cdot)$  AND  $m_{1,2,3}^{\text{PCR6}^+}(\cdot)$  RESULTS.

Focal Elements	$m_{1,2}^{\text{PCR6}^+}(\cdot)$	$m_{1,2,3}^{\text{PCR6}^+}(\cdot)$
$B$	0.054877	0.054485
$A \cup B$	0.406987	0.407174
$C \cup D$	0.312886	0.312660
$A \cup B \cup C \cup D$	0.024917	0.025232
$E$	0.200333	0.200449
$\Theta$	0	0

## V. APPLICATION TO LEVEE CHARACTERIZATION

We now present the advantages of the new PCR6<sup>+</sup> rule for an application on a numerical case study representing a levee section. To do so, we use the geophysical and geotechnical information fusion methodology introduced in [2].

### A. Model and Information Sources

The figure 1 displays the structure of the levee, the location of the different layers and the representation of the study levee section.

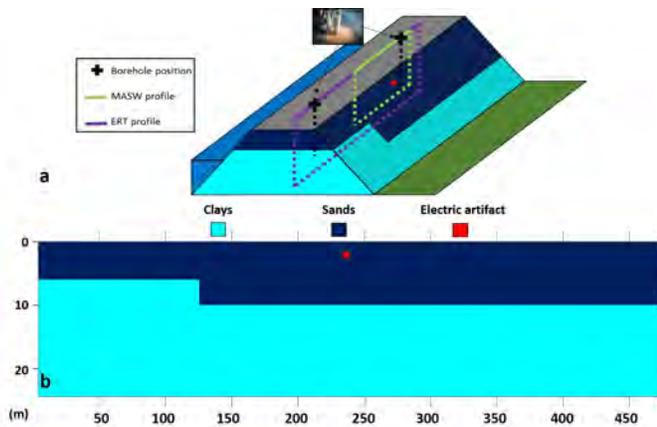


Fig. 1. a) Levee with position of investigation methods and b) materials in the section of interest.

The area is a lengthwise (parallel to the river) vertical section composed of two lithological materials: i) compact clays ( $C$  hypothesis) and ii) soft sands ( $S$  hypothesis). The

<sup>8</sup>We recall that one always has  $m_{1,2}^{\text{PCR6}^+}(\cdot) = m_{1,2}^{\text{PCR6}}(\cdot)$ .

sands are present over 6 meters thick on the first 125 meters of the section and over 10 meters thick after. Clayey materials are positioned below. A small electrically conductive anomaly is located near the surface in the center of the model. Thus, the FoD is defined such that  $\Theta = \{C, S, O\}$ . As required by the fusion method,  $O$  is an additional hypothesis standing for any other material different from the other two known. For this case study, two geophysical methods are used: the Electrical Resistivity Tomography (ERT) and the Multi-channel Analysis of Surface Waves (MASW). Two geotechnical boreholes providing information on the lithology are also considered in this study.

### B. BBA Distribution for Each Source

1) *Electrical Resistivity Tomography*: The basic principle of DC-resistivity methods consists in injecting an electric current of known intensity [A] by means of two "current" electrodes and measuring a voltage [V] between two "potential" electrodes. Such measurements are acquired for several positions of the current and the potential electrodes. Apparent resistivity values can then be computed and inverted to reconstruct a complete section of electrical resistivity [ $\Omega.m$ ]. From these electrical resistivity data, the fusion methodology [2] enables the BBA distribution depicted in Figure 2. The ERT characterization is disturbed by the conductive electrical artifact. Thus, clays are locally characterized in the center of the section while we know that sands are actually present. Also, the interface between clays and sands are not correctly defined.

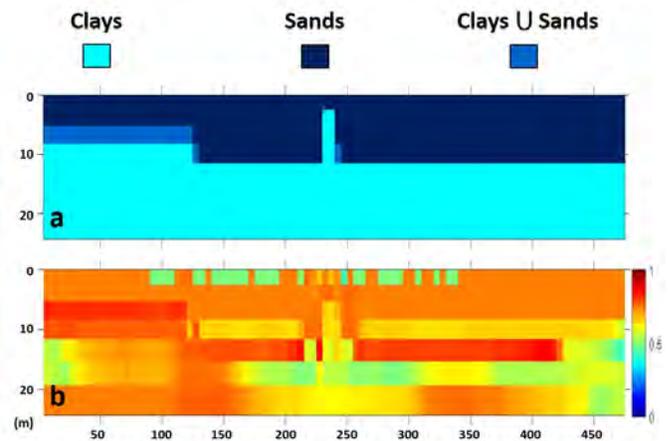


Fig. 2. a) Material with highest mass (from electrical resistivity data) and b) their mass values.

2) *Multichannel Analysis of Surface Waves*: The MASW method consists in studying the surface wave's dispersion (waveform deformation) to determine the shear wave's velocity [ $m.s^{-1}$ ]. A seismic source is generated at various locations and geophones are aligned on the ground surface to record the seismic waves arrival times. The use of this method comprises three stages: (i) the data acquisition, (ii) the determination of the Rayleigh dispersion curve, and (iii) the inversion process

with the determination of the shear velocities. In this work, the seismic acquisition is carried out from  $x = 212$  m to  $x = 428$  m. From the shear wave velocity data, the associated BBA distribution is displayed in Figure 3. The MASW characterization is not disturbed by the electrical artifact. Thus, the method characterizes correctly the two lithological materials as well as the lithological interface position. However,  $\Theta$  is characterized in most part of the section (in black, Figure 3.a), where no data is available.

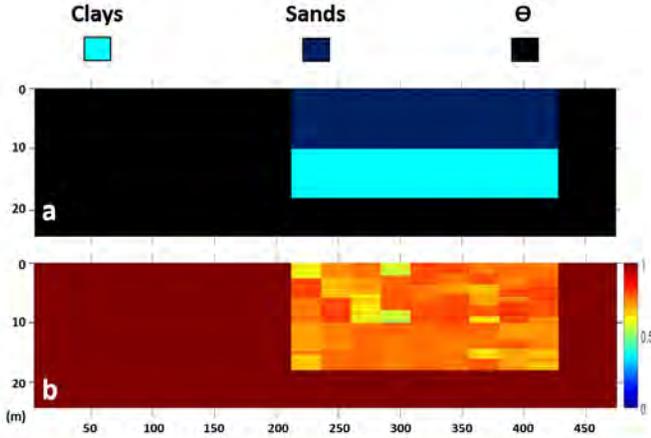


Fig. 3. a) Material with highest mass (from shear wave velocity data) and b) their mass values.

3) *Core Drillings*: Two core drillings with particle size analysis are simulated at  $x = 80$  m and  $x = 350$  m from the surface to 20 m depth. From the simulated geotechnical data, the associated BBA distribution is displayed in Figure 4. The lithological materials are correctly characterized but an important area of uncertainty remains between 6 and 10 m depth. Indeed, since two different materials are identified in both boreholes at such depths, the section is poorly defined between them [2].

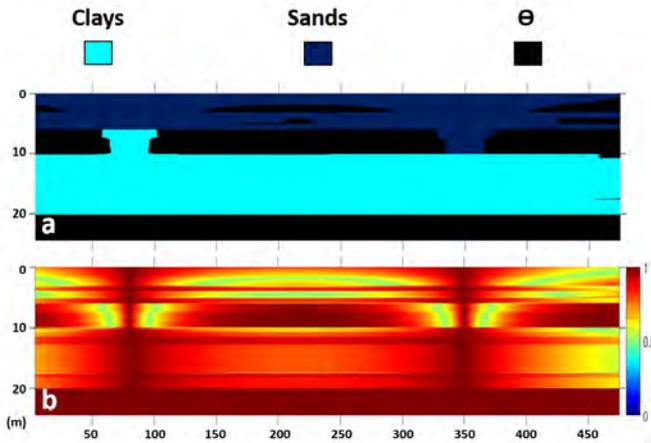


Fig. 4. a) Material with highest mass (from two borehole data) and b) their mass values.

### C. PCR6 and PCR6<sup>+</sup> Fusion Results

The fusion results using PCR6 and PCR6<sup>+</sup> rules are respectively depicted in Figures 5.a-b and Figures 5.c-d. These results highlight the lack of characterization at the center of the model using PCR6 rule (in the red boxes, Figure 5.a). Indeed,  $\Theta$  is characterized while PCR6<sup>+</sup> rule enables to correctly characterize sands. For PCR6, this area is difficult to define since the ERT suggests the presence of clays, the MASW suggests the presence of sands and the geotechnical source of information is ignorant. However, PCR6<sup>+</sup> rule manages to allocate the conflictual masses on the individual hypothesis instead of  $\Theta$ . Furthermore, the global belief mass values are greater with PCR6<sup>+</sup> rule (Figure 5.d) than with PCR6 (Figure 5.c). This improvement in the results could be valuable in the context of an investigation campaign on a real earthen structure. Indeed, knowing the nature of the materials as well as their location is crucial to achieve a good diagnosis and limit the risk of breakage. Since many investigation methods can be ignorant or partially ignorant in the context of levee characterization, this new combination rule would be of great operational interest to give credit to the most informative source and to avoid uncharacterized areas inside the earthen structure.

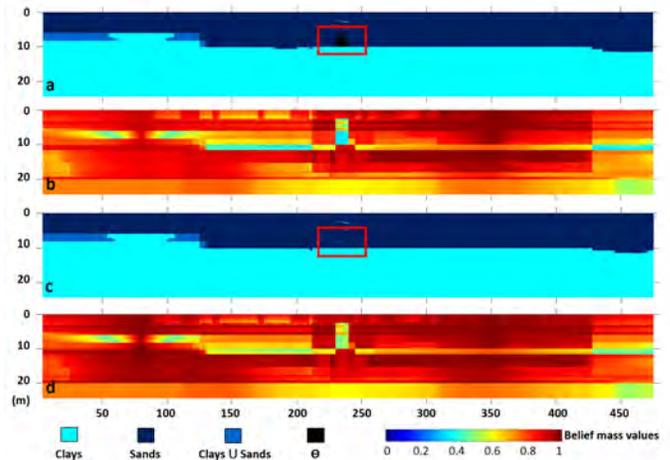


Fig. 5. Material with highest mass (from ERT, MASW and core drillings), using PCR6 (a) and PCR6<sup>+</sup> (c) rules, with area of interest in red box. (b) and (d) mass values associated with the hypothesis depicted in (a) and (c).

## VI. CONCLUSIONS

In this work, after having introduced the belief functions as well as conjunctive, DS and PCR6 rules of combination, we presented the flawed behavior of PCR6 rule. We then described improvements to correct these behaviors, introducing a new PCR6<sup>+</sup> rule. The computation of a keeping index, making it possible to discard ignorant information sources for the calculation of each partial conflict, was detailed. This keeping index has been integrated into the original formulation of PCR6 in order to ensure the neutrality property of the vacuous BBA. The interest of such combination rule has finally been demonstrated for an application on a numerical levee section

with simulated geophysical and geotechnical acquisitions. As a following perspective, we wish to apply this new PCR6<sup>+</sup> rule to risk analysis issues with data fusion acquired from real investigation campaigns.

#### ACKNOWLEDGMENTS

The authors are grateful to Sérgio Palma Lopes and Yannick Fargier for their contributions in the geophysical and geotechnical fusion methodology used in this work.

#### REFERENCES

- [1] M. Sharp et al., *The international levee handbook*, CIRIA, London, 2013.
- [2] T. Dezert, S.P. Lopes, Y. Fargier, P. Côte, *Combination of geophysical and geotechnical data using belief functions: Assessment with numerical and laboratory data*, Journal of Applied Geophysics, Vol. 170, 2019.
- [3] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., Vol. 38, pp. 325–339, 1967.
- [4] A.P. Dempster, *A generalization of Bayesian inference*, J. of Royal Sta. Soc. B30, pp. 205–247, 1968.
- [5] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [6] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion (Collected Works)*, Volumes 1–4. American Research Press, 2004–2015.
- [7] P. Halmos, *Naive Set Theory*, Springer-Verlag, 1974.
- [8] X.-R. Li, *Probability, random signals, and statistics*, CRC Press, 1999.
- [9] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Int. Conf. on Information Fusion (Fusion 2012), Singapore, July 2012.
- [10] A. Tchamova, J. Dezert, *On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory*, in Proc. of 6th IEEE Int. Conf. on Int. Syst. (IEEE IS'2012), Sofia, Bulgaria, Sept. 6–8, 2012.
- [11] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., Vol. 29(3), pp. 223–252, 2014.
- [12] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, in [6], Volume 2, Chap 2, 2006.
- [13] A. Martin, C. Osswald, J. Dezert, F. Smarandache, *General combination rules for qualitative and quantitative beliefs*, Journal of Advances in Information Fusion (JAIF), Vol. 3(2), Dec. 2008.
- [14] F. Smarandache, J. Dezert, *Information fusion based on new proportional conflict redistribution rules*, in Proc. of Int. Conf. on Information Fusion (Fusion 2005), Philadelphia, PA, USA, July 25–29, 2005.
- [15] F. Smarandache, J. Dezert, *Proportional conflict redistribution rules for information fusion*, in [6], Volume 2, Chap 1, 2006.
- [16] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of Int. Conf. on Information Fusion (Fusion 2010), Edinburgh, Scotland, UK, July 26–29, 2010.
- [17] T. Dezert, J. Dezert, F. Smarandache, *Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments*, Journal of Advances in Information Fusion (JAIF), Vol. 16, No. 1, June 2021.



# Multi-Criteria Information Fusion for Storm Prediction Based on Belief Functions

Jean Dezert, Aurélie Bouchard, Magalie Buguet  
 ONERA - The French Aerospace Lab  
 F-91761 Palaiseau, France.

Emails: jean.dezert@onera.fr, aurelie.bouchard@onera.fr, magalie.buguet@onera.fr

Originally published as: J. Dezert, A. Bouchard, M. Buguet, *Multi-Criteria Information Fusion for Storm Prediction Based on Belief Functions*, in Proc. of Int. Conf. on Information Fusion (Fusion 2021), South Africa, November 2021, and reprinted with permission.

**Abstract**—The objective of this paper is to present a general methodology for storm risk assessment and prediction based on several physical criteria thanks to the belief functions framework to deal with conflicting meteorological information. For this, we adapt the Soft ELECTRE TRI (SET) approach to this storm context and we show how to use it on outputs of atmospheric forecast model, given an estimate of the state of the atmosphere in a future time. This work could also serve as a benchmark for other methods dealing with multi-criteria decision-making (MCDM) support and conflicting information fusion.

**Keywords:** storm risk assessment, information fusion, belief functions, decision-making, Soft ELECTRE TRI.

## I. INTRODUCTION

In the context of storm prediction, many sources of observations of the atmosphere may be used. The aim of storm risk assessment is to exploit as best as possible some of these available data to evaluate the risk of thunderstorm at a given location in the surveillance area under concern in a close future. Each type of data is associated to a given source of information called a criterion in our context. In the present paper, the data used are coming from a numerical weather prediction model. These kinds of models allow to simulate the evolution of the state of the atmosphere by solving dynamical and thermodynamical equations, by including data assimilation of observations of the atmosphere (from satellite, rawinsonde or buoys, for instance) and by adding physical parametrization for unresolved processes as convection. The outputs of the Global Forecast System (GFS), developed by the Centers for Environmental Prediction (NCEP) have been used for our study [1]. The estimation of storm risk level is of prime importance for many applications (aeronautical safety, air traffic management, ...). In this work we present a general methodology showing how to use belief functions [2] coupled with the Soft ELECTRE TRI (SET) outranking method [3] to manage efficiently the conflicting sources of information in a multi-criteria decision-making context.

This paper is organized as follows. After a short presentation of the Soft ELECTRE TRI outranking method in Section II for Multi-Criteria Decision-Making (MCDM) support we introduce the storm risk assessment problematic in Section III, and we show how it can fit well with the Soft ELECTRE TRI framework. We also provide an example of our storm assessment methodology based on data set coming from the atmospheric forecast supplied by GFS, and we show the

performances of SET approach with respect to “ground truth” obtained by the World Wide Lightning Location Network (WWLLN) [4]. Conclusions are given in section IV with some perspectives.

## II. SOFT ELECTRE TRI FOR MCDM

### A. A short presentation of SET

The Soft ELECTRE TRI method (SET) proposed in [3] is an evolution of the ELECTRE TRI (ET)<sup>1</sup> method proposed by Roy in [5] for making the outranking of alternatives with respect to profiles of categories. The SET method is based on belief functions calculus [2] (see appendix) and improves the classical ET method because it does not require an arbitrary choice of  $\lambda$ -cut strategy for making the outranking of alternatives with respect to profiles of categories, nor an ad-hoc choice of decisional attitude for making the final assignment. Actually, the SET method solves the assignment problem in a soft manner. The Fig. 1 shows the general MCDM problem that can be addressed by the SET method. More precisely, SET solves an assignment problem in complex situations where a (or several) given alternative has to be assigned to predetermined categories based on multiple criteria values. Each criterion  $G_j$  ( $j = 1, \dots, n_G$ ) is evaluated quantitatively. Each profile is defined by the green points limiting the bounds of each category with respect to each criterion. The red chain represents a “multi-criteria value” (i.e. an alternative  $\mathbf{a}$ ) that one wants to assign to a predefined category.

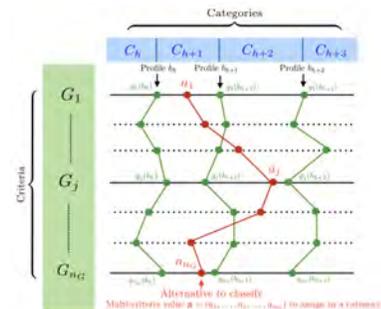


Figure 1: How to assign a category to an alternative?

<sup>1</sup>The acronym ELECTRE stands for “ÉLImination Et Choix Traduisant la Réalité” (Elimination and Choice Expressing the Reality) [7], [8].

In the context of our storm risk assessment application each category  $C_h$  corresponds to a level of risk (low, moderate, strong, very strong, or extreme), and it is defined by its profile lower and upper bounds denoted respectively by  $\mathbf{b}_{h-1}$  and  $\mathbf{b}_h$  (see vertical green plots in Fig. 1). The profiles bounds define ad-hoc categories for the values of each criterion  $G_j$  corresponding to a meteorological parameter which is either a direct measure of a meteorological parameter, or an estimation of the parameter resulting from a sophisticated meteorological forecast model. These meteorological parameters (i.e. criteria) will be described in details in section III. Any alternative corresponds to a multi-criteria value  $\mathbf{a} = (a_1, \dots, a_j, \dots, a_{n_G})$  whose component  $a_j$  is nothing but the instance of the parameter (i.e. criterion)  $G_j$  for this alternative  $\mathbf{a}$ . Each alternative  $\mathbf{a}$  is associated with a (2D) cell of the area of interest on the surface of the Earth.

The SET method allows to take into account the weight of importance of each criterion entering this assignment problem and to give a soft (i.e. probabilistic) assignment solution to commit any multi-criteria value  $\mathbf{a}$  to a category. More precisely, SET calculates the probability that a chosen alternative  $\mathbf{a}$  belongs to a predetermined category  $C_h$  based on all information available (the criteria values, the importances of the criteria, and the bounds of each predetermined category).

### B. SET principle

We present briefly the principle and the steps of the Soft ELECTRE TRI (SET) outranking method developed originally in [3]. SET makes a soft assignment of  $n_a \geq 1$  alternatives  $\mathbf{a}_i$  in predefined ordered categories  $C_h$  ( $h = 1, \dots, n_h$ ) according to criteria<sup>2</sup> measure  $g_j(\cdot)$ ,  $j \in \mathbf{J} = \{1, \dots, n_G\}$ . Each category  $C_h$  is delimited by the set of its lower and upper limits  $b_{h-1}^j$  and  $b_h^j$  with respect to each criterion  $G_j$  measured by  $g_j(\cdot)$ . By convention,  $b_0^j \leq b_1^j \leq \dots \leq b_{n_h}^j$ .  $\mathbf{b}_0 = (b_0^1, \dots, b_0^j, \dots, b_0^{n_G})$  is the lower (minimal) profile bound and  $\mathbf{b}_{n_h} = (b_{n_h}^1, \dots, b_{n_h}^j, \dots, b_{n_h}^{n_G})$  is the upper (maximal) profile bound. The overall profile  $\mathbf{b}_h$  is defined by  $(g_1(b_h), g_2(b_h), \dots, g_{n_G}(b_h))$ , and it is represented by the vertical plot joining the green dots in Fig. 1.

The outranking relations used in SET are based on the calculation of partial concordance and discordance indices from which global concordance and credibility indices are derived based on Basic Belief Assignment (BBA) modeling [2], and on an advanced fusion technique based on Proportional Conflict Redistribution rule no. 6 (PCR6) [9]–[11]. A soft assignment of each alternative  $\mathbf{a}_i$  in a predetermined category is obtained by the calculation of the probabilized outranking relations, from which a final hard assignment can be drawn (if needed) for some action. In the storm risk assessment context, an action for instance could be the broadcast of an alert message to the air traffic management organisms or airports.

The Soft ELECTRE TRI method requires the following four steps:

<sup>2</sup>In our context, a criteria is a meteorological parameter.

**SET-Step 1:** Calculation of partial (local) concordance indices  $c_j(\mathbf{a}_i, \mathbf{b}_h)$ , partial discordance indices  $d_j(\mathbf{a}_i, \mathbf{b}_h)$ , and also partial uncertainty indices  $u_j(\mathbf{a}_i, \mathbf{b}_h)$  between an alternative  $\mathbf{a}_i$  and a profile  $\mathbf{b}_h$  thanks to a smooth sigmoidal model [12]. The partial indices are encapsulated in BBAs  $m_{ih}^j(\cdot)$  for alternative  $\mathbf{a}_i$  versus profile  $\mathbf{b}_h$  (i.e.  $\mathbf{a}_i$  vs.  $\mathbf{b}_h$ ) as follows:

$$\begin{cases} c_j(\mathbf{a}_i, \mathbf{b}_h) \triangleq m_{ih}^j(c) & \text{(local concordance)} \\ d_j(\mathbf{a}_i, \mathbf{b}_h) \triangleq m_{ih}^j(\bar{c}) & \text{(local discordance)} \\ u_j(\mathbf{a}_i, \mathbf{b}_h) \triangleq m_{ih}^j(c \cup \bar{c}) & \text{(local uncertainty).} \end{cases} \quad (1)$$

where  $m_{ih}^j(\cdot)$  is a Basic Belief Assignment (BBA) defined on the frame of discernment  $\Theta \triangleq \{c, \bar{c}\}$ , where  $c$  means that the alternative  $\mathbf{a}_i$  is concordant (i.e. it agrees) with the assertion ‘ $\mathbf{a}_i$  is at least as good as profile  $\mathbf{b}_h$ ’, and  $\bar{c}$  means that the alternative  $\mathbf{a}_i$  is opposed to this assertion (i.e. it is discordant, or it disagrees with this assertion). For each criterion  $G_j$ , a BBA  $m_{ih}^j$  defined on the power-set of  $\Theta$  is obtained by the fusion of two simple BBAs  $m_1^j(\cdot)$  and  $m_2^j(\cdot)$  based on the following sigmoid models, see [12] for justification and details. Similarly, we compute also partial (local) concordance indices  $c_j(\mathbf{b}_h, \mathbf{a}_i)$ , partial discordance indices  $d_j(\mathbf{b}_h, \mathbf{a}_i)$ , and partial uncertainty indices  $u_j(\mathbf{b}_h, \mathbf{a}_i)$  between a profile  $\mathbf{b}_h$  and an alternative  $\mathbf{a}_i$ . This partial indices are encapsulated in BBAs  $m_{hi}^j(\cdot)$  for profile  $\mathbf{b}_h$  versus alternative  $\mathbf{a}_i$  (i.e.  $\mathbf{b}_h$  vs.  $\mathbf{a}_i$ ).

**SET-Step 2:** Calculation of the global (overall) concordance indices  $c(\mathbf{a}_i, \mathbf{b}_h)$ ,  $c(\mathbf{b}_h, \mathbf{a}_i)$ , discordance indices  $d(\mathbf{a}_i, \mathbf{b}_h)$ ,  $d(\mathbf{b}_h, \mathbf{a}_i)$ , and uncertainty indices  $u(\mathbf{a}_i, \mathbf{b}_h)$ ,  $u(\mathbf{b}_h, \mathbf{a}_i)$  by the fusion of local indices. More precisely, one must calculate

$$\begin{cases} m_{ih}(\cdot) = [m_{ih}^1 \oplus m_{ih}^2 \oplus \dots \oplus m_{ih}^{n_G}](\cdot) \\ m_{hi}(\cdot) = [m_{hi}^1 \oplus m_{hi}^2 \oplus \dots \oplus m_{hi}^{n_G}](\cdot) \end{cases} \quad (2)$$

where  $\oplus$  denotes symbolically a chosen fusion operator.

To take into account the weight of importance  $w_j \in [0, 1]$  of each criterion  $G_j$ , we propose two fusion methods:

- 1) Fusion method 1: we use the weighting averaging (WA) fusion rule because it is a very simple rule, and it can be processed very quickly. This is of prime importance in our storm risk assessment context because one can have millions of cells (depending the resolution cell we want to work with) in a wide surveillance areas.
- 2) Fusion method 2: we use a more sophisticated PCR6 fusion rule adapted with importance discounting presented in details in [13] if more computational power is available<sup>3</sup>.

Once the BBAs  $m_{ih}(\cdot)$  and  $m_{hi}(\cdot)$  are obtained, the global indices are defined by

$$\begin{cases} c(\mathbf{a}_i, \mathbf{b}_h) \triangleq m_{ih}(c)\alpha(\mathbf{a}_i, \mathbf{b}_h) \\ d(\mathbf{a}_i, \mathbf{b}_h) \triangleq m_{ih}(\bar{c})\beta(\mathbf{a}_i, \mathbf{b}_h) \\ u(\mathbf{a}_i, \mathbf{b}_h) \triangleq 1 - c(\mathbf{a}_i, \mathbf{b}_h) - d(\mathbf{a}_i, \mathbf{b}_h). \end{cases} \quad (3)$$

<sup>3</sup>Due to the complexity of this fusion rule and computational burden, only problems of relatively small dimensions, say for  $n_G \leq 6$ , can be addressed by this second method.

The discounting factors  $\alpha(\mathbf{a}_i, \mathbf{b}_h)$  and  $\beta(\mathbf{a}_i, \mathbf{b}_h)$  in (3) are defined in [3]. They are not given here due to space limitation restraints.  $c(\mathbf{b}_h, \mathbf{a}_i)$ ,  $d(\mathbf{b}_h, \mathbf{a}_i)$  and  $u(\mathbf{b}_h, \mathbf{a}_i)$  are similarly computed using the dual formula of (3).

The belief and plausibility of the outranking propositions  $X = \mathbf{a}_i > \mathbf{b}_h$  ( $\mathbf{a}_i$  outranks  $\mathbf{b}_h$ ), and  $Y = \mathbf{b}_h > \mathbf{a}_i$  ( $\mathbf{b}_h$  outranks  $\mathbf{a}_i$ ) are then given by

$$\text{Bel}(X) = c(\mathbf{a}_i, \mathbf{b}_h) \text{ and } \text{Bel}(Y) = c(\mathbf{b}_h, \mathbf{a}_i) \quad (4)$$

$$\text{Pl}(X) = 1 - d(\mathbf{a}_i, \mathbf{b}_h) \text{ and } \text{Pl}(Y) = 1 - d(\mathbf{b}_h, \mathbf{a}_i) \quad (5)$$

**SET-Step 3:** Calculation of the probabilized outranking relations. In SET-Step 2 we have characterized the outrankings  $X = \mathbf{a}_i > \mathbf{b}_h$  and  $Y = \mathbf{b}_h > \mathbf{a}_i$  by their imprecise probabilities  $P(X) \in [\text{Bel}(X); \text{Pl}(X)]$  and  $P(Y) \in [\text{Bel}(Y); \text{Pl}(Y)]$ . Solving the outranking problem consists in choosing (deciding) if finally  $X$  dominates  $Y$  (in such case we must decide  $X$  as being the valid outranking), or if  $Y$  dominates  $X$  (in such case we decide  $Y$  as being the valid outranking). This hard assignment problem is difficult in general because  $P(X)$  in  $[\text{Bel}(X); \text{Pl}(X)]$  and  $P(Y)$  in  $[\text{Bel}(Y); \text{Pl}(Y)]$  and these belief intervals can partially overlap. Fortunately, a soft (probabilized) outranking solution is possible by computing the probability that  $X$  dominates  $Y$  (or that  $Y$  dominates  $X$ ) by assuming uniform distribution of unknown probabilities between their lower and upper bounds. To get the probabilized outrankings, we have to calculate  $P_{X>Y} \triangleq P(P(X) > P(Y))$  and  $P_{Y>X} \triangleq P(P(Y) > P(X))$  which are given by the ratio of two polygonal areas, or can be estimated using sampling techniques, as explained and illustrated in [3]. The probabilities of outrankings are denoted  $P_{ih} \triangleq P_{X>Y}$  where  $X \triangleq \mathbf{a}_i > \mathbf{b}_h$  and  $Y \triangleq \mathbf{b}_h > \mathbf{a}_i$ . Reciprocally, we denote  $P_{hi} \triangleq P_{Y>X} = 1 - P_{ih}$ . This probabilization of outrankings is directly obtained by this Step 3 of SET, and thus eliminates the arbitrary  $\lambda$ -cut strategy used in classical ELECTRE TRI method.

**SET-Step 4:** Final soft assignment of  $\mathbf{a}_i$  into a category  $C_h$ . From the probabilized outrankings obtained in SET-Step 3, we can make directly the soft assignment of alternatives  $\mathbf{a}_i$  to categories  $C_h$  defined by their profiles  $\mathbf{b}_h$ . This is easily obtained by the combinatorics of all possible sequences of outrankings taking into account their probabilities  $P_{ih}$  to calculate all the assignment probabilities  $P(\mathbf{a}_i \rightarrow C_h)$ . Moreover, this soft assignment mechanism provides also the probability  $\delta_i \triangleq P(\mathbf{a}_i \rightarrow \emptyset)$  reflecting the impossibility to make a coherent outranking. This soft assignment procedure of the SET method does not require an arbitrary choice of decisional attitude unlike to what is proposed in the classical ET method. A simple detailed example of this SET-Step 4 is given in [3] for convenience.

### III. APPLICATION OF SET TO STORM RISK ASSESSMENT

#### A. Surveillance zone and data set

We apply the SET method briefly presented in the previous section to storm risk assessment problematic. For this, we consider in this study five meteorological parameters (i.e. criteria)

drawn from GFS (Global Forecast System) open data available on the web [1]. We have used GFS data for the 9th May 2016 at 3h UTC. The GFS data used in this study are available in [22]. The wide surveillance area covers Atlantic ocean from  $[-1, 70.5]$  degrees in latitude, and  $[-100, 10.5]$  degrees in longitude. We have 21592 cells of size  $0.5 \times 0.5 \text{ deg}^2$  to evaluate. Each cell corresponds to an alternative  $\mathbf{a}_i$  that must be assigned to a storm risk category  $C_h$  by the SET method. The table I shows the five ( $n_G = 5$ ) meteorological parameters (i.e. criteria) used in our study, their units, their preference ordering, and their qualitative importance chosen for this problematic.

Criteria	Units	Preference ordering	Importance weight
$G_1 = \text{PConv}$	$kg/m^2$	increasing	very high
$G_2 = \text{LI}$	$^\circ K$	decreasing	very high
$G_3 = \text{CAPE}$	$J/kg$	increasing	high
$G_4 = \text{DivB}$	$s^{-1}$	decreasing	low
$G_5 = \text{DivS}$	$s^{-1}$	increasing	low

Table I: Criteria used for SET method.

where

- PConv is the 3-h accumulated precipitation induced by convective process (in  $kg/m^2$ ) [24];
- LI is the lifted index which characterizes the instability of the atmosphere (in  $^\circ K$ ). This parameter, developed by Galway [25], is the gap between the environmental temperature and the temperature of a parcel lifted dry-adiabatically to saturation then moist-adiabatically to 500 hPa;
- CAPE is the convective available potential energy (in  $J/kg$ ). This parameter is the potential energy available to the parcel to lift up beyond the level of free convection. As the lifted index, this parameter relies on the difference between the environmental temperature and the temperature of a parcel lifts adiabatically [26];
- DivB is the low-level wind divergence if there is convective clouds in the cell (in  $s^{-1}$ ). This parameter is derived from horizontal wind component and the pressure level of the bottom of convective cloud [27];
- DivS is the divergence of the wind above the top of the convective clouds (in  $s^{-1}$ ). Indeed, isolated storm cloud is associated with low-level wind convergence and divergence near the top of the cloud [27];

PConv has an increasing preference order which means that bigger the PConv value is, higher is the storm risk. LI has a decreasing preference order meaning that lower the LI value is, higher is the storm risk. To work with quantitative importance weights, we need to transform qualitative labels (low, high, very high) into numerical values. For this we use the following mapping: Low importance  $\mapsto 1$ , Moderate importance  $\mapsto 2$ , High importance  $\mapsto 3$  and Very high importance  $\mapsto 4$ . This mapping is quite ad-hoc, and could be changed/adapted for reflecting a better subjective interpretation of the importance level expressed by the expert who provides these qualitative importance factors. This mapping is specific of the fusion

system design. After the normalization of the numerical importance factors, we get the following normalized weights of importance of each criterion<sup>4</sup>:  $w_1 = 4/12$ ,  $w_2 = 4/12$ ,  $w_3 = 2/12$ ,  $w_4 = 1/12$ , and  $w_5 = 1/12$ .

In our study, we consider five ( $n_h = 5$ ) levels of risk defined qualitatively as: Low, Moderate, Strong, Very strong and Extreme which respectively correspond quantitatively to the levels 1, 2, 3, 4 and 5 that will be shown in on the followings figures. The profile bounds for each category of risk and each criterion are given in Table II.

Criteria\Bounds of categories	$b_1$	$b_2$	$b_3$	$b_4$
$G_1 = \text{PConv}$	0	1.5	7	10
$G_2 = \text{LI}$	0	-2	-6	-10
$G_3 = \text{CAPE}$	0	1000	2000	4000
$G_4 = \text{DivB}$	0	-0.1	-0.5	-1
$G_5 = \text{DivS}$	0	0.1	0.5	1

Table II: Bounds of categories of risk.

For convenience, the figure 2 presents the flow chart of the proposed method related to our storm application.

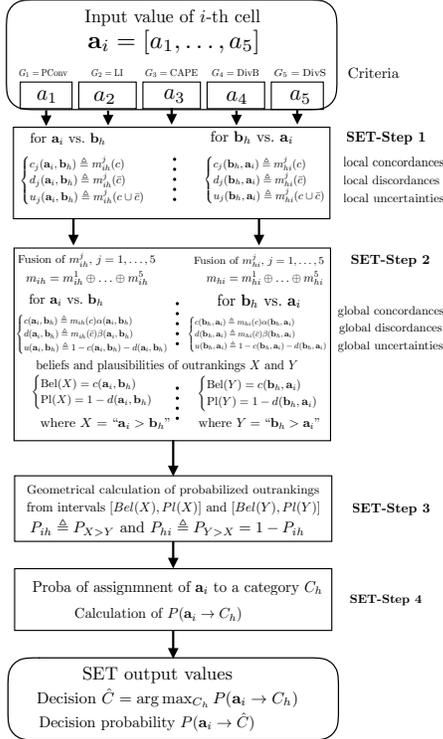


Figure 2: Flow chart of the method related to the application.

The bounds for LI and CAPE criteria are those defined by Wesoleck in [23]. The bounds for PConv criteria have been deduced from the different thresholds used to distinguish light, moderate and heavy rainfall [28] and adapted to our geographical area. Heavy rain, correspondings to accumulated precipitation above 10-30 mm/h, is associated to severe storm

<sup>4</sup>In this work and for simplicity, the importance factor  $w_j$  of each criteria  $G_j$  used to make the importance discounting of BBAs in the SET method is chosen independently of the profile bound values.

risk [29]. The bounds for low-level convergence wind and high-level divergence wind have been chosen larger than usual threshold [30], because the most important information is the sign of the divergence merge with the presence of convective cloud in the cell. Hence, if the PConv value for the cell under analysis is greater than  $b_4^{j=1} = 20$ , then the storm risk for this cell is considered as extreme (risk=5). If the PConv value belongs to  $(b_3^{j=1}, b_4^{j=1}] = (10, 20]$  then the storm risk for this cell is considered as very high (risk=4), etc. If the LI value is lower than  $b_4^{j=3} = -10$  then the storm risk for this cell is considered as extreme (risk=5), but if the LI value is between  $b_4^{j=3} = -10$  and  $b_3^{j=3} = -6$  the storm risk for this cell is considered only as very high (risk=4).

The figures 3–7 show the risk levels (1, 2, 3, 4 and 5) corresponding to each criterion for the 21592 cells covering the Atlantic ocean surveillance area for each criterion considered separately. The Dark blue cells with values -1 correspond to ground cells which are not taken into account in this study. One clearly sees the difference of risks drawn from the five meteorological parameters and the conflicting information between these five maps of risks of storm that illustrate the input data we have to process by the SET method to get the global risk assessment.

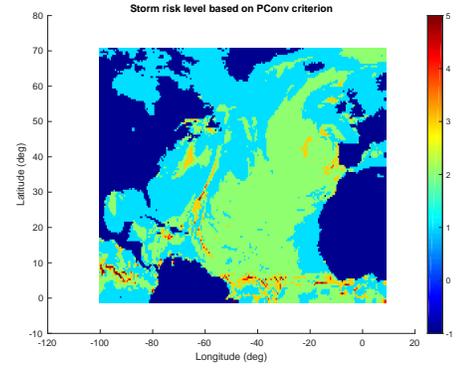


Figure 3: Storm risk levels based on PConv criterion. 0 means no risk; 1, low level of risk; 2, moderate level of risk; 3, strong level of risk; 4, very strong level of risk and 5, extreme risk. Risk are not calculated over earth, fixed to -1 value (dark blue).

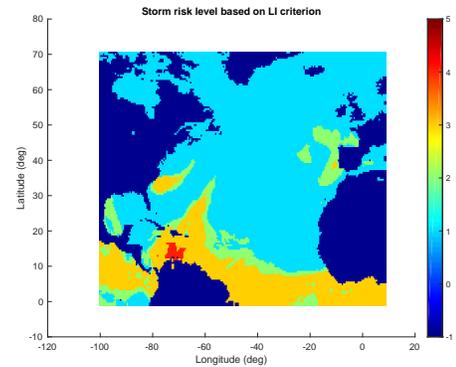


Figure 4: Storm risk levels based on LI criterion.

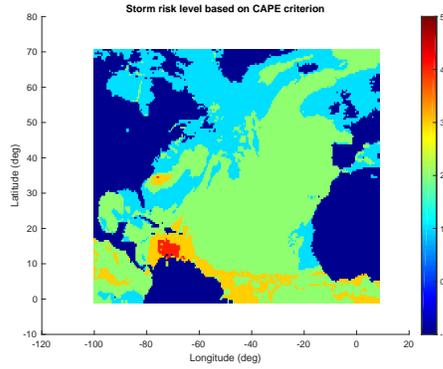


Figure 5: Storm risk levels based on CAPE criterion.

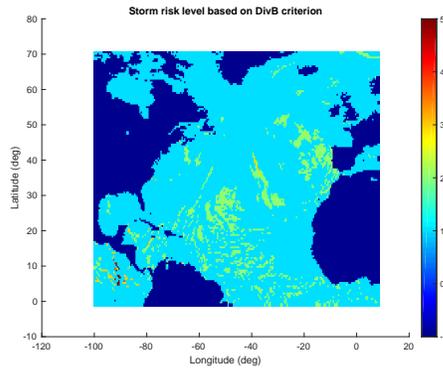


Figure 6: Storm risk levels based on DivB criterion.

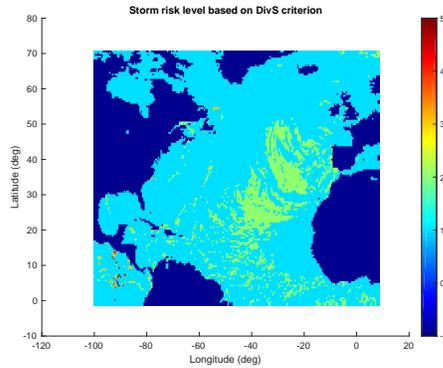


Figure 7: Storm risk levels based on DivS criterion.

To estimate the variability (randomness) of GFS data in each cell, we estimate for each category  $C_h$  of risk the probability  $P(C_h)$  by counting the number of criteria associated with  $C_h$  divided by  $n_G = 5$ . This level of randomness is characterized by Shannon's entropy. Hence, for the cell # $i$ , if we have the probability measure  $\mathbf{p}_i = (p_1, p_2, \dots, p_{n_h})$ , the normalized Shannon entropy<sup>5</sup> is given by

$$H(\mathbf{p}_i) = -\frac{1}{\log_2 n_h} \cdot \sum_{h=1}^{n_h} p_h \log_2 p_h \quad (6)$$

<sup>5</sup>One takes  $p_h \log_2 p_h = 0$  if  $p_h = 0$ .

$H(\mathbf{p}_i) = 0$  when all meteorological parameters agree with the same storm risk category, and  $H(\mathbf{p}_i)$  is maximum if  $\mathbf{p}_i$  is the uniform pmf. One defines the mean entropy  $\bar{H}$  of the GFS data by averaging the entropy values of the  $N = 21592$  cells of the surveillance area by

$$\bar{H} = \frac{1}{N} \sum_{i=1}^N H(\mathbf{p}_i) \quad (7)$$

Figure 8 shows the normalized entropies of the meteorological GFS data we have used in this study. The mean entropy of these GFS data is  $\bar{H} = 0.2989$ , and only 32% of the data are totally in agreement on the same risk level (shown in green color on Fig 8). As we observe on this figure most of the data are conflicting because the entropies values are much bigger than zero. The high level of randomness of these data justifies a sophisticated MCDM method able to deal efficiently with conflicting sources of information. This motivates the use of SET approach proposed in this work.

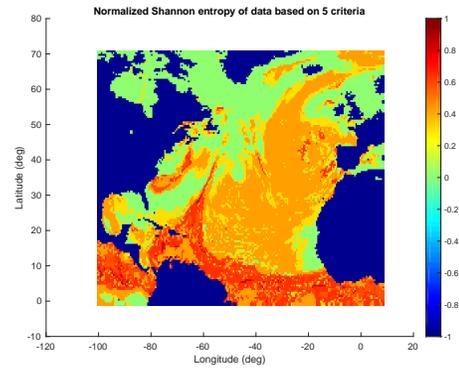


Figure 8: Normalized Shannon entropy of GFS data.

### B. Results based on the weighted averaging rule

The figure 9 shows the storm risk levels based on the weighted average<sup>6</sup> of risk levels shown figures 3–7, with the weights of importance  $w_1 = 4/12$ ,  $w_2 = 4/12$ ,  $w_3 = 2/12$ ,  $w_4 = 1/12$ , and  $w_5 = 1/12$ .

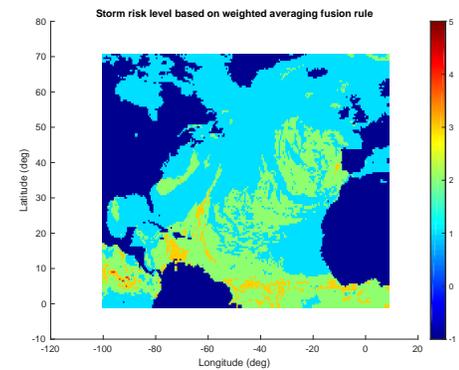


Figure 9: Storm risk levels based on weighted average of risks.

<sup>6</sup>For representation convenience and comparison with SET results, the risk values of Fig. 9 have been rounded to their closest integer value.

Based on this simple fusion rule one observes that the strong (and higher) risks of thunderstorms are located mainly in the intertropical convergence zone (around the equatorial line), on the Caribbean Sea, and aside the Portugal coast. However the method of fusion does not provide a measure of the trustfulness (confidence) of this result, and it does not manage precisely the level of conflict between the different sets of data.

C. Results based on the SET approach

The figure 10 shows the map of storm risk levels based on weighted averaging fusion rule used in SET-step 2, whereas the figure 11 shows the result when the PCR6 fusion rule<sup>7</sup> is used in SET-step 2. These two resulting maps of risk levels can be interpreted as the SET-combination of maps shown in figures taking into account the importance and contradiction of the five meteorological criteria.

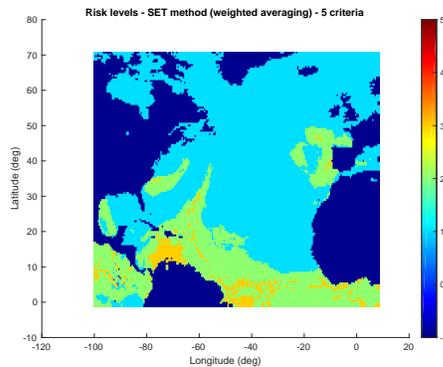


Figure 10: Storm risk levels based on SET (averaging rule).

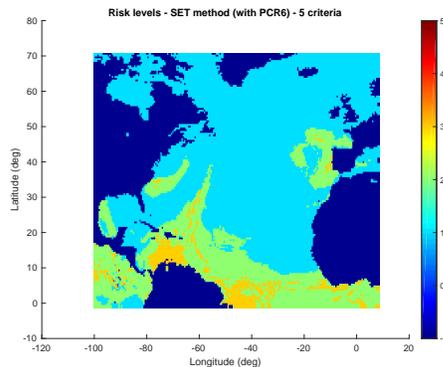


Figure 11: Storm risk levels based on SET (PCR6 rule).

The confidence in the resulting storm risk maps of figures 10 and 11 are shown in figures 12 and 13 respectively.

D. Performances analysis

To measure the performance of our method of storm risk assessment we need to compare our SET results with some ground truth. For this, we consider as ground truth the

<sup>7</sup>with importance discounting of BBAs, as explained in [13].

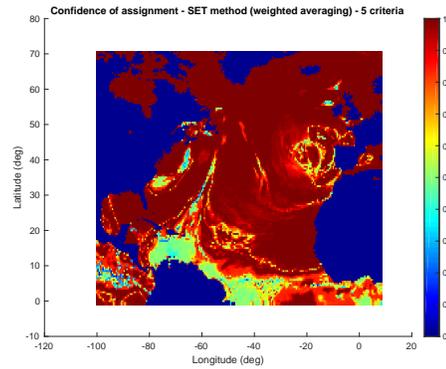


Figure 12: Confidence in decision (SET with averaging rule).

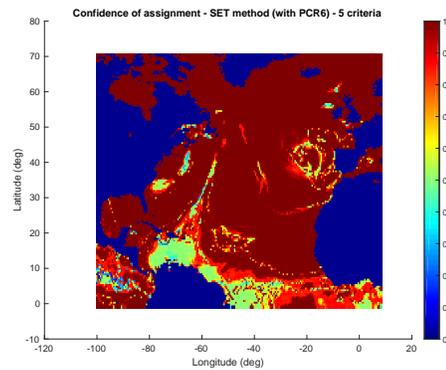


Figure 13: Confidence in decision (SET with PCR6 rule).

information of location of strokes supplied by the World Wide Lightning Location Network (WWLLN) [4]. WWLLN archival data are copyrighted by the University of Washington and are available to the public at nominal cost. For a given date, at a time T, all cells where strokes impacts have been detected by WWLLN network, in the time interval [T +/- 1h30], have been tagged. These data consist of the  $N_L = 223$  locations of all the detected lightnings on May 9th, 2016 in the time interval [1h30 - 4h30 UTC] which are shown as red dots in Fig 14.

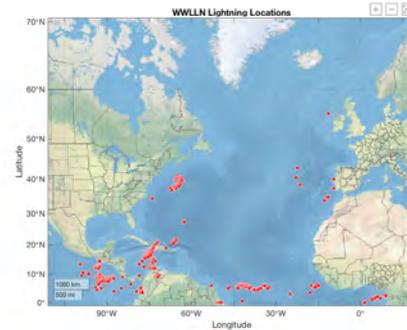


Figure 14: WWLLN lightning detections.

The performance of the method are evaluated by the estimation of the detection probability  $\hat{P}_d = P(\hat{C} > C_1 | d_W = 1)$

of lightnings, and the false alarm probability  $\hat{P}_{fa} = P(\hat{C} > C_1 | d_W = 0)$ , where  $\hat{C}$  denotes the decision (i.e. the category of the risk) taken for a chosen confidence threshold,  $d_W = 1$  indicates that a lightning flash has been detected by the WWLLN for the cell under analysis, and  $d_W = 0$  indicates no detection. These probabilities are empirically estimated by

$$\hat{P}_d = P(\hat{C} > C_1 | d_W = 1) \approx \frac{n(\hat{C} > C_1, d_W = 1)}{n(d_W = 1)} \quad (8)$$

$$\hat{P}_{fa} = P(\hat{C} > C_1 | d_W = 0) \approx \frac{n(\hat{C} > C_1, d_W = 0)}{n(d_W = 0)} \quad (9)$$

where  $n(\hat{C} > C_1, d_W = 1)$  is the number of cells for which the joint event  $\hat{C} > C_1$  and  $d_W = 1$  has occurred, and  $n(d_W = 1)$  is the number of cells for which one has got a WWLLN detection  $d_W = 1$ . Similarly,  $n(\hat{C} > C_1, d_W = 0)$  is the number of cells where events  $\hat{C} > C_1$  and  $d_W = 0$  have occurred, and  $n(d_W = 0)$  is the number of cells having no WWLLN detection (i.e.  $d_W = 0$ ).

The tables III indicates the estimations of the detection probability  $\hat{P}_d$  of lightnings, and of the false alarm probability  $\hat{P}_{fa}$  obtained by the three methods tested based on WWLLN set of detections.

Methods	$\hat{P}_d$	$\hat{P}_{fa}$
Weighted Averaging Rule	0.9775	0.3601
SET with averaging rule	0.9462	0.2945
SET with PCR6 rule	0.9507	0.2954

Table III:  $\hat{P}_d$  and  $\hat{P}_{fa}$  performances.

Our results show that SET approach (with averaging rule, or with PCR6 rule) provides interesting results because it allows to identify and predict the areas with high risk of storm that are coherent with the real location of lightnings detected by the WWLLN. One sees that the direct weighted averaging fusion of the risk maps of the five criteria shown in Fig. 9 produces notably more false alarms than with SET method, and only a little increase detection probability. In this work there is no clear advantage of using the PCR6 rule with respect the weighted averaging rule in step 2 of SET method because the performances in term of probability of detection and false alarms are very close. In terms of computational time, the direct weighted averaging fusion of the risk maps is the fastest method which takes few seconds<sup>8</sup> with MatLab (R2108a version) running with a MacBookPro laptop computer (2.8 GHz Intel Core i7), then the second fastest method is the SET method using weighted averaging rule taking 1mn12sec, and the slowest (and most complicate) method is the SET method based on PCR6 rule which takes approximately 26mn to produce the results. One important advantage of the SET method (aside its aforementioned performances) is its ability to provide the confidence map of the solutions obtained by SET (i.e. the predicted risk levels) as shown in Fig. 12 and 13. These confidence maps are useful to identify areas of risks

<sup>8</sup>Once the five maps of risks have been computed.

where the confidences are low and thus very uncertain if some important decision but be taken based on these solutions (for instance the diverting of the flight of an aircraft, etc). Such type of useful confidence map cannot be drawn from the direct weighted averaging fusion of the risk maps. Due to space restraint, we did not include the decision inconsistency maps<sup>9</sup> (i.e. the map of the probabilities)  $P(\mathbf{a}_i \rightarrow \emptyset)$  reflecting the impossibility to make a coherent outranking (see SET step 4), but these maps obtained by SET (with weighted averaging or with PCR6 rule) reveal actually only very few cells located mainly at the west of Panama. This very small number of cells yielding decision inconsistency indicates that SET has provided solutions in good decision-making conditions in general.

#### IV. CONCLUSIONS

In this paper we have presented an application of belief functions for storm prediction based on multi-criteria analysis and the Soft Electre Tri (SET) methodology. We have shown that SET allows to reduce notably the false alarms rate with respect to a simple weighted averaging fusion method without sacrificing much the detection of lightnings, and to provide the confidence map of the solutions obtained. The SET method based on PCR6 rule of combination performs well but it has a high computational burden which prevent it to use it for quasi-real time applications, and for working with multi-criteria problems involving many more criteria. This work, we hope, could serve as a benchmark problem for testing many MCDM methods in future. More investigations are currently done to apply this type of new methodology using more meteorological criteria on other type of real data sets, and more refined parameter settings that will be reported if possible in a forthcoming publication.

#### ACKNOWLEDGEMENTS

The authors wish to thank the World Wide Lightning Location Network (<http://wwlln.net>), a collaboration among over 50 universities and institutions, for providing the lightning location data used in this paper. The authors would like to acknowledge the National Centers for Environmental Information of the National Oceanic and Atmospheric Administration (NOAA) for providing the data.

#### APPENDIX

##### *Basic definitions of belief functions*

In this appendix we provide basics of belief functions (BF) introduced by Shafer [2] to model epistemic uncertainty to reason about uncertainty. We assume that the answer of the problem under concern belongs to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are exhaustive and exclusive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . A (normal) basic

<sup>9</sup>They have been included with the data set files in [22] for convenience.

belief assignment (BBA) associated with a given source of evidence is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The number  $m(A)$  is called the mass of  $A$  committed by the source of evidence. The subset  $A \in 2^\Theta$  is called a focal element (FE) of the BBA  $m(\cdot)$  if and only if  $m(A) > 0$ . The set of all the focal elements of the BBA  $m(\cdot)$  is noted by  $\mathcal{F}_\Theta(m) = \{X \in 2^\Theta | m(X) > 0\}$ . The belief of  $A$  denoted  $Bel(A)$  and the plausibility of  $A$  denoted  $Pl(A)$  are usually interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$ . They are respectively defined for any  $A \in 2^\Theta$  from the BBA  $m(\cdot)$  by

$$Bel(A) = \sum_{X \in 2^\Theta | X \subseteq A} m(X) \quad (10)$$

and

$$Pl(A) = \sum_{X \in 2^\Theta | A \cap X \neq \emptyset} m(X) = 1 - Bel(\bar{A}). \quad (11)$$

where  $\bar{A}$  represents the complement of  $A$  in  $\Theta$ , that is  $\bar{A} \triangleq \Theta - \{A\} = \{X | X \in \Theta \text{ and } X \notin A\}$ . The symbol  $\triangleq$  means *equal by definition*, and the minus symbol denotes the set difference operator. The vacuous BBA (VBBA for short) representing a totally ignorant source is defined as  $m_v(\Theta) = 1$ .

#### PCR6 rule of combination

The PCR6 rule proposed in [10], [11] is an interesting alternative of original PCR rule of combination no. 5 (PCR5) proposed in [9], [19]. PCR6 and PCR5 rules coincide if we combine only two BBAs defined on the same FoD. The PCR6 fusion of  $S > 2$  BBAs is obtained by  $m_{1,2,\dots,S}^{PCR6}(\emptyset) = 0$ , and for all  $A \in 2^\Theta \setminus \{\emptyset\}$  by

$$\begin{aligned} m_{1,2,\dots,S}^{PCR6}(A) &= m_{1,2,\dots,S}^{Conj}(A) \\ &+ \sum_{j \in \{1,\dots,S\} | A \in \mathbf{X}_j \wedge \pi_j(\emptyset)} \\ &\left[ \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = A} m_i(X_{j_i}) \right) \right. \\ &\cdot \left. \frac{\pi_j(\emptyset)}{\sum_{X \in \mathbf{X}_j} \left( \sum_{i \in \{1,\dots,S\} | X_{j_i} = X} m_i(X_{j_i}) \right)} \right] \quad (12) \end{aligned}$$

where  $m_{1,2,\dots,S}^{Conj}(A) = \sum_{\substack{\mathbf{x}_j \in \mathcal{F}(m_1,\dots,m_S) \\ X_{j_1} \cap \dots \cap X_{j_S} = A}} \prod_{i=1}^S m_i(X_{j_i})$  is the conjunctive fusion rule, and where  $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S}) \triangleq \prod_{i=1}^S m_i(X_{j_i})$ , and  $\pi_j(\emptyset)$  in (12) is the concise notation of  $\pi_j(X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S})$  when  $X_{j_1} \cap X_{j_2} \cap \dots \cap X_{j_S} = \emptyset$ . The  $x \wedge y$  is the logical conjunction operator meaning that conditions  $x$  and  $y$  must be satisfied. PCR6 rule is quasi-associative and it offers a more precise conflict redistribution than DS rule but it requires a higher computational burden. PCR6 does not preserve the neutrality of the vacuous BBA however. PCR6 is simpler to implement than PCR5. Very basic Matlab™ codes of PCR5 and PCR6 rules can be found in [13], [20], and also from the BFAS (Belief Functions and Applications Society) repository [21].

#### REFERENCES

- [1] G. White, F. Yang, V. Tallapragada, *The Development and Success of NCEP's Global Forecast System*, NCEP Office Note, 2019, AMS, 99th Annual Meeting, 6-10 January 2019. see <https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forecast-system-gfs>
- [2] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [3] J. Dezert, J.-M. Tacnet, *Soft ELECTRE TRI outranking method based on belief functions*, Proc. Of Fusion 2012, Singapore, July 2012.
- [4] A.R. Jacobson, R. Holzworth, J. Harlin, R. Dowden, E. Lay, *Performance assessment of the world wide lightning location network (WWLLN) using the Los Alamos Sferic Array (LASA) as ground truth*, J. of Atmos. and Ocean. Tech., Vol. 23, pp. 1082-1092, 2006.
- [5] B. Roy, *Main sources of inaccurate determination, uncertainty and imprecision in decision models*, Math. and Comput. Modelling, Vol. 12, No. 10-11, pp. 1245-1254, 1989.
- [6] J. Figueira, S. Greco, M. Ehrgott (Editors), *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer Verlag, 2005.
- [7] J. Figueira, V. Mousseau, B. Roy, *ELECTRE methods*, in [6], pp. 133-162.
- [8] LAMSADE web site: <http://www.lamsade.dauphine.fr/mcda/biblio/>
- [9] J. Dezert, F. Smarandache, *Proportional Conflict Redistribution Rules for Information Fusion*, in [20], Vol. 2, pp. 3-68, 2006.
- [10] A. Martin, C. Osswald, *A new generalization of the proportional conflict redistribution rule stable in terms of decision*, Chap. 2 of [20], Vol.2, pp. 69-88, 2006.
- [11] A. Martin, C. Osswald, J. Dezert, F. Smarandache, *General Combination Rules for Qualitative and Quantitative Beliefs*, Journal of Advances in Information Fusion (JAIF), 3(2), pp. 67-89, Dec. 2008.
- [12] J. Dezert, J.-M. Tacnet, *Sigmoidal Model for Belief Function-based ELECTRE TRI Method*, Belief 2012 Proc., Compiègne, May 2012.
- [13] F. Smarandache, J. Dezert, J.-M. Tacnet, *Fusion of sources of evidence with different importances and reliabilities*, in Proc. of Fusion 2010 Conf., Seattle, USA, July 2010.
- [14] A. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., Vol. 38, pp. 325-339, 1967.
- [15] A. Dempster, *A generalization of Bayesian inference*, J. of Royal Sta. Soc., B30, pp. 205-247, 1968.
- [16] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Fusion 2012.
- [17] A. Tchamova, J. Dezert, *On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory*, in Proc. of 6th IEEE Int. Conf. on Intell. Syst., 2012.
- [18] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., Vol. 29(3), pp. 223-252, 2014.
- [19] F. Smarandache, J. Dezert, *Information fusion based on new proportional conflict redistribution rules*, in Proc. of Int. Fusion 2005 Conf., Philadelphia, PA, USA, 2005.
- [20] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion (Collected works)*, Vol. 1-4, ARP, 2004-2015.
- [21] <https://www.bfasociety.org/>
- [22] <https://www.onera.fr/sites/default/files/297/Fusion2021SETDataSet.zip>
- [23] E. Wesolek, *Le modèle WRF et la prévision des orages. Guide d'usage des champs diffusés*, Keraunos Editions, June 2012 (in French).
- [24] J. Han, H.-P. Pan, *Revision of Convection and Vertical Diffusion Schemes in the NCEP Global Forecast System*, Weather and Forecasting, Vol. 26, pp. 520-533, 2011.
- [25] J.G. Galway, *The lifted index as a predictor of latent instability*, Bull. Amer. Meteor. Soc., pp. 528-529, 1956.
- [26] J.R. Holton, *An introduction to dynamic meteorology (4th edition)*, Elsevier Academic Press, pp. 535, 2004.
- [27] S. Malardel, *Fondamentaux de Météorologie, 2nd édition à l'école du temps*, Météo France, Cépadués edition, pp. 711, 2009.
- [28] M.C. Llasat, *An objective classification of rainfall events on the basis of their convective features. Application to rainfall intensity in the north-east of Spain*, Int. J. of Climatology, Vol. 21, pp. 1385-1400, 2001.
- [29] P.M. James, B.K. Reichert, D.Heizenreider, *NowCastMIX: Automatic Integrated Warnings for Severe Convection on Nowcasting Time Scales at the German Weather Service*, Weather and Forecasting, Vol. 33, pp. 1413-1433, 2018.
- [30] E. Weller, K. Shelton, M.J. Reeder, C. Jakob, *Precipitation Associated With Convergence Lines*, Vol. 30, pp. 3169-3183, 2017.

# Evidential Data Association Based on Dezert-Smarandache Theory

Mohammed Boumediene<sup>a</sup>, Hossni Zebiri<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Signals and Images Laboratory, USTO Mohamed Boudiaf, Oran, Algeria.

<sup>b</sup>The French Aerospace Lab (ONERA), Palaiseau, France.

Emails: mohammed.boumediene@enstic.dz, hossni.zebiri@gmail.com, jean.dezert@onera.fr

Originally published as: M. Boumediene, H. Zebiri, J. Dezert, *Evidential Data Association Based on Dezert-Smarandache Theory*, Int. Journal of Intelligent Robotics and Applications, Vol. 7, pp. 91–102, 2023, and reprinted with permission.

**Abstract**—Data association has become pertinent task to interpret the perceived environment for mobile robots such as autonomous vehicles. It consists in assigning the sensor detections to the known objects in order to update the obstacles map surrounding the vehicle. Dezert-Smarandache Theory (DSMT) provides a mathematical framework for reasoning with imperfect data like sensor’s detections. In DSMT, data are quantified by belief functions and combined by the Proportional Conflict Redistribution (PCR6) rule in order to obtain the fusion of evidences to make a decision. However, this combination rule has an exponential complexity and that is why DSMT is rarely used for real-time applications. This paper proposes a new evidential data association based on DSMT techniques. The proposed approach focuses on the significant pieces of information when combining and removes unreliable and useless information. Consequently, the complexity is reduced without degrading substantially the decision-making. The paper proposes also a new simple decision-making algorithm based on a global optimization procedure. Experimental results obtained on a well-known KITTI dataset show that this new approach reduces significantly the computation time while preserving the association accuracy. Consequently, the new proposed approach makes DSMT framework applicable for real-time applications for autonomous vehicle perception.

**Keywords:** Data Association, Belief Functions, Dezert-Smarandache Theory, Proportional Conflict Redistribution 6, Dezert-Smarandache Probability.

## I. INTRODUCTION

Multi-Target Tracking (MTT) is a fundamental system to interpret the perceived environment of mobile robots such as autonomous vehicles [1], [2]. These cars require precise knowledge of their surrounding environment in order to ensure safe and comfortable driving [3]–[5]. The MTT system estimates the status of detected objects surrounding the vehicle at different times by single or multiple sensors. Data Association is a central problem in MTT which assigns *targets* to the predicted *tracks* in order to update their status. Targets refer to the *detected objects* at the current time and tracks refer to the *known objects* in the scene. A dynamic environment, like the road environment, makes the object association more difficult because of the appearance/disappearance of objects in the perceived scene.

Usually, the assignment problem is resolved by the probability theory. Several methods have been proposed as the well-known Global Nearest Neighbour (GNN) method and the Joint

Probability Data Association Filter (JPDAF) [6]–[8]. GNN provides the optimal pairing by minimizing the global distance between detections and known objects. JPDAF is based on a weighted linear combination of all detections to estimate status of known objects. More details about these methods can be found in [8]–[10].

Recently, the belief function theory has also been used to cope with the association problem [4], [11]. This theory, also called Dempster-Shafer Theory (DST) [12], [13] allows to reason about uncertainty thanks to the belief functions that are often interpreted as lower and upper bound of unknown probability measures. In fact, sensor’s detections can be inaccurate and incomplete. However, the DST models these imperfect information through a distribution of belief masses which quantify the confidence granted. Thereafter, these masses are combined by Dempster’s rule to make decisions. Because Dempster’s rule has been used and promoted by Shafer in his mathematical theory of evidence, it is also often denoted as DS rule in the literature.

Rombaut in [14] formalizes the association problem by DST to reconstruct the environment of intelligent vehicles. This approach measures the confidence of the association hypotheses between perceived and known obstacles by combining belief masses using DS rule. This approach is extended in [11], [15] to track vehicles where the association process is based on the Transferable Belief Model (TBM) [16]. This latter is a subjective and non-probabilistic interpretation of the Belief theory. In TBM, the decision-making is based on the pignistic probabilities derived from the belief quantities. Several alternative probabilistic transformations have been proposed in the literature. Our previous work [17] evaluates some of them on real-data in the context of the DST framework. In [11], the decision is performed by maximizing the joint pignistic probability. However, this probability is computed for all possible associations which grows the computation time exponentially with the objects number. To tackle this problem, the decision is made by selecting associations corresponding to local maxima of pignistic probabilities [4], [18]. More recently, Denœux et al. [19] express DS rule in terms of contour functions and plausibility functions which reduces the complexity and makes this approach applicable for real-time applications.

All those aforementioned approaches use Dempster's rule which provides a counter-intuitive behavior specially in high and low conflicting situations [20], [21]. In fact, DS rule redistributes the conflicting mass on all elements which can cause the lost of the information specificity and then generates unacceptable results. In addition, serious mistakes have been shown in logical fundamentals of the DST framework [22]–[24]. To overcome those drawbacks, a more sophisticate rule has been proposed and defined in the framework of Dezert-Smarandache Theory (DSMT) [21]. Based on the Proportional Conflict Redistribution (PCR) process, PCR6 rule preserves the information specificity by transferring the conflicting mass only to the elements involved in the conflict and proportionally to their individual masses. However, PCR6 has an exponential complexity and that is why it is rarely used for real-time applications.

In this paper, we propose a new evidential data association based on the DSMT framework. The first contribution is to reduce the complexity of the combination step based on PCR6 rule developed originally in the framework of Dezert-Smarandache Theory. The proposed approach focuses on the significant pieces of information when combining and removes unreliable and useless information. Consequently, the complexity is reduced without degrading substantially the decision-making. The second contribution is to propose a new simple decision-making algorithm based on a global optimization. Experimental results obtained on a well-known intelligent transportation systems dataset show the benefits of this new approach in terms of computation time reduction and association accuracy.

The rest of this paper is organized as follows. In section II, few basics of the DSMT are presented. Section III details the new proposed evidential data association approach and its experimental validation is presented in Section IV. Finally, Section V concludes this paper.

## II. FUNDAMENTALS OF DSMT

In the Belief theory context, a problem is modelled by a finite set of hypotheses  $H_i$  likely to be the solutions, called *Frame of Discernment* (FoD). In the general DSMT framework, the elements of the FoD do not need to be mutually exhaustive as in the DST framework, but in the particular context of our application presented in this paper, we work with Shafer's model of the FoD where all elements of the FoD are mutually exclusive and exhaustive, that is:

$$\Theta = \bigcup_{i=1}^k \{H_i\} \text{ with } H_i \cap H_j = \emptyset \quad (1)$$

where  $H_i$  are denoted as *singletons*, the lowest piece of discernible knowledge in the FoD.

### A. Basic Belief Assignment

A *basic belief assignment* (bba) or mass function associated to a given source is defined as a function  $m : 2^\Theta \rightarrow [0, 1]$  satisfying:

$$\sum_{A \in 2^\Theta} m(A) = 1 \quad (2)$$

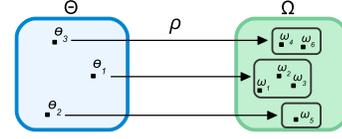


Figure 1. Illustration of the refinement function  $\rho$  [11].

where  $m(A)$  is the mass of belief that supports  $A$ . The source is totally ignorant if  $m(\Theta) = 1$  and so the bba is considered as vacuous function. Whether  $m(A) > 0$ ,  $A$  is called a focal element of the bba  $m(\cdot)$ . Thus  $\mathcal{F}(m) = \{A \in 2^\Theta / m(A) > 0\}$  defines the set of focal elements.

### B. Vacuous Extension

Some sources of information can express on different FoDs but related. However, in order to combine them, it is necessary to work with the same common frame. For that, it can be defined a finer FoD [13]. Let  $\Omega$  a finer frame of  $\Theta$  where every element of  $\Theta$  is mapped into one or more elements of  $\Omega$  (Cf. Fig. 1). Therefore, the refinement function  $\rho$  matches proposition  $A$  from  $2^\Theta$  to  $2^\Omega$  according to:

$$\begin{cases} \{\rho(\{\theta\}), \theta \in \Theta\} \text{ is a partition of } \Omega \\ \forall A \subseteq \Theta, \rho(A) = \bigcup_{\theta \in A} \rho(\{\theta\}). \end{cases} \quad (3)$$

The vacuous extension  $m^{\Theta \uparrow \Omega}$  defines the bba on  $\Omega$  from the bba  $m^\Theta$  defined on  $\Theta$  and the refinement  $\rho$ :

$$m^{\Theta \uparrow \Omega}(\rho(A)) = \begin{cases} m^\Theta(A), & \forall A \subseteq \Theta \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

### C. Belief Combination

The belief combination consists in merging the measures of evidence  $m_i^\Theta$  of  $M$  distinct sources  $S_i$ , defined on the same frame  $\Theta$ , to a new distribution of evidence. For that, the Proportional Conflict Redistribution rule 6 (PCR6) have been proposed in [25] and theoretically justified in [21]. In fact, PCR6 rule overcomes the drawbacks of the Dempster rule [13] by redistributing proportionally the partial conflict only on elements involved in this conflict. The formula of PCR6 is defined by  $m_{\text{PCR6}}(\emptyset) = 0$  and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$  by [26], [27]:

$$\begin{aligned} m_{\text{PCR6}}(A) = & m_{\text{Conj}}(A) \\ & + \sum_{j \in \{1, \dots, M\} | A \in \mathbf{A}_j \wedge \pi_j(\emptyset)} \\ & \left[ \left( \sum_{i \in \{1, \dots, M\} | A_{j_i} = A} m_i^\Theta(A_{j_i}) \right) \right. \\ & \cdot \left. \frac{\pi_j(\emptyset)}{\sum_{A \in \mathbf{A}_j} \left( \sum_{i \in \{1, \dots, M\} | A_{j_i} = A} m_i^\Theta(A_{j_i}) \right)} \right], \quad (5) \end{aligned}$$

where  $\wedge$  is the logical conjunction<sup>1</sup> and  $\mathbf{A}_j$  is a possible  $M$ -uple of focal elements with  $A_{j_i} \in \mathcal{F}(m_i^\Theta)$ , that

<sup>1</sup>i.e.  $x \wedge y$  means that conditions  $x$  and  $y$  are both true.

is  $\mathbf{A}_j \triangleq (A_{j_1}, A_{j_2}, \dots, A_{j_M})$ .  $\mathcal{F}$  is the cardinality of  $\mathcal{F}(m_1^\ominus, m_2^\ominus, \dots, m_M^\ominus)$  which is the set of all possible  $M$ -uple.

And where  $\pi_j(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_M}) \triangleq \prod_{i=1}^M m_i^\ominus(A_{j_i})$ , and  $\pi_j(\emptyset) = \pi_j(A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_M})$  defines the *conflicting mass product* of  $\mathbf{A}_j$  if  $A_{j_1} \cap A_{j_2} \cap \dots \cap A_{j_M} = \emptyset$  and the conjunctive rule  $m_{\text{Conj}}$  is given by:

$$m_{\text{Conj}}(A) = \sum_{A_{j_1} \cap \dots \cap A_{j_M} = A} \prod_{i=1}^M m_i^\ominus(A_{j_i}). \quad (6)$$

#### D. Probabilistic Transformation

Decision-making consists of selecting a solution among all possible hypotheses. Usually, the decision must be made among elements of the frame. However, the belief combination also generates masses for disjunctive propositions. Therefore, it is necessary to redistribute the masses of these unions on elements of  $\Theta$  in order to make a decision. For that, Dezert-Smarandache Probability (DSmP) transformation is defined [28] where  $DSmP(\emptyset) = 0$  and  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ :

$$DSmP_\epsilon(A) = \sum_{Y \in 2^\Theta} \frac{\sum_{\substack{Z \subset A \cap Y \\ c(Z)=1}} m(Z) + \epsilon \cdot \mathcal{C}(A \cap Y)}{\sum_{\substack{Z \subset Y \\ c(Z)=1}} m(Z) + \epsilon \cdot \mathcal{C}(Y)} \quad (7)$$

Where  $\epsilon \geq 0$  is used to adjust the effect of element's cardinality ( $\mathcal{C}(\cdot)$ ) in the proportional redistribution. In addition,  $\epsilon$  permits to compute  $DSmP$  when encountering zero masses. Typically,  $\epsilon = 0.001$  because with a smaller  $\epsilon$  the Probabilistic Information Content (PIC) [29] is higher. The PIC indicates the level of the available knowledge to make a correct decision.  $PIC = 0$  indicates that no knowledge exists to make a correct decision.

### III. DATA ASSOCIATION USING DSMT

Four steps are needed to solve the data association problem: modeling, estimation, combining, and decision-making. However, PCR6 rule combination has an exponential complexity which makes it not appealing for real-time applications. This is why in this paper, only  $k$ -significant sources are combined (with  $k$  lesser than the original number or sources available). Thereafter, a simple global optimization is used to make association decisions.

#### A. Data Modelling

Let us consider  $n$  detected objects at time  $t$  and  $m$  known objects at previous time  $t - 1$ . In this context, data association aims at matching the  $n$  detected objects  $X_i$  to the  $m$  known ones  $Y_j$  under certain conditions:

- multiple associations are not accepted, a detected object is associated with only one known object at most and *vice versa*,
- multiple new objects can appear,
- multiple known objects can disappear.

The distances between the attributes of objects (position, velocity, etc.) are considered as pieces of evidence. For a given distance, its belief will be expressed on the elementary FoD  $\theta_{i,j} = \{\text{yes}_{(i,j)}, \text{no}_{(i,j)}\}$  which models the relevance of the association between  $X_i$  and  $Y_j$ . Therefore, three *bba* masses are constructed for each pairwise objects ( $X_i, Y_j$ ):

- $m^{\theta_{i,j}}(\text{yes}_{(i,j)})$  : degree of belief that  $X_i$  is associated with  $Y_j$ ,
- $m^{\theta_{i,j}}(\text{no}_{(i,j)})$  : degree of belief that  $X_i$  is not associated with  $Y_j$ ,
- $m^{\theta_{i,j}}(\theta_{i,j})$  : represents the ignorance.

#### B. Belief Estimation

The estimation of belief masses is related to the considered application. The most suitable model for data association applications [30] is the non-antagonist model [14], [15] defined by:

$$m_j^{\Theta_{i,\cdot}}(Y_{(i,j)}) = \begin{cases} 0 & , I_{i,j} \in [0, \tau] \\ \Phi_1(I_{i,j}) & , I_{i,j} \in [\tau, 1] \end{cases} \quad (8)$$

$$m_j^{\Theta_{i,\cdot}}(\bar{Y}_{(i,j)}) = \begin{cases} \Phi_2(I_{i,j}) & , I_{i,j} \in [0, \tau] \\ 0 & , I_{i,j} \in [\tau, 1] \end{cases} \quad (9)$$

$$m_j^{\Theta_{i,\cdot}}(\Theta_{i,\cdot}) = \begin{cases} 1 - m_j^{\Theta_{i,\cdot}}(\bar{Y}_{(i,j)}) & , I_{i,j} \in [0, \tau] \\ 1 - m_j^{\Theta_{i,\cdot}}(Y_{(i,j)}) & , I_{i,j} \in [\tau, 1] \end{cases}, \quad (10)$$

where  $I_{i,j} \in [0, 1]$  is an index of similarity between  $X_i$  and  $Y_j$ .  $\Phi_1(\cdot)$  and  $\Phi_2(\cdot)$  are two cosine functions defined as follows:

$$\begin{cases} \Phi_1(I_{i,j}) = \frac{\alpha}{2} \left[ 1 - \cos\left(\pi \frac{I_{i,j} - \tau}{\tau}\right) \right] \\ \Phi_2(I_{i,j}) = \frac{\alpha}{2} \left[ 1 + \cos\left(\pi \frac{I_{i,j}}{\tau}\right) \right] \end{cases} \quad (11)$$

where  $0 < \alpha < 1$  is the reliability factor of the data source and  $0 < \tau < 1$  represents the impartiality of the association process.

#### C. $k$ -Significant sources combination

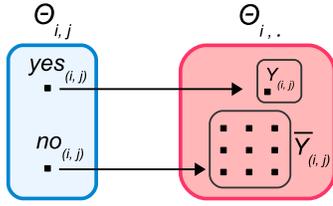
Before decision-making, sources should be combined which is possible only if they express on the same FoD. Hence, to determine who is associated to the detected object  $X_i$ , a new FoD is defined  $\Theta_{i,\cdot}$  (12). This new frame is composed of the  $m$  possible  $X_i$ -to- $Y_j$  associations denoted  $Y_{(i,j)}$  and the appearance hypothesis of object  $X_i$  denoted by  $Y_{(i,*)}$ :

$$\Theta_{i,\cdot} = \{Y_{(i,1)}, Y_{(i,2)}, \dots, Y_{(i,m)}, Y_{(i,*)}\}. \quad (12)$$

Therefore,  $\Theta_{i,\cdot}$  is a refinement frame of the previous FoDs  $\theta_{i,j}$  in which the belief is initially expressed (Cf. Fig. 2). Based on a vacuous extension (3), initial belief functions  $m^{\theta_{i,j}}$  are expressed on  $\Theta_{i,\cdot}$  as follows:

$$\begin{cases} m_j^{\Theta_{i,\cdot}}(Y_{(i,j)}) = m^{\theta_{i,j}}(\text{yes}_{(i,j)}) \\ m_j^{\Theta_{i,\cdot}}(\bar{Y}_{(i,j)}) = m^{\theta_{i,j}}(\text{no}_{(i,j)}) \\ m_j^{\Theta_{i,\cdot}}(\Theta_{i,\cdot}) = m^{\theta_{i,j}}(\theta_{i,j}) \end{cases} \quad (13)$$

where  $\bar{Y}_{(i,j)}$  represents the hypothesis "X<sub>i</sub> is not associated to Y<sub>j</sub>" which corresponds to the union of all association hypotheses expect the  $Y_{(i,j)}$ , i.e.  $\bar{Y}_{(i,j)} =$


 Figure 2. The refinement frames of  $\theta_{i,j}$ :  $\Theta_{i,*}$ .

$\{Y_{(i,1)}, \dots, Y_{(i,j-1)}, Y_{(i,j+1)}, \dots, Y_{(i,m)}, Y_{(i,*)}\}$ . It should be noted that no information is initially considered on  $Y_{(i,*)}$ . This information appears during combination step.

Once the sources are expressed on the same frame, the *bbas* are combined with the PCR6 rule. However, combining all sources increases the time-consuming and can reach an exponential complexity when the number of sources is important. To overcome this drawback, this paper proposes a new method to reduce the combination complexity without sacrificing too much the decision quality.

The proposed approach selects only information having belief in top  $k$  highest masses. Formally, for each  $X_i$  object, initial masses on association hypotheses are sorted:

$$\begin{cases} b_1 \geq b_2 \geq \dots \geq b_z \geq \dots \geq b_m \\ b_z = m^{\theta_{i,j}}(yes_{(i,j)}), \text{ and } z, j \in \{1, \dots, m\} \end{cases} \quad (14)$$

where  $b_1$  is highest mass of belief, so the source that generated it is the most significant for matching  $X_i$ . On other hand, the least important source is that which generates the lowest belief  $b_m$ .

Now, only  $k$  most significant sources are selected for their combination. Therefore, for each  $X_i$  assignment,  $\Theta_{i,*}$  is defined as follows:

$$\Theta_{i,*} = \{Y_{(i,z)}/b_z \geq b_k, Y_{(i,*)}\} \quad (15)$$

with  $z \in \{1, \dots, m\}$  and  $k < m$ . Consequently,  $\Theta_{i,*}$  contains only the most relevant hypotheses and ignores others ( $b_z < b_k$ ). By this simple selection procedure one reduces the computation complexity of the combination process.

If  $b_k = 0$ ,  $b_{k-1}$  is used to select significant sources. In the case where no  $b_k > 0$ , the object  $X_i$  is considered as an appearance and is associated directly to  $Y_{(i,*)}$ . Thereafter, initial mass functions  $m^{\theta_{i,j}}(\cdot)$  is hence transferred to  $\Theta_{i,*}$  by the refinement defined in (13) and the PCR6 rule of combination (5) is applied.

#### D. Decision-Making

The assignment decision is based on the  $DSmP_{i,*}$  matrix which is the probabilistic approximation of the combined masses. Table I presents the  $DSmP_{i,*}$  of the detected-to-known objects association. Each line defines the association probabilities of the detected object  $X_i$  with all known ones  $Y_j$ .  $DSmP_{i,*}(Y_{(i,*)})$  defines the appearance probability of  $X_i$ . It is useful to note that multiple objects can appear/disappear.

Different decision-making strategies have been proposed according to the desired objectives [11], [18]. There are two

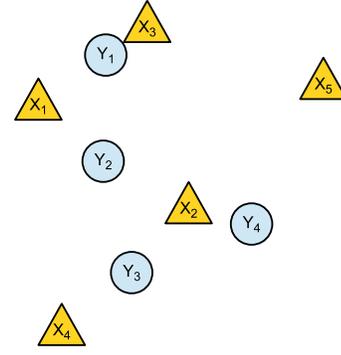


Figure 3. Scenario showing 5 detected objects (triangle) and 4 known objects (circle).

approaches depending on the type of optimization: global or local. The first approach selects the “best” associations optimizing a global cost function [31], [32]. The Joint Pignistic Probability (JPP)  $BetP_{\prod_{i=1}^n}$  is defined as the cost function in [11]:

$$BetP_{\prod_{i=1}^n} = BetP_{1,*}(Y_{(1,j_1)}) \times \dots \times BetP_{n,*}(Y_{(n,j_n)}) \quad (16)$$

with  $j_i \in \{1, 2, \dots, m, *\}$ . Among all possible solutions for the detected-to-known association, the *best* is that maximizing  $BetP_{\prod_{i=1}^n}$ . However, when the number of possible associations is important, this optimization generates a high computational complexity. To cope with this inconvenience, another approach consists of resolving the assignment problem by a local optimization. The Local Pignistic Probability (LPP) [18] makes the association decisions according to local maxima of the pignistic matrix ( $BetP_{i,*}$ ). The LPP method performs a successive selection of  $n$  local maxima while respecting the association constrains (Cf. Section III-A). However, local optimization is considered as a sub-optimal solution.

In this paper, a new simple global optimization is applied/proposed. Firstly, the last column ( $Y_{(i,*)}$ ) of the  $DSmP$  matrix is removed in order to select “best” associations by using the well-known Munkres algorithm [33]. The complexity of this algorithm is only  $O(n^3)$  [33]. Secondly, for each selected association  $Y_{(i,j)}$ , if  $DSmP_{i,*}(Y_{(i,j)}) < DSmP_{i,*}(Y_{(i,*)})$  the association  $Y_{(i,j)}$  is removed and the object  $X_i$  is considered to be a new object ( $Y_{(i,*)}$ ).

#### IV. ILLUSTRATIVE EXAMPLE

Let us consider the simulated example presented in Fig. 3. The scenario shows 5 detected objects and 4 known objects. By observing the corresponding initial *bba* presented in Table II, one can already assume some associations. For instance, with  $m^{\Theta_{3,*}}(yes_{(3,1)}) = 0.85$  and  $m^{\Theta_{2,*}}(yes_{(2,4)}) = 0.75$ ,  $X_3$  and  $X_2$  are most likely to be associated respectively to  $Y_1$  and  $Y_4$ . As for the detected object  $X_5$ , no source supports its association with a known objects, so it can be an appearance.

Therefore, it is possible to make decisions by combining only some information? To answer, the proposed method is

Table I  
DSmP PROBABILITIES OF DETECTED-TO-KNOWN OBJECT ASSOCIATIONS.

$\Theta_{i..}$	$Y_{(i,1)}$	...	$Y_{(i,m)}$	$Y_{(i,*)}$
$DSmP_{1..}(\cdot)$	$DSmP_{1..}(Y_{(1,1)})$	...	$DSmP_{1..}(Y_{(1,m)})$	$DSmP_{1..}(Y_{(1,*)})$
$DSmP_{2..}(\cdot)$	$DSmP_{2..}(Y_{(2,1)})$	...	$DSmP_{2..}(Y_{(2,m)})$	$DSmP_{2..}(Y_{(2,*)})$
$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
$DSmP_{n..}(\cdot)$	$DSmP_{n..}(Y_{(n,1)})$	...	$DSmP_{n..}(Y_{(n,m)})$	$DSmP_{n..}(Y_{(n,*)})$

Table II  
INITIAL MASS FUNCTIONS FOR THE SCENARIO IN FIG. 3.

$S_{1,1} \begin{cases} m^{\theta_{1,1}}(yes_{(1,1)}) = \mathbf{0.45} \\ m^{\theta_{1,1}}(no_{(1,1)}) = 0.35 \\ m^{\theta_{1,1}}(\theta_{1,1}) = 0.20 \end{cases}$	$S_{1,2} \begin{cases} m^{\theta_{1,2}}(yes_{(1,2)}) = \mathbf{0.48} \\ m^{\theta_{1,2}}(no_{(1,2)}) = 0.32 \\ m^{\theta_{1,2}}(\theta_{1,2}) = 0.20 \end{cases}$	$S_{1,3} \begin{cases} m^{\theta_{1,3}}(yes_{(1,3)}) = 0.00 \\ m^{\theta_{1,3}}(no_{(1,3)}) = 0.95 \\ m^{\theta_{1,3}}(\theta_{1,3}) = 0.05 \end{cases}$	$S_{1,4} \begin{cases} m^{\theta_{1,4}}(yes_{(1,4)}) = 0.00 \\ m^{\theta_{1,4}}(no_{(1,4)}) = 0.99 \\ m^{\theta_{1,4}}(\theta_{1,4}) = 0.01 \end{cases}$
$S_{2,1} \begin{cases} m^{\theta_{2,1}}(yes_{(2,1)}) = 0.00 \\ m^{\theta_{2,1}}(no_{(2,1)}) = 0.99 \\ m^{\theta_{2,1}}(\theta_{2,1}) = 0.01 \end{cases}$	$S_{2,2} \begin{cases} m^{\theta_{2,2}}(yes_{(2,2)}) = 0.32 \\ m^{\theta_{2,2}}(no_{(2,2)}) = 0.58 \\ m^{\theta_{2,2}}(\theta_{2,2}) = 0.10 \end{cases}$	$S_{2,3} \begin{cases} m^{\theta_{2,3}}(yes_{(2,3)}) = \mathbf{0.47} \\ m^{\theta_{2,3}}(no_{(2,3)}) = 0.43 \\ m^{\theta_{2,3}}(\theta_{2,3}) = 0.10 \end{cases}$	$S_{2,4} \begin{cases} m^{\theta_{2,4}}(yes_{(2,4)}) = \mathbf{0.75} \\ m^{\theta_{2,4}}(no_{(2,4)}) = 0.15 \\ m^{\theta_{2,4}}(\theta_{2,4}) = 0.10 \end{cases}$
$S_{3,1} \begin{cases} m^{\theta_{3,1}}(yes_{(3,1)}) = \mathbf{0.85} \\ m^{\theta_{3,1}}(no_{(3,1)}) = 0.05 \\ m^{\theta_{3,1}}(\theta_{3,1}) = 0.10 \end{cases}$	$S_{3,2} \begin{cases} m^{\theta_{3,2}}(yes_{(3,2)}) = 0.00 \\ m^{\theta_{3,2}}(no_{(3,2)}) = 0.90 \\ m^{\theta_{3,2}}(\theta_{3,2}) = 0.10 \end{cases}$	$S_{3,3} \begin{cases} m^{\theta_{3,3}}(yes_{(3,3)}) = 0.00 \\ m^{\theta_{3,3}}(no_{(3,3)}) = 0.90 \\ m^{\theta_{3,3}}(\theta_{3,3}) = 0.10 \end{cases}$	$S_{3,4} \begin{cases} m^{\theta_{3,4}}(yes_{(3,4)}) = 0.00 \\ m^{\theta_{3,4}}(no_{(3,4)}) = 0.99 \\ m^{\theta_{3,4}}(\theta_{3,4}) = 0.01 \end{cases}$
$S_{4,1} \begin{cases} m^{\theta_{4,1}}(yes_{(4,1)}) = 0.00 \\ m^{\theta_{4,1}}(no_{(4,1)}) = 0.99 \\ m^{\theta_{4,1}}(\theta_{4,1}) = 0.01 \end{cases}$	$S_{4,2} \begin{cases} m^{\theta_{4,2}}(yes_{(4,2)}) = 0.00 \\ m^{\theta_{4,2}}(no_{(4,2)}) = 0.90 \\ m^{\theta_{4,2}}(\theta_{4,2}) = 0.10 \end{cases}$	$S_{4,3} \begin{cases} m^{\theta_{4,3}}(yes_{(4,3)}) = \mathbf{0.50} \\ m^{\theta_{4,3}}(no_{(4,3)}) = 0.40 \\ m^{\theta_{4,3}}(\theta_{4,3}) = 0.10 \end{cases}$	$S_{4,4} \begin{cases} m^{\theta_{4,4}}(yes_{(4,4)}) = 0.00 \\ m^{\theta_{4,4}}(no_{(4,4)}) = 0.99 \\ m^{\theta_{4,4}}(\theta_{4,4}) = 0.01 \end{cases}$
$S_{5,1} \begin{cases} m^{\theta_{5,1}}(yes_{(5,1)}) = 0.00 \\ m^{\theta_{5,1}}(no_{(5,1)}) = 0.90 \\ m^{\theta_{5,1}}(\theta_{5,1}) = 0.10 \end{cases}$	$S_{5,2} \begin{cases} m^{\theta_{5,2}}(yes_{(5,2)}) = 0.00 \\ m^{\theta_{5,2}}(no_{(5,2)}) = 0.85 \\ m^{\theta_{5,2}}(\theta_{5,2}) = 0.15 \end{cases}$	$S_{5,3} \begin{cases} m^{\theta_{5,3}}(yes_{(5,3)}) = 0.00 \\ m^{\theta_{5,3}}(no_{(5,3)}) = 0.90 \\ m^{\theta_{5,3}}(\theta_{5,3}) = 0.10 \end{cases}$	$S_{5,4} \begin{cases} m^{\theta_{5,4}}(yes_{(5,4)}) = 0.00 \\ m^{\theta_{5,4}}(no_{(5,4)}) = 0.90 \\ m^{\theta_{5,4}}(\theta_{5,4}) = 0.10 \end{cases}$

applied with  $k = 2$ . The selected information for the detected-to-known association are represented by (17):

$$\left\{ \begin{array}{l} \Theta_{1..} = \{Y_{(1,1)}, Y_{(1,2)}, Y_{(1,*)}\} \\ \Theta_{2..} = \{Y_{(2,3)}, Y_{(2,4)}, Y_{(2,*)}\} \\ \Theta_{3..} = \{Y_{(3,1)}, Y_{(3,*)}\} \\ \Theta_{4..} = \{Y_{(4,3)}, Y_{(4,*)}\} \\ \text{direct decision: } X_5 \text{ appears.} \end{array} \right. \quad (17)$$

Regarding the association of  $X_1$ , the two highest belief masses (0.48 and 0.45) are respectively related to the  $Y_{(1,2)}$  and  $Y_{(1,1)}$  hypotheses which makes them relevant for decision-making. Thus, we work with the frame  $\Theta_{1..} = \{Y_{(1,1)}, Y_{(1,2)}, Y_{(1,*)}\}$  instead the set of all hypotheses  $\{Y_{(1,1)}, Y_{(1,2)}, Y_{(1,3)}, Y_{(1,4)}, Y_{(1,*)}\}$  which decreases the complexity of combination. In the same way,  $\Theta_{2..} = \{Y_{(2,3)}, Y_{(2,4)}, Y_{(2,*)}\}$  because the highest beliefs (0.75 and 0.47) are related to the  $Y_{(2,4)}$  and  $Y_{(2,3)}$  hypotheses. In this case,  $Y_{(2,1)}$  and  $Y_{(2,2)}$  are ignored because their beliefs are less significant than those of  $Y_{(2,3)}$  and  $Y_{(2,4)}$  ( $0.75 > 0.47 > 0.32 > 0.00$ ). For  $X_3$  and  $X_4$ , there is only one piece of information with a non-null belief for their association. Therefore,  $\Theta_{3..} = \{Y_{(3,1)}, Y_{(3,*)}\}$  and  $\Theta_{4..} = \{Y_{(4,3)}, Y_{(4,*)}\}$ . Concerning  $X_5$ , no source believes on its association, so  $X_5$  is a new detected object which means an appearance  $Y_{(5,*)}$ . In this case, the decision is directly made without combination. Consequently, the cardinality of each  $\Theta_{i..}$  (17) is reduced which means less computation time when combining.

To make decision, the selected information are combined by (5) and transformed to  $DSmP$  probabilities by (7). Table III represents  $DSmP_{i..}(\cdot)$  based on the two most significant mass functions. The dimension of each  $DSmP_{i..}$  vector is smaller than usual (Cf. Table IV) and corresponds to the number of relevant associations. In this context, the complexity of decision-making can be reduced too. In addition, it can be observed that the proposed approach preserves the relevant association probabilities. Therefore, the same decisions (18) are made through Tables III and IV.

Table III  
 $BetP_{i..}$ , BASED ON 2-SIGNIFICANT MASS FUNCTIONS.

$\Theta_{i..}$	$Y_{(i,1)}$	$Y_{(i,2)}$	$Y_{(i,3)}$	$Y_{(i,4)}$	$Y_{(i,*)}$
$DSmP_{1..}$	0.40	0.45	-	-	0.15
$DSmP_{2..}$	-	-	0.27	0.66	0.07
$DSmP_{3..}$	0.94	-	-	-	0.06
$DSmP_{4..}$	-	-	0.56	-	0.44

Table IV  
 $BetP_{i..}$ , BASED ON ALL MASS FUNCTIONS.

$\Theta_{i..}$	$Y_{(i,1)}$	$Y_{(i,2)}$	$Y_{(i,3)}$	$Y_{(i,4)}$	$Y_{(i,*)}$
$DSmP_{1..}$	<b>0.39</b>	<b>0.43</b>	0.00	0.00	<b>0.18</b>
$DSmP_{2..}$	0.00	0.11	<b>0.22</b>	<b>0.61</b>	<b>0.06</b>
$DSmP_{3..}$	<b>0.95</b>	0.00	0.00	0.00	<b>0.05</b>
$DSmP_{4..}$	0.00	0.00	<b>0.56</b>	0.00	<b>0.44</b>
$DSmP_{5..}$	0.00	0.00	0.00	0.00	1.00

Table V  
KITTI IMAGE SEQUENCE CHARACTERISTICS.

	Seq. 2	Seq. 4	Seq. 6	Seq. 7	Seq. 8	Seq. 13	Seq. 14	Seq. 16	Seq. 18	Seq. 19	Seq. 20
Number of frames	233	314	270	800	390	340	106	209	339	1059	837
Number of associations	668	545	474	2083	492	617	744	1872	1130	4968	4673
Max vehicle speed (km/h)	43	56	33	34	62	26	35	0	55	21	54
Min vehicle speed (km/h)	0	20	0	1	38	8	1	0	0	0	0
Speed < 30 km/h (%)	66	15	93	75	0	100	87	100	66	100	51
Speed > 30 km/h (%)	34	85	7	25	100	0	13	0	34	0	49

$$\left\{ \begin{array}{l} X_1 \rightarrow Y_2 \\ X_2 \rightarrow Y_4 \\ X_3 \rightarrow Y_1 \\ X_4 \rightarrow Y_3 \\ X_5 \text{ appears.} \end{array} \right. \quad (18)$$

## V. EXPERIMENTAL RESULTS

This section evaluates the proposed approach on real data coming from the well-known KITTI dataset [34]. First, the dataset description is presented, followed by the experimental setting. Secondly, the obtained results are analyzed and commented. It is noted that this evaluation focuses only on data association, so no tracking is done.

### A. Datasets

The KITTI vision dataset provides data recorded from different sensors mounted on a moving vehicle on urban roads [34]. It contains camera images, laser scans, and GPS/IMU data. The dataset also includes object labels classified in 8 categories. For this evaluation, only image data have been used where detections are defined by 2D bounding box tracklets. Four object classes have been considered: pedestrian, cyclist, car, and van. Table V presents a part of these sequences according to their different road context and the number of detections. On some sequences, the vehicle mainly moving at a speed less than 30 km/h which is common in urban areas, e.g. sequences 6, 13, 14, and 19. Sequence 16 was recorded when the vehicle stopped at a crosswalk, i.e. *speed* = 0 km/h. On other sequences, the vehicle was moving at a speed sometimes exceeding 50 km/h, e.g. sequences 4 and 8. Fig. 4 illustrates the number of objects per image and their proportion on each of the sequences where more than 30000 associations have been evaluated. To the best of our knowledge, no study has been evaluated on so many real data. These latter cover different road scenarii containing various objects as shown in Fig. 5.

### B. Experimental Setting

The matching process is based on the distance between objects attributes. In this work, only 2D position in the image plane is considered as pieces of evidence. Thus, the distance  $d_{i,j}$  is defined as follows:

$$d_{i,j} = 0.5 \times (d_{i,j}^{left} + d_{i,j}^{right}) \quad (19)$$

Table VI  
COMPUTATION TIME (ms) OF THE COMBINATION STEP FOR 24 FRAMES CONTAINING  $(n, m)$  OBJECTS.

$(n, m)$	All Sources	4-Sig. Src.	3-Sig. Src.	2-Sig. Src.
(4, 4)	1.33	1.49	0.60	0.39
(7, 7)	> 0.1s	2.27	0.92	0.59
(10, 10)	> 5s	3.54	1.35	0.89
(13, 13)	$\simeq 4min$	5.28	2.20	1.33

where  $d_{i,j}^{left}$  ( $d_{i,j}^{right}$ ) is the Euclidean distance between top-left (bottom-right) points of the bounding boxes of objects  $X_i$  and  $Y_j$  as illustrated in Fig. 6.

The critical parameters to estimate belief masses are:  $\alpha = 0.9$ ,  $\tau = 0.5$  and  $\epsilon = 0.001$  for *DSmP* transformation. The proposed approach is written in C++ and runs on Intel core i7 2.20 GHz with 8 GB RAM.

### C. Results and Analysis

The performance of the k-significant sources combination refers to its capacity to reduce complexity while maintaining a high decision quality. Therefore, the evaluation focuses on the Computation Time (*CT*) and the *recall* which are defined as follows:

$$\left\{ \begin{array}{l} CT = \sum_t ET_t \\ recall = \frac{\sum_t TA_t}{\sum_t GT_t} \end{array} \right. \quad (20)$$

where  $ET_t$  is the execution time of the frame  $t$ ,  $TA_t$  and  $GT_t$  are the numbers of true associations and ground truth associations respectively.

Table. VI compares the running time of the combination step using two approaches according to the number of objects. The first is to combine all the sources and the second combines the k-significant sources where  $k \in [2, 4]$ . To show the real-time aspect of the proposed approach, the association process is applied for 24 frames. The results confirm that the proposed approach needs low computation time than combining all sources. The smaller the number of combined sources, the shorter the computation time. With  $n = m = 13$ , the proposed approach ( $k = 2$ ) needs 1.33ms on 24 frames while combining all sources takes  $\simeq 4$  minutes which is not acceptable for real-time applications. In addition, combining all sources grows exponentially the computation cost with  $(n, m)$  while the time complexity of the proposed approach is polynomial which makes it well-suited for real-time applications (Cf. Fig. 7).

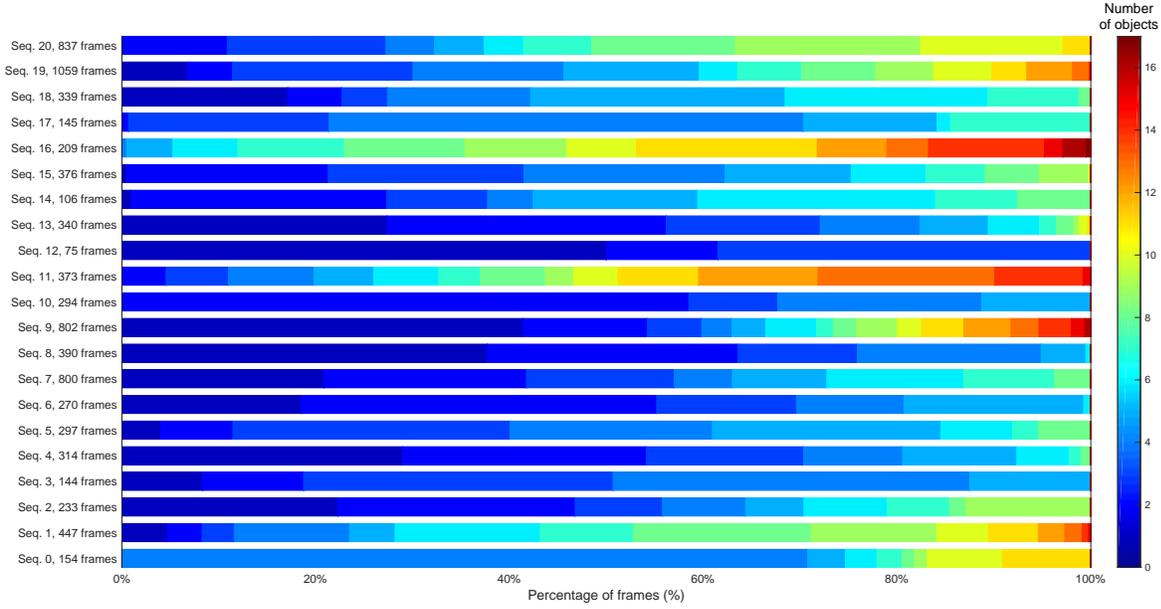


Figure 4. The number of objects per frame vs. percentage of frames.

Table VII  
COMPUTATION TIME ( $ms$ ) OF THE DECISION-MAKING  
STEP FOR 24 FRAMES CONTAINING  $(n, m)$  OBJECTS.

$(n, m)$	JPP	Our method	Comp. time gain
(2, 2)	0.21	0.16	23.91%
(3, 3)	1.2	0.16	86.66%
(4, 4)	9	0.21	97.66%
(5, 5)	104	0.27	99.74%
(6, 6)	> 9s	0.33	99.99%
(7, 7)	> 46min	0.90	99.99%

Table. VII compares the complexity of the proposed decision-making algorithm with the JPP method according to the number of objects. Both of these methods are based on a global optimization. The results show that the proposed algorithm needs low computation time than JPP to make association decisions. With more than 4 perceived/detected objects, the complexity is reduced by more than 97%. For instance, with  $n = m = 7$ , our proposed algorithm needs less than  $1ms$  to assign perceived objects on 24 frames while JPP takes too large time, more than  $\simeq 46$  minutes. Fig. 8 confirms that our algorithm is characterized by a polynomial complexity while JPP has a high exponential complexity which makes impossible its application on the KITTI sequences. For this reason, the rest of the results presented in this section are obtained by our simple decision-making algorithm.

To measure the gain on complexity, the variation in the computation time of a system without  $(CT_w^i)$  and with the  $k$ -significant sources combination  $(CT_k^i)$  is computed for each sequence ( $i$ ) (21). The higher gain, the better complexity reduction we get. In the same manner, the recall gain is computed (22). The higher  $Gain_{recall}^i$ , the better decision-

quality we get. A higher  $Gain_{recall}^i$  preserves well the decision-quality.

$$Gain_{CT}^i = \frac{(CT_w^i - CT_k^i)}{CT_w^i} 100. \quad (21)$$

$$Gain_{recall}^i = \frac{(recall_k^i - recall_w^i)}{recall_w^i} 100. \quad (22)$$

The weighted average of gain based on all sequences is given by:

$$\begin{cases} Gain_{CT}^{avg} &= \sum_{i=0}^{20} w_i Gain_{CT}^i \\ Gain_{recall}^{avg} &= \sum_{i=0}^{20} w_i Gain_{recall}^i \end{cases} \quad (23)$$

where the weight  $w_i$  is  $w_i = n_i / \sum_{i=0}^{20} n_i$  and  $n_i$  being the number of associations of the  $i$ -th sequence.

Fig. 9 presents the weighted average of the computation time gain versus  $k$ . These results are obtained by varying the number of significant sources selected, i.e.  $k$ . For all dataset, more than 30000 associations, the gain exceeds 99.90% which is well-suited for real-time applications. This gain is explained by the fact that our approach has a polynomial complexity while combining all sources is characterized by an exponential complexity (Cf. Fig. 7). In addition, the obtained results show that the computation time reduction is inversely proportional to the  $k$  parameter as shown in Table. VI. Indeed, by reducing the number of significant sources, the combination complexity decreases which allows a more important gain. Although if the gain, which is expressed as a percentage, seems small between the different values of  $k \in [2, 7]$ , it remains important for real-time constrain.

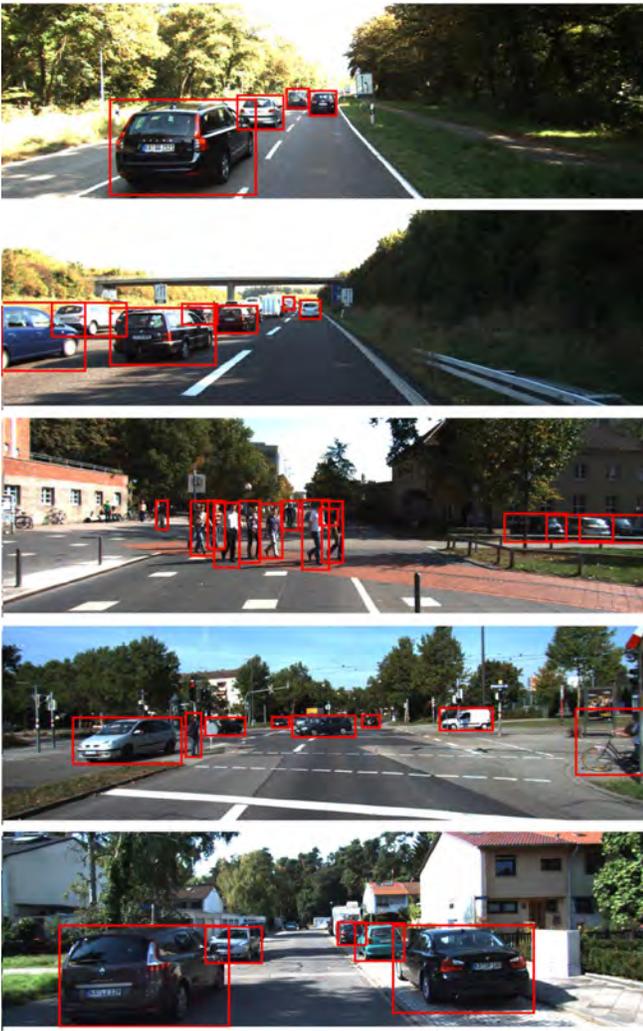


Figure 5. Examples of images provided by KITTI [30].

The gain depends also on the number of perceived objects. In fact, contrary to our approach, combining all sources increases exponentially the computation time with perceived/detected objects ( $n, m$ ). Therefore, the more objects in the scene, the greater the gain will be (Cf. Fig. 10). That is why for sequences 3, 6, 8, 10, and 12 where the number of detections is mostly less than 4, the gain is less than 40% while for other sequences is more than 80%. Therefore, the obtained results lead to conclude that the more complex is the sequence, the larger is the computation time reduction.

Now, how about the decision quality? Combine just the significant sources, affects the decisions or not? Fig. 11 presents the weighted average of the recall gain versus  $k$ . It is clear that the gain is insignificant,  $-0.1\% < Gain_{recall} < 0.05\%$ . This result proves that focusing only on significant information does not necessary affect the decision quality. Furthermore, the obtained results also show that ignoring the useless information can improve slightly the quality of decisions. For instance, on sequences 11, 17, and 18 the association decisions are improved by more than 4% (Cf. Fig. 12). Therefore, the

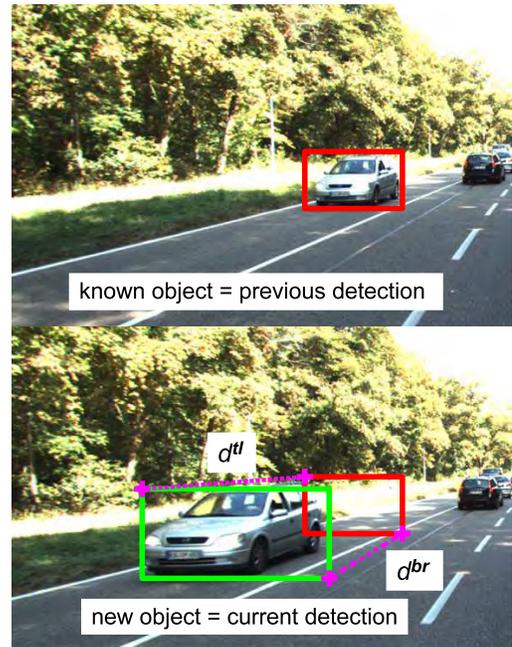


Figure 6. The illustration of the distance between a detected and a known object [30].

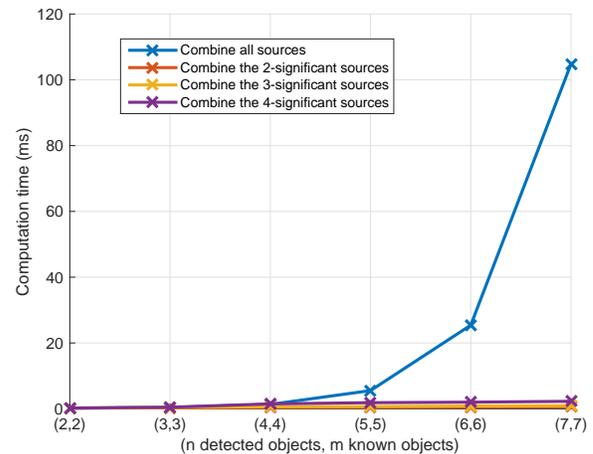


Figure 7. Computation time of the combination step as a function of the number of objects.

solution proposed provides good performances by reducing significantly the computation time while preserving the association decisions.

The choice of parameter  $k$  depends on the application context and on the desired performances. For the object association in road environment and based on our tests,  $k = 3$  appears to be a good setting threshold parameter.

## VI. CONCLUSION

This paper presented a new evidential data association based on significant sources combination and a simple decision-making algorithm. The main objective of the proposed approach is to reduce the complexity and time consumption of

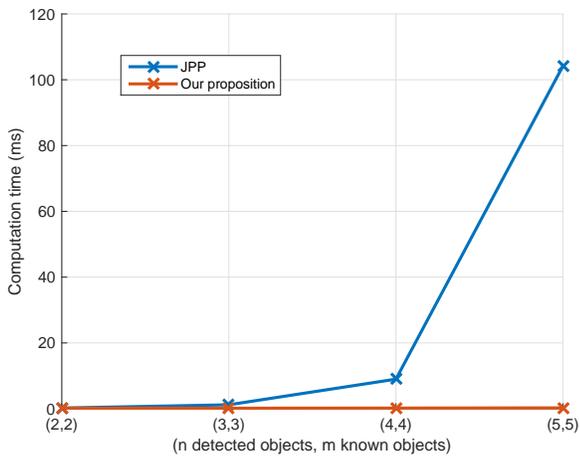


Figure 8. Computation time of the decision-making step as a function of the number of objects.

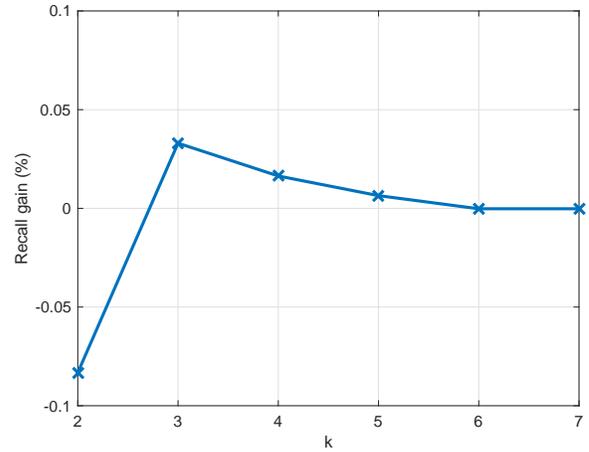


Figure 11. Recall gain as a function of the parameter  $k$  new.

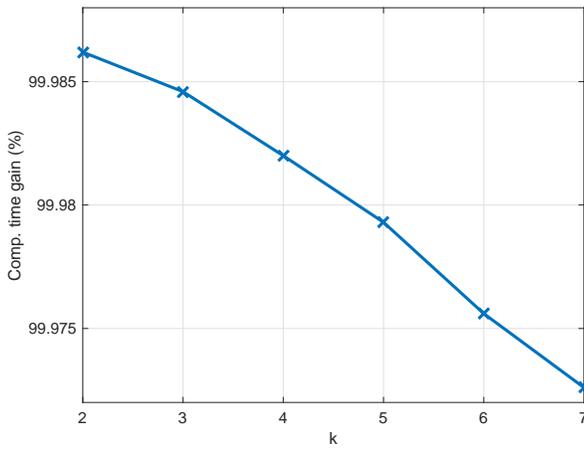


Figure 9. Computation time gain as a function of the parameter  $k$ .

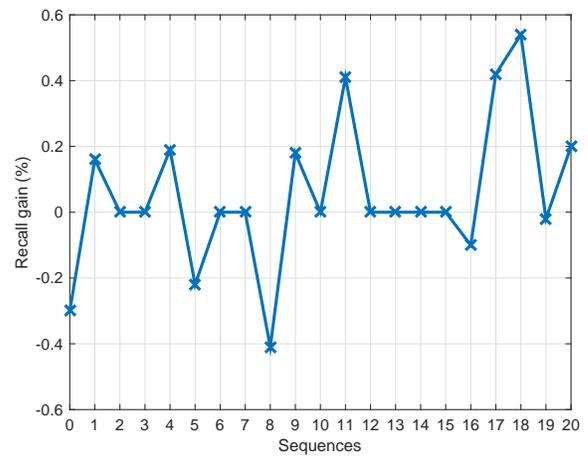


Figure 12. Recall gain of 3-sig. sources approach on each sequence.

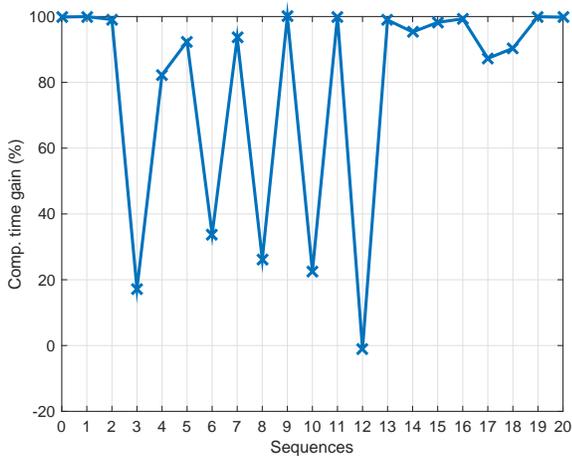


Figure 10. Computation time gain of 3-sig. sources approach on each sequence.

data fusion based on DS<sub>m</sub>T techniques (PCR6 and DS<sub>m</sub>P). This approach focuses only on information having belief in top  $k$  highest masses and removes useless information. Therefore, only  $k$ -significant sources are combined to deal with the association problem.

Applied to intelligent vehicles perception, the experimental results show the effectiveness of the proposed approach in the reduction of the complexity by more than 99% in dense scenes. Besides, experimental results show that the proposed solution preserves well the decision-quality. It can be noted that the  $k$ -significant sources combination is not intended only for road environment perception. It can be applied to any data association process based on these DS<sub>m</sub>T techniques.

Future work should combine heterogeneous sensor data to enhance the object association. Also, we plan to evaluate if an improvement of PCR6 rule of combination would be helpful for the data association problems.

## REFERENCES

- [1] J. M. Armingol, J. Alfonso, N. Aliane and al., “Chapter 2 - Environmental Perception for Intelligent Vehicles, Intelligent Vehicles: Enabling Technologies and Future Developments”, Elsevier, 2018.
- [2] J. V. Brummelen, M. O’Brien, D. Gruyer and H. Najjaran, “Autonomous vehicle perception: The technology of today and tomorrow”, *Transp. Res. Part C: Emerg. Technol.*, Vol. 89, 384–406, 2018.
- [3] M. Boumediene, J.-P. Lauffenburger, J. Daniel, C. Cudel, “Coupled detection, association and tracking for traffic sign recognition”, in *IEEE Intell. Vehicle Symposium*, Dearborn, Michigan, USA, pp. 1402–1407, 2014.
- [4] M. Boumediene, J.-P. Lauffenburger, J. Daniel, C. Cudel, A. Ouamri, “Multi-roi association and tracking with belief functions: Application to traffic sign recognition”, *IEEE Trans. on Intell. Transp. Syst.*, Vol. 15, pp. 2470–2479, 2014.
- [5] S. Steyer, G. Tanzmeister, D. Wollherr, “Grid-based environment estimation using evidential mapping and particle tracking”, *IEEE Trans. on Intell. Vehicles*, Vol. 3, pp. 384–396, 2018.
- [6] S. Blackman, “Multiple-target Tracking with Radar Applications”, *Radar Library*, Artech House, 1986.
- [7] T. E. Fortmann, Y. Bar-Shalom, M. Scheffe, “Sonar tracking of multiple targets using joint probabilistic data association”, *IEEE J. of Ocean. Eng.*, Vol. 8, pp. 173–184, 1983.
- [8] Y. Bar-Shalom, P.K. Willett, X. Tian, “Tracking and Data Fusion: A Handbook of Algorithms”, YBS Publishing, 2011.
- [9] Y. Bar-Shalom, X. Li, “Multitarget-multisensor tracking: principles and techniques”, YBS Publishing, Brandford, U.K., 1995.
- [10] S. Blackman, R. Popoli, “Design and analysis of modern tracking systems”, Artech House, Norwood, MA, USA, 1999.
- [11] D. Mercier, E. Lefèvre, D. Jolly, “Object association with belief functions, an application with vehicles”, *Inf. Sci.*, Vol. 181, pp. 5485–5500, 2011.
- [12] A. P. Dempster, “A generalization of bayesian inference”, *J. R. Stat. Soc., Ser. B*, Vol. 30, pp. 205–247, 1968.
- [13] G. Shafer, “A mathematical theory of evidence”, *Princeton University Press*, Princeton, NJ, USA, 1976.
- [14] M. Rombaut, “Decision in multi-obstacle matching process using Dempster-Shafer’s theory”, in *Int. Conf. on Advances in Vehicle Control and Safety*, pp. 63–68, Amiens, France, 1998.
- [15] D. Gruyer, S. Demmel, V. Magnier, R. Belaroussi, “Multi-hypotheses tracking using the Dempster-Shafer theory, application to ambiguous road context”, *Inf. Fus.*, Vol. 29, pp. 40–56, 2016.
- [16] P. Smets, R. Kennes, “The transferable belief model”, *Artif. Intell.*, Vol. 66, pp. 191–234, 1994.
- [17] M. Boumediene, J. Dezert, “Evaluation of Probabilistic Transformations for Evidential Data Association”, in *Proc. of IPMU 2020 Int. Conf.*, pp. 312–326, Lisbon, Portugal, June 15–19, 2020.
- [18] J. Daniel, J. P. Lauffenburger, “Multi-object association decision algorithms with belief functions”, in *Proc. of Fusion 2012 Int. Conf.*, Singapore, July 9–12, 2012.
- [19] T. Denœux, N. El Zoghby, V. Cherfaoui, A. Jouglet, “Optimal object association in the Dempster-Shafer framework”, *IEEE Trans. on Cybern.*, Vol. 44, pp. 2521–2531, 2014.
- [20] L. A. Zadeh, “On the validity of Dempster’s rule of combination”, *Memo M79/24*, Univ. of California, Berkeley, USA
- [21] F. Smarandache, J. Dezert (Editors), “Advances and applications of DSMT for information Fusion (Collected works)”, Vol. 1-4, American Research Press, 2004–2015.
- [22] J. Dezert, P. Wang and A. Tchamova, “On the validity of Dempster-Shafer theory”, in *Proc. of Fusion 2012 Int. Conf.*, Singapore, July 9–12, 2012.
- [23] A. Tchamova, J. Dezert, “On the behavior of Dempster’s Rule of combination and the foundations of Dempster-Shafer theory”, (Best paper awards), *6th IEEE Int. Conf. on Int. Syst.*, Sofia, Bulgaria, Sept. 6–8, 2012.
- [24] F. Smarandache, V. Kroumov, J. Dezert, “Examples where the conjunctive and Dempster’s rules are insensitive”, in *Proc. of Advanced Mechatronic Systems Int. Conf.*, Luoyang, China, Sept. 25–27, 2013.
- [25] A. Martin, C. Osswald, “A new generalization of the proportional conflict redistribution rule stable in terms of decision”, in *Florentin Smarandache Jean Dezert. Adv. Appl. DSMT Inf. Fusion, Am. Res., Press, Rehoboth, USA. pp. 69–88, 2006.*
- [26] T. Dezert, J. Dezert, “Improvement of Proportional Conflict Redistribution Fusion Rules for Levee Characterization”, in *Proc. of ESREL 2021 Int. Conf.*, Angers, France, 19–23, Sept. 2021.
- [27] T. Dezert, J. Dezert, F. Smarandache, “Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments”, in *Journal of Advances in Information Fusion (JAIF)*, Vol. 16, pp. 48–73, 2021.
- [28] J. Dezert, F. Smarandache, “A new probabilistic transformation of belief mass assignment”, *Proc. of Fusion 2012 Int. Conf.*, Cologne, Germany, 2008.
- [29] J. Sudano, “The system probability information content (PIC) relationship to contributing components, combining independent multi-source beliefs, hybrid and pedigree pignistic probabilities”, *International Conference on Information Fusion*, Annapolis, USA, 2002.
- [30] M. Boumediene, “Evidential Data Association: Benchmark of Belief Assignment Models”, in *Proc. of International Conference on Advanced Electrical Engineering*, Algeria, November 2019.
- [31] D. Gruyer, V. Berge-Cherfaoui, “Multi-objects association in perception of dynamical situation”, in *Proc. of the Fifteenth Conf. on Uncertainty in Artificial Intelligence*, pp. 255–262, San Francisco, CA, USA, 1999.
- [32] C. Royère, D. Gruyer, V. Cherfaoui, “Data association with believe theory”, in *Proc. of Fusion 2000 Int. Conf.*, Paris, France, 2000.
- [33] J. Munkres, “Algorithms for the assignment and transportation problems”, *J. Society for Industrial & Applied Mathematics*, Vol. 5, 1957.
- [34] A. Geiger, P. Lenz, R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite”, in *Proc. of CVPR 2012 Int. Conf.*, Rhode Island, USA, 2012.

# Data Fusion of In Situ Geophysical and Geotechnical Information for Levee Characterization

Théo Dezert<sup>a,b</sup>, Sérgio Palma Lopes<sup>a</sup>, Yannick Fargier<sup>b,c</sup>, Lucile Saussaye<sup>b</sup>, Philippe Côte<sup>a</sup>

<sup>a</sup>Gustave Eiffel University, GERS, GeoEND, F-44344 Bouguenais, France.

<sup>b</sup>CEREMA Direction Territoriale Normandie-Centre, F-41000 Blois, France.

<sup>c</sup>Gustave Eiffel University, GERS, RRO, F-69675 Bron, France.

Emails: theo.dezert@univ-eiffel.fr, sergio.lobes@univ-eiffel.fr, yannick.fargier@univ-eiffel.fr,  
Lucile.saussaye@cerema.fr, philippe.cote@univ-eiffel.fr

Originally published as: T. Dezert, S. Palma Lopes, Y. Fargier, L. Saussaye, P. Cote, *Data Fusion of In Situ Geophysical and Geotechnical Information for Levee Characterization*, Bulletin of Engineering Geology and the Environment, Vol. 80, pp. 5181–5197, 2021 (with correction in Vol. 80, p.5199), and reprinted with permission.

**Abstract**—Due to the lack of knowledge concerning their construction and their history (breaks and repairs, extensions...), fluvial levees are often badly characterized. Breaks of work are likely to lead to disastrous consequences such as loss of lives and economic disasters. In order to prevent the risk of breakage, special supervision of the protection levee is required. Recognized methodologies for the assessment of hydraulic structures include complementary geotechnical and geophysical reconnaissance methods. This work presents a new way of mathematically combining data from these two types of information sources, taking into account the specificities of each kind of method (level of imperfection associated with the data, spatial distribution of the information). This new methodology considers the framework fixed by the theory of belief masses and improves the characterization of lithological sets within levees. It provides information on the level of conflict between information sources while proposing a confidence index associated with the results. The methodology is implemented through a subsoil section characterized by a real earthen levee investigation campaign. This campaign involves electrical resistivity tomography as well as particle size distribution from laboratory testing and on-site cone penetrometer test. The results highlight the ability of this fusion methodology to characterize the considered materials as well as to specify the positions of the interfaces and the associated levels of confidence.

**Keywords:** geophysics, geotechnics, protection levee, belief functions.

## I. INTRODUCTION

Fluvial levees are elevated manmade structures, built up for flood protection, between channels and floodplains [1]. Unfortunately, some hydraulic earthworks cannot ensure their role in flood episode, and are likely to break, leading to catastrophic events (human, material or economic damages). There is therefore a real need to prevent the risks of rupture by characterizing these complex human structures. To answer this need, investigation campaigns are set up for subsoils characterization and weak zone identification. These campaigns usually involve the use of geophysical and geotechnical methods [2] to make a diagnosis and assess the levee stability.

Geophysical investigation methods are non-intrusive and provide physical information over a large volume of subsoil. This information, however, is potentially tainted with important uncertainties, notably due to the indirect and integrating

aspects of the methods as well as to the limited resolution of the inverse problems. In a complementary way, geotechnical methods are intrusive but provide spatially more punctual and more precise information, being directly in contact with the material to be identified. An important outcome for the characterization of the investigation campaigns is to be able to combine the information acquired by these two sets of methods, while taking advantage of their specificities and their respective uncertainties, inaccuracies and spatial distributions [3]. The complementarity of these methods is rarely used to its full potential and the results are often simply graphically superimposed instead of being mathematically merged [4].

To characterize a levee and its possible weak areas, it is necessary to distinguish the different geological materials in place. The positions of interfaces must also be located, as well as the presence of any anomalies (low-density zone, presence of pipe,...). It is at these interfaces or anomalies that the internal erosion is likely to be initiated, eventually leading to the rupture of the structure [5]. A characterization of these geological sets and interfaces with associated confidence indexes could be of great help if they were included in failure hazard models.

In this work, we propose a novel methodology for the fusion of information based on the use of belief functions [6], [7]. We compare two different combination rules to merge geophysical and geotechnical data while taking into account the specificities of each method (spatial distribution, inaccuracy, uncertainty and incompleteness). We use belief functions (BFs) theory since it does not require learning periods or having very large data sets as the use of artificial neural networks would require [8]. The theory of the BFs makes it possible to quantify the conflict between the sources of information and to quantify uncertainty where the probabilistic theory only considers equiprobabilities. Finally, the strong point of the BFs theory with respect to Zadeh's theory of possibilities [9] is that it is possible to merge disjoint intervals. In the field of geosciences, some works use the BFs to provide results for slope instability [10], [11], ground water [12] or flood susceptibility mapping [13]. To our knowledge, no work has been published on the merging of geophysical

and geotechnical data for an investigation campaign of a river embankment levee.

Our fusion methodology has already been tested and validated considering two sources of information, as part of numerical simulations [14], [15] and of a laboratory test bench experiment [16]. In this work, we present the results of our approach applied to data acquired by three sources of information, on a real levee of the Loire River, located in Saint-Clément-des-Levées (France). It is the first in-situ validation of this merging process. The geotechnical methods used are the cone penetration test (CPT) and the particle size distribution from laboratory testing, after on-site drilling. The considered geophysical method is the electrical resistivity tomography (ERT). The objective of this study is to highlight the ability of our methodology to distinguish the major geological sets constitutive of the levee by suggesting their distribution and proposing the location of the interfaces. The presented results are associated with confidence indexes.

This article is organized as follows: in section II, we give a presentation of the Investigated site and of the three investigation methods, both geophysical and geotechnical, used in this campaign. In section III, we describe the fusion methodology by first introducing the BF's concept and the considered combination rules. We then show how the geophysical and geotechnical data are respectively processed, and we propose two brief parametric studies on different belief mass allocation methods. We finally present the final fusion results in section IV and discuss them in section V to highlight the interests, drawbacks and perspectives of such a fusion methodology.

## II. INVESTIGATED LEVEE AND INVESTIGATION METHODS

### A. Saint-Clément-des-Levées fluvial levee

The studied structure is a fluvial levee located in Saint-Clément-des-Levées near Saumur (France) along the River Loire in the Val d'Authion area (Fig. 1-b). It is a clayous-sandy embankment with Turonian bedrock overlaid by alluvial materials (Fig. 1-a). This levee has been the subject of many studies presenting both geophysical (ERT, radio-magnetotelluric, Slingram, ground penetrating radar) and geotechnical (laboratory tests on collected samples: particle size distribution, clay content, moisture content, density) investigation campaigns to detect possible water circulation within the structure [17], [18], [19]. The investigation campaigns were carried out during the day between June 26 and July 5, 2018. On this earthen hydraulic structure, we carried out a geophysical campaign using the ERT method and a geotechnical campaign with coring and CPT tests carried out on the levee crest. The positions of the tests and the electrode line are displayed in Fig. 1-c.

### B. Electrical resistivity tomography (ERT)

The basic principle of DC-resistivity methods consist in injecting an electric current of known intensity ( $A$ ) by means of two "current" electrodes and measuring a voltage ( $V$ ) between two "potential" electrodes. Such measurements are acquired for several stations (positions of the current and the

potential electrodes). Depending on some parameters such as electrode layout and topography, apparent resistivity values can be computed. A two dimensional (2D) ERT, such as the one considered in this study, consists in aligning a series of electrodes and acquiring a large number of measurements based on four-electrode configuration. The apparent resistivity data acquired are then inverted using an inversion software to reconstruct a complete 2D-section of electrical resistivity ( $\Omega \cdot m$ ). In this work, we used the *Res2Dinv* procedure (ver 3.71.118) [20].

In 2008, as part of the French ERINOH Project [21], the levee was instrumented using two electrode lines. Each line is composed of 48 electrodes with an inter-electrode spacing of 2 meters and buried at 1.10 meters deep below the roadway. The embedded electrode lines are borehole resistivity cables that had been laid horizontally in the levee subsoil at installation time. Each take-out on these cables is a molded stainless steel cylinder with a length of 60 mm and a diameter of 12 mm. These molded take-outs are directly in contact with the soil and act as electrodes without using any rods or so. In this study, we used the electrode profile located on the land side of the crest (Fig. 1-c). The acquisition was carried out in a Wenner-Schlumberger array configuration and three data points (standard deviation greater than 5%) were removed. The acquisition system is a Syscal Pro Switch 96 (Iris Instruments) multi-channel resistivity meter. The current transmission signal is a regular step function. A step duration (half-period) of 250 ms was used, and each received potential measurement was stacked 6 times (except very few that were stacked 9 times) which yielded relative standard deviation values of less than 2% for most of the 1591 data points (and up to about 10-15% for only a couple of data points). All ground to electrode contact resistances are smaller than about 2 k $\Omega$ ·m, enabling sufficient current transmission and high quality potential measurement. Indeed, received potential drops range from 12 mV to 1100 mV, which allows the implemented resistivity meter to deliver data with high signal to noise ratios. Since the ERT method is an integrative method and the electrodes are buried, we considered a 1.10 meters thick layer above the electrode line in the inversion process, with an associated resistivity of about 65  $\Omega$ ·m. This value was determined after a first inversion process, considering the average resistivity value over the first 50 centimeters of the subsoil. The proposed section of the inversion results, which will be used for the rest of our study, is displayed in Figure 2. We do not display inverted resistivity values below 12m since they cannot be merged with geotechnical data due to limited borehole depth. The inversion was carried out on 4 iterations, since the change in RMS error is below 0.3 % between iteration 3 and 4, considering a  $L_1$  norm regularization [22]. It corresponds to a robust inversion allowing to emphasize the contrast of resistivities between the geological sets but also to limit the effects of too noisy data. In a geoelectrical acquisition, the electrical current flowing in the extra-trapezium zones is a minority compared to the intra-trapezium zones. Even though, we use an extended model discretization. This choice is made because the geometry of a

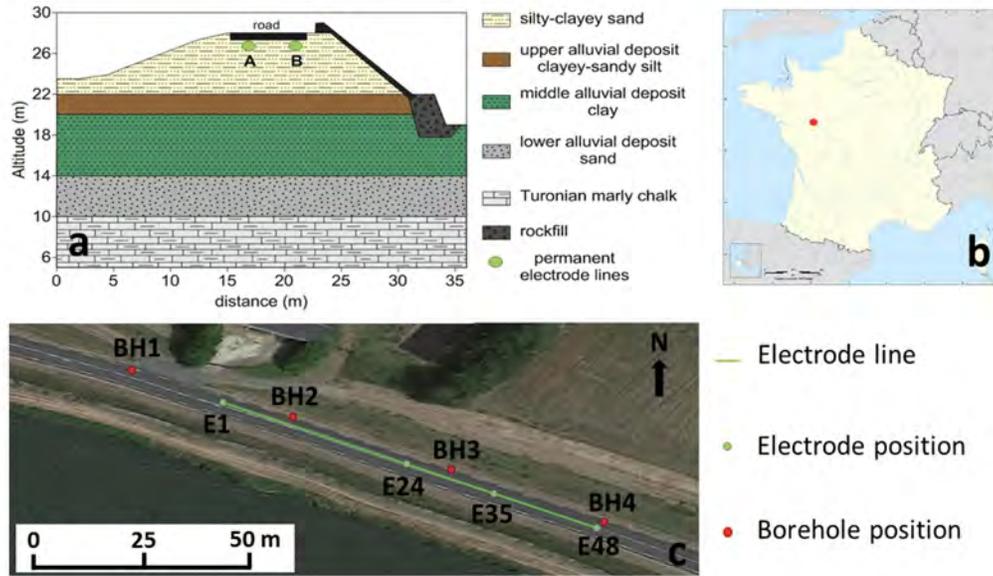


Fig. 1. (a) Saint-Clément-des-Levées' levee cross section displaying the geological materials and the installed monitoring devices [18]. Only electrode line A was used for our study. (b) Map of France.

rectangular section makes it possible to simplify the processing of the data as well as their fusion. It enables to work with cells of identical surface at a fixed depth. In addition, this avoids having to work with large extra-trapezium meshes having very important sensitivity values. However, to take into account the difference in reliability of the results between intra and extra-trapezium cells, the resistivity imprecision values resulting from the inversion are taken into account during the belief masses attribution stage. To go further, in future works, it could be pertinent to integrate sensitivity values.

Looking at the results, a more resistive upper part of about 3 meters thick stands out. The underlying part seems more conductive with an area that seems even more conductive from horizontal position  $x = 46$  m to  $x = 70$  m and from vertical position  $z = 3.86$  m to  $z = 10.7$  m.

#### C. Drilling cores and Particle size distribution

Drilling cores were carried out on the levee crest, in four locations, displayed in Fig. 1-c. These drilling were made down to 7.40 meters deep, using a Texoma machine producing 10 cm diameter cores. Once returned to the laboratory, the cores were visually identified in order to delineate sections of material that could be considered as belonging to the same particle size class. Some samples were collected to perform particle size distribution analysis following the NF P94-056 French standard [23]. The results show the existence of two major particle size classes according to the NF P11 300 [24] soil classification. The two major classes characterized are the fine materials (designated as “A materials”) and the sandy to gravelly materials with presence of fines (designated as “B materials”). In this study,  $D_{\max}$  value [23] is always lower than 50 mm, thus we take into account the value of the cumulative sieve under 80  $\mu\text{m}$  to characterize A from B

materials. When this value is greater than 35%, the geological material is considered to belong to A class, otherwise it is considered to belong to B class. The results of the particle size distribution tests with the associated material classes are shown in Figure 3, with the depth 0 m corresponding to the position of the buried electrodes. The horizontal black lines stand for the delimitation of the materials made by the visual inspection. These results point out that B materials seem present in the upper part of the section (from 0 to 3.40 m deep for borehole 2 and 3) while finer materials A tend to be located below.

#### D. Cone penetrometer test (CPT)

The CPT method consists of pushing rods into the soil with a conical tip at the end at a controlled rate in order to record tip resistance,  $q_c$  [MPa], and friction sleeve,  $f_s$  [MPa], values. Four tests were carried out at the same locations as the drilling tests (Fig. 1-c) using a Gouda machine with a tip of 3.6 cm diameter and with an acquisition rate of 10 cm, following the French standard NF P94-113 [25]. The tests were carried out on a vertical length of 8.80 m. Using the two measured parameters, it is possible to determine ISBT, the Soil Behavior Type Index proposed by Robertson [26] and presented Eq. (1):

$$I_{\text{SBT}} = \sqrt{(3.47 - \log(\frac{q_c}{P_a}))^2 + (1.22 - \log(\frac{f_s}{q_c \cdot 10^2}))^2} \quad (1)$$

With  $P_a = 0.1$  MPa. The  $I_{\text{SBT}}$  index provides information on the nature of the soil in terms of particle size class. The results obtained are shown in Figure 4, with the depth 0 m corresponding to the position of the buried electrodes. Most of the materials appear to be sand mixtures thanks to the computation of the  $I_{\text{SBT}}$ , which appears to be in contradiction with the particle size distribution below 3.40 m depth (Figure 3).

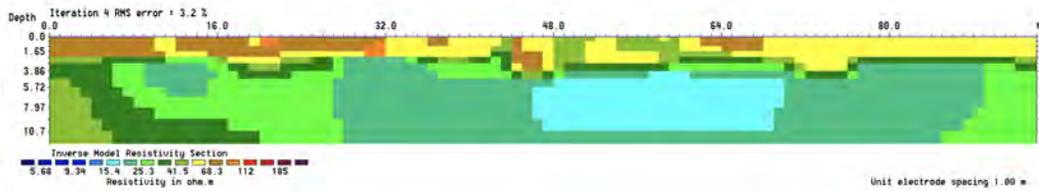


Fig. 2. Levee modeled resistivity longitudinal section obtained by inverting Wenner-Schlumberger apparent resistivity data acquired with electrode line A shown in Fig. 1-a. The depth 0 m corresponds to the position of the electrodes.

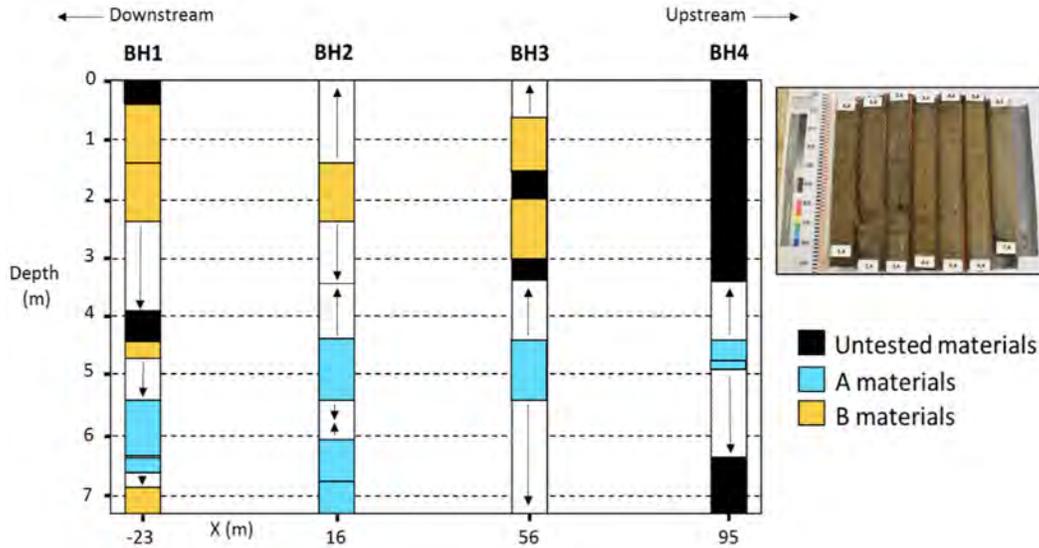


Fig. 3. Photograph of extracted material from BH1 and results of the particle size distribution analysis with the associated soil classes (A and B) in the four drilling cores, locations specified in Fig. 1-c. Arrows symbolize the vertical extension of the information thanks to the visual characterization of a technician.

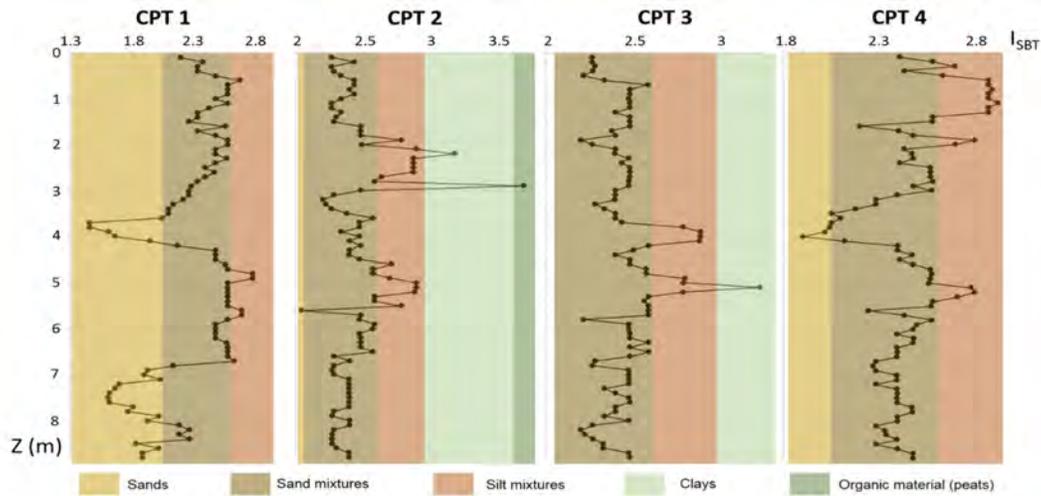


Fig. 4.  $I_{SBT}$  vertical profiles for each CPT test and associated soil classes.

### III. FUSION METHODOLOGY

#### A. Belief functions and combination rules

Shafer [7] introduced the BFs by developing the mathematical theory of evidence inspired by earlier works of

Dempster [6]. Hence, Shafer's theory is often referred to as the Dempster-Shafer theory (DST). This theory allows the computation of the belief and the plausibility of a hypothesis (corresponding to soil material classes in this work) from distinct sources of information (measured data). The practical

benefit of using BFs lies in its ability to manage information from different sources, associated with their respective levels of uncertainties and inaccuracies. In this work, we will be considering three sources of information: two geotechnical (CPT and particle size distribution) and one geophysical (ERT). Another feature of the BFs theory is its ability to assess the level of conflict ( $\emptyset$ ) between sources, i.e. when the information given by one source is contradictory to the information given by another one. Following Smets [27], we consider that uncertainties correspond to degrees of confidence that are given to a value, whereas inaccuracies correspond to intervals of values that can be directly associated with measurement errors related to the investigation method. For example, the uncertainty on measuring the value of a geotechnical parameter identical to the one measured in a borehole increases with the distance to that borehole. The inaccuracy can for its part be associated with the error bar of the corresponding measured datum. In addition, the BFs allow taking into consideration the ignorance and incompleteness of the information. It is indeed possible to grant credit on all the possible results in order to quantify our ignorance, whereas the probabilistic theory would simply assign an equiprobability to each single hypothesis. For the reader eager to learn more, more details concerning the theory can be found in [28].

To define and to use the BFs, it is required (i) to set a Frame of Discernment (FoD), (ii) to assign belief mass values to the hypothesis of this set and for each source of information, (iii) to implement a fusion rule for merging the information; and (iv) to provide a representation of the combined information. The FoD  $\Theta$  consists of all the possible hypothesis within the problem under concern. The elements of the FoD are exhaustive and exclusive, such as for  $n$  hypothesis:

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}. \quad (2)$$

In our problematic, the possible hypothesis of the FoD correspond to classes of geological materials that can be associated with intervals of values of geophysical and geotechnical parameters. Here, we consider that  $\theta_1$  stands for fine-grained materials and  $\theta_2$  stands for coarser-grained materials. We also consider a third hypothesis  $\theta_3$  that will be associated with intervals of values of geophysical and geotechnical parameters that are not included in the two first sets. Thus, we can qualify  $\theta_3$  as being “another” material and we have:

$$\Theta = \{\theta_1, \theta_2, \theta_3\}. \quad (3)$$

The set of all subsets of  $\Theta$  (including the conflict hypothesis,  $\emptyset$ ), is named as “powerset” and written  $2^\Theta$ . In our case, we get:

$$2^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_1 \cup \theta_2, \theta_3, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3\}. \quad (4)$$

The belief mass function  $m_j$  is defined for a source of information  $S_j$  (for  $j = 1, 2$  or  $3$  in our study) and is attributed to a subset  $X$  (defined on  $2^\Theta$ ) in  $[0, 1]$  such that, as in the

probability theory, the more  $m(X)$  tends to 1 and the more the confidence in  $X$  is important :

$$\sum_{X \in 2^\Theta} m(X) = 1. \quad (5)$$

The main difference with the probability theory is that  $X$  can represent the union of many hypotheses. For example if belief mass is attributed to  $\theta_1 \cup \theta_2$ , it means that either  $\theta_1$  OR  $\theta_2$  are possible. Thus, it is possible to model uncertainty and lack of knowledge. Belief and plausibility functions,  $Bel$  and  $Pl$  respectively, are considered as upper and lower bounds of an unknown probability  $P$  such that for any  $2^\Theta$ ,  $Bel(X) \leq P(X) \leq Pl(X)$ . Belief and plausibility functions are in one-to-one relation with the belief mass,  $m(\cdot)$ , and defined by:

$$Bel(X) = \sum_{Z \in 2^\Theta | Z \subseteq X} m(Z) \quad (6)$$

$$Pl(X) = \sum_{Z \in 2^\Theta | Z \cap X \neq \emptyset} m(Z) \quad (7)$$

In our study, we only use belief mass functions since fusion rules are set directly from the allocated belief masses from each information source. The approach developed by Smets [29] in his Transferable Belief Model (TBM) (i.e. conjunctive fusion, so called “open-world assumption”) allows the assignment of a belief mass to the conflict represented by the empty set, so that one considers:

$$m_{1,2,\dots,S}(\emptyset) > 0, \quad (8)$$

where  $m_{1,2,\dots,S}(\cdot)$  denotes the merged belief mass resulting from the combination of information from the different sources. The belief mass  $m_{1,2,\dots,S}(X)$  resulting from the conjunctive fusion of information from all sources (from  $j = 1$  to  $s$ ) is written:

$$m_{1,2,\dots,S}(X) = \sum_{\substack{X_1, X_2, \dots, X_S \in 2^\Theta \\ X_1 \cap X_2 \cap \dots \cap X_S = X}} \prod_{j=1}^S m_j(X_j), \quad (9)$$

with  $m_j(X_j)$  the belief mass respectively attributed to hypothesis  $X_j$  by information source  $j$ .

The conflict level between the  $s$  considered sources of information can therefore be written as:

$$m_{1,2,\dots,S}(\emptyset) = \sum_{\substack{X_1, X_2, \dots, X_S \in 2^\Theta \\ X_1 \cap X_2 \cap \dots \cap X_S = \emptyset}} \prod_{j=1}^S m_j(X_j). \quad (10)$$

According to Shafer’s approach and unlike Smets’ rule, DS rule does not allow the attribution of a belief mass to the conflict. Thus, in DST (which uses the “closed-world assumption”), one has by definition:

$$m_{1,2,\dots,S}^{DS}(\emptyset) = 0. \quad (11)$$

The conflict mass is then reallocated through a normalization factor. The mass of belief in  $X$ ,  $m_{1,2,\dots,S}^{DS}(X)$ , resulting from the fusion of information from  $S$  sources is written:

$$m_{1,2,\dots,S}^{DS}(X) = \frac{1}{1 - m_{1,2,\dots,S}^{DS}(\emptyset)} \sum_{\substack{X_1, X_2, \dots, X_S \in 2^\Theta \\ X_1 \cap X_2 \cap \dots \cap X_S = X}} \prod_{j=1}^S m_j(X_j). \quad (12)$$

The drawback of this combination rule is that the conflict between the sources is no longer displayed. Furthermore, it is possible to obtain counterintuitive results when the conflict level is important. However, the PCR6 (Proportional Conflict Redistribution No. 6) combination rule [30] allows the redistribution of all partial conflicts, in proportion to the masses of the subset concerned by these conflicts, such as  $m_{1,2,\dots,S}^{PCR6}(\emptyset) = 0$  and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$ :

$$\begin{aligned} m_{1,2,\dots,S}^{PCR6}(X) &= m_{1,2,\dots,S}(X) \\ &+ \sum_k^{S-1} \sum_{\substack{X_{i_1}, X_{i_2}, \dots, X_{i_k} \in 2^\Theta \setminus \{X\} \\ (\cap_{j=1}^k X_{i_j}) \cap X = \emptyset}} [m_{i_1}(X) + m_{i_2}(X) + \dots + m_{i_k}(X)] \\ &\cdot \frac{\prod_{l=1}^k m_{i_l}(X) \prod_{l=k+1}^S m_{i_l}(X_{i_l})}{\sum_{l=1}^k m_{i_l}(X) + \sum_{l=k+1}^S m_{i_l}(X_{i_l})} \end{aligned} \quad (13)$$

where  $P^S$  is the set of all permutations of the elements  $\{1, 2, \dots, S\}$ . It should be emphasized that  $X_{i_1}, X_{i_2}, \dots, X_{i_s}$  may be different from each other, equal, or some equal and some different, etc. In this paper, for the sake of conciseness and since PCR6 and DS rules provide quite similar results, we will only be displaying fusion results using Smets and PCR6 rules of combination.

### B. Attribution of belief masses from geophysical data

To attribute belief masses from electrical resistivity data, it is first necessary to define the limits of the resistivity intervals corresponding to the hypothesis of the FoD. To do so, we use a representation in modal classes of the number of cells of the 2D section resulting from the inversion (Figure 2) according to the resistivity values represented in log scale. The accuracy, to two decimal places, of the classes' values shown in Figure 5 make little difference in the characterization methodology of the boundaries we use. We are aware that this precision is superfluous for a geophysical interpretation. What matters is the general trend of the values' distribution. It enables to highlight the large sets of materials constitutive of the subsoil section. This subjective division of classes (abscissa axis, Figure 5) comes from the computation of the upper and lower bound values, considering the following geometric sequence:

$$b_{n+1} = 1.1b_n, \quad (14)$$

with  $b_n$  the lower bound of the interval  $n$ ,  $b_{n+1}$  the upper bound and  $b_0 = 2 \Omega \cdot m$ .

The representation implemented as modal classes (Figure 5), associated with the reading of Figure 2, suggests the presence of two materials of different kinds: i) a material of lower resistivity (in blue, Figure 5) that can be associated with  $\theta_1$  hypothesis of a fine-grained material and ii) a material with higher resistivity (in orange, Figure 5) that can be associated with the hypothesis of a coarser-grained material ( $\theta_2$ ).

The delimitation between these two classes, however, is not straightforward. We propose to associate the intermediate values of resistivities with the hypothesis  $\theta_1 \cup \theta_2$ , suggesting that these resistivity values can be related to the hypothesis “ $\theta_1$  or  $\theta_2$ ”. Thus the bounds of the electrical resistivity classes ( $\Omega \cdot m$ ) are fixed such that:

$$\begin{aligned} [16.28, 31.73] &\text{ is associated with } \theta_1, \\ [51.10, 99.57] &\text{ is associated with } \theta_2, \\ [31.73, 51.10[ &\text{ is associated with } \theta_1 \cup \theta_2, \\ [5.10, 16.28[ \cup ]99.57, 312.45] &\text{ is associated with } \theta_3. \end{aligned} \quad (15)$$

To limit the possible biases imputed to computation of the distance between intervals (approach detailed below), we consider the intervals associated with the hypothesis  $\theta_3$  (i.e.  $[5.10; 16.28[$  and  $]99.57; 312.45]$ ) of same width in log scale (i.e. same ratio between upper and lower bound values) to the intervals associated with  $\theta_1$  and  $\theta_2$ . Once the DC-resistivity intervals are defined, it is required to associate masses of belief to each of the considered hypothesis. This belief masses attribution process has to be carried out for each cell of the section mesh resulting from the inversion (Figure 2). We propose three alternative approaches for the attribution of the belief masses.

The first one, referred as  $D_g$ , consists in considering a Gaussian probability distribution Eq. (16) centered on the inverted resistivity value:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (16)$$

with  $\mu$  the value of inverted resistivity in the considered cell,  $\sigma$  the inaccuracy provided by the inversion process resulting from the computation of the covariance matrix. Given that the area under the Gaussian distribution is equal to 1, the mass is assigned to the hypotheses according to the proportion of the area intersecting the defined resistivity intervals. This can only be done in accordance with Eq. (5), so that each cell is associated with a standardized belief mass distribution. The results of this approach are presented in Figure 6. Figure 6-a highlights the hypotheses having the highest belief mass for each cell section while Figure 6-b displays their associated belief mass values. We find that the belief masses associated with this approach ( $D_g$ ) are very large (often close to 1), suggesting that the ERT method is completely reliable and able to characterize the subsoil materials. Such level of confidence may seem exaggerated. Thus, we propose a second belief masses assignment method.

This other approach, referred as  $D_w$ , relies on the calculation of Wasserstein's distances [31] as previously used

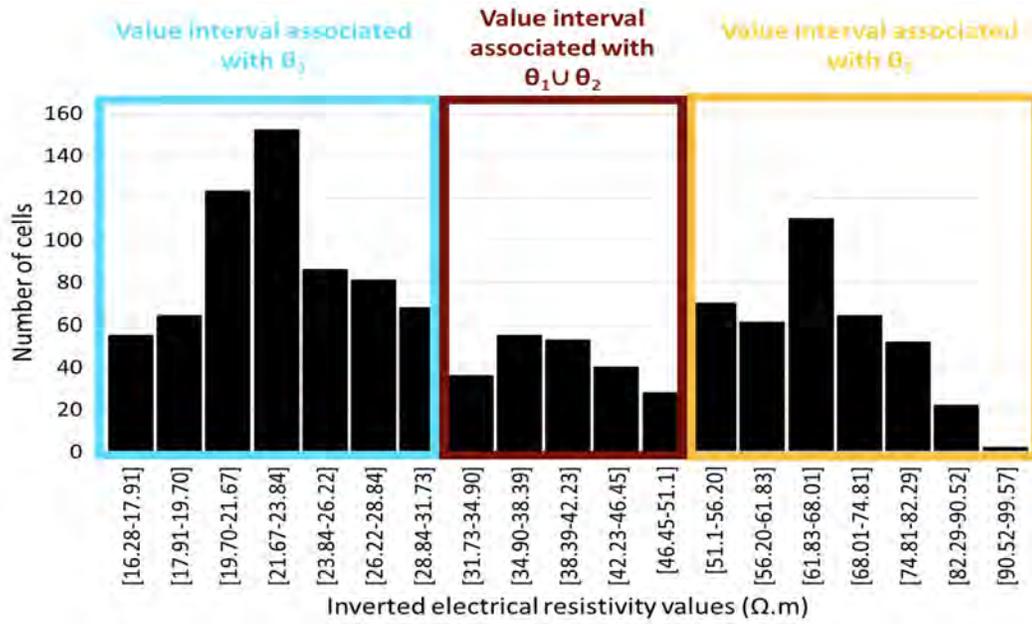


Fig. 5. Model classes' distribution of the cells displayed in Figure 2, according to the inverted electrical resistivity values ( $\Omega \cdot m$ ) and intervals associated to the soil classes.

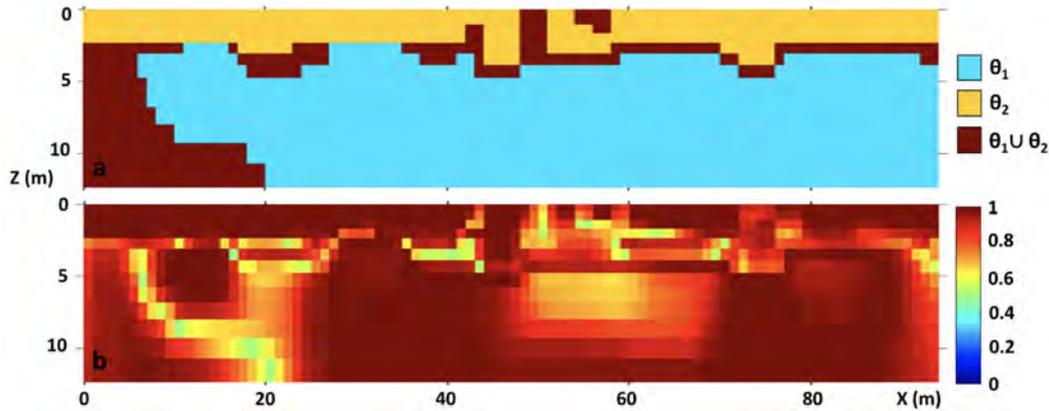


Fig. 6. a) Representation of the hypotheses having the highest belief mass according to the masses attribution from electrical resistivity data considering a Gaussian probability distribution centered on the inverted resistivity value, and b) the associated belief mass values.

in [14], [15]. To associate the belief masses with the FoD hypotheses, we consider the intervals of inverted resistivity values with their associated inaccuracies (example in red in Figure 7). The belief masses are issued from the computation of the Wasserstein distances, considering two resistivity intervals  $A = [a_1, a_2]$  and  $B = [b_1, b_2]$ ,  $A$  being the interval corresponding to a defined hypothesis Eq. (15) and  $B$  being an interval of inverted values. We take into consideration the imprecision level resulting from the inversion for each cell of the mesh, such that  $b_1 = \mu - \sigma$  and  $b_2 = \mu + \sigma$ :

$$D_{\text{Wass}} = \left[ \left[ \frac{\log(a_1 a_2)}{2} - \frac{\log(b_1 b_2)}{2} \right]^2 + \frac{1}{3} \left[ \left( \frac{\log(a_2/a_1)}{2} \right)^2 + \left( \frac{\log(b_2/b_1)}{2} \right)^2 \right] \right]^{1/2} \quad (17)$$

This computation gives the Wasserstein distance between

two intervals taking their size and the distance between them into account. The Wasserstein distances are computed between the inverted values with estimated inaccuracies, and the intervals associated with each hypothesis Eq. (15). In the example illustrated in Figure 7, the Wasserstein distance would be computed between  $[b_1, b_2]$  and each of the other intervals  $[a_0, a_1]$ ,  $[a_1, a_2]$ ,  $[a_2, a_3]$ ,  $[a_3, a_4]$  and  $[a_4, a_5]$ .

Each cell is then associated with a standardized belief mass distribution in accordance with Eq. (5) and inversely proportional to the distance value. This way, the more the distance of an electrical resistivity interval resulting from the inversion is "small" to one hypothesis of the FoD, the more the mass of belief associated is large, and reciprocally. The results of this belief mass assignment approach,  $D_w$ , are displayed in Figure 8. The belief masses derived from this approach

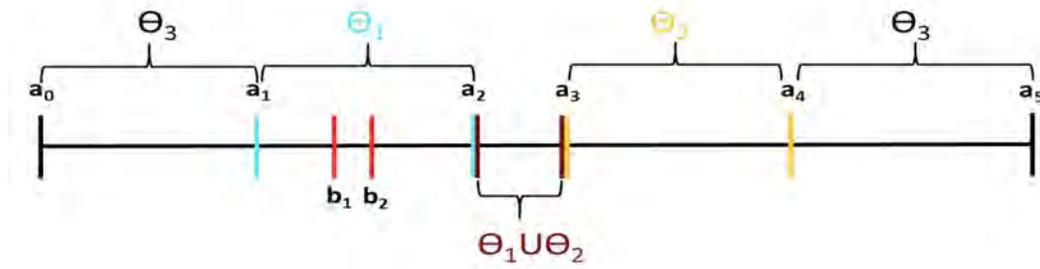


Fig. 7. Diagram displaying the classes described in Eq. (15) with the red interval  $[b_1; b_2]$  corresponding to an interval of inverted electrical resistivity values from one cell of the 2D section of subsoil, used for Wasserstein distances' computation.

$(D_w)$  are lower than those displayed in Figure 6 ( $D_g$ ) and more  $\theta_1 \cup \theta_2$  emerges. This approach ( $D_w$ ) is therefore more cautious.

We finally propose the use of a third approach, referred as  $D_{wg}$ , combining the first two previously described. This approach ( $D_{wg}$ ) is similar to the previous one ( $D_w$ ) in that it considers the distribution of a mass on the defined hypotheses, using Wasserstein distances. However, the allocated mass is here equal to  $m = 1/2$  instead of 1 as previously defined. The remaining mass ( $m = 1/2$ ) is allocated proportionally to the area under the Gaussian distribution, on the hypotheses associated with the implied resistivity intervals, as described above for the first approach,  $D_g$ . The results are displayed in Figure 9 and are intermediate to the results of the first two methods displayed in Figures 6 and 8.

#### C. Attribution of belief masses from geotechnical data

For the two geotechnical information sources, belief masses must be associated with the different hypotheses of the FoD for each cell of the vertical boreholes. To do so, we consider the geotechnical parameter values available at each depth (Figure 10-a) with their respective associated inaccuracies. Thus, we obtain intervals of values as for the attribution of belief masses from geophysical data. We generate a mesh for each geotechnical source (particle size distribution and CPT) consisting of as many cells in depth as the number of geotechnical measurements in each borehole (Figure 10-b). The cells are of same dimensions. At each borehole position a belief mass of a given value (see details below) is assigned, in the borehole points, to the hypothesis corresponding to the geotechnical parameter value. We then construct a new mesh (Figure 10-c), covering the full section of the subsoil, according to the depth of the boreholes. In order to characterize the entire section of the model, as does the ERT method, and to associate belief mass values to each newly generated cell, we impose an exponential lateral decay of the belief mass from the borehole point to the nearby one so that the decay rate is a function of the values proposed by the nearby borehole. Thus, we get for a specific depth:

$$M(x) = M(0) \cdot e^{-C_v x}, \quad (18)$$

with  $x$  being the horizontal distance from the considered cell to the reference borehole in meters ( $x = 0$  in the borehole),

$M(x)$  the belief mass values assigned to each hypothesis in the FoD for a position  $x$ , with  $M(0)$  the belief mass value assigned in the borehole.  $C_v$  corresponds to the coefficient of variation expressed in Eq. (19), such as used in Phoon and Kulhawy [32]:

$$C_v = \frac{1}{Q} \sqrt{\frac{1}{n_{\text{mesh}} - 1} \sum_{i=1}^{n_{\text{mesh}}} (Q - Q_i)^2}, \quad (19)$$

where  $Q$  is the geotechnical parameter value of the considered cell in the borehole and  $Q_i$  the geotechnical parameter values in the nearby borehole centered on the same depth. For Figure 10-b, and more broadly in this study, we considered  $n_{\text{mesh}} = 3$ , so that the computation of  $C_v$  takes into account 3 cells in the nearby borehole. Indeed, for two consecutive boreholes with similar values at fixed depth, we consider the soil to be less variable laterally and the decay of the confidence to be slower than for two consecutive boreholes displaying drastically different values. This decrease of belief mass is carried out to the left and to the right, from each borehole point. If the belief mass associated with a hypothesis  $X$  is less than 1 ( $m(X) < 1$ ), then the remainder of belief mass to be allocated to satisfy Eq. (5), is reallocated on the hypothesis "any type of geological material" symbolized by the union of all hypotheses, such as:

$$m(\theta_1 \cup \theta_2 \cup \theta_3) = 1 - m(X). \quad (20)$$

In each cell of the section, the information on the belief masses coming from the borehole on the left is merged with the information coming from the borehole on the right, using the DS rule, in respect with Eq. (12). Considering the 4 boreholes of this study, from 1 to 4 from left to right: borehole 1 cannot be compared to any borehole to its left neither can borehole 4 be compared to any borehole to its right. Indeed, for a given depth, we consider an equal  $C_v$  for left and right directions for boreholes located at the beginning and at the end of the section. In our case, the sections to the left of the first borehole and to the right of the last borehole are of little interest since they are not covered by the ERT investigation. The chosen hypotheses (geological material) at the borehole positions for each geotechnical method are displayed in Figure 11 and detailed below.

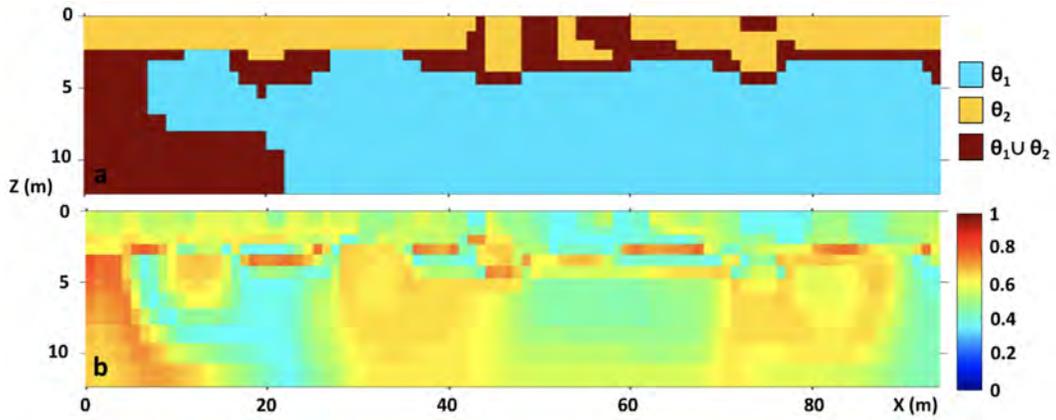


Fig. 8. a) Representation of the hypotheses having the highest belief mass according to the masses attribution based on Wasserstein distances applied to the inverted resistivity data (Figure 2), and b) the corresponding belief mass values.

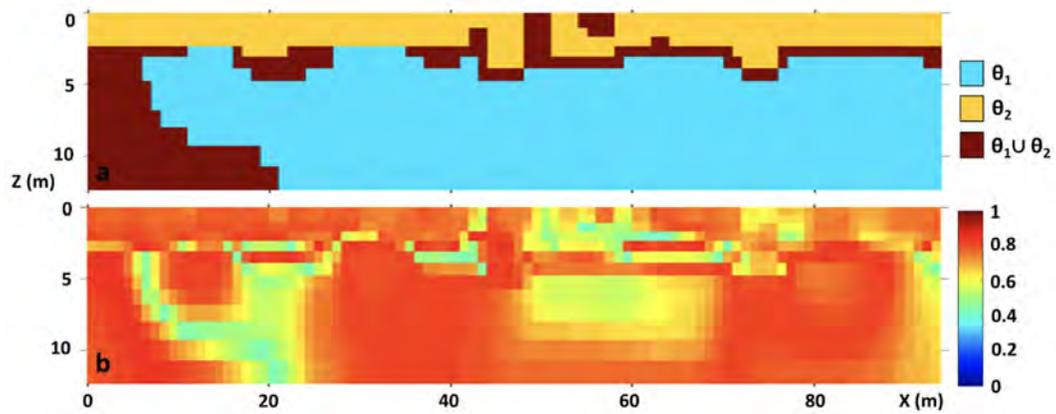


Fig. 9. a) Representation of the hypotheses having the highest belief mass according to the masses attribution based on Wasserstein distances and considering a Gaussian probability distribution applied to the inverted resistivity data (Figure 2) and b) the corresponding belief mass values.

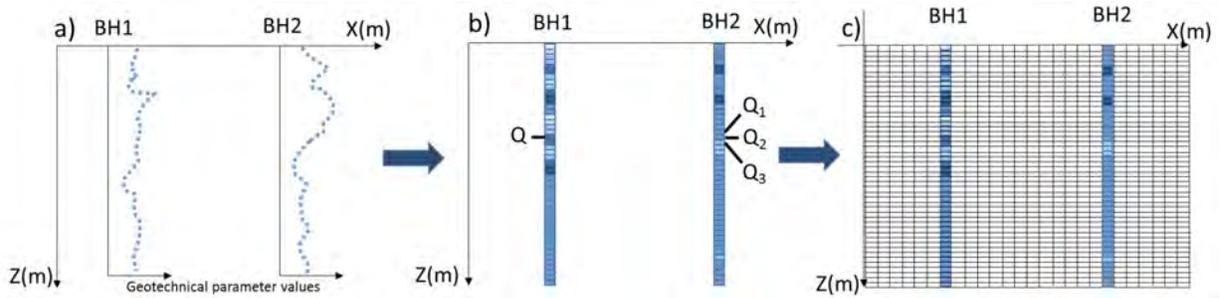


Fig. 10. Construction of a geotechnical discretization mesh from two vertical borehole acquisitions (BH1 and BH2). a) Representation of the geotechnical parameter values for BH1 and BH2 with depth. b) The boreholes are split in cells of same thickness associated with belief mass equal to a given value for the considered hypothesis. c) Construction of a full section mesh according to the depth of the boreholes.

*D. Attribution of belief masses from particle size distribution*

For the particle size distribution, materials described as *A* materials in Section II are considered as belonging to  $\theta_1$  and *B* materials as belonging to  $\theta_2$ . The inaccuracies taken into account correspond to 0.1% of the weighed value as indicated in the French standard NF P94-056 [23]. If the value of the cumulated sieve under 80  $\mu\text{m}$  cannot be characterized

as being greater or less than 35%, taking into account the inaccuracies, then the selected hypothesis is considered to be  $\theta_1 \cup \theta_2$ . This points out our inability to choose. Where materials have not been collected from the core for analysis (black areas in Figure 3), we consider that we have no information. Therefore, the belief mass is attributed to the union of all hypotheses:  $\theta_1 \cup \theta_2 \cup \theta_3$ , which represents the highest possible uncertainty (lowest knowledge level).

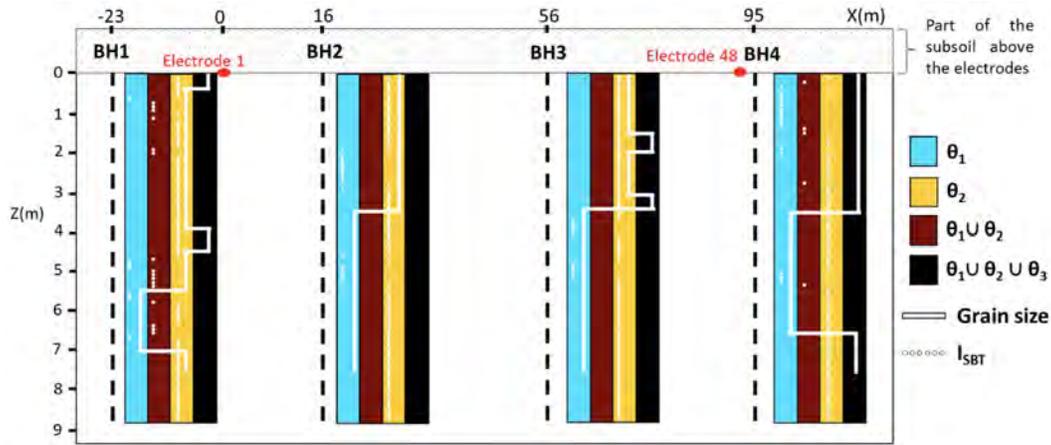


Fig. 11. 2D representation of the levee section displaying borehole positions in dashed lines and associated  $I_{SBT}$  (white dotted line) and particle size distribution (white solid line) corresponding classes.

In the boreholes, we consider a belief mass  $m(\cdot) = 0.99$  on the characterized hypothesis, at the depths for which the soil samples were analyzed and a belief mass of  $m(\cdot) = 0.01$  on the union of all other hypotheses, in accordance with Eq.(5). A mass of  $M(0) = 0.99$  was chosen because the established hypotheses are soil particle size distribution classes by definition. Thus, we consider that particle size information is the most appropriate kind of information that one can obtain. However, a mass of  $M(0) = 1$ , has not been set in order to avoid any total conflict that may arise in the fusion process. For depths at which the materials have not been analyzed, but still belonging to the same geological set (limits established by a geotechnical engineer and displayed in Figure 3) as the analyzed materials, we consider a vertical extension of the information.

A vertical decrease of the confidence level associated with the hypothesis  $\theta_i$  is carried out from the limit depth ( $p = 0$  m) of the collected sample, up to the limit between the two geological sets established visually by the geotechnical engineer (to  $p = 1$  m in the example, Figure 12). The distance between these two depths is  $d$ . The vertical decay of the belief mass on the considered hypothesis  $\theta_i$  is expressed as follows:

$$m(\theta_i; p) = 0.99(1 - e^{-(p-d)}). \quad (21)$$

So that  $m(\theta_i; p) = 0$  at the boundary between the two lithological sets characterized. The complementary belief mass is allocated to  $\theta_1 \cup \theta_2 \cup \theta_3$  in accordance with Eq. (5).

Figure 12 shows the results of the particle size distribution tests carried out on Borehole 2 (Figure 3). The colored areas correspond to the depths at which the collected samples have been analyzed while the arrows symbolize the vertical extension of the information where the materials have not been analyzed. On this example, the values of  $p$  and  $d$  apply to the vertical extension of the information from the lower bound of the analyzed sample, at a depth of 2.40 m, to the boundary between the two materials, established visually by the technician, at a depth of 3.40 m.

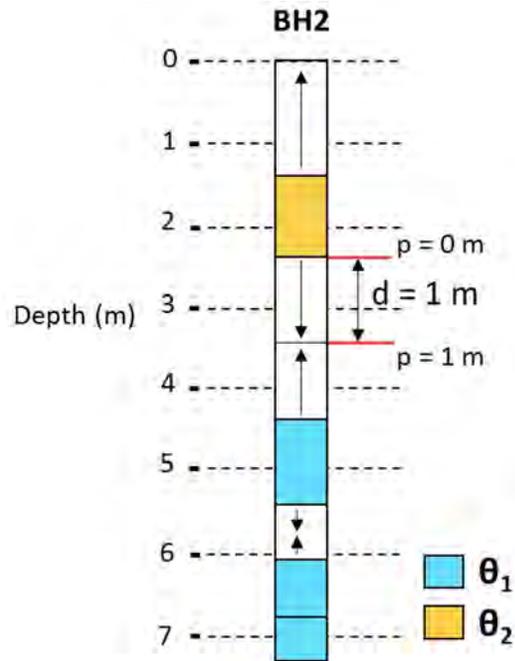


Fig. 12. Representation of the vertical extension of particle size distribution information. Example of results from particle size distribution carried out on extracted materials from BH2.

Then, the information is extended laterally, as detailed above, and the 2D section is cut so that it corresponds to the dimensions and coordinates of the ERT section (Figure 2). The results are displayed in Figure 13, with Figure 13-a highlighting the hypotheses having the highest belief mass for each cell section while Figure 13-b is displaying their associated belief mass values.

#### E. Attribution of belief masses from CPT

For the characterization of geological materials by the CPT method, we consider them to be fine-grained materials and

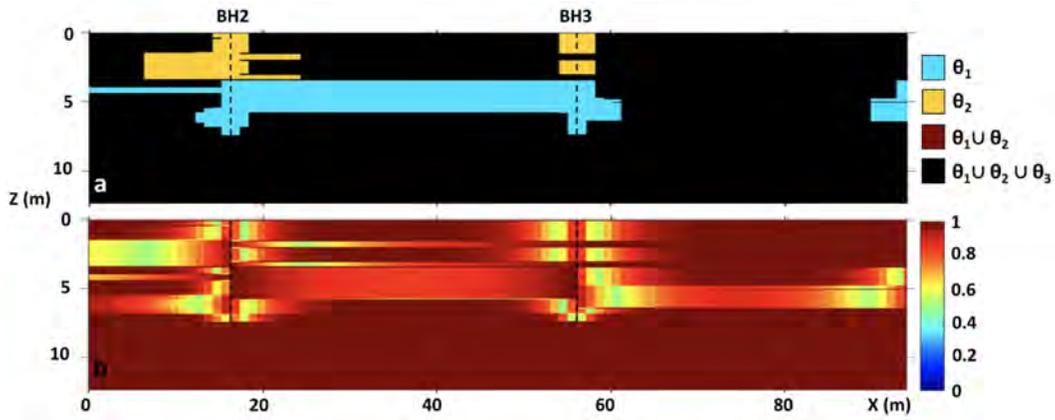


Fig. 13. a) Representation of the hypotheses having the highest belief mass according to the belief masses attribution from particle size distribution data considering  $M(0) = 0.99$  in the borehole points and b) the associated belief mass values.

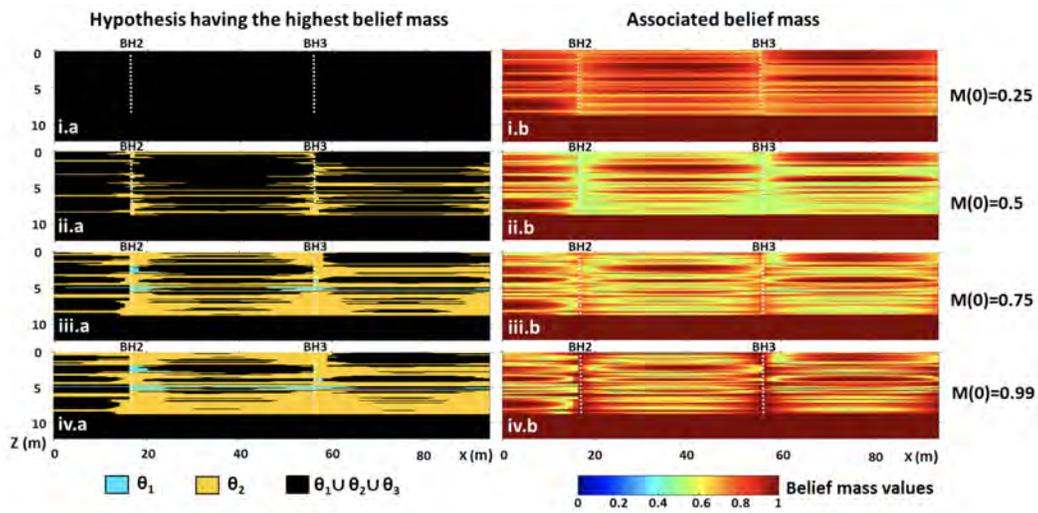


Fig. 14. Considering i)  $M(0) = 0.25$ , ii)  $M(0) = 0.5$ , iii)  $M(0) = 0.75$  and iv)  $M(0) = 0.99$ , a) representation of the hypotheses having the highest belief mass according to the belief masses attribution from CPT data in the borehole points and b) the associated belief mass values.

belonging to  $\theta_1$  when  $I_{SBT} > 2.6$ . When  $I_{SBT} < 2.6$ , the materials are deemed to be coarser-grained and therefore belong to  $\theta_2$ . As recommended in the NF P94-113 French standard [25], we consider a maximum inaccuracy on the computation of  $I_{SBT}$ , the maximum tolerated inaccuracy being the smallest of the following values:

- 5% of the measured value (for  $q_c$  and  $f_s$ ),
- 1% of the maximum value of the measuring range (for  $q_c$  and  $f_s$ ).

If the value of  $I_{SBT}$  cannot be characterized as greater or less than 2.6, then the selected hypothesis is considered to be the hypothesis  $\theta_1 \cup \theta_2$ , highlighting our disability to select a geological material. The attribution of a belief mass of  $M(0) = 0.99$  at the borehole positions, used for the particle size distribution, is questionable for the use of the  $I_{SBT}$  index. Indeed, the characterization of geological sets in terms of particle size distribution is less reliable by the use of such an index. The reliability is decreased by the fact that the value

of  $I_{SBT}$  is obtained following a computation involving the two recorded parameters and also by the fact that no sample is extracted. Therefore, we propose a brief parametric study with the results of the attribution of the belief masses for the CPT method, using a value of  $M(0) = 0.25$  (Figure 14-i),  $M(0) = 0.5$  (Figure 14-ii),  $M(0) = 0.75$  (Figure 14-iii) and  $M(0) = 0.99$  (Figure 14-iv) on the hypothesis concerned at the borehole positions. As in the case of the particle size distribution, after horizontal extension of the information and cutting the 2D section in accordance with the dimensions and coordinates of the ERT section (Figure 2), we obtain the results shown in Figure 14.

#### F. Dimensioning of the mesh prior to the fusion

Each investigation method has its specific mesh. In order to merge the belief masses from the geophysical information source (ERT) and the two geotechnical sources (particle size distribution and CPT), it is necessary to have a common mesh containing the belief masses from the three sources

for each cell. We chose to consider a superimposition of the three discretization grids. In order to avoid alteration of the quality of the information, no interpolation nor extrapolation is carried out. We thus obtain an irregular mesh but without any approximation of the cells and associated values (Figure 15). The bounds of the  $x$ -axis (right/left) and  $z$ -axis are imposed by the ERT section even though the data of boreholes 1 and 4 (out of the electrode line) were taken into account in the attribution of belief masses from geotechnical data (as displayed in section III-C).

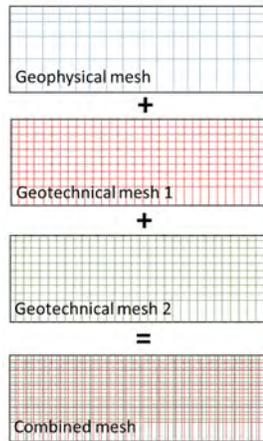


Fig. 15. Example of a geophysical mesh (in blue) and two geotechnical meshes (in red and in green) superimposed to provide a new irregular mesh used for the fusion computation and the fusion result representation.

#### IV. DATA FUSION RESULTS

The results issued from the merging of the data of the three considered information sources are displayed in Figures 16 and 17. First, let us compare the results obtained using the two different rules of combination. Unlike the PCR6 method for which a normalization process was carried out (Figures 16-c and 17-c), the Smets rule of combination makes it possible to highlight the conflict zones (Figures 16-a and 17-a). Thus, it appears that the conflict is greater close to the boreholes, from the interface between  $\theta_1$  and  $\theta_2$ , that seems present at 3.40 m depth, down to the maximum depth of geotechnical investigation for all  $M(0)$  values and for both belief masses attribution approach ( $D_w$  and  $D_{wg}$ ). This is due to the fact that the  $I_{SBT}$  index essentially considers a coarse-grained material in the levee (Figure 14) where the particle size distribution (Figure 13) and the ERT (Figures 8 and 9) consider fine-grained materials (from about 3.40 m depth). The conflict level decreases when deviating from the borehole positions since the confidence level on the geotechnical information decreases with the distance to these geotechnical testing points. Comparatively, the influence of the geophysical information is gradually becoming more important. The geotechnical information made it possible to characterize the  $\theta_1/\theta_2$  interface at 3.40 m depth, unlike the 2.5 m interface proposed by the ERT (Figures 8 and 9). This interface is also defined more precisely. The geotechnical information also allowed

the discrimination between  $\theta_1$  and  $\theta_2$  materials where the geophysics brought out the doubt between  $\theta_1$  and  $\theta_2$  (i.e.  $\theta_1 \cup \theta_2$ ). Some doubts about the nature of the material, however, remain present between  $x = 0$  m to  $x = 7$  m. The areas with the highest associated confidence level are those close to the borehole points and between boreholes BH2 and BH3 between  $z = 3.5$  m and  $z = 5.5$  m. This is explained by the concordance between geophysical information (Figures 8 and 9) and particle size information (Figure 13). Indeed, Figure 13 indicates a significant belief mass on  $\theta_1$  because the particle size distribution values at such depths are quite similar between boreholes BH2 and BH3.

Now let us compare the results obtained with two different approaches of belief masses attribution for the geophysical source of information,  $D_w$  and  $D_{wg}$  (Figure 16). The belief masses associated with the selected hypotheses, using  $D_{wg}$  approach, are more significant since this approach grants more mass initially (Figures 16-ii.b and ii.d). The horizontal interface between  $\theta_1$  and  $\theta_2$  is well characterized and the doubt lies on the left part of the subsoil section. Using Smets rule of combination, there is no significant difference concerning the distribution of the conflict for  $D_w$  and  $D_{wg}$  approaches (Figures 16-i.a and ii.a). In comparison with  $D_{wg}$  approach, using  $D_w$  approach, a noticeable difference lies in the appearance of sets of coarser-grained materials for PCR6. This material is identified close to the borehole points between  $z = 6$  m and  $z = 9$  m (Figure 16-i.c). This is due in particular to the results of the CPT (Figure 14) and to the lower confidence level brought by the geophysical information at these locations (Figure 8-b). When geophysical belief masses increase using the  $D_{wg}$  belief masses attribution approach,  $\theta_1$  material becomes dominant again and erases the presence of  $\theta_2$  material. In contrast, Figure 16-ii.d displays that these areas remain areas of lesser confidence.

Keeping the  $D_{wg}$  approach of geophysical belief masses attribution, let us compare the variation of  $M(0)$  values for CPT characterization. It appears that the conflict level increases along with the value of  $M(0)$ . Indeed, for a low value of  $M(0) = 0.25$  (Figure 17-i.a), there is almost no conflict while conflict covers nearly a third of the subsoil section for a high value of  $M(0) = 0.99$  (Figure 17-iii.a). Using PCR6 rule of combination, the presence of  $\theta_2$  increases along with  $M(0)$  value (Figures 17-i.c, 16-ii.c, 17-ii.c, 17-iii.c) below 5 m depth and the associated levels of confidence decrease (Figures 17-i.d, 16-ii.d, 17-ii.d, 17-iii.d). This can be explained by the fact that the CPT characterization goes against the ERT and the particle size distribution characterizations. Thus, putting more confidence in the CPT characterization by increasing  $M(0)$  value can only increase the conflict level and potentially change the proposed characterization after normalization of the conflictual masses (PCR6 rule). However, these areas characterized as coarse-grained materials between 6 and 9 m depth close to boreholes BH2 and BH3 (Figures 17-ii.c and iii.c) are associated with low belief mass values (Figures 17-ii.d and iii.d).

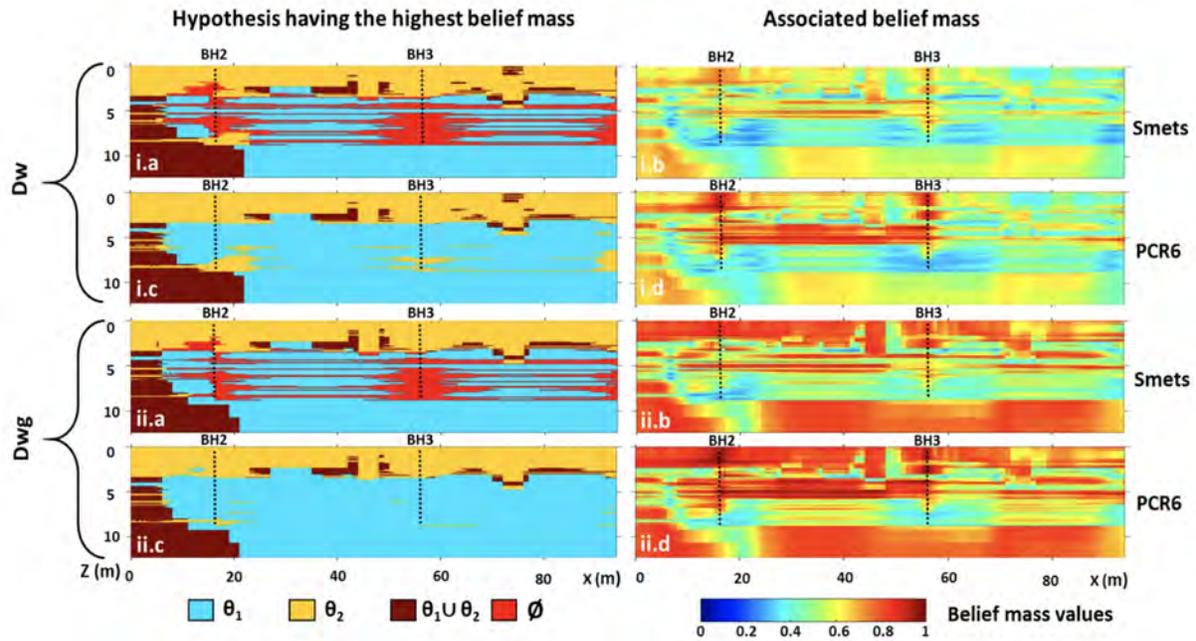


Fig. 16. Hypotheses having the highest belief mass in a) and c) and their associated mass values in b) and d). i) Belief mass attribution from electrical resistivity data considering only the computation of Wasserstein distance and ii) belief mass attribution from electrical resistivity data considering the computation of Wasserstein distance and also a Gaussian probability distribution centered on the inverted resistivity value. (a,b) figures are results of Smets rule of combination, (c,d) figures are results of PCR6. The boreholes positions are in dashed lines with  $M(0) = 0.5$ .

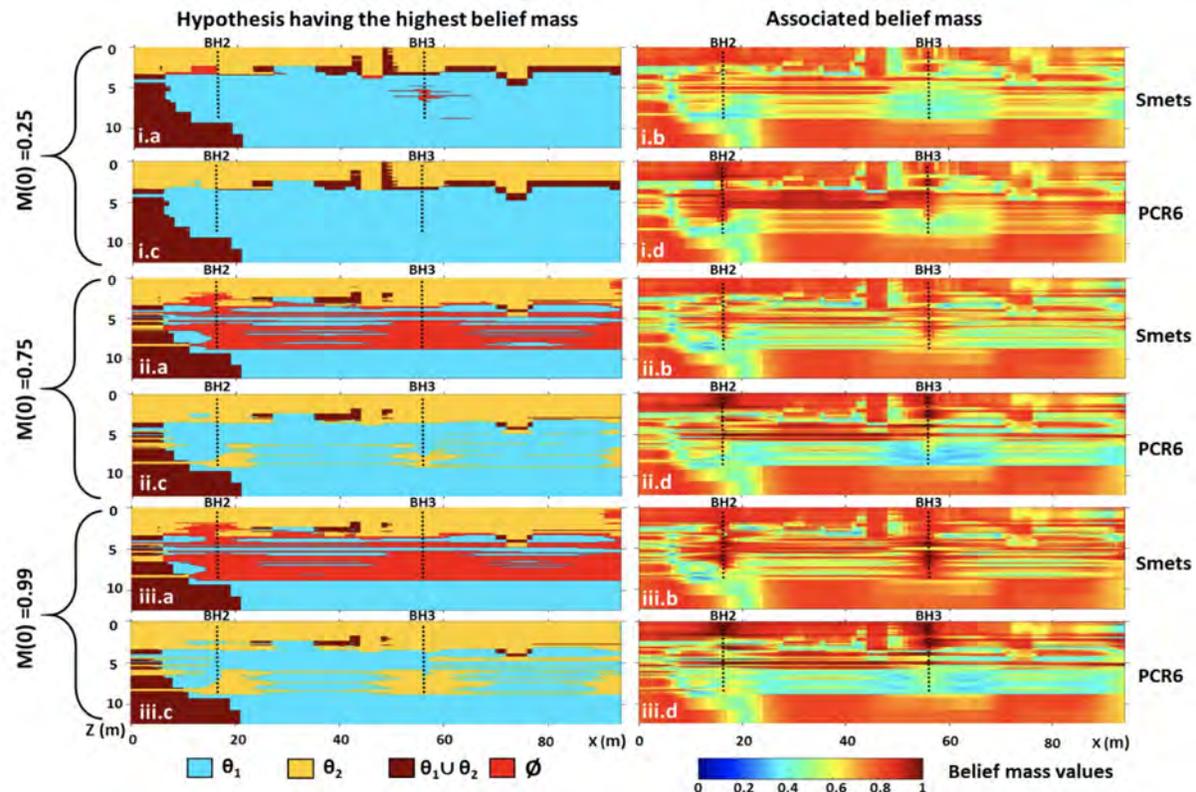


Fig. 17. Hypotheses having the highest belief mass in a) and c) and their associated mass values in b) and d), considering i)  $M(0) = 0.25$ , ii)  $M(0) = 0.75$  and iii)  $M(0) = 0.99$ . We consider a belief mass attribution from electrical resistivity data considering the computation of Wasserstein distance and also a Gaussian probability distribution centered on the inverted resistivity value. (a,b) figures are results of Smets rule of combination, (c,d) figures are results of PCR6. The boreholes positions are in dashed lines.

## V. DISCUSSION

The results of our fusion methodology suggest the presence of two material layers. First, a coarse-grained material layer about 3.5 m thick, and then, a fine-grained material layer below. It is possible to associate the first layer to the embankment (levee) and the second layer to alluvium. These results are in agreement with the internal representation of the levee that was proposed before the investigation campaign (Figure 1-a) [18]. The drawback of conducting an investigation campaign on a real levee is that there is no “true” model to which we can compare our results.

Most of the conflict present in the results is related to the characterization made by the  $I_{SBT}$  index from CPT data. This characterization essentially considers a coarse-grained material in almost the whole levee section, contrary to the ERT and the particle size analysis that dissociate quite clearly two sets of materials. It appears that the characterization of materials by the  $I_{SBT}$  index is not ideal in this case. This is why we consider that a mass of  $M(0) = 0.5$  should be retained in the CPT boreholes, contrary to a mass of  $M(0) = 0.99$  for granulometric analyzes. In the future, it could be relevant to find a more appropriate index for particle size characterization of materials from the parameters measured with the CPT.

We proposed a brief parametric study bringing to light the results associated with four belief mass attribution values in the sampling boreholes ( $M(0)$ ). This parametric study highlights the fact that increasing the value of  $M(0)$  extends the lateral extension of the selected hypothesis. It seems that a large value of  $M(0)$  cannot be equally attributed to both the values of  $I_{SBT}$  and particle size distribution results when these two sources of information seem contradictory in many locations (Figure 11). We believe that for a real investigation campaign, the value associated with each geotechnical method and at each borehole point could be adjusted by a well-informed geotechnical engineer, relying on an elicitation process [33] and based on the ability of the method to provide information on the nature of the investigated material.

In this study, the water table height and its time variations were ignored. However, we believe that this choice has no significant impact on the proposed results. First, water level variations are very low on the dates of the investigation campaign [18], so there should be no problem in considering the data as if they had been acquired at the same time. Secondly, the water table height is more than 8 m below the surface, where fine-grained materials are deemed to be present (low resistivities). The hydraulic conductivity of these soils (alluvium) is probably very low and their hydric state is potentially quite insensitive to seasonal changes of the water table.

However, for other cases, it may be relevant to take into account the water table level and its variations (especially for a long-term monitoring of the levee and for winter investigation campaigns). To do this, it would be interesting to consider more hypotheses within the FoD  $\Theta$ . For example, the hypothesis  $\theta_1$  could be associated to a fine material in the dry state

and the hypothesis  $\theta_2$  to this same material in a saturated state. The number of hypotheses constitutive of  $\Theta$  is not limited and does not require any modification of our methodology. However, the computational cost could vary significantly.

Recently, many equipment to use ERT for long-term monitoring have been developed [34], [35]). Thus, it would be interesting to combine such use of ERT with this fusion methodology. As soon as the fusion parameters are fixed, the use of the methodology could be automated in order to propose a daily update of a levee section for a long-term monitoring system. It would then be a matter of integrating data from time-domain reflectometry moisture sensor as well as piezometric surveys, making it possible to monitor the fluctuation of the water table as a function of rainfalls and periods of irrigation. It is possible to draw inspiration from works such as [36] which focus on the effects of environmental perturbations (variations in water table heights, temperatures and rainfalls) and propose calibration curves from ERT data for a specific investigation site.

For this study, we considered data from three investigation methods, usually used for levee characterization [37], [38], [39], each one having its own belief mass attribution method. A limitation of our methodology is our ability to define the values of the upper and lower bounds of the physical parameters characterizing the constitutive hypotheses of the FoD for each investigation method. We currently rely on bibliographic references [26], standards [24] or visual distribution of the data (Figure 5). Our ability to define these interval values is a function of the type of investigation method. If these intervals are poorly defined, the fusion results may not be able to display any coherent solution. Automating the procedure of characterization of the intervals of the FoD would be a great contribution to this methodology (e.g., by means of data mining [40]).

## VI. CONCLUSION

In this work, we present a novel fusion methodology based on the use of belief functions to optimize the combination of data from three different sources of information composed of a geophysical method (ERT) and two geotechnical methods (CPT and particle size distribution) applied to levee characterization. These sources of information are considered complementary, each one having its own spatial distribution and associated level of uncertainties and inaccuracies. This methodology was tested on real datasets acquired on an earthen fluvial levee located in St-Clément-des-Levéés (France) and has proved to be efficient. We compared the results obtained with different geophysical belief mass attribution approaches (Wasserstein distance computation and Gaussian probability distribution centered on the inverted resistivity value), as well as for different belief mass values in the boreholes for CPT investigations and different combination rules (Smets and PCR6).

A representation of the merged information associated with degrees of belief was proposed. We advocate that this representation is more relevant and informative than a simple

superposition of the different physical parameters recorded. The ability of our methodology to distinguish two geological sets of different nature in terms of particle size distribution (fine-grained material and coarser-grained material) as well as its ability to accurately characterize a horizontal interface (at about 3.40 m depth below the electrode line) was demonstrated. A coarser-grained geological material was identified at the top of the structure and a fine-grained material is present underneath. The characterization seems more reliable between boreholes 2 and 3 and the results highlight a doubt about the nature of the material at the left part of the section. Two zones of lesser confidence are also located in the lower part of the section (below 7.5 m depth) near the geotechnical points intersecting the ERT section.

Finally, in the proposed results (Figures 16 and 17), the areas of lesser confidence level indicate where the investigation could be strengthened. Moreover, the conflict zones inform us where at least two sources of information disagree. These two types of outcome are believed to be precious for a real investigation campaign. The level of confidence would also be important for decision support (e.g., models of failure hazards).

This methodology would be perfectly transferable to other applications (landslide, pedology, pollutant tracking, mining...) when at least two sources of information (geotechnical and geophysical) are involved [41].

#### ACKNOWLEDGMENTS

We thank all the technicians at the CEREMA French institute involved in the investigation campaign and laboratory tests for their precious help throughout the experimental phase. We also thank the *Région Pays de la Loire* for their financial support.

#### REFERENCES

- [1] G.J. Brierley, R.J. Ferguson, K.J. Woolfe, *What is a fluvial levee?*, Sedimentary Geology, Vol. 114(1-4), pp. 1–9, 1997.
- [2] C. Fauchard, P. Mériaux, *Geophysical and geotechnical methods for diagnosing flood protection dikes: Guide for implementation and interpretation*, Quae, 2007L.
- [3] T. Dezert, Y. Fargier, S. Palma Lopes, P. Côte, *Geophysical and geotechnical methods for fluvial levee investigation: A review*, Engineering Geology, Vol. 260, pp. 105–206, 2019.
- [4] F. Shaaban, A. Ismail, U. Massoud, H. Mesbah, A. Lethy, A.M. Abbas, *Geotechnical assessment of ground conditions around a tilted building in Cairo-Egypt using geophysical approaches*, Journal of the Association of Arab Universities for Basic and Applied Sciences, Vol. 13(1), pp. 63–72, 2013.
- [5] M. Foster, R. Fell, M. Spannagle, *The statistics of embankment dam failures and accidents*, Canadian Geotechnical Journal, Vol. 37(5), pp. 1000–1024, 2000.
- [6] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The Annals of Mathematical Statistics, Springer, Berlin, pp. 325–339, 1967.
- [7] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [8] D.F. Specht, *A general regression neural network*, IEEE Trans on Neural Networks, Vol. 2(6), pp. 568–576, 1991.
- [9] L.A. Zadeh, *Fuzzy sets as a basis for a theory of possibility*, Fuzzy Sets and Systems, Vol. 1(1), pp. 3–28, 1978.
- [10] E. Binaghi, L. Luzi, P. Madella, F. Pergalani, A. Rampini, *Slope instability zonation: a comparison between certainty factor and fuzzy Dempster–Shafer approaches*, Natural hazards, Vol. 17(1), pp. 77–97, 1998.
- [11] O.F. Althuwaynee, B. Pradhan, S. Lee, *Application of an evidential belief function model in landslide susceptibility mapping*, Computers & Geosciences, Vol. 44, pp. 120–135, 2012.
- [12] K.A. Mogaji, H.S. Lim, K. Abdullah, *Regional prediction of groundwater potential mapping in a multifaceted geology terrain using GIS-based Dempster–Shafer model*, Arabian Journal of Geosciences, Vol. 8(5), pp. 3235–3258, 2015.
- [13] M.S. Tehrani, L. Kumar, *The application of a Dempster–Shafer-based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods*, Environmental Earth Sciences, Vol. 77(13), p. 490, 2018.
- [14] T. Dezert, Y. Fargier, S. Palma Lopes, P. Côte, *Application of Belief Functions to Levee Assessment*, in Proc. of Int. Conf. on Belief Functions, Springer, Cham, 2018.
- [15] T. Dezert, S. Palma Lopes, Y. Fargier, P. Côte, *Geophysical and Geotechnical Data Fusion for Levee Assessment - Interface Detection with Biased Geophysical Data*, in 24th European Meeting of Environmental and Engineering Geophysics, 2018.
- [16] T. Dezert, S. Palma Lopes, Y. Fargier, Q. Forquenot de la Fortelle, P. Côte, R. Tourment, *Merging geophysical and geotechnical data acquired on a test bed for levee diagnosis application*, in Dignes 2019 - 3ème Colloque National sur les digues maritimes et fluviales de protection contre les inondations, 2019.
- [17] C. Jodry, S. Palma Lopes, Y. Fargier, P. Côte, M. Sanchez, *A cost-effective 3D electrical resistivity imaging approach applied to dike investigation*, Near Surface Geophysics, Vol. 15(1), pp. 27–41, 2017.
- [18] C. Jodry, S. Palma Lopes, Y. Fargier, M. Sanchez, P. Côte, *2D-ERT monitoring of soil moisture seasonal behaviour in a river levee: A case study*, Journal of Applied Geophysics, Vol. 167, pp. 140–151, 2019.
- [19] S. Hervé, R. Bénot R (2013) Digue de la Loire à St Clément des Levées (49) : Suivi par méthodes géophysiques, Compte-rendu des mesures 2012, IFSTTAR – Opération 11R103 DOFEAS, 2013.
- [20] M.H. Loke, *Tutorial: 2-D and 3-D electrical imaging surveys*, 2013. <http://www.geotomosoft.com>
- [21] D. Franois, P. Mériaux, J. Monnet, *Méthodologie de reconnaissance et de diagnostic de l'érosion interne des ouvrages hydrauliques en remblai*, Presses des Ponts: Publications IREX, Paris. OCLC: 966418274, 2016.
- [22] M.H. Loke, I. Acworth, T. Dahlin, *A comparison of smooth and blocky inversion methods in 2D electrical imaging surveys*, Exploration Geophysics, Vol. 34(3), pp. 182–187, 2003.
- [23] AFNOR NF P94-056, *Sols : reconnaissance et essais - Analyse granulométrique - Méthode par tamisage à sec après lavage*, 1995.
- [24] AFNOR NF P11-300, *Exécution des terrassements - Classification des matériaux utilisables dans la construction des remblais et des couches de forme d'infrastructures routières*, 1992.
- [25] AFNOR NF P94-113, *Sols : reconnaissance et essais - Essai de pénétration statique*, 1996.
- [26] P.K. Robertson, *Soil Behavior Type: An Update*, in Proc. of 2nd Int. Symposium on Cone Penetrating Testing, Huntington Beach, CA, USA, 2010.
- [27] P. Smets, *Imperfect information: Imprecision and uncertainty*, in Uncertainty Management in Information Systems, Springer, pp. 225–254, 1997.
- [28] A. Martin, C. Osswald, J. Dezert, F. Smarandache, *General combination rules for qualitative and quantitative beliefs*, Journal of Advances in Information Fusion, Vol. 3(2), pp. 67–89, Dec 2008.
- [29] P. Smets, *The combination of evidence in the transferable belief model*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12(5), pp. 447–458, 1990.
- [30] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion (Collected works)*, Vol. 3, American Research Press, Gallup, NM, USA, 2009.
- [31] L. Tran, L. Duckstein, *Comparison of fuzzy numbers using a fuzzy distance measure*, Fuzzy Sets and Systems, Vol. 130(3), pp. 331–341, 2002.
- [32] K.K. Phoon, F.H. Kulhawy, *Characterization of geotechnical variability*, Canadian Geotechnical Journal, Vol. 36(4), pp. 612–624, 1999.
- [33] M. Vuillet, *Élaboration d'un modèle d'aide à la décision probabiliste pour l'évaluation de la performance des digues fluviales*, P.D. Thesis, Univ. Paris-Est, 2012.

- [34] D. Arosio, S. Munda, G. Tresoldi, M. Papini, L. Longoni, L. Zanzi, *A customized resistivity system for monitoring saturation and seepage in earthen levees: installation and validation*, Open Geosciences, Vol. 9, pp. 457–467, 2017.
- [35] A. Hojat, D. Arosio, L. Longoni, M. Papini, G. Tresoldi, L. Zanzi, *Installation and validation of a customized resistivity system for permanent monitoring of a river embankment*, EAGE-GSM 2nd Asia Pacific Meeting on Near Surface Geoscience & Engineering, April 22–26, Kuala Lumpur, 2019.
- [36] G. Tresoldi, D. Arosio, A. Hojat, L. Longoni, M. Papini, L. Zanzi, *Long-term hydrogeophysical monitoring of the internal conditions of river levees*, Engineering Geology, Vol. 259, pp. 105–139, 2019.
- [37] D. Cazanaceli, N.D. Smith, *A study of morphology and texture of natural levees - Cumberland Marshes, Saskatchewan, Canada*, Geomorphology, Vol. 25(1-2), pp. 43–55, 1998.
- [38] J.M. Lorenzo, J. Hicks, E.E. Vera, *Integrated seismic and cone penetration test observations at a distressed earthen levee: Marrero, Louisiana, USA*, Engineering Geology, Vol. 168, pp. 59–68, 2014.
- [39] P. Royet, S. Palma Lopes, C. Fauchard, P. Mériaux, L. Auriou, *Rapid and Cost-Effective Dike Condition Assessment Methods: Geophysics and Remote Sensing*, FloodProBE Project, 2013.
- [40] K. Sabor, D. Jougnot, R. Guerin, L. Apffel, B. Steck, *Integrated Analysis of Geophysical Data Using a Data Mining Approach*, in Proc. of 25th European Meeting of Environmental in Engineering Geophysics, 2019.
- [41] B. Pasierb, M. Grodecki, R. Gwóźdz, *Geophysical and geotechnical approach to a landslide stability assessment: a case study*, Acta Geophysica, pp. 1–12, 2019.

# Canal Dike Characterization by Means of Electrical Resistivity, Shear Wave Velocity and Particle Size Data Fusion

Théo Dezert<sup>a,b</sup>, Yannick Fargier<sup>b</sup>, Sérgio Palma Lopes<sup>a</sup>, Vincent Guihard<sup>c</sup>

<sup>a</sup>Gustave Eiffel University, GERS, GeoEND, F-44344 Bouguenais, France.

<sup>b</sup>Gustave Eiffel University, GERS, RRO, F-69675 Bron, France.

<sup>c</sup>EDF R&D, PRISME, Chatou Cedex 78401, France.

Emails: theo.dezert@ntnu.no, yannick.fargier@univ-eiffel.fr, sergio.palma-lobes@univ-eiffel.fr, vincent.guihard@edf.fr

Originally published as: T. Dezert, Y. Fargier, S. Palma Lopes, V. Guihard, *Canal Dike Characterization by Means of Electrical Resistivity, Shear Wave Velocity and Particle Size Data Fusion*, Journal of Applied Geophysics, Vol. 204, September 2022, and reprinted with permission.

**Abstract**—A reliable lithological characterization of earthen dikes constitutes an important asset by virtue of enhancing a good diagnosis, which is of immense value in preventing dike breakage. Ruptures of hydraulic works can lead to disastrous consequences (loss of life, severe environmental and economic impacts). Recognized methodologies for characterizing earthen dikes include complementary geophysical and geotechnical investigation methods. This article explores a fusion methodology to combine data from these two types of information sources in considering actual datasets from a canal dike investigation campaign. This campaign involves electrical resistivity tomography as well as a multi-channel analysis of surface waves and a particle size analysis derived from laboratory testing. Our fusion methodology is based on the use of belief masses to enhance the characterization of lithological sets within earthen structures. While taking into consideration the particularities of each method (spatial distribution, data imperfection), this approach provides information on the conflict level between information sources and moreover displays a confidence index associated with the results. This work contributes several improvements to the fusion methodology (including the fusion of two distinct geophysical datasets and the implementation of K-means clustering algorithms) in addition to new application possibilities (larger area of investigation, more complex structure and lithological variability). It also offers fusion results and dike characterization whether considering zero, four or seven boreholes. Fusion results highlight the ability of this enhanced methodology to identify the position of lithological sets (fine and coarse fill materials with limestone breccia, marls and limestones) as well as specify the interface positions and associated levels of confidence, ensuring consistency with available knowledge on the geological setting and presence of a fault. These results also display good consistency between the geoelectrical and seismic characterizations for this specific investigation site despite the inability to characterize each material individually.

**Keywords:** Canal dike, Belief functions, Data fusion, Electrical Resistivity Tomography, Multi-channel analysis of surface waves, Particle size analysis.

## I. INTRODUCTION

Hydraulic works such as river and canal dikes are built to maintain a given flow of water. To prevent eventual breakage of these works, which could lead to catastrophic events (i.e. casualties, property damage, economic impacts), an effective

characterization of these complex structures is required. For this purpose, investigation campaigns involving geophysical and geotechnical methods [1] are typically employed for dike characterization and weak zone identification. The objective of these campaigns is to generate a hazard assessment.

Whereas geophysical methods are non-intrusive and provide information on a large volume of subsoil, geotechnical methods are intrusive and yield detailed spatial information. In addition, the quality of the information derived from the two methods differs. The uncertainties associated with geophysical information are significant, especially owing to the integrative and indirect aspects of the method as well as to the resolution step for inverse problems [2]. On the other hand, the information acquired by geotechnical methods is much more reliable due to a direct contact with the material. It is therefore implied that these two types of techniques are mutually beneficial. Unfortunately, too few of the methodologies available actually consider a mathematical combination of the acquired data, opting instead for a simple graphic superposition of the results [3], [4], [5].

In order to facilitate the diagnosis of hydraulic embankments, it is important to identify the lithological materials present within the structure and distinguish them. The interfaces as well as the presence of any anomalies must also be located. This information is in fact likely to provide indications on the subsequent development of internal erosion zones or areas of physical instability [6]. Such a characterization, associated with confidence indices, would be useful in producing failure hazard models.

An information fusion methodology based on the use of belief functions [7], [8] was developed and proposed to merge data from geophysical and geotechnical information sources in [9]. Furthermore, bibliographic research on earthen dike properties and failure modes are available in [10], and parametric studies on several parameters involved in the fusion methodology are available in [11], [12], [13]. Conclusive studies were also carried out for numerical models as well as on an experimental test bench in [14] and then on an actual fluvial earthen dike in [15]. This methodology is unique by

virtue of taking into account the various types of imperfections associated with information (uncertainty, imprecision, incompleteness), as well as the respective spatial expressions of information and representations of the inconsistency between information sources (investigation methods). It could be applied well beyond dike characterization (e.g. liquefaction risk, landslide, pedology, pollutant tracking, mining) when at least two sources are involved [16]. In the field of geosciences, some research works have applied the use of belief functions to propose results for: slope instability [17], [18], groundwater [19], and flood susceptibility mapping [20]. Also, many works on joint inversion [21], [22], [23], [24] have proposed utilizing large amounts of geophysical information through joint inversion in order to avoid some of the ambiguity inherent in the methods when applied individually. However, these works differ from our approach, which considers the information sources to be mutually independent.

We are proposing herein the application of this belief function-based methodology to a new type of structure, namely a canal dike owned by the French EDF electric utility company. This structure includes the presence of two distinct substratum materials, two fill materials and a fault. In comparison with earlier research, this study features numerous advances. First, the research area is substantially larger than previously: around 1,800 meters long by 24 meters high, as compared to 100 meters long by 12 meters high in the Dezert et al. study [15]. The structure under investigation is also more complicated, offering more lithological variation. In addition to the ERT (Electrical Resistivity Tomography) approach, the MASW (Multi-channel Analysis of Surface Waves) method is employed as a second source of geophysical information. Also taken into account are the sensitivities associated with resistivity values, thus indicating the extent to which a change in resistivity will influence the potential measured by the array. Moreover, the possibility of associating physical parameter values with lithological materials has been integrated through use of the K-means clustering method. The purpose here is to automate a procedure that had previously been conducted by means of simple expert opinion.

This article is organized as follows. First, the studied canal dike and geology will be introduced along with the three investigation methods considered (ERT, MASW, core drillings). Then, the fusion methodology will be described by use of the two combination rules (Smets and Proportional Conflict Redistribution Rule No. 6), with an analysis of the data acquired for each method. This section will also present clustering and the belief masses computation. In the next section, the belief mass results for each individual information source as well as the overall fusion results will be displayed. The fusion process is operated by initially considering only the geophysical information sources and then adding the geotechnical information source in two situations: one including the information from four boreholes, and the other composed of seven boreholes. Lastly, the results of this work will be discussed in terms of their advantages, limitations and perspectives.

## II. INVESTIGATED DIKE AND INVESTIGATION METHODS

### A. Hydraulic embankment and geological context

The studied hydraulic embankment is a canal dike owned by the EDF electric utility company and located in the south of France. Since for reasons of confidentiality it is not authorized to disclose the precise geographic location of this canal dike, the Kilometric Point (KP) notation is being used to identify the geophysical and geotechnical investigation positions. These KPs (denoted in kilometers) correspond to the dike length, along the crest moving from the upstream part to the downstream part. The stretch selected for our study is located on the right bank and extends from KP 10.35 to KP 12.13. Five geological formations have been identified throughout the whole structure. For the present study, this particular section has the benefit of intersecting two distinct geological formations, with the presence of a fault oriented NE-SW that lowers the western compartment. Up to KP 10.80 approximately, the canal is essentially rock-based on more or less marly limestone terrain from the Lower Cretaceous (shown in yellow, Figure 1). Beyond that point, the substratum is generally formed of more or less clayey and indurated marls from the Oligocene (purple, Figure 1). Between KP 11.50 and KP 12.13, the presence of more cohesive materials, most likely corresponding to Cretaceous limestones, is suspected. The respective positions of the three distinct investigation methods implemented (two geophysical and one geotechnical) are displayed in Figure 1.

### B. Electrical Resistivity Tomography

The basic principle behind DC-resistivity methods consists of transmitting direct electrical current (DC) of known intensity [A] by means of two “current” electrodes and then measuring the voltage drop [V] between two “potential” electrodes. Depending on parameters such as electrode layout and acquisition array, apparent resistivity values can be computed. Such a measurement yields indirect information though due to the integrative aspect of this geophysical method. Some available forward modeling software can integrate the electrode layout (inter-electrode spacing) and topography in order to simulate a measurement (e.g. COMSOL Multiphysics, BERT, CESAR-LCPC). Maressot [25] recalled that apparent resistivity is the measured transfer resistance divided by the simulated transfer resistance of a model (with topography), at a resistivity of  $1 \Omega\cdot\text{m}$ . A two-dimensional (2D) ERT, like the one considered in this work, consists of an alignment of electrodes and an inversion of a large number of measurements based on a four-electrode configuration (two current electrodes for electrical current injection and two potential electrodes to measure an electric potential drop).

A geoelectric campaign was performed in 2014 on the dike crest. The array consisted of 48 electrodes with an inter-electrode spacing of 5 meters and a dipole-dipole configuration array set up with an ABEM Terrameter LS resistivity meter. Cables were rolled so that the information covered the entire stretch of dike with a constant theoretical depth of investigation and without any blind areas. All measured contact

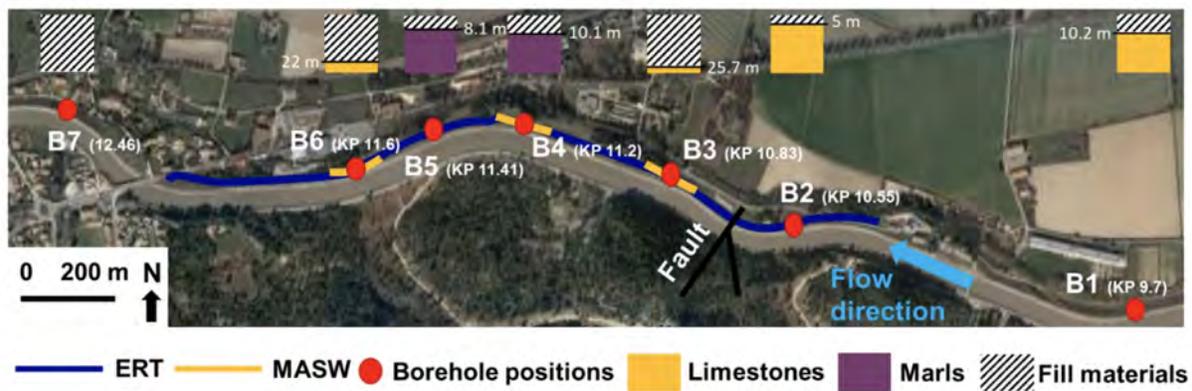


Fig. 1. Locations of the geotechnical borings and geophysical profiles on the crest of the studied dike section. The depths of contact between fill materials and substratum at each borehole are displayed on top.

resistances were below 5 k $\Omega$ . This electrode layout extended from KP 10.35 to KP 12.13 (blue profile, Figure 1).

The apparent resistivity data acquired (Figure 2a) were then inverted, in considering an L1 norm regularization on both the data and the model [26] and using an inversion software [27] to reconstruct a complete 2D-section of electrical resistivity [ $\Omega$ -m] compatible with the data. This step corresponded to a robust inversion that allowed accentuating resistivity contrast between lithological formations, in addition to limiting the effects of overly noisy data. This work uses the Res2Dinv software (version 3.71.118) [28].

For the inversion process, a flat topography is considered since the elevation variation is negligible all along the dike crest (only 11 cm). To simplify the processing of data as well as their fusion, the extended model option in Res2Dinv has been used. This option extends the model cells' vertical division to the edges of the survey line. However, to account for the difference in reliability of the resistivity values between meshes located at the center, bottom or sides, the resistivity imprecision values [28] resulting from the inversion are taken into account during the belief mass attribution stage. Moreover, in this new work, sensitivity values [28] are considered in the procedure. The sensitivity function indicates the extent to which a change in resistivity of a section of the subsurface influences the potential measured by the array. The higher the value of the sensitivity function, the greater the influence of the subsurface region on the measurement. Mathematically, the sensitivity function is given by the Fréchet derivatives [29], as detailed in [28].

### C. Multi-channel Analysis of Surface Waves

The MASW method consists of studying the dispersion of seismic surface waves (waveform deformation) in order to determine the shear wave velocity. As described in [30], use of this method comprises three stages: (i) data acquisition, (ii) determination of the Rayleigh dispersion curve, and (iii) the inversion process with a determination of shear wave velocities. The seismic campaign using MASW was carried out in 2017. A device with towed streamers shooting every 24 m was

activated, and the acquisition was performed with a Geode device from Geometric, allowing for the characterization of three sections within the considered dike area (Figure 1).

These three sections (yellow lines, Figure 1) extend respectively from KP 10.74 to KP 10.9, KP 11.1 to KP 11.3 and KP 11.5 to KP 11.7, with a geophone spacing of 2 meters. The MASW method assumes a laterally invariant medium. This assumption is verified by comparing the dispersion curves obtained for the recordings of the direct and reverse shots, as well as for the recordings of ambient vibrations. A non-horizontally layered medium will generally produce different dispersion curves on both the forward and reverse shots.

Velocity profiles are computed by means of: i) selecting one dispersion curve per section, and ii) carrying out an inversion to obtain a  $V_s$  profile that correctly explains the data using the Surfseis 5.0 software from the Kansas Geological Survey [31]. Unlike the electrical resistivity data, no imprecision values associated with the shear wave velocities are found after the inversion. This issue remains complex and one that has yet to reach a consensus opinion in the community [32], [33].

### D. Core drillings with particle size analysis testing

This work considers the geotechnical information from seven core drillings carried out on the crest of the studied area in 2016, five of which being located on the ERT profile (shown in red, Figure 1). These core drillings feature variable investigation depths (from 13 to 24 m), unlike those considered in [15]. The vertical resolution of information contained in the core is very fine (0.1 m). Particle size analysis tests were conducted in the laboratory on a large portion of the extracted samples; moreover, the classification outlined in French standard AFNOR NF P 11-300 [34] has been applied. The cohesive materials unable to undergo particle size analysis have been identified thanks to technician observations. Taking these lithological basements into account constitutes an innovation in the methodology and was not previously introduced in [15].

### III. DATA ANALYSIS AND PROCESSING

#### A. Acquired and processed data

1) *Electrical Resistivity Tomography*: The inverse model resistivity section, which will be used throughout this paper, is displayed in Figure 2c, and both the measured and calculated apparent resistivity (a) pseudo-sections are shown respectively in Figures 2a and 2b. The inversion was carried out over 4 iterations, yielding a final RMS error of 5.3%. A greater number of iterations tend to overstructure, geologically speaking, our imaging result, as explained by Descloitre et al. [35]. Figure 3 highlights the strong correlation existing in our study between measured and calculated apparent resistivity values.

The resistivity values (Figure 2c) suggest the local presence of resistive materials: from 8 to 16 m deep between KP 10.4 and KP 10.52; near the crest surface from KP 10.55 to KP 10.7; from 8 to 24 m deep between KP 10.8 and KP 10.95; and from 6 to 14 m deep at KP 12 to KP 12.1. Lower resistivities are observed over an area extending more than 500 meters in length from KP 11.1 to KP 11.65. The exact positions of the interfaces between the lithological sets cannot be accurately determined.

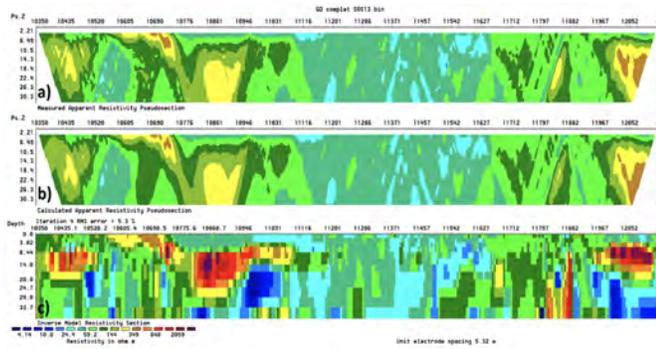


Fig. 2. Resistivities of the longitudinal dike section displayed in Figure 1: a) measured apparent resistivities, b) calculated apparent resistivities and c) model resistivity section by inverting dipole-dipole apparent resistivity data.

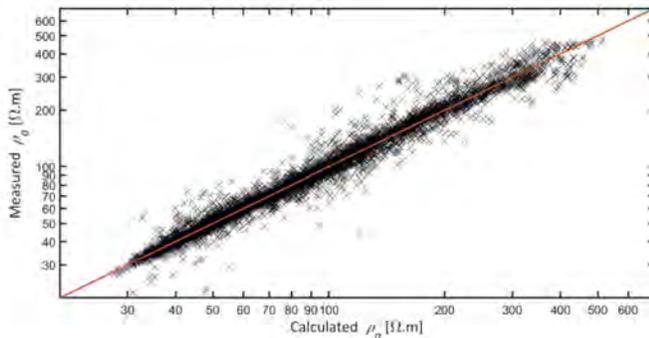


Fig. 3. Measured apparent resistivity vs. calculated apparent resistivity from data acquired on the ERT profile displayed in Figure 1.

2) *Multi-channel Analysis of Surface Waves*: From the data acquisition, the Rayleigh dispersion curves can be used

to plot the phase velocities vs. frequency. The maximum amplitude values are then picked from such a plot (Figure 4). After verifying the laterally invariant medium assumption, 20 seismic velocity profiles ( $V_s$  in  $m \cdot s^{-1}$ ) were obtained after an inversion process from the picked values of the dispersion curve. Each profile was representative of a 24-m long section with variable depths and a vertical discretization of 0.1 m. These velocity profiles are displayed in Figure 5; they all indicate lower velocities near the surface and increasing values at depths below 10 meters for the first (KP 10.74 to KP 10.9) and third (KP 11.5 to KP 11.7) sections. The second section (KP 11.1 to KP 11.3) primarily involves low  $V_s$  values.

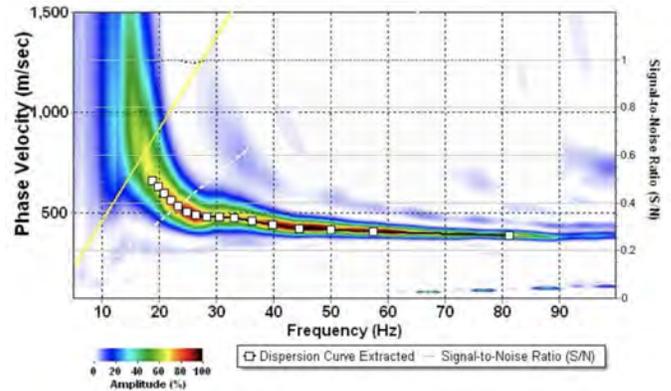


Fig. 4. Example of a Rayleigh dispersion curve with the maximum extracted amplitude values. This curve corresponds to data acquired on the profile extending from KP 11.572 to KP 11.596, with an offset from the source equal to 54 m.

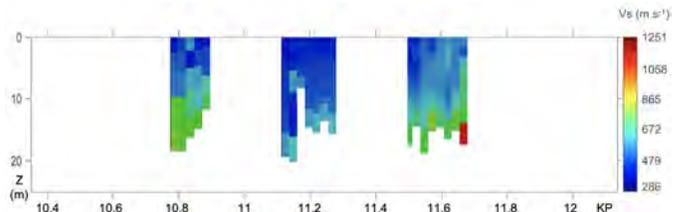


Fig. 5. The 20  $V_s$  profiles acquired by applying the MASW method within the studied dike area.

3) *Core drillings with particle size analysis testing*: The core drillings provide information on the presence of both fill materials (fine or coarse-grained) and the basement (marls or limestones). Figure 6 shows the presence of fine fill materials near the surface from boreholes B3 to B7 as well as marls below for the B4 and B5 drillings. Limestone is identified below the fill materials in boreholes B1, B2 and B6. The lithological basement seems particularly high in B2. Let's note that the B3, B4 and B6 core drillings are located in MASW investigation areas.

#### B. Belief functions and combination rules

Belief functions (BFs) were introduced by Shafer [7] in the mathematical theory of evidence inspired by the works

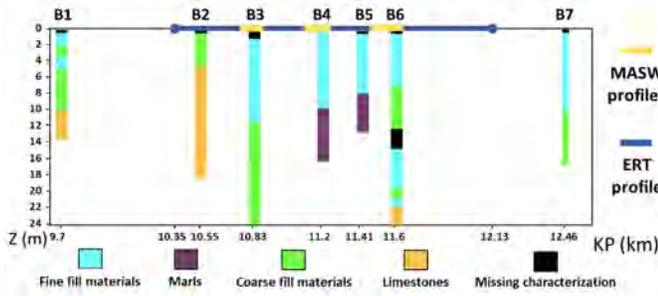


Fig. 6. Representative diagram of the studied dike section, with the positions of the seven core drillings and the identified lithological materials.

of Dempster [8], which is why belief function theory is often referred to as the Dempster-Shafer theory (DST). It serves to compute the belief and plausibility of a hypothesis (corresponding, in this article, to lithological materials) from various sources of information (ERT, MASW, core drillings).

The main advantage of BFs is their ability to manage information from various sources, in association with their respective imperfections (uncertainties and imprecisions). BF theory is also able to assess the level of conflict between sources, i.e. when information given by one source is inconsistent with that given by another source. According to the definition offered by Smets [36], it can be considered that uncertainties correspond to degrees of confidence related to a physical value, whereas imprecisions correspond to value intervals directly associated with measurement errors related to the investigation method. For example, the uncertainty of measuring a geotechnical parameter value identical to the one measured in a borehole increases with the distance to that borehole. However, imprecision may be associated with the error bar of the measured datum. In our methodology, given that uncertainties correspond to the belief masses associated with the various defined hypotheses, the imprecisions associated with each type of data will be detailed in the corresponding sections.

Furthermore, BFs can take into account ignorance and the incompleteness of information. It is possible to grant a credit on all possible results (all possible types of lithological materials) in order to quantify our ignorance, while probability theory would simply assign an equiprobability to each single hypothesis. Martin et al. [?] provided a detailed explanation of BF theory for the interested reader.

The implementation of BFs has been divided into four stages, namely: (i) define a Frame of Discernment (FoD), denoted  $\Theta$ ; (ii) assign belief mass values to the hypotheses of this FoD for each information source; (iii) select and use a combination rule for the information fusion step; and (iv) provide a representation of the merged information.  $\Theta$  consists of all possible hypotheses within the problem under concern. The elements of the FoD are exhaustive and exclusive, e.g. for  $n$  hypotheses:

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}. \quad (1)$$

For the problem under consideration in this article, the possible

hypotheses of the FoD correspond to lithological materials potentially found in the studied dike section. In light of available knowledge, let's set an FoD of five hypotheses, common for all sources, such that:

$$\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}, \quad (2)$$

where

- $\theta_1$  corresponds to fine fill materials,
- $\theta_2$  corresponds to marls,
- $\theta_3$  corresponds to coarse fill materials with limestone breccia,
- $\theta_4$  corresponds to limestones,
- $\theta_5$  corresponds to any material different from the four listed above.

Now, let's introduce an additional hypothesis ( $\theta_5$  here) representing one or more unexpected materials, in order to cover the entire field of possibilities. The space of belief mass functions is the set of all subsets of  $\Theta$ , written  $2^\Theta$ . It is determined by all the disjunctions and the conflict between information sources (denoted  $\emptyset$ ), yielding in the present case:

$$\begin{aligned} 2^\Theta = \{ & \emptyset, \theta_1, \theta_2, \theta_1 \cup \theta_2, \theta_3, \theta_1 \cup \theta_3, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3, \\ & \theta_4, \theta_1 \cup \theta_4, \theta_2 \cup \theta_4, \theta_1 \cup \theta_2 \cup \theta_4, \theta_3 \cup \theta_4, \\ & \theta_1 \cup \theta_3 \cup \theta_4, \theta_2 \cup \theta_3 \cup \theta_4, \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4, \\ & \theta_5, \theta_1 \cup \theta_5, \theta_2 \cup \theta_5, \theta_1 \cup \theta_2 \cup \theta_5, \theta_3 \cup \theta_5, \\ & \theta_1 \cup \theta_3 \cup \theta_5, \theta_2 \cup \theta_3 \cup \theta_5, \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_5, \\ & \theta_4 \cup \theta_5, \theta_1 \cup \theta_4 \cup \theta_5, \theta_2 \cup \theta_4 \cup \theta_5, \theta_1 \cup \theta_2 \cup \theta_4 \cup \theta_5, \\ & \theta_3 \cup \theta_4 \cup \theta_5, \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_5, \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5, \\ & \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5\}. \end{aligned} \quad (3)$$

The belief mass function  $m_j$  in  $[0, 1]$  is defined for each information source  $S_j$  (with  $j = 1, 2$  or  $3$  in this study) and attributed to a subset  $X$  (defined on  $2^\Theta$ ). Like in probability theory, the more  $m_j(X)$  tends to 1 the higher the confidence in  $X$ . Furthermore, the definition of a belief mass function implies that the sum of the masses (over all subsets) of a given source of information equals 1:

$$\sum_{X \in 2^\Theta} m_j(X) = 1. \quad (4)$$

The essential difference with probability theory is that  $X$  can represent the union of two or more hypotheses. For example, if a belief mass is allocated to  $\theta_1 \cup \theta_2$ , this means that either  $\theta_1$  OR  $\theta_2$  is a possible solution, thus making it possible to model uncertainty and lack of knowledge. Belief and plausibility functions,  $Bel$  and  $Pl$  respectively, are considered as upper and lower bounds of an unknown probability  $P$  such that for any  $X \in 2^\Theta$ ,  $Bel(X) \leq P(X) \leq Pl(X)$ . Belief and plausibility functions are in a one-to-one relationship with the belief mass function,  $m(\cdot)$ , and defined by:

$$Bel(X) = \sum_{\substack{Z \subseteq X \\ Z \in 2^\Theta}} m(Z), \quad (5)$$

$$Pl(X) = \sum_{\substack{Z \cap X = \emptyset \\ Z \in 2^\Theta}} m(Z). \quad (6)$$

However, in our methodology, only belief mass functions have been used since the combination rules are set directly from the allocated belief masses for each information source. Two combination rules have been applied herein. First, Smets [38], in the Transferable Belief Model, enables the assignment of a belief mass to the conflict,  $\emptyset$ , such that:

$$m_{1,2,\dots,S}(\emptyset) > 0, \quad (7)$$

where  $m_{1,2,\dots,S}(\cdot)$  denotes the combined belief mass resulting from the merging of information stemming from  $S$  different sources. The belief mass  $m_{1,2,\dots,S}(X)$  resulting from the conjunctive fusion of information from all sources (from  $j = 1$  to  $S$ ) is written:

$$m_{1,2,\dots,S}(X) = \sum_{\substack{X_1, \dots, X_S \in 2^\Theta \\ X_1 \cap \dots \cap X_S = X}} \prod_{j=1}^S m_j(X_j). \quad (8)$$

The level of inconsistency between the  $S$  information sources is then expressed as:

$$m_{1,2,\dots,S}(\emptyset) = \sum_{\substack{X_1, \dots, X_S \in 2^\Theta \\ X_1 \cap \dots \cap X_S = \emptyset}} \prod_{j=1}^S m_j(X_j). \quad (9)$$

$m_j(X_j)$  stands for the belief mass assigned to hypothesis  $X_j$  by information source  $j$ . The second combination rule to be applied is the Proportional Conflict Redistribution Rule No. 6 (PCR6) [39]. According to Shafer's approach and unlike Smets' rule, the PCR6 rule does not allow assigning any belief mass to the conflict. Thus, in PCR6, one has by definition:

$$m_{1,2,\dots,S}^{PCR6}(\emptyset) = 0 \quad (10)$$

Hence, PCR6 allows for the reallocation of all partial conflicts, in proportion to the masses of the subset concerned by these conflicts, so that the specificity of the information is fully preserved during the fusion process. The belief mass in  $X$ ,  $m_{1,2,\dots,S}^{PCR6}(X)$ , resulting from the fusion of information from  $s$  sources, is:

$$m_{1,2,\dots,S}^{PCR6}(X) = m_{1,2,\dots,S}(X) + \sum_{k=1}^{S-1} \sum_{\substack{X_{i_1}, X_{i_2}, \dots, X_{i_k} \in 2^\Theta \setminus \{X\} \\ \cap_{j=1}^k X_{i_j} = \emptyset}} [m_{i_1}(X) + \dots + m_{i_k}(X)] \\ \cdot \frac{m_{i_1}(X) \dots m_{i_k}(X) m_{i_{k+1}}(X_{i_{k+1}}) \dots m_{i_S}(X_S)}{m_{i_1}(X) + \dots + m_{i_k}(X) + m_{i_{k+1}}(X_{i_{k+1}}) + \dots + m_{i_S}(X_S)} \quad (11)$$

where  $P^S(\{1, \dots, S\})$  is the set of all permutations of elements  $\{1, 2, \dots, S\}$ . It should be emphasized that  $X_{i_1}, X_{i_2}, \dots, X_{i_S}$  may be different from each other, equal, or some equal and some different, etc.

For example, let's consider two hypotheses  $A$  and  $B$ , with  $2^\Theta = \{\emptyset, A, B, A \cup B\}$  and two information sources,

such that  $m_1(A) = 0.6$  and  $m_2(B) = 0.3$ . With PCR6, the partial conflicting mass  $m_1(A)m_2(B) = 0.6 \cdot 0.3 = 0.18$  is redistributed to  $A$  and  $B$  only with respect to the following proportions:  $x_A = 0.12$  and  $x_B = 0.06$  because:

$$\frac{x_A}{m_1(A)} = \frac{x_B}{m_2(B)} = \frac{m_1(A)m_2(B)}{m_1(A) + m_2(B)} = \frac{0.18}{0.9} = 0.2$$

More numerical examples along these lines can be found in [39].

### C. Geophysical definition of the FoD using the K-means clustering classification

Once the geophysical data have been acquired, the next step is to determine the belief mass distributions associated with the various hypotheses of the FoD  $\Theta$ . These sets of belief masses are specific to each information source and associated with each grid cell representative of the dike section. Before the fusion stage, the methodology indeed requires all information sources to have sets of belief masses defined on the same section and on a common mesh. The dimensions of this section are to be fixed by the source covering the largest area, which here would be the ERT method.

The hypotheses of the FoD (Eq. 2) must be associated with physical values (electrical resistivity, shear wave velocity). Previously, in Dezert et al. [14], [15], a representation of the distribution of geophysical values, in the form of modal classes, was employed. Such a representation enables highlighting the minima and maxima, under expert interpretation, that would be used to set the bounds of the intervals associated with the FoD hypotheses. Obviously, the general trend in the value distribution serves to identify the lithological material groups making up the studied section.

A new procedure is being proposed herein to determine the geophysical parameter intervals associated with the FoD hypotheses, based on  $K$ -means clustering [40]. This clustering method allows classifying the geophysical parameter values (electrical resistivity and shear wave velocity, respectively) into  $K$  clusters, as derived following the resolution of a combinatorial optimization problem. The  $K$ -means clustering algorithm is an iterative minimization of the sum of distances between each data point and a fixed centroid (initialization). This algorithm modifies the affiliation of the points of each cluster until the sum of the distances can no longer decrease (least squares method), resulting in a set of compact and delimited clusters. The use of this type of clustering algorithm rather than another seems relevant for our problem set-up since our data do not display clearly delimited sets; however, it remains easily possible for the user to indicate the desired number of clusters.

Even though clustering methods are typically performed on multi-variate parameters [41],  $K$ -means clustering cannot be run on co-located resistivity and seismic velocity data in our methodology since the information sources are assumed to be independent in belief function theory and unable to interact with each other. Such use of  $K$ -means clustering would belie that definition. Therefore, the geophysical methods are to be

considered individually, with clustering applied on individual parameters.

The number of clusters is a subjective choice that relies on an interpretation of the geophysical acquisitions (Figures 2 and 5) and modal class distribution of the physical parameters (Figures 7 and 8). A more objective manner would be to use existing indices, such as the Davies-Bouldin index, to fix the number of clusters  $K$ . However, such an index would require a large number of clusters that may prove impossible to associate with specific hypotheses. Our opinion is that a reading and interpretation of the results from geophysical acquisitions is still valuable to fixing the number of clusters to be associated with the FoD hypotheses. This  $K$ -means clustering classification then serves as an aid in selecting precise values for the physical intervals.

1) *Electrical Resistivity Tomography*: From an observation of Figure 2c, as previously described, three sets of resistivities emerge. It has been decided to associate low resistivity values with materials  $\theta_1 \cup \theta_2$  (more conductive materials: fine-grained fill or marl basement) and then high resistivity values with materials  $\theta_3 \cup \theta_4$  (more resistive materials: coarse-grained fill or limestone basement). Since the intermediate values do not provide information on the exact nature of the lithological material, they will be associated with the union of the four hypotheses,  $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$ . The creation of three clusters by means of the  $K$ -means clustering classification technique will therefore be considered.

A modal class representation of the number of cells of the 2D-ERT section (Figure 2c), with respect to the resistivity values depicted in log scale, is displayed in Figure 7. This figure highlights the clustering proposed by the  $K$ -means algorithm for the three defined clusters. The intervals of values associated with the FoD hypothesis for the ERT method are thus defined in  $\Omega \cdot m$  for the characterization as:

- $[2.5, 75]$  associated with  $\theta_1 \cup \theta_2$  (fine-grained fill or marl),
- $[354, 10^4]$  associated with  $\theta_3 \cup \theta_4$  (coarse-grained fill with breccia or limestone),
- $]75, 354[$  associated with  $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$  (one of the four materials described),
- $[0.1, 2.5[ \cup ]10^4, 3 \cdot 10^5]$  associated with  $\theta_5$  (none of the previously described materials).

By definition, the intervals associated with the  $\theta_5$  hypothesis (other lithological materials) do not contain any resistivity value present in the section. However, the association of a resistivity interval with this hypothesis is needed in order to provide it with a “physical reality” in terms of resistivity and compute the associated belief masses to  $\theta_5$ . As described hereafter, the computation of belief masses associated with each hypothesis does require the computation of distances between physical intervals of values.

2) *Multi-channel Analysis of Surface Waves*: On the seismic velocity profiles available (Figure 5), like for the ERT method, three sets emerge: low, intermediate, and high shear wave velocities. Low velocities are associated with finer materials  $\theta_1 \cup \theta_2$  and high velocities with coarser materials

$\theta_3 \cup \theta_4$ . Since the intermediate values of  $V_s$  do not provide information on the exact nature of the lithological material, they can be associated with the union of the four hypotheses  $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$ . Low speeds are initially associated with less cohesive materials (fine and coarse fill) and higher speeds with more cohesive materials (marl and limestone). However, this characterization of FoD generated a significant conflict between the source of seismic information and other information sources. The FoD characterization used herein agrees more closely with both the ERT characterization and particle size analysis. The high velocities associated with  $\theta_3$  could be attributed to the numerous cohesive limestone breccia present in the coarse-grained fill materials.

Figure 8 displays the same type of representation as that proposed in Figure 7; it highlights the clustering proposed for the three defined clusters. The velocity value intervals associated with the FoD hypothesis for the MASW method are therefore defined in  $m \cdot s^{-1}$  as:

- $[180, 450]$  associated with  $\theta_1 \cup \theta_2$  (fine-grained fill or marl),
- $[670, 1.3 \cdot 10^3]$  associated with  $\theta_3 \cup \theta_4$  (coarse-grained fill with breccia or limestone),
- $]450, 670[$  associated with  $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$  (one of the four materials described),
- $[1, 180[ \cup ]1.3 \cdot 10^3, 2 \cdot 10^3]$  associated with  $\theta_5$  (none of the previously described materials).

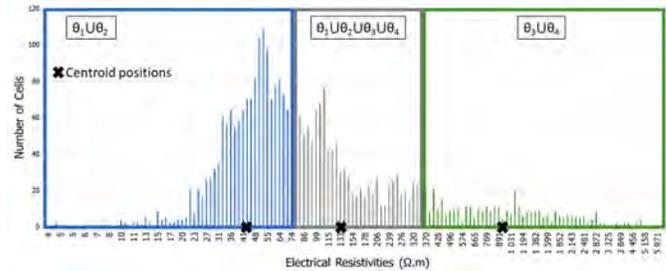


Fig. 7. Modal class distribution of electrical resistivities in the form of three clusters using the  $K$ -means clustering classification and the positions of the respective centroids.

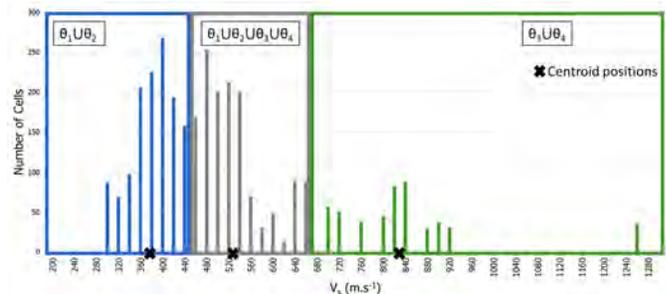


Fig. 8. Modal class distribution of shear wave velocities divided into three clusters using the  $K$ -means clustering classification and the positions of the respective centroids.

#### D. Belief mass computations from single sources of information

1) *Electrical Resistivity Tomography*: Once the FoD,  $\Theta$ , has been characterized with resistivity values, belief masses need to be assigned to each hypothesis of  $2^\Theta$ , for each of the common meshes. It is therefore necessary, from the standpoint of a resistivity value, to assign a distribution of belief masses on all hypotheses. This step entails taking into account the imprecision on the inverted resistivities, stemming from inaccuracies and sensitivities provided by the inversion process. More specifically, the imprecision on resistivity values equals the ratio of the inaccuracy to the sensitivity, expressed as a percent. Thus, instead of considering a simple resistivity value, it is possible to consider an interval of resistivities with a lower bound (resistivity value minus its associated imprecision) and an upper bound (calculated resistivity plus its associated imprecision), whose central value is the inverted resistivity value. Thus, the greater the imprecision, the wider the interval obtained.

Once the interval has been determined, the ascribed belief mass is calculated as a function of the “distance” between this interval and the intervals associated with the FoD hypothesis. These distances are computed by considering “Wasserstein distances” [42], whereby the shorter the distance of a resistivity interval to an FoD hypothesis, the greater the belief mass and vice-versa. This procedure is explained in detail in [9] as well as in [15].

Belief masses are assigned on all hypotheses with defined resistivity intervals (here  $\theta_1 \cup \theta_2$ ,  $\theta_3 \cup \theta_4$ ,  $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$  and  $\theta_5$ ), as the other masses are set to 0. In accordance with the definition of a belief mass function, the sum of these masses equals 1 (Eq. 4). Each mesh is thus associated with normalized belief mass functions.

2) *Multi-channel Analysis of Surface Waves*: Like for the ERT method, once the FoD,  $\Theta$ , has been characterized with  $V_s$  values, belief masses must be assigned to each hypothesis of  $2^\Theta$ , for each cell (sized  $24 \times 0.1 \text{ m}^2$ ). It is therefore necessary, from the standpoint of a shear wave velocity value associated with a cell, to assign a belief mass distribution on all hypotheses. The process is identical to that introduced for the ERT. However, unlike the ERT method, imprecision values associated with the available shear wave velocities are not available. Hence, imprecision values are randomly simulated [43] according to a normal distribution (for a mean imprecision value of 10% and associated standard deviation of 5%).

Belief masses are thus assigned to all hypotheses with defined shear wave velocity intervals (here  $\theta_1 \cup \theta_2$ ,  $\theta_3 \cup \theta_4$ ,  $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$  and  $\theta_5$ ), with the other masses being set to 0. As mentioned above, by virtue of the definition of a belief mass function, the sum of these masses equals 1 (Eq. 4).

The dike section covered by the ERT method is larger than that covered by the MASW method. However, the ERT section dimensions serve as a reference in our methodology. As such, belief mass distributions need to be proposed for the second source of geophysical information (i.e. MASW), where no seismic velocity information is available. For this step, a

belief mass of  $m_2(\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5) = 1$  is chosen. This mass represents the “complete” uncertainty, in recognition of the absence of information in those areas not covered by the MASW campaign.

3) *Core analysis*: Like for the two geophysical methods previously discussed, it is necessary, for each mesh of the section, to associate the results of particle size analyses or other geotechnical observations with the FoD hypothesis,  $\Theta$ . Unlike the ERT and MASW methods, the materials extracted by coring serve to discriminate the four hypotheses  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  and  $\theta_4$ . Imprecisions associated with weighing of the materials are taken into account in the procedure but do not alter our ability to discriminate the hypotheses.

As mentioned above, the classification of French standard AFNOR NF P 11-300 [34] is used to distinguish fine-grained from coarse-grained fill materials. Moreover, this standard allows associating fine-grained fill materials with hypothesis  $\theta_1$  and coarse-grained fill materials with hypothesis  $\theta_3$ . The cohesive materials, i.e. marly and limestone basements, that have not undergone particle size analysis are respectively associated with hypothesis  $\theta_2$  (marl) and  $\theta_4$  (limestone) thanks to the visual characterization carried out by the technicians.

A belief mass close to 1, i.e.  $m_3(\cdot) = 0.99$ , associated with the characterized hypothesis is fixed at the sampling points. In a complementary manner, a mass of  $m_3(\cdot) = 0.01$  is then attributed to the union of all hypotheses ( $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5$ ), in agreement with the definition of the belief mass function, which requires that the sum of the belief masses assigned by an information source equal 1 (Eq. 4). The other belief masses are all set to 0.

This mass value of 0.99 can theoretically be modified depending on the ability of the geotechnical method to characterize the lithological material under investigation. Given our belief that core extraction and observation along with particle analysis constitute the best means for characterizing the lithology of a material, the high level of confidence in boreholes finds its justification. If the information provided by the geotechnical observation is missing or incomplete and no particle size analysis has been carried out, then the entire belief mass is allocated to the absolute uncertainty such that  $m_3(\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5) = 1$  (e.g. B6 borehole depth between 13 and 15 m, Figure 6).

Core drillings provide spatially specific information compared to the output of the ERT and MASW methods, both of which cover larger areas. Since the dimensions of the ERT section serve as a geometrical reference in our methodology, belief masses must be assigned to the geotechnical source of information where no core drilling is carried out. A discretization of the whole section is thus performed to cover the entire inter-borehole space.

An exponential decrease of the belief masses is imposed laterally from each borehole point to the adjacent borehole, to both the left and right. The rate of decrease is a function of two parameters: i) the lateral decay coefficient  $k$ , and ii) the coefficient of variation of particle size values  $C_v$ . As a result,

for a given depth, we obtain:

$$M(x) = 0.99 \cdot e^{-kC_v x}, \quad (12)$$

with  $x$  being the horizontal distance from the considered mesh to the reference borehole, in meters (with  $x = 0$  in the borehole),  $M(x)$  the belief mass values assigned to each hypothesis in the FoD for a position  $x$ , with 0.99 as the belief mass value assigned to the hypothesis identified in borehole ( $M(0)$ ).

The coefficient  $k$  is set by the user of the methodology and depends on the lateral variability of the investigated medium. The value of  $k$  should increase with subsoil variability. A parametric study on the influence of this lateral decay coefficient has been proposed in [14], in suggesting a value of  $k = 0.1$ . The coefficient of variation of particle size  $C_v$  is computed for a considered borehole, at a fixed depth, by using values from the borehole as well as from the adjacent borehole. As such, for two consecutive boreholes displaying similar particle size values at the same depth, the decrease in confidence is smaller than for two consecutive boreholes with radically different values. The expression of  $C_v$  is shown in Eq. 13.

$$C_v = \sqrt{\frac{1}{n_{\text{mesh}} \sum_{i=1}^{n_{\text{mesh}}} (Q - Q_i)^2}}, \quad (13)$$

where  $Q$  is the geotechnical parameter value of the considered cell in the borehole (cumulative sieve less than  $80 \mu\text{m}$ ), and  $Q_i$  the value in the adjacent borehole centered at the same depth. This study considers  $n_{\text{mesh}} = 7$ , so that the computation of  $C_v$  takes into account 7 cells in the adjacent borehole (i.e. 70 cm thick when assuming a vertical resolution of 10 cm).

Consequently, in our studied section, boreholes B1 and B7 are of interest, even though they are absent from the ERT profile (Figure 2c). These two boreholes make it possible to compute the coefficients of variation and therefore the decrease rate of B2 belief masses to the left and the decrease rate of B6 belief masses to the right. For a given cell, when the belief mass associated with a hypothesis is less than 1, the remainder of the mass to be allocated is assigned to the “any material” hypothesis ( $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5$ ). Beyond the maximum depth of geotechnical investigation, since no information is available,  $m_3(\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5) = 1$ .

When the materials are characterized visually by a technician but not analyzed in the laboratory, the  $C_v$  value cannot be computed due to the absence of particle size data. In this case, two  $C_v$  values are set by the user of the methodology, i.e.: i) a low value making it possible to extend the information widely when two identical materials are present at the same depth for adjacent boreholes (e.g. between 10 and 12 m deep, to the right of borehole B4, Figure 6); and ii) a high value serving to limit the extension of information locally when two different materials are present at the same depths for two adjacent boreholes (e.g. between 10 and 12 m deep to the left of borehole B4, Figure 6).

The geotechnical information being extended from a borehole to adjacent boreholes implies double information at the

level of the inter-borehole meshes (information originating from both the left and right). In order to have just one distribution of geotechnical belief masses in each cell, this double information is processed into a single one. For this step, the information originating from the left is considered as an initial source while that from the right as a second source. A preliminary fusion process is then conducted between these two belief mass distributions using the PCR6 rule (Eq. 11), in each cell of the section.

#### IV. RESULTS

Before displaying the fusion results, since belief mass distributions have been computed for all individual sources, they will first be displayed in Figures 9 to 12. It is essential to mention that this fusion process can take place between all information sources or else by considering them in pairs. This work thus presents the results for the fusion of information acquired solely by the geophysical ERT and MASW methods (Figure 13) and then by all three methods (ERT, MASW and core drillings) in considering respectively four boreholes (Figure 14) and seven boreholes (Figure 15).

##### A. Belief mass distributions for individual information sources

1) *Electrical Resistivity Tomography*: Let’s start by a display consisting of two complementary figures (9a and 9b) that highlight two distinct types of information. Figure 9a shows, for each cell of the ERT section, the hypotheses of  $\Theta$  having the greatest belief mass, while the associated mass values are provided in Figure 9b. The presentation of these results serves to highlight the ability of our methodology to represent the uncertainty (union of hypotheses with associated confidence indices) while taking imprecision into account. The materials most plausibly present in the dike section according to the ERT method are thus depicted in Figure 9a, with the associated confidence index in Figure 9b. The four areas of high resistivities described in Figure 2c appear in green in Figure 9a.

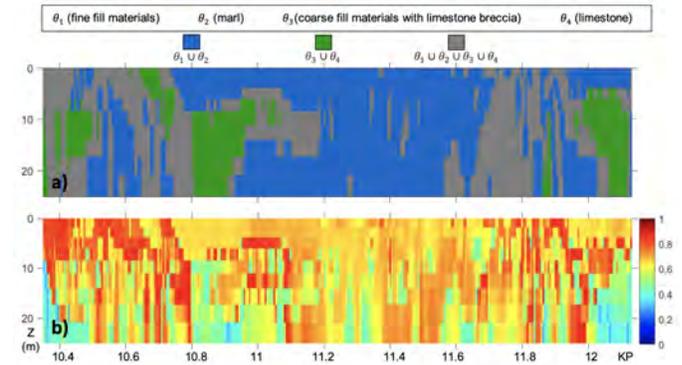


Fig. 9. a) lithological section of the dike based on the hypotheses having the highest belief mass according to the mass attribution from electrical resistivity data, with an FoD characterization by means of clustering; b) dike section of the mass values associated with the hypotheses shown in a).

The high uncertainty (low belief masses) is due to a lack of real measured data in the extended ERT model, as well as to

the resolution of ERT data intrinsically decreasing with depth. This decrease in confidence with depth could not be demonstrated in previous works [14], [15] since sensitivity values had not been taken into consideration. Figure 9 indicates that the characterization of the section using the ERT method fails to discriminate the four hypotheses (fine fill, coarse fill, marly basement, limestone) featuring common resistivity values.

2) *Multi-channel Analysis of Surface Waves*: The lithological materials most plausibly present in the subsoil according to the MASW method are displayed in Figure 10a, along with the associated confidence index in Figure 10b. It appears that these results are in close agreement with the characterization proposed by the ERT method in Figure 9, in particular with both the characterization of  $\theta_1 \cup \theta_2$  at shallow depth for the three sections covered and the characterization of  $\theta_3 \cup \theta_4$  beyond a depth of 10 m for the first section, i.e. around KP 10.8.

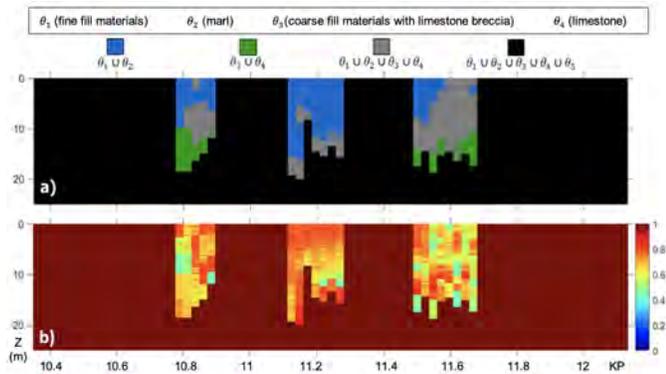


Fig. 10. a) representation of the hypotheses having the highest belief mass according to the mass attribution from shear wave velocity data, with an FoD characterization by clustering; b) representation of the mass values associated with the hypothesis presented in a).

3) *Core analysis*: The lithological materials most plausibly present in the subsoil according to the core drilling method are displayed in Figures 11a and 12a, with the associated confidence indices in Figures 11b and 12b. Figure 11 shows the results when considering four boreholes (B1, B2, B5 and B7), while the results in Figure 12 consider all available boreholes (B1 through B7).

As expected, these two figures highlight the strong confidence (belief masses close to 1) near the borehole locations (Figures 11b and 12b) as well as their variable lateral decrease depending on the materials encountered in adjacent boreholes. For example, when looking at borehole B4 (Figure 12), while the confidence associated with the hypothesis  $\theta_1$  (fine fill materials) extends widely to the right over the first 8 meters because B5 characterizes the same material at similar depths, the extent of confidence in this hypothesis is much more restricted between  $z = 8$  and  $z = 10$  m, since  $\theta_2$  is characterized in B5 at these depths. As for the MASW method, these figures substantiate the ability of the methodology to represent the lack of information (incompleteness). Note that

the complete uncertainty ( $\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5$ ) is displayed in black in Figures 11a and 12a.

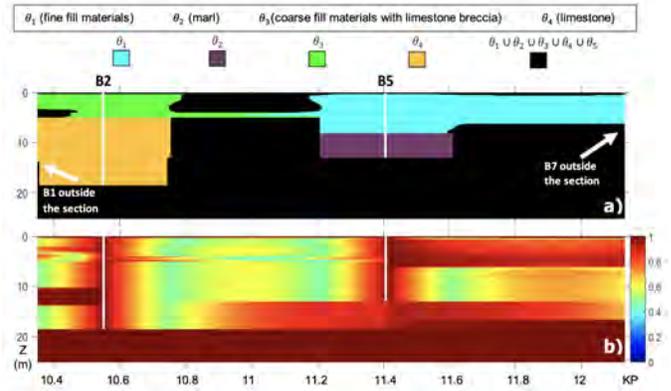


Fig. 11. a) representation of the hypothesis with the highest belief mass according to the mass attribution from four core drilling data; b) representation of the mass values associated with the hypothesis presented in a).

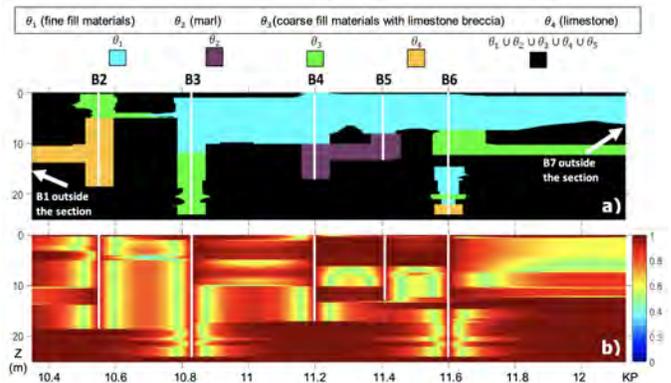


Fig. 12. a) representation of the hypothesis with the highest belief mass according to the mass attribution from seven core drilling data; b) representation of the mass values associated with the hypothesis presented in a).

### B. ERT and MASW fusion results

The fusion results of ERT and MASW presented in Figures 13a and 13c do not allow characterizing the lithological materials individually given the inability of either method (see Figures 9 and 10). Figures 13b and 13d reveal that the confidence level is enhanced when the same materials are characterized by the two geophysical methods, especially near the dike crest in the case of fine materials. Two conflict zones appear, with the larger one being located below the 3rd section of MASW (KP 11.5-11.7) 15 meters deep. The MASW method characterizes the  $\theta_3 \cup \theta_4$  hypothesis (Figure 10a) while the ERT method characterizes the  $\theta_1 \cup \theta_2$  hypothesis (9a). On the whole, the results proposed here are very close to the characterization produced by ERT alone (Figure 9) since information is provided on a much larger area than the MASW method. A large part of the dike section is characterized with  $m_2(\theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5) = 1$  for the MASW method (in black, Figure 10a), which yields zero information. Also, the

masses associated with the PCR6 rule are higher and closer to 1 (Figure 13d) than those associated with Smets' rule (13b). This outcome stems from the fact that the conflict mass  $m(\emptyset)$  in Smets' rule is reallocated on the other FoD hypotheses with the PCR6 rule, as described in Eq. (11), hence  $m^{PCR6}(\emptyset) = 0$ .

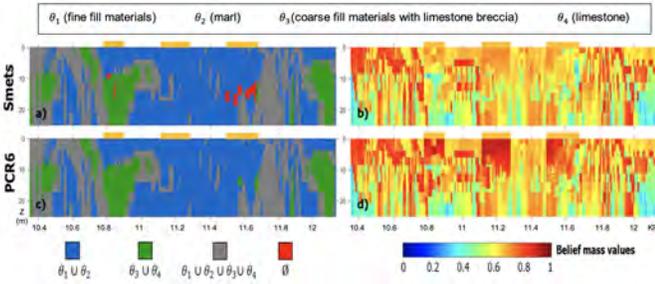


Fig. 13. a,c) representations of the hypotheses with the highest belief mass according to the mass attribution from ERT and MASW fusion using Smets' and PCR6 rules; b,d) representation of the mass values associated with the hypotheses presented in a) and c) respectively.

### C. ERT, MASW and core drillings fusion results

In comparison with the results displayed above, the contribution of core drillings is helpful, as they allow dissociating the lithological materials individually and proposing precise interface positions, as opposed to the smooth ones shown in Figure 2. It can be observed in Figures 14b, 14d, 15b and 15d that the conflict level decreases when deviating from the borehole positions since the confidence level on the geotechnical information decreases with distance to these borehole positions. Comparatively speaking, the influence of the geophysical information becomes more significant.

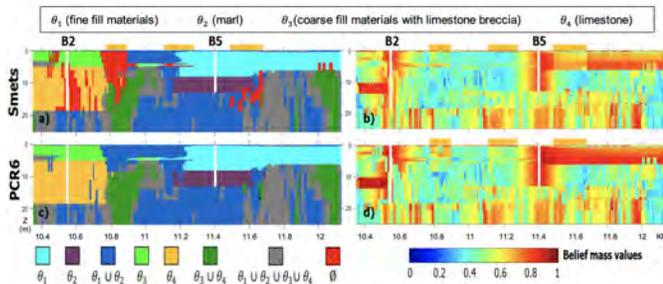


Fig. 14. a,c) representation of the hypotheses with the highest belief mass according to the mass attribution from the fusion of ERT, MASW and core drillings (four boreholes), using the Smets' and PCR6 rules respectively; b,d) representation of the mass values associated with the hypotheses presented in a) and c) respectively.

Overall, the belief masses of the characterized hypothesis seem to be lower with three sources of information (Figures 14b, 14d, 15b and 15d) than with two (13b and 13d). However, this finding does not imply that the results are of lower quality. Indeed, it is necessary to state that the materials characterized are individual hypotheses (Figures 14a, 14c, 15a and 15c), while a union of hypothesis was depicted in Figure 13. Therefore, the characterization using three sources of

information is more precise than that proposed by considering just the ERT and MASW methods.

It is interesting to compare the fusion results considering four (Figure 14) and seven boreholes (Figure 15) by examining the contribution of the three additional core drillings (B3, B4, B6). Overall, confidence is higher with seven geotechnical investigation boreholes (Figures 15b and 15d) than with four (14b and 14d). Some conflict zones present with four boreholes are reduced when integrating the three additional ones (e.g. on the first 10 meters, KP 10.8), yet new conflict zones can also appear when the particle size information contradicts the geophysical characterization (e.g. B3, 10 meters deep). As for the fusion of two information sources, the PCR6 results indicate higher belief masses and remove the conflict displayed by results using Smets' rule. On the other hand, they also show that the zones where conflict is present with Smets' rule exhibit low confidence with PCR6 (e.g. KP 10.8 - 11 at 10 meters in depth, conflict represented in Figure 15a, and associated low masses following use of the PCR6 rule in Figure 15d).

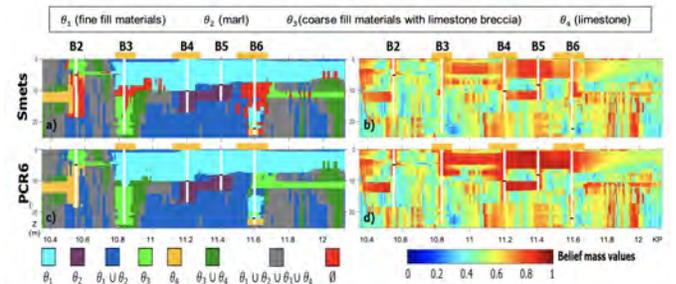


Fig. 15. a,c) representation of the hypotheses with the highest belief mass according to the mass attribution from the fusion of ERT, MASW and core drillings (seven boreholes in white, with the black lines indicating major contact between cohesive and non-cohesive materials), using the Smets' and PCR6 rules respectively; b,d) representation of the mass values associated with the hypotheses presented in a) and c) respectively.

## V. DISCUSSION

### A. Lithological characterization of the dike section

The results obtained by our fusion methodology are consistent with the geological description of the dike section established in the beginning of this article. Figure 15c reveals the presence of a limestone basement up to KP 10.6, with coarse fill materials above. Then, between KP 10.6 and KP 10.8, the NE-SW fault is most likely being detected, lowering the western compartment. This finding explains why coarse fill materials are characterized at such depths at KP 10.8, whereas they are much closer to the surface at KP 10.55. This fault zone between KPs 10.55 and 10.8 is poorly characterized (union of hypotheses displayed in Figure 15c, and low associated belief masses in Figure 15d), which suggests the possibility of a displaced region. Moreover, fine fill materials are located up to a depth of about 10 m, with the presence of an underlying marly basement. From KP 11.6, the characterization is less obvious with certainly the alternation of finer and coarser fill materials. The uncertainty associated with geophysical information and

the lack of geotechnical information (with boreholes B6 and B7 being very far apart) leaves this area beyond 10 m deep poorly characterized. The available knowledge suggests that Cretaceous Limestone-type materials could be present between KP 11.5 and KP 12.13; this characterization thus appears to be consistent with our fusion results.

These fusion results allow proposing more precise lithological interface positions thanks to the geotechnical characterization. In addition to being able to individually characterize lithological materials, it is very helpful to have areas with no doubt existing between two or even four materials. This information is valuable in order to determine where it would be pertinent to strengthen the geotechnical investigations. The displayed results highlight the ability to take advantage of complementary information from geophysical and geotechnical methods. While the ERT and MASW methods did not serve to distinguish the four described materials in Figure 13, the core drilling results did clarify the section by both distinguishing the materials and specifying the interface locations. This step is performed, for example, at the level of borehole B3 (around 11.5 m deep) between fine and coarse fill materials. While the ERT and MASW methods suggested a higher interface (Figure 13c), the information contained in the core drillings makes it possible to readjust this position and express this contradiction between sources thanks to Smets' combination rule, with a representation of the conflict (in red, Figure 15a). This notion of conflict is valuable to understanding both the data and results; furthermore, it seems that no other existing work exposes this type of information.

Since the conflict is not a solution in itself, the PCR6 rule is also essential in order to expose the material most likely present despite this high level of conflict. It is also required to observe closely and simultaneously the fusion results and associated belief masses. Although these mass values should not be considered as values serving as absolute indicators for a hazard study, they do provide a great deal of information when compared in relative terms. Thus, in Figure 15c, the presence of fine fill materials is observed over the first 8 meters of thickness, from KP 10.8 to the end of the section. However, Figure 15d makes it possible to qualify this characterization with a drastic drop in confidence to the left of borehole B3, as well as a gradual decrease in confidence to the right of borehole B6. The characterization of  $\theta_1$  is therefore less reliable at KP 12 than at KP 11.3.

At some positions of the section, the union of two hypotheses is represented after the fusion process (e.g.  $\theta_1 \cup \theta_2$  below a depth of 15 m at KP 11.2). As such, it seems impossible to characterize a single material, yet an expert's observation should be sufficient to propose the most plausible solution. In this example, although borehole B4 does not extend deeper than 15 m, it seems rational that the materials present beyond are also marls and not fine fill materials. Thus, according to the positioning of the union of characterized hypotheses and thanks to expert observation, even if the geotechnical investigation does not extend to the base of the section, it is still possible to make reasonable suggestions as to the

materials present beyond the geotechnical investigation depths.

The fusion results presented for four (Figure 14) and seven core drillings (Figure 15) underscore how the section characterization differs depending on the number and position of geotechnical investigations. It may be relevant to use this fusion methodology during geotechnical acquisitions in order to ascertain where to strengthen the investigation campaign and where it would be valuable to acquire information. Specifically, the campaign should be reinforced as a priority where no material is precisely characterized (in gray, Figures 14a, 14c, 15a and 15c) and where the confidence level associated with the hypotheses is low (in blue, Figures 14b, 14d, 15b and 15d).

### B. Methodology improvements, limitations and outlook

Compared to the previous works of Dezert et al. [14], [15], several improvements have been introduced in this article. First, the bounds of the intervals associated with the FoD hypotheses were previously fixed by the user under expert interpretation. Here, the  $K$ -means clustering method makes it possible to objectify these bound values. It appears that the results obtained are consistent with the interpretation an expert could issue of the geoelectrical and seismic model inversion results (Figures 2 and 5). It also seems that the choices of bound values are not aberrant when placed into perspective with the distributions of physical values in the form of modal classes (Figures 7 and 8). However, it is important to keep in mind that an expert's interpretation is still essential to determine the number of desired clusters. In this study, three clusters have been set for both the ERT and MASW methods, but other studies may require a different number of clusters depending on the methods used. It is also the expert's responsibility to know which FoD hypotheses to associate with each cluster, based on their knowledge.

The second improvement of this work consists of having integrated data acquired by the MASW method. Although the integration of this geophysical method into the fusion methodology was made feasible, the information provided by the method for this case study is not extremely valuable compared to the characterization produced by the ERT method. On the one hand, the area covered by the seismic investigation is much smaller than that covered by the ERT (approx. 20% of the section covered), while on the other, poor complementarity exists between these two geophysical methods when it comes to discriminating various hypotheses. This finding is certainly due to the presence of limestone blocks in the coarse fill materials, associating high seismic velocities in these materials like in the limestones from the Cretaceous. One point of discussion regarding the MASW method concerns the imprecisions associated with the calculated shear wave velocities. The use of random values has been proposed in this study, yet it would be relevant to quantify these imprecisions in an alternative manner. To this point however, no research works have seemingly reached a consensus [32], [33].

Another improvement pertains to the inclusion of lithological materials extracted, yet their cohesion does not allow for

laboratory analysis (marl and limestone). This consideration is valuable since it allows the geotechnical information to be considered more in depth than the case of focusing solely on fill materials. Without this consideration, several areas after the fusion process would still be poorly characterized, e.g.  $\theta_1 \cup \theta_2$  or  $\theta_3 \cup \theta_4$ . The final improvement of this paper is the inclusion of ERT sensitivities into the mass assignment process, thus displaying a less confident characterization at the bottom and sides of the section. It would be worthwhile to integrate information of such a nature in the future for other geophysical sources (e.g. the MASW method).

Finally, since this study focuses on a hydraulic work, transversal topography and the presence of water on one side of the studied object may have an impact on the resistivity values. However, these effects are not majority and relatively homogeneous all along the acquisition profile [44]. We are aware that softwares such as BERT or pyGIMLi could provide an advantage to investigate 3D structures. Though, in this work, we focused on the fusion of data classically available for the project manager (longitudinal ERT inverted in 2D). Also, for the seismic characterization, the MASW data are 1D (almost 2D because of a longitudinal distribution). Since the objective of this work is to look for a 2D model of the dike, the inversion of ERT data in 3D would not appear consistent with the complete approach.

## VI. CONCLUSION

This paper has presented several improvements to the fusion methodology, based on the use of belief functions first proposed by Dezert et al. [14], [15], along with a detailed application of the methodology to a section of canal dike with large dimensions and complex lithology using two combination rules (Smets and PCR6). This methodology serves to take into consideration the various forms of imperfections associated with information (uncertainty, imprecision, incompleteness), as well as the spatial expressions specific to each information source. The level of contradiction (conflict) between the information sources has also been quantified. In terms of methodological improvements, in addition to integrating data from the MASW method, the sensitivity values associated with electrical resistivities have now been taken into account. An automated procedure for associating physical values with lithological materials using the  $K$ -means clustering method has been proposed; moreover, materials extracted but not analyzed in the laboratory are included in the characterization of the studied dike section.

The results obtained thanks to this fusion process make it possible to highlight the significant variability in lithology as well as the location of the fault between KP 10.6 and KP 10.8. They also provide information on the position of fine and coarse fill materials, marls and limestones. The presence of a limestone substratum up to KP 10.6 with coarse fill materials above has been successfully characterized. The fine fill materials then appear to be present from the top to a depth of approximately 10 m, i.e. from KP 10.8 to the end of the section. Below 10 m of depth, while coarse fill materials

are located at KP 10.83, the marl substratum is present in the center of the studied section. Below the fine fill material layer and beyond KP 11.6, an alternation of finer and coarser fill materials seems to be the most plausible characterization. The results displayed in the figures herein help locate areas of: high confidence level (high belief masses), doubt between two (union of two characterized hypotheses) or four materials (union of four characterized hypotheses), and conflict between the information sources (high belief masses associated with  $\emptyset$ ). The fine fill materials close to the surface are characterized with a strong confidence level, while the area where the fault is assumed to be present has not been well constrained. These results also highlight consistency between the characterization made by the ERT and the MASW methods at this specific investigation site with a very low level of conflict.

Areas of a lower confidence level could indicate where investigation should be strengthened in the future and might also be valuable for decision support in failure hazards models.

## ACKNOWLEDGMENTS

The authors are grateful to the EDF-CIH (Électricité de France - Centre Ingénierie Hydraulique) facility for their financial contribution to this work.

## REFERENCES

- [1] C. Fauchard, P. Mériaux, *Geophysical and geotechnical methods for diagnosing flood protection dikes: Guide for implementation and interpretation*, Quae, 124 pp., 2007.
- [2] W. Menke, *Geophysical data analysis: Discrete inverse theory*, Academic Press, 2018.
- [3] G. Sauvin, I. Lecomte, S. Bazin, J.S. L'Heureux, M. Vanneste, I.L. Solberg, E. Dalsegg, *Towards geophysical and geotechnical integration for quick-clay mapping in Norway*, Near Surface Geophysics, Vol. 11(6), pp. 613–624, 2013.
- [4] F. Shaaban, A. Ismail, U. Massoud, H. Mesbah, A. Lethy, A.M. Abbas, *Geotechnical assessment of ground conditions around a tilted building in Cairo-Egypt using geophysical approaches*, Journal of the Association of Arab Universities for Basic and Applied Sciences, Vol. 13(1), pp. 63–72, 2013.
- [5] G. Bièvre, P. Lacroix, L. Oxarango, D. Goutaland, G. Monnot, Y. Fargier, *Integration of geotechnical and geophysical techniques for the characterization of a small earth-filled canal dyke and the localization of water leakage*, Journal of Applied Geophysics, Vol. 139, pp. 1–15, 2017.
- [6] M. Foster, R. Fell, M. Spannagle, *A method for assessing the relative likelihood of failure of embankment dams by piping*, Canadian Geotechnical Journal, Vol. 37(5), pp. 1025–1061, 2000.
- [7] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 296 pp., 1976.
- [8] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The Annals of Mathematical Statistics, pp. 325–339, 1967.
- [9] T. Dezert, *Combinaison d'informations ponctuelles et volumiques pour le diagnostic d'ouvrages en terre soumis à des risques hydrauliques*, PhD. Thesis, Université de Nantes-Faculté des Sciences et Techniques), 277 pp, tel.archives-ouvertes.fr/tel-02459100, 2019.
- [10] T. Dezert, Y. Fargier, S.P. Lopes, P. Côte, *Geophysical and geotechnical methods for fluvial levee investigation: A review*, Engineering Geology, Vol. 260, 105206, 2019.
- [11] T. Dezert, Y. Fargier, S.P. Lopes, P. Côte, *Application of Belief Functions to Levee Assessment*, in International Conference on Belief Functions, pp. 73–76. Springer, Cham, 2018.
- [12] T. Dezert, S.P. Lopes, Y. Fargier, P. Côte, *Geophysical and geotechnical data fusion for levee assessment-interface detection with biased geophysical data*, in 24th European Meeting of Environmental and Engineering Geophysics (Vol. 2018, No. 1, pp. 1–5), European Association of Geoscientists & Engineers, 2018.

- [13] T. Dezert, S.P. Lopes, Y. Fargier, Y. *Levee Characterization by Means of Geophysical and Geotechnical Data Fusion: Improvements in Methodology*, in NSG2020 26th European Meeting of Environmental and Engineering Geophysics (Vol. 2020, No. 1, pp. 1–5), European Association of Geoscientists & Engineers, 2020.
- [14] T. Dezert, S.P. Lopes, Y. Fargier, P. Côte, *Combination of geophysical and geotechnical data using belief functions: Assessment with numerical and laboratory data*, Journal of Applied Geophysics, Vol. 170, 103824, 2019.
- [15] T. Dezert, S.P. Lopes, Y. Fargier, L. Saussaye, P. Côte, *Data fusion of in situ geophysical and geotechnical information for levee characterization*, Bulletin of Engineering Geology and the Environment, pp. 1–17, 2021.
- [16] B. Pasierb, M. Grodecki, R. Gwóźdz, *Geophysical and geotechnical approach to a landslide stability assessment: a case study*, Acta Geophysica, Vol. 67(6), pp. 1823–1834, 2019.
- [17] E. Binaghi, L. Luzzi, P. Madella, F. Pergalani, A. Rampini, *Slope instability zonation: a comparison between certainty factor and fuzzy Dempster-Shafer approaches*, Natural hazards, Vol. 17(1), pp. 77–97, 1998.
- [18] O.F. Althuwaynee, B. Pradhan, S. Lee, *Application of an evidential belief function model in landslide susceptibility mapping*, Computer Geosciences, Vol. 44, pp. 120–135, 2012.
- [19] K.A. Mogaji, H.S. Lim, K. Abdullah, *Regional prediction of ground-water potential mapping in a multifaceted geology terrain using GIS-based Dempster-Shafer model*, Arab J. Geosci., Vol. 8(5), pp. 3235–3258, 2015.
- [20] M.S. Tehrani, L. Kumar, *The application of a Dempster-Shafer based evidential belief function in flood susceptibility mapping and comparison with frequency ratio and logistic regression methods*, Environmental Earth Science, Vol. 77(13), 490, 2018.
- [21] J.A. Huisman, J. Rings, J.A. Vrugt, J. Sorg, J., H. Vereecken, *Hydraulic properties of a model dike from coupled Bayesian and multi-criteria hydrogeophysical inversion*, Journal of Hydrology, Vol. 380 (1-2), pp. 62–73, 2010.
- [22] M. Moorkamp, *Integrating Electromagnetic Data with Other Geophysical Observations for Enhanced Imaging of the Earth: A Tutorial and Review*, Surveys in Geophysics, Vol. 38(5), pp. 935–962, 2017.
- [23] M. Moorkamp, *Joint inversion of gravity and magnetotelluric data from the Ernest-Henry IOCG deposit with 30 a variation of information constraint*, First International Meeting for Applied Geoscience & Energy, pp. 1711–1715, Denver, CO, USA, 2021.
- [24] J.B. Rittgers, A. Revil, M.A. Mooney, M. Karaoulis, L. Wodajo, C.J. Hickey, *Time-lapse joint inversion of geophysical data with automatic joint constraints and dynamic attributes*, Geophysical Journal International, Vol. 207(3), pp. 1401–1419, 2016.
- [25] L. Maescot, *An introduction to electrical imaging*, Bulletin de la Société Vaudoise des Sciences Naturelles, Vol. 90(1), pp. 23–40, 2006.
- [26] M.H. Loke, I. Acworth, T. Dahlin, *A comparison of smooth and blocky inversion methods in 2D electrical imaging surveys*, Exploration geophysics, Vol. 34(3), pp. 182–187, 2003.
- [27] M.H. Loke, *Manual for RES2DINV*, Advanced Geosciences Inc., Austin, Texas, 1996.
- [28] M.H. Loke, *Tutorial 2-D and 3-D electrical Imaging Surveys*, Technical Report, 2011.
- [29] P.R. McGillivray, D.W. Oldenburg, *Methods for calculating Fréchet derivatives and sensitivities for the non-linear inverse problem: A comparative study*, Geophysical Prospecting, Vol. 38, pp. 499–524, 1990.
- [30] C.B. Park, R.D. Miller, J. Xia, J. Ivanov, *Multichannel analysis of surface waves (MASW)-active and passive methods*, Lead. Edge Vol. 26(1), pp. 60–64, 2007.
- [31] Surfseis 5.0., *Software for Multichannel Analysis of Surface Waves*, Kansas Geological Survey, 2016. [www.kgs.ku.edu/software/surfSeis/publications.html](http://www.kgs.ku.edu/software/surfSeis/publications.html)
- [32] S.C. Griffiths, B.R. Cox, E.M. Rathje, D.P. Teague, *Mapping dispersion misfit and uncertainty in Vs profiles to variability in site response estimates*, J. Geotech Geoenviron Eng, doi:10.1061/(ASCE)GT. 1943-5606.0001553, 2016.
- [33] B.R. Cox, D.P. Teague, *Layering ratios: a systematic approach to the inversion of surface wave data in the absence of a priori information*, Geophys J. Int., Vol. 207, pp. 422–438, 2016.
- [34] AFNOR NF P11-300, *Exécution des terrassements - Classification des matériaux utilisables dans la construction des remblais et des couches de forme d'infrastructures routières*, 1992.
- [35] M. Descloitres, O. Ribolzi, Y. Le Troquer, J.P. Thiébaux, *Study of water tension differences in heterogeneous sandy soils using surface ERT*, Journal of Applied Geophysics, Vol. 64(3-4), pp. 83–98, 2008.
- [36] P. Smets, *Imperfect information : Imprecision and uncertainty*, in Uncertainty Management in Information Systems, pp. 225–254. Springer, 1997.
- [37] A. Martin, C. Osswald, J. Dezert, F. Smarandache, *General Combination Rules for Qualitative and Quantitative Beliefs*, J. of Advances in Information Fusion, Vol. 3, No. 2, 2008.
- [38] P. Smets, *The combination of evidence in the transferable belief model*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12(5), pp. 447–458, 1990.
- [39] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion - Collected works*, American Research Press, Volume 3, 760 pp., 2009.
- [40] A. Likas, N. Vlassis, J.J. Verbeek, *The global k-means clustering algorithm*, Pattern recognition, Vol. 36(2), pp. 451–461, 2003.
- [41] K. Sabor, D. Jougnot, R. Guerin, L. Apffel, B. Steck, *Integrated analysis of geophysical data using a data mining approach*, in 25th European Meeting of Environmental and Engineering Geophysics, Vol. 2019, No. 1, pp. 1–5, European Association of Geoscientists & Engineers, 2019.
- [42] L. Tran, L. Duckstein, *Comparison of fuzzy numbers using a fuzzy distance measure*, Fuzzy Sets and Systems, Vol. 130(3), pp. 331–341, 2002.
- [43] F. Hollender, et al., *Characterization of site conditions (soil class, V S30, velocity profiles) for 33 stations from the French permanent accelerometric network (RAP) using surface-wave methods*, Bulletin of Earthquake Engineering, Vol. 16(6), pp. 2337–2365, 2018.
- [44] Y. Fargier, S.P. Lopes, C. Fauchard, D. François, P. Côte, *DC-Electrical Resistivity Imaging for embankment dike investigation: A 3D extended normalisation approach*, Journal of Applied Geophysics, Vol. 103, pp. 245–256, 2014.

# Application of Dezert-Smarandache Theory Rules in Fusion of Identification Information From ESM Sensors and Radars

Tadeusz Pietkiewicz

Institute of Radioelectronics, Military University of Technology

Gen. S. Kaliskiego str. 2, 00-908 Warszawa, Poland.

Email: tadeusz.pietkiewicz@wat.edu.pl

**Abstract**—This paper presents a method of fusion of identification (attribute) information provided by two types of sensors: combined primary and secondary (IFF) surveillance radars and ESM (Electronic Support Measures). In the first section, the basic taxonomy of attribute identification is adopted in accordance with the standards of STANAG 1241 ed. 5 and STANAG 1241 ed. 6 (draft). These standards provide the following basic values of the attribute identifications: FRIEND, HOSTILE, NEUTRAL, UNKNOWN and additional values: ASSUMED FRIEND and SUSPECT. The basis of the theoretical considerations is Dezert-Smarandache theory (DSMT) of inference. The paper presents and practically uses for combining identification information from different ESM sensors and radars six information fusion rules proposed by DSMT - the Proportional Conflict Redistribution rules (PCR1, PCR2, PCR3, PCR4, PCR5 and PCR6). In the paper, rules of determining attribute information by ESM sensor equipped with the data base of radar emitters are presented. It was proposed that each signal vector sent by the ESM sensor contained an extension specifying a randomized identification declaration (hypothesis). This declaration specifies the reliability of the identification information - basic belief assignment (BBA) for the identification information set. The paper also presents a model for determining the basic belief assignment for a combined primary and secondary radar. Each sensor report sent to the fusion information center contains a vector of belief mass of attribute identification. Results of the PCR rules of sensor information combining for different scenarios of radio-electronic situation (deterministic and Monte Carlo) are presented in the final part of the paper. At the end of the paper conclusions are given. They confirm the legitimacy of the use of Dezert-Smarandache theory into information fusion for primary radars, secondary radars and ESM sensors.

**Keywords:** information fusion, Dezert-Smarandache theory (DSMT) of inference, conflict redistribution rules, radar emitters recognition, electronic support measures (ESM), primary and secondary radars.

## I. INTRODUCTION

The paper is devoted to the fusion of identification information from ESM sensors and combined primary and secondary radar (IFF) using the rules of Dezert-Smarandache theory (DSMT) called proportional conflict redistribution rules. The first part of the paper presents the applied interpretation of attribute identification in accordance with the NATO STANAG 1241 standard. It should be noted that this is one of the possible interpretations of the adopted definitions. It leads to the Bayesian model of the basic belief assignment.

The identification classification method depends on the organization that operates the ESM sensors. In the paper, one assumes that the sensor identification classification is consistent with the NATO STANAG 1241 standard [1], [2]. In addition, one assumes that five identification classes are used - three primary and two secondary ones. Sensors can transmit identification information in the form of a hard decision, sometimes determined as non-randomized, or a soft decision, sometimes determined as a randomized decision. In the paper, one assumes that the sensors send identification information to the system in a randomized form, i.e. in the form of basic belief assignment on the set of identification classes. This assignment determines the sensor's belief that the detected emitter belongs to separate identification classes.

The next part of the paper presents the mathematical form of the DSMT conflict proportional redistribution rules PCR1, PCR2, PCR3, PCR4, PCR5 and PCR6 [3], [4] for two sensor inputs and PCR5 and PCR6 for three sensor inputs, assuming the Bayesian model of the basic belief assignment of hypothesis. The next two sections show how to determine the basic belief assignment for combined primary and secondary (IFF) radar and ESM sensors.

Combined primary and secondary (IFF) radars are the main source of identification information about air and maritime objects. A primary radar allows only to detect an object in a supervised area. The detection of the object is the precondition for sending a request to the object by the secondary radar (interrogator). Interpretation of the object response is dependent on the type of request. The so-called civilian modes allow only to determine whether the detected object replies to an interrogation or not. The paper presents a method for determining the basic belief assignment of airborne targets moving in observation space of combined primary and secondary (IFF) radars sensor.

ESM (electronic intelligence - electronic support measures) electronic surveillance sensors consist of passive receivers and direction-finders, which allows them to capture emitter signals coming from certain directions. In this way, the electronic recognition system can receive, among others, information on radar emitters mounted on air or maritime platforms. Reports sent from the ESM sensors include, among others, the characteristics of the intercepted signal, the emitter's azimuth

and the so-called identification information. The paper also assumes that sensors are equipped with specialized databases called the databases of emitter signal patterns, in which information about previously captured, processed, analyzed, recognized and described radar emitter signals is stored along with additional information about the type and mode of the emitter work, the platform on which these emitters can be installed, and the national or organizational affiliation of these platforms. The detected signals are the subject of an analysis procedure, which allows to determine the so-called distinctive features of the signal and then assigning this information to a specific electronic entity (already existing or created ad hoc) [5]. The basis for assigning distinctive information to an electronic entity is the azimuth angle of the incoming signal.

In the case of a high density of targets, identification information may fluctuate due to incorrect assignment of signal information to the electronic entity [6]. The impact of this negative phenomenon can be significantly reduced by an efficient estimation of the emitter positions [7]. Assuming that sensors send all reports on the tracked electronic entities to the superior operation center in the electronic recognition system, such a center (in the paper called the information fusion center (IFC) can perform the fusion function of the identification information. The fusion of identification information ensures the greater stability of this information - resistance to accidental changes in sensor decisions. Each sensor report sent to the fusion information center contains a vector of belief mass for all attribute identification values. Results of the Proportional Conflict Redistribution sensor information combining rules for selected deterministic and Monte Carlo scenarios are presented in the final part of the paper. The identification information fusion can be realized based on three basic theories - Bayesian theory of inference, Dempster-Shafer theory - called the theory of evidence and Dezert-Smarandache theory. The methods of Dezert-Smarandache information fusion are used in this paper. In addition, their effectiveness is compared with the Dempster's rule of inference.

At the end of the paper conclusions are given. They confirm the legitimacy of the use of Dezert-Smarandache theory into information fusion for primary radars, secondary radars and ESM sensors.

## II. INTERPRETATION OF ATTRIBUTE IDENTIFICATION ACCORDING TO STANAG 1241

The set of possible values of attribute identifications used by sensors can be adopted based on standardization documents of organizations that exploit these sensors [1], [2], [8]–[10]. This paper assumes a basic taxonomy of identification in accordance with the draft of STANAG 1241 ed. 6 [2]. To other similar documents one may include the following standards: STANAG 4420 and STANAG 1241 ed. 5, which provide the following basic values of the attribute identifications:

- FRIEND (F),
- HOSTILE (H),
- NEUTRAL (N),
- UNKNOWN (U).

Each of these documents contain their own definitions of the declarations.

The following definitions of these basic values of the attribute identification are used in the paper (in accordance with [2]):

- FRIEND - an allied/coalition military track, object or entity; a track, object or entity, supporting friendly forces and belonging to an allied/coalition nation or a declared or recognized friendly faction or group,
- HOSTILE - a track, object or entity whose characteristics, behavior or origin indicate that it belongs to opposing forces or poses a threat to friendly forces or their mission,
- NEUTRAL - a military or civilian track, object or entity, neither belonging to allied/coalition military forces nor to opposing military forces, whose characteristics, behavior, origin or nationality indicates that it is neither supporting nor opposing friendly forces or their mission,
- UNKNOWN - an evaluated track, object or entity, which does not meet the criteria for any other standard identity.

These standards bring additional values of the attribute identification:

- ASSUMED FRIEND,
- SUSPECT.

One should pay attention on these two recent identities contained in [1] as well as their definitions [2]:

- ASSUMED FRIEND - a track, object or entity which is assumed to be friend or neutral because of its characteristics, behavior or origin,
- SUSPECT - a track, object or entity whose characteristics, behavior or origin indicate that it potentially belongs to opposing forces or potentially poses a threat to friendly forces or their mission.

The identification definitions in [1], [2] can lead to different interpretations. This paper adopts the interpretation, the graphical form of which is shown in Figure 1.

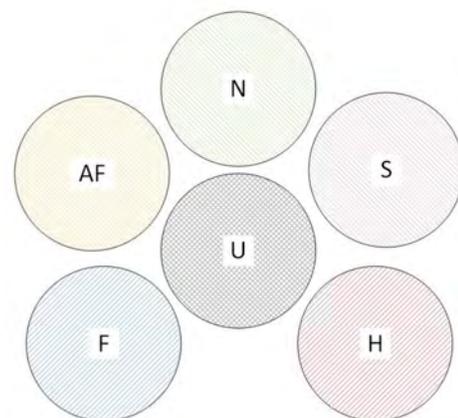


Figure 1. The interpretation of STANAG 1241 using the Venn diagram.

### III. FUSION OF INFORMATION FROM ESM SENSORS AND RADARS IN THE INFORMATION FUSION CENTER (IFC)

#### A. Diagram of the process of information fusion for two sensors in the information fusion center

In this work, it is assumed that ESM sensors send messages asynchronously to the information fusion center. These reports contain sensor decisions regarding the identification of objects emitting detected signals. The set of possible identifications is following:

$$\Theta = \{\theta_i, i = 1, \dots, 6\} \quad (1)$$

wherein the following interpretation is used:

- $\theta_1$  - FRIEND (F),
- $\theta_2$  - HOSTILE (H),
- $\theta_3$  - NEUTRAL (N),
- $\theta_4$  - ASSUMED FRIEND (AF),
- $\theta_5$  - SUSPECT (S),
- $\theta_6$  - UNKNOWN (U).

According to Figure 1, the hypotheses are mutually exclusive, i.e.

$$\theta_i \cap \theta_j = \begin{cases} \theta_i, & \text{if } i = j, \\ \emptyset, & \text{if } i \neq j. \end{cases} \quad (2)$$

Each sensor with the number  $i$  ( $i \in \mathbb{N}$ ) sends its decisions as so-called soft decisions, i.e. as BBA measure vectors (BBA - basic belief assignment)

$$\mathbf{m}_i = [m_i(\theta_1), \dots, m_i(\theta_6)]. \quad (3)$$

One should also introduce a vector of generalized BBA measures for the information fusion center

$$\mathbf{m}_F = [m_F(\theta_1), \dots, m_F(\theta_6)]. \quad (4)$$

The paper adopts the Bayesian BBA model due to the fact that this model has been adopted as valid in the STANAG 4162 standard [9]. This means that equation (5) applies in addition to (1) and (2).

$$\sum_{i=1}^6 m_F(\theta_i) = \sum_{i=1}^6 m_i(\theta_i) = 1. \quad (5)$$

The first case will be considered when two sensors send, asynchronously in one cycle, one report each containing decisions regarding the BBA related to the target. The IFC system, after receiving the report from the sensor, fuses the information contained in the two vectors: in the current generalized BBA vector  $\mathbf{m}_F = [m_F(\theta_1), \dots, m_F(\theta_6)]$ , and in the BBA vector  $\mathbf{m}_1$  from sensor 1 or in the BBA vector  $\mathbf{m}_2$  from sensor 2.

The information fusion procedure performed in the IFC is carried out in accordance with the following formula:

$$\mathbf{m}'_F = R_F(\mathbf{m}_F, \mathbf{m}_i), \quad (i = 1, \text{ or } i = 2) \quad (6)$$

wherein  $\mathbf{m}'_F$  is a vector of the generalized BBA measure determined by the  $R_F$  rule based on the previous generalized BBA measure vector  $\mathbf{m}_F$  and the new BBA measure vector  $\mathbf{m}_i$  sent by the  $i$ -th sensor. The diagram of the identification

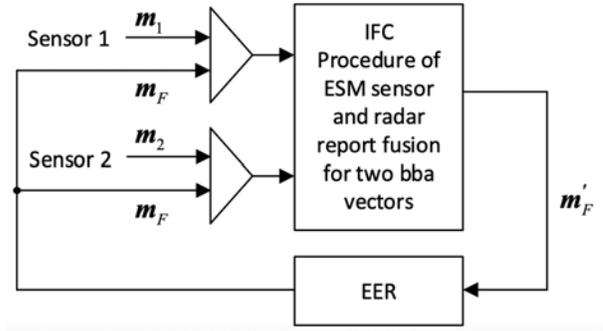


Figure 2. The diagram of the information fusion process in the information fusion center IFC for two sensors. Explanations:  $\mathbf{m}_i$  - BBA measure vector of  $i$ -th sensor,  $\mathbf{m}_F$  - generalized BBA measure vector that is a part of the electronic entity record in IFC, EER - electronic entity record in IFC database.

information fusion from the ESM sensors is shown in Figure 2.

The second case will be considered when two sensors send, asynchronously in one cycle, one report each containing decisions regarding the BBA related to the target. The IFC system waits for reports from both sensors in one cycle, using registers. Only when both registers are full, the IFC system performs a fusion of the information contained in three vectors: BBA vector  $\mathbf{m}_F = [m_F(\theta_1), \dots, m_F(\theta_6)]$ , and BBA vector  $\mathbf{m}_1$  from sensor 1, and BBA vector  $\mathbf{m}_2$  from sensor 2. It should be noted that this method has a drawback - the information stored in registers are losing credibility.

In this case, the information fusion procedure performed in the IFC is carried out in accordance with the following formula:

$$\mathbf{m}'_F = R_F(\mathbf{m}_F, \mathbf{m}_1, \mathbf{m}_2) \quad (7)$$

wherein  $\mathbf{m}'_F$  is a vector of the generalized BBA measure determined by the  $R_F$  rule based on the previous generalized BBA measure vector  $\mathbf{m}_F$  and the new BBA measure vectors  $\mathbf{m}_1$  and  $\mathbf{m}_2$  sent by both sensors. The diagram of the identification information fusion from the ESM sensors is shown in Figure 3.

In the further part of the paper, the combination rules of the BBA vector from the  $i$ -th sensor and the generalized BBA vector in the CFI are described.

#### B. The rules of combination of BBA measures vectors

This section presents formulas defining various combination rules for calculating basic belief assignments for the system shown in Figure 2 and Figure 3. The general forms are described in details in [4], [11], [12]. The information fusion rules of the DSMT are presented below with the following constraints:

- the properties of a set of hypotheses are described by the formulas (1) and (2),
- for the first scheme (Figure 2), the information fusion procedure handles two information inputs: on one input,

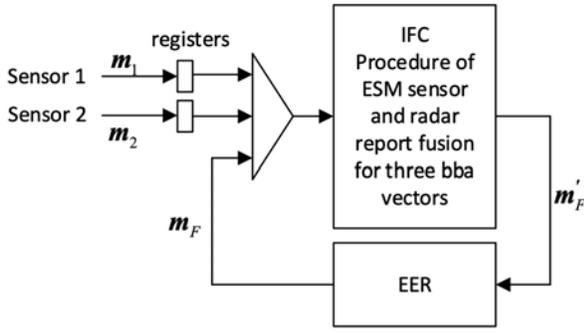


Figure 3. The diagram of the information fusion process in the information fusion center IFC for two sensors and electronic entity record from IFC database. Explanations:  $\mathbf{m}_i$  - BBA measure vector of  $i$ -th sensor,  $\mathbf{m}_F$  - generalized BBA measure vector that is a part of the electronic entity record in IFC, EER - electronic entity record in IFC database.

reports from two ESM sensors appear alternately, on the second input, electronic entity records from IFC database appear,

- for the second scheme (Figure 3), the information fusion procedure handles three information inputs: on the first input, reports from a combined primary and secondary surveillance radar appear, on the second input, reports from an ESM sensor appear, on the third input, electronic entity records from IFC database appear.

### Dempster's rule

Dempster's rule [13], [14] of the BBA measure vector  $\mathbf{m}_i$  sent by the  $i$ -th sensor and the generalized BBA measure vector  $\mathbf{m}_F$  in IFC is described for each  $\theta_i \in \Theta$  by the following formula:

$$\begin{aligned} m'_F(\theta_j) &= m_D(\theta_j) \\ &= \frac{\sum_{\substack{k=1, \dots, 6 \\ l=1, \dots, 6 \\ \theta_k \cap \theta_l = \theta_j}} m_F(\theta_k) m_i(\theta_l)}{1 - \sum_{\substack{k=1, \dots, 6 \\ l=1, \dots, 6 \\ \theta_k \cap \theta_l = \emptyset}} m_F(\theta_k) m_i(\theta_l)} \\ &= \frac{m_F(\theta_j) m_i(\theta_j)}{1 - \sum_{k=1}^6 \sum_{\substack{l=1 \\ l \neq k}}^6 m_F(\theta_k) m_i(\theta_l)} \\ &= \frac{m_{Fi}(\theta_j)}{1 - k_{Fi}} \end{aligned} \quad (8)$$

wherein the  $k_{Fi}$  degree of conflict is defined by the formula:

$$k_{Fi} = \sum_{\substack{k=1, \dots, 6 \\ l=1, \dots, 6 \\ \theta_k \cap \theta_l = \emptyset}} m_F(\theta_k) m_i(\theta_l) = \sum_{k=1}^6 \sum_{\substack{l=1 \\ l \neq k}}^6 m_F(\theta_k) m_i(\theta_l), \quad (9)$$

while

$$m_{Fi}(\theta_j) = m_F(\theta_j) m_i(\theta_j). \quad (10)$$

One could notice that

$$\sum_{k=1}^6 \sum_{l=1}^6 m_F(\theta_k) m_i(\theta_l) = 1. \quad (11)$$

$$\begin{aligned} &\sum_{\substack{k=1, \dots, 6 \\ l=1, \dots, 6 \\ \theta_k \cap \theta_l = \emptyset}} m_F(\theta_k) m_i(\theta_l) + \sum_{\substack{k=1, \dots, 6 \\ l=1, \dots, 6 \\ \theta_k \cap \theta_l \neq \emptyset}} m_F(\theta_k) m_i(\theta_l) = \\ &\sum_{k=1}^6 \sum_{\substack{l=1 \\ l \neq k}}^6 m_F(\theta_k) m_i(\theta_l) + \sum_{k=1}^6 \sum_{\substack{l=1 \\ l=k}}^6 m_F(\theta_k) m_i(\theta_l) = \\ &\sum_{k=1}^6 \sum_{\substack{l=1 \\ l \neq k}}^6 m_F(\theta_k) m_i(\theta_l) + \sum_{k=1}^6 m_F(\theta_k) m_i(\theta_k) = 1 \end{aligned} \quad (12)$$

From (12) it follows that if

$$\sum_{k=1}^6 m_F(\theta_k) m_i(\theta_k) = 1, \text{ i.e. } \sum_{k=1}^6 \sum_{\substack{l=1 \\ l \neq k}}^6 m_F(\theta_k) m_i(\theta_l) = 0 \quad (13)$$

then the degree of conflict is full.

If

$$\sum_{k=1}^6 m_F(\theta_k) m_i(\theta_k) = 0, \text{ i.e. } \sum_{k=1}^6 \sum_{\substack{l=1 \\ l \neq k}}^6 m_F(\theta_k) m_i(\theta_l) = 1 \quad (14)$$

then there is no conflict.

$m_D(\cdot)$  is Dempster-Shafer fusion result if and only if the denominator of the expression (8) is non-zero, i.e. the degree of conflict  $k_{Fi}$  is less than 1.

### The Proportional Conflict Redistribution rule PCR1

PCR1 rule is the simplest and the easiest version of proportional conflict redistribution rule. The concept of the PCR1 rule assumes the calculation of the total conflicting mass (not worrying about the partial conflicting masses). The total conflicting mass is redistributed to all non-empty sets of hypotheses proportionally with respect to their corresponding non-empty column sum of the associated mass matrix. The PCR1 rule is defined for every non-empty hypothesis in the following way:

$$\begin{aligned} m'_{Fi}(\theta_j) &= m_{PCR1}(\theta_j) \\ &= \left[ \sum_{\substack{k=1, \dots, 6 \\ l=1, \dots, 6 \\ \theta_k \cap \theta_l = \theta_j}} m_F(\theta_k) m_i(\theta_l) \right] + \frac{c_{Fi}(\theta_j)}{d_{Fi}} \cdot k_{Fi} \\ &= m_F(\theta_j) m_i(\theta_j) + \frac{c_{Fi}(\theta_j)}{d_{Fi}} \cdot k_{Fi} \\ &= m_{Fi}(\theta_j) + \frac{c_{Fi}(\theta_j)}{d_{Fi}} \cdot k_{Fi} \end{aligned} \quad (15)$$

where  $c_{Fi}(\theta_j)$  is the non-zero sum of the column corresponding to the hypotheses  $\theta_j$  in the mass matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}_F \\ \mathbf{m}_i \end{bmatrix} \quad (16)$$

specified by the formula

$$c_{Fi}(\theta_j) = m_F(\theta_j) + m_i(\theta_j) \quad (17)$$

where:

- $\mathbf{m}_i$  ( $i = 1, 2$ ) is a row vector of the basic belief assignments masses of the  $i$ -th sensor's hypotheses  $\theta_j$ ,
- $\mathbf{m}_F$  is a row vector of the basic belief assignments masses of the IFC system's hypotheses,
- $k_{Fi}$  is the degree of mass conflict specified by the formula

$$k_{Fi} = \sum_{\substack{k=1,\dots,6 \\ l=1,\dots,6 \\ \theta_k \cap \theta_l = \emptyset}} m_F(\theta_k) m_i(\theta_l) = \sum_{k=1}^6 \sum_{\substack{l=1 \\ l \neq k}}^6 m_F(\theta_k) m_i(\theta_l), \quad (18)$$

- $d_{Fi}$  is the sum of all non-zero column sums of all non-empty sets

$$d_{Fi} = \sum_{j=1}^6 [m_F(\theta_j) + m_i(\theta_j)] = \sum_{j=1}^6 c_{Fi}(\theta_j). \quad (19)$$

In our case  $d_{Fi} = 2$  because

$$\sum_{j=1}^6 m_F(\theta_j) = \sum_{j=1}^6 m_i(\theta_j) = 1. \quad (20)$$

In addition

$$m_{Fi}(\theta_j) = m_F(\theta_j) m_i(\theta_j). \quad (21)$$

### The Proportional Conflict Redistribution rule PCR2

In PCR2 rule, the total conflicting mass  $k_{Fi}$  is distributed only to the non-empty sets involved in the conflict (not to all non-empty sets) and taken proportionally with respect to their corresponding non-empty column sum.

A non-empty set  $\theta_k \in \Theta$  is considered involved in the conflict if there exists another set  $\theta_l \in \Theta$  which is neither included in  $\theta_k$  nor includes  $\theta_k$  such that  $\theta_k \cap \theta_l = \emptyset$  and  $m_{Fi}(\theta_k \cap \theta_l) > 0$ . The PCR2 rule is defined for every non-empty hypothesis  $\theta_j \in \Theta$  in the following way:

$$\begin{aligned} m'_{Fi}(\theta_j) &= m_{\text{PCR2}}(\theta_j) \\ &= \left[ \sum_{\substack{k=1,\dots,6 \\ l=1,\dots,6 \\ \theta_k \cap \theta_l = \theta_j}} m_F(\theta_k) m_i(\theta_l) \right] + C(\theta_j) \frac{c_{Fi}(\theta_j)}{e_{Fi}} \cdot k_{Fi} \\ &= m_F(\theta_j) m_i(\theta_j) + C(\theta_j) \frac{c_{Fi}(\theta_j)}{d_{Fi}} \cdot k_{Fi} \\ &= m_{Fi}(\theta_j) + C(\theta_j) \frac{c_{Fi}(\theta_j)}{d_{Fi}} \cdot k_{Fi} \end{aligned} \quad (22)$$

where

$$C(\theta_j) = \begin{cases} 1, & \text{if } \theta_j \text{ is involved in the conflict,} \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

Formula (23) can be written differently in the form (25), taking into account the definition of involvement in a conflict and formula (24) [12]:

$$m_{Fi}(\theta_j \cap \theta_k) = m_F(\theta_j) \cdot m_i(\theta_k) + m_F(\theta_k) \cdot m_i(\theta_j) \quad (24)$$

$$C(\theta_j) = \begin{cases} 1, & \text{if } \exists \theta_k \in \Theta, k \neq j \text{ such that } m_{Fi}(\theta_j \cap \theta_k) > 0, \\ 0, & \text{otherwise.} \end{cases} \quad (25)$$

$c_{Fi}(\theta_j)$  is the non-zero sum of the column corresponding to the hypotheses  $\theta_j$  in the mass matrix  $\mathbf{M}$  (16) specified by the formula

$$c_{Fi}(\theta_j) = m_F(\theta_j) + m_i(\theta_j), \quad (26)$$

where:

- $\mathbf{m}_i$  ( $i = 1, 2$ ) is a row vector of the basic belief assignments masses of the  $i$ -th sensor's hypotheses  $\theta_j$ ,
- $\mathbf{m}_F$  is a row vector of the basic belief assignments masses of the IFC system's hypotheses,
- $k_{Fi}$  is the degree of mass conflict specified by (18),
- $e_{Fi}$  is the sum of all non-zero column sums of all non-empty sets only involved in the conflict

$$\begin{aligned} e_{Fi} &= \sum_{j \in \text{CF}} [m_F(\theta_j) + m_i(\theta_j)] = \sum_{j=1}^6 \sum_{j \in \text{CF}} c_{Fi}(\theta_j) \\ &= \sum_{j=1}^6 C(\theta_j) [m_F(\theta_j) + m_i(\theta_j)] = \sum_{j=1}^6 C(\theta_j) c_{Fi}(\theta_j) \end{aligned} \quad (27)$$

where

$$\text{CF} = \{j = 1, \dots, 6 : \forall \theta_k \in \Theta | m_{Fi}(\theta_j \cap \theta_k) > 0\} \quad (28)$$

and  $m_{Fi}(\theta_j \cap \theta_k)$  is defined by (24).

In addition

$$m_{Fi}(\theta_j) = m_F(\theta_j) m_i(\theta_j) \quad (29)$$

It will be shown below that in the case of data used in numerical experiments (section VI)  $e_{Fi} = 2$ , this means that the PCR2 rule is equivalent to the PCR1 rule. The BBA vectors used there contain values less than 1, which means that

$$\forall j = 1, \dots, 6 : m_F(\theta_j) < 1 \wedge m_i(\theta_j) < 1 \quad (30)$$

It follows that each BBA vector contains at least two non-zero components, that is  $\forall j = 1, \dots, 6, \exists k = 1, \dots, 6$  with  $k \neq j$  such that

$$0 < m_F(\theta_j) < 1 \wedge 0 < m_F(\theta_k) < 1, \quad (31)$$

$$0 < m_i(\theta_j) < 1 \wedge 0 < m_i(\theta_k) < 1. \quad (32)$$

From (31) and (32) it follows that if  $m_F(\theta_j) > 0$ , then there exists at least one value  $k \neq j$  such that  $m_i(\theta_k) > 0$ , which can be written in the following form

$$\begin{aligned} \forall j = 1, \dots, 6 : 0 < m_F(\theta_j) < 1 \\ \Rightarrow \exists k = 1, \dots, 6, k \neq j : 0 < m_i(\theta_k) < 1. \end{aligned} \quad (33)$$

From (33) it follows that

$$\forall j = 1, \dots, 6 : 0 < m_F(\theta_j) < 1 \\ \Rightarrow \exists k = 1, \dots, 6, k \neq j : m_F(\theta_j)m_i(\theta_k) > 0. \quad (34)$$

The same applies:

$$\forall j = 1, \dots, 6 : 0 < m_i(\theta_j) < 1 \\ \Rightarrow \exists k = 1, \dots, 6, k \neq j : m_i(\theta_j)m_F(\theta_k) > 0. \quad (35)$$

Taking into account (34), (35) and (25) one can obtain

$$\forall j = 1, \dots, 6 : 0 < m_F(\theta_j) < 1 \Rightarrow \exists k = 1, \dots, 6, k \neq j \\ \text{such that } m_F(\theta_j)m_i(\theta_k) + m_i(\theta_j)m_F(\theta_k) > 0, \quad (36)$$

$$\forall j = 1, \dots, 6 : 0 < m_i(\theta_j) < 1 \Rightarrow \exists k = 1, \dots, 6, k \neq j \\ \text{such that } m_i(\theta_j)m_F(\theta_k) + m_F(\theta_j)m_i(\theta_k) > 0. \quad (37)$$

From (36) and (37) it follows that

$$\forall j = 1, \dots, 6 : 0 < m_F(\theta_j) < 1 \Rightarrow C(\theta_j) = 1, \quad (38)$$

$$\forall j = 1, \dots, 6 : 0 < m_i(\theta_j) < 1 \Rightarrow C(\theta_j) = 1. \quad (39)$$

This means that any hypothesis with a non-zero BBA value for any of the two sensors is involved in a conflict. From (27) it follows that

$$e_{F_i} = \sum_{j=1}^6 C(\theta_j)[m_F(\theta_j) + m_i(\theta_j)] \\ = \sum_{j=1}^6 C(\theta_j)m_F(\theta_j) + \sum_{j=1}^6 C(\theta_j)m_i(\theta_j). \quad (40)$$

Using (36), (37) and (40), the value  $e_{F_i}$  will be determined. Because, one has

$$\sum_{j=1}^6 C(\theta_j)m_F(\theta_j) = 1, \quad (41)$$

$$\sum_{j=1}^6 C(\theta_j)m_i(\theta_j) = 1, \quad (42)$$

we get

$$e_{F_i} = \sum_{j=1}^6 C(\theta_j)m_F(\theta_j) + \sum_{j=1}^6 C(\theta_j)m_i(\theta_j) = 2. \quad (43)$$

Considering (43), it can be said that in this case the PCR2 rule is equivalent to the PCR1 rule. For this reason, the results of the PCR2 rule are not presented in section VI, as they would be identical to the results of the PCR1 rule because we work only with Bayesian BBAs in this application.

### The Proportional Conflict Redistribution rule PCR3

In PCR3 rule, one distributes the partial conflicting masses, instead of the total conflicting mass  $k_{F_i}$ , to the non-empty sets involved in the partial conflict. If an intersection is empty, for instance  $\theta_k \cap \theta_l = \emptyset$ , then the mass  $m(\theta_k \cap \theta_l)$  of the

partial conflict is transferred to the non-empty sets  $\theta_k$  and  $\theta_l$  proportionally with respect to the non-zero sum of masses assigned to  $\theta_k$  and respectively to  $\theta_l$  by the BBAs  $m_F(\cdot)$  and  $m_i(\cdot)$ . The PCR3 rule works if at least one set between  $\theta_k$  and  $\theta_l$  is non-empty and its column sum is non-zero.

The PCR3 rule is defined for every non-empty hypothesis  $\theta_j \in \Theta$  in the following way:

$$m'_F(\theta_j) = m_{\text{PCR3}}(\theta_j) \\ = [ \sum_{\substack{k=1, \dots, 6 \\ l=1, \dots, 6 \\ \theta_k \cap \theta_l = \theta_j}} m_F(\theta_k)m_i(\theta_l) ] \\ + [c_{F_i}(\theta_j) \sum_{\substack{k=1, \dots, 6 \\ \theta_k \cap \theta_l = \emptyset}} S_{F_i}^{\text{PCR3}}(\theta_j, \theta_k)] \\ = m_{F_i}(\theta_j) + [c_{F_i}(\theta_j) \sum_{\substack{k=1, \dots, 6 \\ k \neq j}} S_{F_i}^{\text{PCR3}}(\theta_j, \theta_k)] \quad (44)$$

where

$$S_{F_i}^{\text{PCR3}}(\theta_j, \theta_k) = \begin{cases} 0, & \text{for } c_{F_i}(\theta_j) + c_{F_i}(\theta_k) = 0, \\ \frac{m_F(\theta_k)m_i(\theta_j) + m_F(\theta_j)m_i(\theta_k)}{c_{F_i}(\theta_j) + c_{F_i}(\theta_k)}, & \text{otherwise.} \end{cases} \quad (45)$$

$c_{F_i}(\theta_j)$  is the non-zero sum of the column corresponding to the hypotheses  $\theta_j$  in the mass matrix  $\mathbf{M}$  (16) specified by the formula

$$c_{F_i}(\theta_j) = m_F(\theta_j) + m_i(\theta_j), \quad (46)$$

### The Proportional Conflict Redistribution rule PCR4

The PCR4 rule redistributes the partial conflicting masses only to the sets involved in the partial conflict in proportion to the non-zero mass sum assigned to  $\theta_k$  and  $\theta_l$  by the conjunction rule according to the following formula:

$$m'_F(\theta_j) = m_{\text{PCR4}}(\theta_j) \\ = m_{F_i}(\theta_j) + [m_{F_i}(\theta_j) \cdot \sum_{\substack{k=1, \dots, 6 \\ \theta_k \cap \theta_j = \emptyset}} S_{F_i}^{\text{PCR4}}(\theta_j, \theta_k)] \\ = m_{F_i}(\theta_j) + [m_{F_i}(\theta_j) \cdot \sum_{\substack{k=1, \dots, 6 \\ k \neq j}} S_{F_i}^{\text{PCR4}}(\theta_j, \theta_k)], \quad (47)$$

where

$$S_{F_i}^{\text{PCR4}}(\theta_j, \theta_k) = \begin{cases} 0, & \text{for } c_1 = 0, \\ \frac{m_{F_i}(\theta_j \cap \theta_k)}{m_{F_i}(\theta_j) + m_{F_i}(\theta_k)}, & \text{for } c_1 \neq 0, \text{ and } c_2 \neq 0, \\ \frac{m_{F_i}(\theta_j \cap \theta_k)}{c_{F_i}(\theta_j) + c_{F_i}(\theta_k)}, & \text{for } c_1 \neq 0, \text{ and } c_2 = 0, \end{cases} \quad (48)$$

where  $c_1 \triangleq c_{F_i}(\theta_j) + c_{F_i}(\theta_k)$  and  $c_2 \triangleq m_{F_i}(\theta_j)m_{F_i}(\theta_k)$ , and wherein

$$m_{F_i}(\theta_j \cap \theta_k) = m_F(\theta_k)m_i(\theta_j) + m_F(\theta_j)m_i(\theta_k) \quad (49)$$

$$m_{F_i}(\theta_j) = m_F(\theta_j)m_i(\theta_j) \quad (50)$$

$$c_{F_i}(\theta_j) = m_F(\theta_j) + m_i(\theta_j) \quad (51)$$

If at least one of BBAs  $m_F(\cdot)$  or  $m_i(\cdot)$  is zero, the fraction is discarded and the mass  $m_{F_i}(\theta_j \cap \theta_k)$  is transferred to  $\theta_j$  and  $\theta_k$  proportionally with respect to their non-zero column sum of masses  $c_{F_i}(\theta_j)$ .

### The Proportional Conflict Redistribution rule PCR5 for two BBAs (two sources)

Similarly to PCR2–PCR4 rules, PCR5 redistributes the partial conflicting mass to the hypothesis involved in the partial conflict. PCR5 provides the most mathematically precise [4], [11], [12] redistribution of conflicting mass to non-empty sets in accordance with the logic of the conjunctive rule. However, it is more difficult to implementation. The PCR5 rule is defined for every non-empty hypothesis  $\theta_j \in \Theta$  in the following way:

$$\begin{aligned} m'_F(\theta_j) &= m_{\text{PCR5}}(\theta_j) \\ &= m_{F_i}(\theta_j) + \sum_{\substack{k=1,\dots,6 \\ \theta_k \cap \theta_j = \emptyset}} S_{F_i}^{\text{PCR5}}(\theta_j, \theta_k) \\ &= m_{F_i}(\theta_j) + \sum_{\substack{k=1,\dots,6 \\ k \neq j}} S_{F_i}^{\text{PCR5}}(\theta_j, \theta_k), \end{aligned} \quad (52)$$

where

$$S_{F_i}^{\text{PCR5}}(\theta_j, \theta_k) = \begin{cases} 0, & \text{for } c_3 = 0 \text{ or } c_4 = 0, \\ \frac{m_F(\theta_j)^2 m_i(\theta_k)}{m_F(\theta_j) + m_i(\theta_k)}, & \text{for } c_3 \neq 0 \text{ and } c_4 \neq 0, \end{cases} \quad (53)$$

where  $c_3 \triangleq m_F(\theta_j) + m_i(\theta_k)$  and  $c_4 \triangleq m_i(\theta_j) + m_F(\theta_k)$ , and wherein

$$m_{F_i}(\theta_j) = m_F(\theta_j) m_i(\theta_j). \quad (54)$$

In the formula (52), the component  $S_{F_i}^{\text{PCR5}}$  is equal to zero if both denominators are equal to zero. In the formula (53), if a denominator is zero, then component is discarded.

### The Proportional Conflict Redistribution rules PCR5 and PCR6 for three BBAs (three sources)

In [4] improved proportional conflict redistribution rules of combination of basic belief assignments PCR6, PCR5<sup>+</sup> and PCR6<sup>+</sup> are presented. The authors point out that these rules should be applied if and only if we are to combine more than two BBAs. If we have only two BBAs to combine ( $s = 2$ ) we always get  $m_{\text{PCR5}} = m_{\text{PCR5}^+} = m_{\text{PCR6}} = m_{\text{PCR6}^+}$  because in this case the PCR5, PCR5<sup>+</sup>, PCR6, and PCR6<sup>+</sup> rules coincide. Below are the formulas that define the PCR5 and PCR6 rules for 3 BBAs.

$$m'_F(\theta_j) = m_{\text{PCR5}}(\theta_j) = \frac{m''(\theta_j)}{\sum_{i=1}^6 m''(\theta_i)} \quad (55)$$

wherein

$$\begin{aligned} m''(\theta_j) &= m_{F_{12}}(\theta_j) \\ &+ \sum_{\substack{k=1,\dots,6 \\ l=1,\dots,6 \\ \theta_k \cap \theta_l \cap \theta_j = \emptyset}} S_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k, \theta_l) \\ &+ \sum_{\substack{k=1,\dots,6 \\ \theta_k \cap \theta_j = \emptyset}} S1_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) + \sum_{\substack{k=1,\dots,6 \\ \theta_k \cap \theta_j = \emptyset}} S2_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) \\ &= m_{F_{12}}(\theta_j) \\ &+ \sum_{\substack{k=1,\dots,6 \\ k \neq j}} \sum_{\substack{l=1,\dots,6 \\ l \neq j \wedge l \neq k}} S_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k, \theta_l) \\ &+ \sum_{\substack{k=1,\dots,6 \\ k \neq j}} S1_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) + \sum_{\substack{k=1,\dots,6 \\ k \neq j}} S2_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) \\ &= m_{F_{12}}(\theta_j) \\ &+ \sum_{\substack{k=1,\dots,6 \\ k \neq j}} \left[ \sum_{\substack{l=1,\dots,6 \\ l \neq j \wedge l \neq k}} S_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k, \theta_l) \right. \\ &\quad \left. + S1_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) + S2_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) \right], \end{aligned} \quad (56)$$

with

$$\begin{aligned} S_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k, \theta_l) &= \frac{m_F(\theta_j)^2 m_1(\theta_k) m_2(\theta_l)}{m_F(\theta_j) + m_1(\theta_k) + m_2(\theta_l)} \\ &+ \frac{m_F(\theta_l) m_1(\theta_j)^2 m_2(\theta_k)}{m_F(\theta_l) + m_1(\theta_j) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_l) m_2(\theta_j)^2}{m_F(\theta_k) + m_1(\theta_l) + m_2(\theta_j)}, \end{aligned} \quad (57)$$

$$\begin{aligned} S1_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) &= \frac{m_F(\theta_j)^2 m_1(\theta_k) m_2(\theta_k)}{m_F(\theta_j) + m_1(\theta_k) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_j)^2 m_2(\theta_k)}{m_F(\theta_k) + m_1(\theta_j) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_k) m_2(\theta_j)^2}{m_F(\theta_k) + m_1(\theta_k) + m_2(\theta_j)}, \end{aligned} \quad (58)$$

$$\begin{aligned} S2_{F_{12}}^{\text{PCR5}}(\theta_j, \theta_k) &= \frac{m_F(\theta_j)^2 m_1(\theta_j)^2 m_2(\theta_l)}{m_F(\theta_j) + m_1(\theta_j) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_j)^2 m_2(\theta_j)^2}{m_F(\theta_k) + m_1(\theta_j) + m_2(\theta_j)} \\ &+ \frac{m_F(\theta_j)^2 m_1(\theta_k) m_2(\theta_j)^2}{m_F(\theta_j) + m_1(\theta_k) + m_2(\theta_j)}, \end{aligned} \quad (59)$$

and

$$m_{F_{12}}(\theta_j) = m_F(\theta_j) m_1(\theta_j) m_2(\theta_j). \quad (60)$$

In the formulas (57)–(59), if a denominator is zero, then component is discarded.

The quotient in the formula (55) ensures the normalization of the BBA vector  $\mathbf{m}'_F$ , which ensures that

$$\sum_{j=1}^6 m'_F(\theta_j) = \sum_{j=1}^6 m_{\text{PCR5}}(\theta_j) = 1.$$

The PCR6 rule for three BBAs (three sources) is defined for every non-empty hypothesis  $\theta_j \in \Theta$  in the following way:

$$m'_F(\theta_j) = m_{\text{PCR6}}(\theta_j) = \frac{m''(\theta_j)}{\sum_{i=1}^6 m''\theta_i} \quad (61)$$

wherein

$$\begin{aligned} m''(\theta_j) &= m_{F12}(\theta_j) \\ &+ \sum_{\substack{k=1,\dots,6 \\ l=1,\dots,6 \\ \theta_k \cap \theta_l \cap \theta_j = \emptyset}} S_{F12}^{\text{PCR6}}(\theta_j, \theta_k, \theta_l) \\ &+ \sum_{\substack{k=1,\dots,6 \\ \theta_k \cap \theta_j = \emptyset}} S1_{F12}^{\text{PCR6}}(\theta_j, \theta_k) \\ &+ \sum_{\substack{k=1,\dots,6 \\ \theta_k \cap \theta_j = \emptyset}} S2_{F12}^{\text{PCR6}}(\theta_j, \theta_k) \\ &= m_{F12}(\theta_j) \\ &+ \sum_{\substack{k=1,\dots,6 \\ k \neq j}} \sum_{\substack{l=1,\dots,6 \\ l \neq j \wedge l \neq k}} S_{F12}^{\text{PCR6}}(\theta_j, \theta_k, \theta_l) \\ &+ \sum_{\substack{k=1,\dots,6 \\ k \neq j}} S1_{F12}^{\text{PCR6}}(\theta_j, \theta_k) \\ &+ \sum_{\substack{k=1,\dots,6 \\ k \neq j}} S2_{F12}^{\text{PCR6}}(\theta_j, \theta_k) \\ &= m_{F12}(\theta_j) \\ &+ \sum_{\substack{k=1,\dots,6 \\ k \neq j}} \left[ \sum_{\substack{l=1,\dots,6 \\ l \neq j \wedge l \neq k}} S_{F12}^{\text{PCR6}}(\theta_j, \theta_k, \theta_l) \right. \\ &\quad \left. + S1_{F12}^{\text{PCR6}}(\theta_j, \theta_k) + S2_{F12}^{\text{PCR6}}(\theta_j, \theta_k) \right], \quad (62) \end{aligned}$$

with

$$\begin{aligned} S_{F12}^{\text{PCR6}}(\theta_j, \theta_k, \theta_l) &= \frac{m_F(\theta_j)^2 m_1(\theta_k) m_2(\theta_l)}{m_F(\theta_j) + m_1(\theta_k) + m_2(\theta_l)} \\ &+ \frac{m_F(\theta_l) m_1(\theta_j)^2 m_2(\theta_k)}{m_F(\theta_l) + m_1(\theta_j) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_l) m_2(\theta_j)^2}{m_F(\theta_k) + m_1(\theta_l) + m_2(\theta_j)}, \quad (63) \end{aligned}$$

$$\begin{aligned} S1_{F12}^{\text{PCR6}}(\theta_j, \theta_k) &= \frac{m_F(\theta_j)^2 m_1(\theta_k) m_2(\theta_k)}{m_F(\theta_j) + m_1(\theta_k) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_j)^2 m_2(\theta_k)}{m_F(\theta_k) + m_1(\theta_j) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_k) m_2(\theta_j)^2}{m_F(\theta_k) + m_1(\theta_k) + m_2(\theta_j)}, \quad (64) \end{aligned}$$

$$\begin{aligned} S2_{F12}^{\text{PCR6}}(\theta_j, \theta_k) &= \frac{m_F(\theta_j)^2 m_1(\theta_j) m_2(\theta_k)}{m_F(\theta_j) + m_1(\theta_j) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_j) m_1(\theta_j)^2 m_2(\theta_k)}{m_F(\theta_j) + m_1(\theta_j) + m_2(\theta_k)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_j)^2 m_2(\theta_j)}{m_F(\theta_k) + m_1(\theta_j) + m_2(\theta_j)} \\ &+ \frac{m_F(\theta_k) m_1(\theta_j) m_2(\theta_j)^2}{m_F(\theta_k) + m_1(\theta_j) + m_2(\theta_j)} \\ &+ \frac{m_F(\theta_j)^2 m_1(\theta_k) m_2(\theta_j)}{m_F(\theta_j) + m_1(\theta_k) + m_2(\theta_j)}, \\ &+ \frac{m_F(\theta_j) m_1(\theta_k) m_2(\theta_j)^2}{m_F(\theta_j) + m_1(\theta_k) + m_2(\theta_j)}, \quad (65) \end{aligned}$$

and

$$m_{F12}(\theta_j) = m_F(\theta_j) m_1(\theta_j) m_2(\theta_j). \quad (66)$$

In the formulas (63)–(65), if a denominator is zero, then component is discarded. The quotient in the formula (61) ensures the normalization of the BBA vector  $\mathbf{m}'_F$ , which ensures that

$$\sum_{j=1}^6 m'_F(\theta_j) = \sum_{j=1}^6 m_{\text{PCR6}}(\theta_j) = 1.$$

Comparing the two fusion schemes (Figures 2 and 3), it should be noted that sequential and global information fusion generally produces different results [4], i.e.

$$\begin{aligned} \text{PCR5}(m_F, m_1, m_2) &\neq \text{PCR5}(\text{PCR5}(m_F, m_1), m_2) \\ &\neq \text{PCR5}(\text{PCR5}(m_F, m_2), m_1). \quad (67) \end{aligned}$$

In addition, the article experimentally verified the theorem on the inequality of the results of both PCR5 and PCR6 rules for three BBAs (three sources) presented in [4]:

$$\text{PCR5}(m_F, m_1, m_2) \neq \text{PCR6}(m_F, m_1, m_2). \quad (68)$$

#### IV. BASIC BELIEF ASSIGNMENT FOR COMBINED PRIMARY AND SECONDARY SURVEILLANCE RADARS

The paper assumes that the analyzed radar sensor consists of two radars: primary and secondary. Therefore, the probability of correct detection and correct identification of a target is expressed by the following formula:

$$P_{pi} = P_d P_{IFF}. \quad (69)$$

where  $P_d$  is the probability of correct detection of target by a primary radar, and  $P_{IFF}$  is the probability of correct reply for interrogation. If a target has been detected by the primary radar and there is a lack of proper identification by the secondary radar one can assume that the target has a value of attribute identification of UNKNOWN - U. So, one can write the following relation:

$$m(U) = P_d(1 - P_{IFF}), \quad (70)$$

where  $m(U)$  is the mass of probability for a value of UNKNOWN identification attribute.

A method of calculating the probabilities  $P_d$  and  $P_{IFF}$  is presented in [5], [15], [16].

The way of allocation of the remaining mass of probability  $(1 - m(U))$  will be described in this section. One assumes there that every simulated target should have a base value of attribute identification from the set

$$\mathbf{Z}_{BI} = \{N_B, F_B, H_B\}, \quad (71)$$

- $N_B$  – base NEUTRAL identity,
- $F_B$  – base FRIEND identity,
- $H_B$  – base HOSTILE identity.

STANAG 1241 introduces in addition to the basic set of attribute identification values also secondary (additional) attribute identification values: SUSPECT (S) and ASSUMED FRIEND (AF). According to Fig. 1 one can introduce a table of possible attribute values transitions between set (71) and the set of secondary attribute identification values:

$$\mathbf{Z}_{SI} = \{N_S, F_S, H_S, AF, S\}, \quad (72)$$

Belief mass values contained in the Table I determine how the mass of the base belief assignment has been transformed into the mass of the secondary belief assignment. They can be estimated as empirical frequencies based on recorded archive events.

Table I  
TRANSFORMATION OF THE BASE BELIEF ASSIGNMENT MASS INTO THE SECONDARY BELIEF ASSIGNMENT MASS

Base identification $\rightarrow$	$F_B$	$N_B$	$H_B$
$F_S$	$m(F_S F_B)$	0	0
$N_S$	0	$m(N_S N_B)$	0
$H_S$	0	0	$m(H_S H_B)$
$AF$	$m(AF F_B)$	$m(AF N_B)$	0
$S$	0	$m(S N_B)$	$m(S H_B)$

Of course, have the normalization conditions satisfied:  $\sum_{x \in \mathbf{Z}_{SI}} m(x|F_B) = 1$ ,  $\sum_{x \in \mathbf{Z}_{SI}} m(x|N_B) = 1$ , and  $\sum_{x \in \mathbf{Z}_{SI}} m(x|H_B) = 1$ .

The final values of the belief mass of secondary attribute identification values are calculated according to the formulas:

- 1) For a target with the FRIEND base value of an attribute identification

$$\begin{aligned} m(U) &= P_d(1 - P_{IFF}), \\ m(AF) &= m(AF|F_B)(1 - m(U)), \\ m(F_S) &= m(F_S|F_B)(1 - m(U)). \end{aligned}$$

- 2) For a target with the NEUTRAL base value of an attribute identification

$$\begin{aligned} m(U) &= P_d(1 - P_{IFF}), \\ m(AF) &= m(AF|N_B)(1 - m(U)), \\ m(S) &= m(S|N_B)(1 - m(U)), \\ m(N_S) &= (1 - m(AF|N_B) - m(S|N_B))(1 - m(U)). \end{aligned}$$

- 3) For a target with the HOSTILE base value of an attribute identification

$$\begin{aligned} m(U) &= P_d(1 - P_{IFF}), \\ m(H_S) &= m(H_S|H_B)(1 - m(U)), \\ m(S) &= m(S|H_B)(1 - m(U)). \end{aligned}$$

Other final values of the belief mass of secondary attribute identification values are equal zero.

## V. BASIC BELIEF ASSIGNMENT FOR ESM SENSORS

An ESM sensor is a passive sensor that captures incoming electromagnetic signals generated first of all by radar emitters mounted on air or maritime platforms. This sensor recognizes radar signals determining values of their distinctive features. In this paper we will not deal with methods of radar signals recognizing in details. However, we will use information about these methods to identify platforms generating the signals according to STANAG 1241 - NATO Standardization Agreement and Dezert-Smarandache theory. As previously it was stated, we are interested in three basic values of identification: friend, hostile and neutral, and two secondary values: suspicious and assumed friendly. In addition, we will assume that in some situations it is not possible to determine the identity of the emitter carrier platform. To clarify this issue, we should briefly describe the method of determining the identification of the emitter carrier platform that generated the captured signal. The sensor recognition system is equipped with a database that can be divided into three components: a platform database, an emitter list and a geopolitical list [10]. The platform database (**PDB**) contains information about platforms that can be met in the area of interest along with their equipment with emitters. The emitter name list (**ENL**) includes all emitters corresponding to each platform of the **PDB** and contains the values of the signal distinctive features for each emitter. The values of distinctive features are the basis for the procedure of recognizing a captured signal. The geopolitical list (**GPL**) provides the allegiance of various countries and platforms and allows to identify them in accordance with STANAG 1241.

The algorithm of signal recognition is realized in two stages:

- 1) Verification at the level of signal quality features. The second stage is executed after a positive assessment of the conformity of quality features.
- 2) The signal recognition procedure determines the distances between the distinctive features of the recognized signal and the distinctive features of all pattern signals stored within the emitter list.

Let us introduce the following notation:

$\mathbf{x}_s$  - vector of distinctive features of the recognized signal,  
 $\mathbf{x}_i$  - vector of distinctive features of  $i$ -th pattern signal ( $i$  – the number of the pattern signal,  $i \in \{1, \dots, M\}$ ),  
 $d_{s,i} = d(\mathbf{x}_s, \mathbf{x}_i)$  - the distance between the distinctive features vector of the recognized signal and the distinctive features vector of  $i$ -th pattern signal; the distance  $d_{s,i}$  is the Mahalanobis distance taking into account the correlations of the distinctive features.

The signal recognition classifier compares the distance  $d(\mathbf{x}_s, \mathbf{x}_i)$  with the acceptable positive distance of the classification  $\delta$ . The distance  $\delta$  is the limit that we will interpret as a boundary of emitter pattern recognition. We will divide the set of pattern signals into two subsets: the patterns satisfying the positive classification condition in relation to the recognized signal  $s$  –  $\mathbf{D}_s^+$  and the patterns that do not satisfy the positive classification condition –  $\mathbf{D}_s^-$ . The formal definition is as follows:

$$\mathbf{D}_s^+ = \{i \in \{1, \dots, M\} | d_{s,i} \leq \delta\} \quad (73)$$

$$\mathbf{D}_s^- = \{i \in \{1, \dots, M\} | d_{s,i} > \delta\} \quad (74)$$

In the paper we propose the following method of determining the basic belief assignment on a set of pattern signals, which is related to the distance between a signal and a pattern in the distinctive features space:

$$m_s(i) = e^{-d(\mathbf{x}_s, \mathbf{x}_i)} \quad (75)$$

As one can see from the formula (75) if  $d(\mathbf{x}_s, \mathbf{x}_i) = 0$  then  $m_s(i) = 1$ , whereas if  $d(\mathbf{x}_s, \mathbf{x}_i) > 0$  then  $0 < m_s(i) < 1$ . The above measure is not normalized, hence we will normalize it

$$\tilde{m}_s(i) = m_s(i) / \sum_{i=1}^M m_s(i) \quad (76)$$

The sum of the measures assigned to all the emitters, whose distinctive features lie outside the limit  $\delta$ , will be treated as a measure assigned to the base hypothesis “unknown” (U)

$$\tilde{m}_s(U) = \sum_{i \in \mathbf{D}_s^-} \tilde{m}_s(i) \quad (77)$$

To determine the belief measure of other base hypotheses ( $H, F, N$ ) and secondary hypotheses ( $AF$  and  $S$ ), we should introduce formal definitions of sets contained in the sensor database and used for recognition of captured signals. As it was mentioned above the set of all the necessary data for platform identification can be divided into three sets: **PDB** – a platform database, **ENL** – an emitter name list and **GPL** – a geopolitical list:

- **PDB** – the platform database contains information about all platforms observed in the area of interest, including information on all emitters mounted on each platform; we assume that one platform can have many emitters and the same type of emitters can be installed on many platforms; the **PDB** contains also information on the national affiliation of each platform,
- **ENL** – the emitter name list is a set of information about all recognized emitters in the area of interest; this set contains the mean values of the distinctive features of emitter signals (so-called signal patterns) and their standard deviations,
- **GPL** – the geopolitical list contains base values of identification attributes ( $H, F, N$ ) assigned to the various countries.

We will also introduce additional notations used in this paper:

- **PDBL** – the list of platform numbers that are stored in the **PDB**,
- **PL**( $i$ ) – the set of numbers of platforms which have the emitter with number “ $i$ ”,
- **IPL**( $j$ ) – the base identification attribute of the platform with number “ $j$ ” determined on the basis of the information contained in **PDB** and **ENL** ( $IPL(j) \in \{F, H, N\}$ ).

The set of signal patterns satisfying the positive classification condition in relation to the recognized signal  $s$  denoted as  $\mathbf{D}_s^+$  can be divided into disjunctive subsets according to the values of the carrier platform identification features:

$$\mathbf{D}_s^+ = \mathbf{D}_s^{+F} \geq \mathbf{D}_s^{+H} \geq \mathbf{D}_s^{+N} \geq \mathbf{D}_s^{+AF} \geq \mathbf{D}_s^{+S}, \quad (78)$$

$$\mathbf{D}_s^{+k} \cap \mathbf{D}_s^{+l} = \emptyset, k \neq l, k, l \in \{F, H, N, AF, S\}. \quad (79)$$

Each subset of the set  $\mathbf{D}_s^+$  for the base identification is defined as follows:

$$\mathbf{D}_s^{+F} = \{i \in \mathbf{D}_s^+ | \forall j \in \mathbf{PL}(i), IPL(j) = F\}, \quad (80)$$

$$\mathbf{D}_s^{+H} = \{i \in \mathbf{D}_s^+ | \forall j \in \mathbf{PL}(i), IPL(j) = H\}, \quad (81)$$

$$\mathbf{D}_s^{+N} = \{i \in \mathbf{D}_s^+ | \forall j \in \mathbf{PL}(i), IPL(j) = N\}. \quad (82)$$

In a similar way, one can define subsets of the set  $\mathbf{D}_s^+$  for the secondary identification ( $AF, S$ ):

$$\mathbf{D}_s^{+AF} = \{i \in \mathbf{D}_s^+ | \exists j \in \mathbf{PL}(i), IPL(j) = F \wedge \exists j \in \mathbf{PL}(i), IPL(j) = N\}, \quad (83)$$

$$\mathbf{D}_s^{+S} = \{i \in \mathbf{D}_s^+ | \exists j \in \mathbf{PL}(i), IPL(j) = H \wedge \exists j \in \mathbf{PL}(i), IPL(j) = N\}. \quad (84)$$

One can notice we assume in this paper that no emitter type can be installed simultaneously on platforms with identifications  $F$  and  $H$ :

$$\{i \in \mathbf{D}_s^+ | \exists j \in \mathbf{PL}(i), IPL(j) = F \wedge \exists j \in \mathbf{PL}(i), IPL(j) = H\} = \emptyset. \quad (85)$$

It should be emphasized that the method presented here is different than in [6], [17]. These papers assume that ESM sensors can only generate basic declarations with attribute values FRIEND, HOSTILE and NEUTRAL but in this paper, we assume, that ESM sensors can generate declarations from an extended set of attribute values (additionally ASSUME FRIEND, SUSPECT and UNKNOWN).

## VI. NUMERICAL EXPERIMENTS OF FUSION OF IDENTIFICATION IN INFORMATION FROM ESM SENSORS

### A. Simulation scenarios

The paper [6] presents a typical simulation scenario for testing the identification information fusion. The authors formulated several requirements that should be met by such a scenario. It should:

- 1) adequately represent the known ground truth of the emitter identification,

- 2) include sufficient numbers of incorrect associations to be realistic and to test the robustness of the rules to temporary incorrect sensor decisions,
- 3) provide only partial knowledge about the ESM sensor declarations, and thus contain uncertainty,
- 4) allow to show stability in case of countermeasures,
- 5) allow to switch identification when the ground truth changes.

The authors [6] propose the following parameters of the scenario:

- 1) ground truth of identification is FRIEND ( $F$ ) for the first 50 iterations of the scenario and HOSTILE ( $H$ ) for the last 50 iterations,
- 2) the number of correct associations is 80% of all iterations, the number of incorrect associations caused by countermeasures is 20% of all iterations in a randomly selected moments of time,
- 3) ESM sensor declarations have a mass of 0.7 for the most credible identification and 0.3 for the identification of UNKNOWN ( $U$ ).

The assumption 5) is not considered in this paper, assuming that the real object does not change its real identity while performing the mission. Therefore, assumption 1) regarding scenario parameters becomes obsolete. The following assumptions concerning the parameters of the scenario are made in this paper:

- 1) the real value of identification is constant in each scenario and is equal to FRIEND ( $F$ ) - in the scenarios 1, 2 & 5, and HOSTILE ( $H$ ) - in the scenarios 3, 4 & 6;
- 2) the above declarations are transmitted by sensor number 1 with the real identification mass equal to 0.7 and the mass of complementary identification (UNKNOWN) equal to 0.3;
- 3) the second sensor shall transmit its declarations in accordance with the tables II and III for the scenarios 1 and 2 and with the tables IV and V for the scenarios 3 and 4.

Table II

BELIEF MASSES FOR THE SECOND SENSOR FOR THE SCENARIOS 1 AND 5.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0.6	0.1	0	0.2	0	0.1
Incorrect identif. (20% of events)	0	0.1	0.6	0	0.2	0.1

Table III

BELIEF MASSES FOR THE SECOND SENSOR FOR THE SCENARIO 2.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0.7	0.1	0	0.1	0	0.1
Incorrect identif. (20% of events)	0	0.1	0.7	0	0.1	0.1

One should note that scenario 2 differs from scenario 1 with a greater belief mass assigned to incorrect identification

Table IV

BELIEF MASSES FOR THE SECOND SENSOR FOR THE SCENARIO 3.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0	0.1	0.6	0	0.2	0.1
Incorrect identif. (20% of events)	0.6	0.1	0	0.2	0	0.1

Table V

BELIEF MASSES FOR THE SECOND SENSOR FOR THE SCENARIOS 4 AND 6.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0	0.1	0.7	0	0.1	0.1
Incorrect identif. (20% of events)	0.7	0.1	0	0.1	0	0.1

of the recognized emitter. The scenarios 3 and 4 are similarly different.

Scenarios 1–6 for the sensor 1 have been presented in Figures 4 and 5.

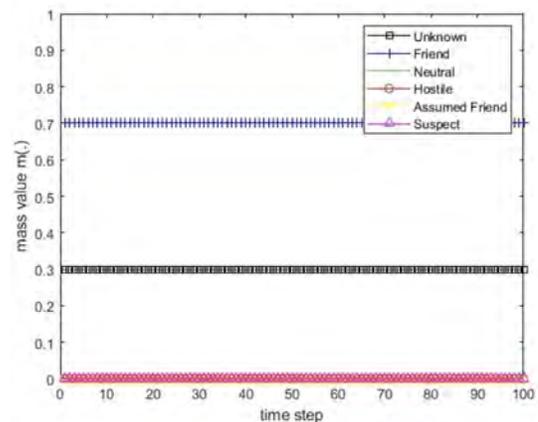


Figure 4. The course of scenarios number 1, 2 and 5 for sensor 1.

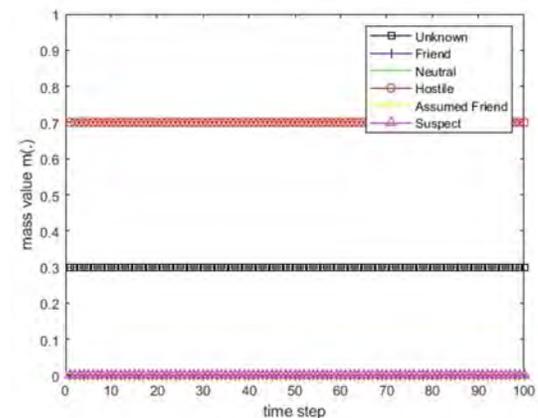


Figure 5. The course of scenarios number 3, 4 and 6 for sensor 1.

All deterministic scenarios for the sensor 2 have been presented in Figures 6–9.

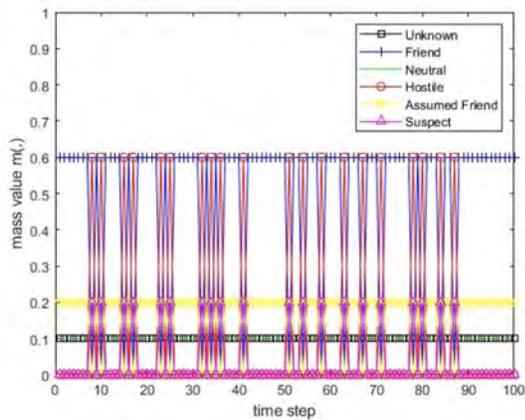


Figure 6. The course of scenarios number 1 for sensor 2.

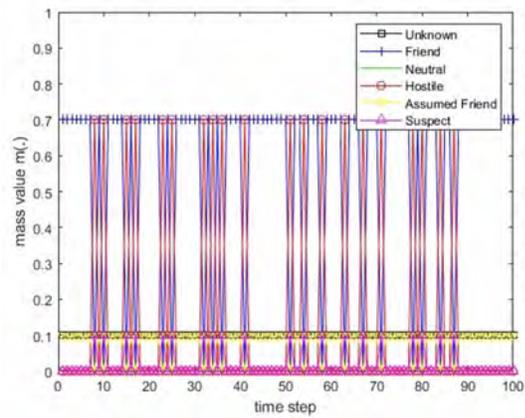


Figure 7. The course of scenarios number 2 for sensor 2.

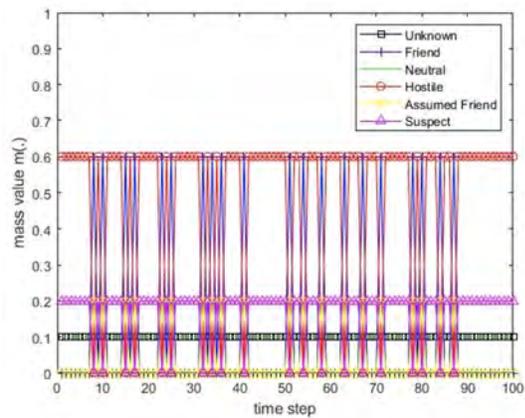


Figure 8. The course of scenarios number 3 for sensor 2.

The Monte Carlo method of generating the scenario for the sensor 2 is also used in this paper. Moments in which incorrect identifications occurred are generated by

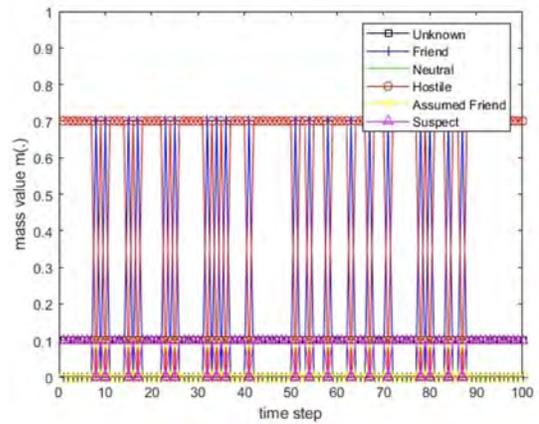


Figure 9. The course of scenarios number 4 for sensor 2.

the pseudo-random integer number generator from the range  $[0, 100]$ . Examples of scenarios are shown in Figures 10 and 11.

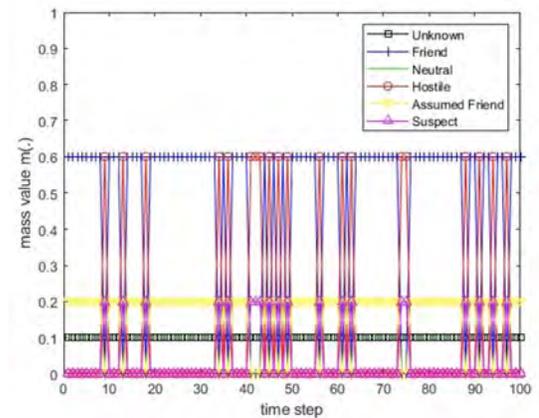


Figure 10. The course of Monte Carlo scenarios number 5 for sensor 2.

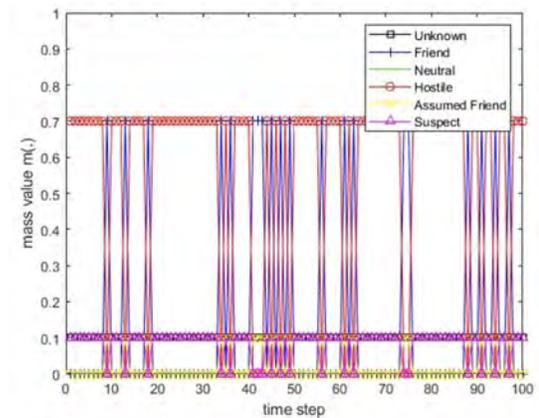


Figure 11. The course of Monte Carlo scenario number 6 for sensor 2.

B. Calculation results for deterministic scenarios

Dempster’s rule

Dempster’s rule is not resistant to a situation when the degree of conflict  $k_{Fi} = 1$ . This means the total conflict between the mass vector sent by the sensor and the mass vector of the information fusion center, which occurs when each non-zero belief mass value sent by the sensor corresponds to zero belief mass value of the vector determined by the information fusion center and vice versa.

The simulation results of the identification information fusion using Dempster’s rule have been presented for the deterministic scenarios 1 and 3 in Figures 12 and 13.

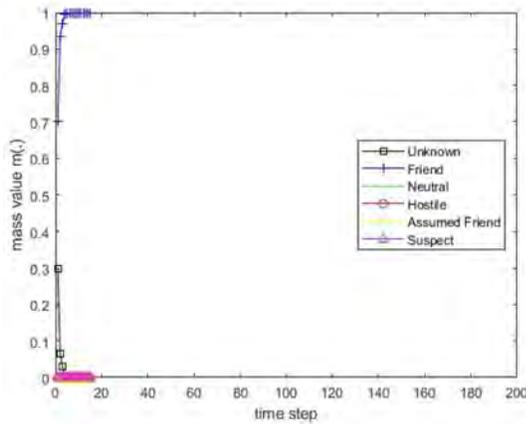


Figure 12. The values of the resulting belief mass for scenario 1 and Dempster’s rule.

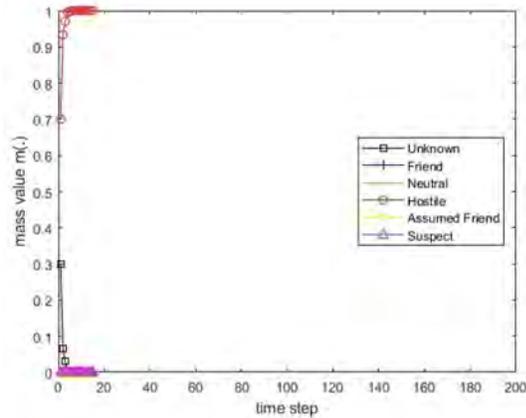


Figure 13. The values of the resulting belief mass for scenario 3 and Dempster’s rule.

The PCR1 rule

The simulation results of the identification information fusion using the PCR1 rule for the deterministic scenarios 1, 2, 3 and 4 are presented in Figures 14–17.

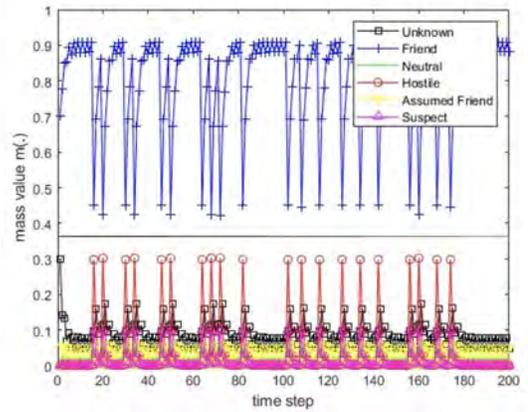


Figure 14. The values of the resulting belief mass for scenario 1 and the PCR1 rule.

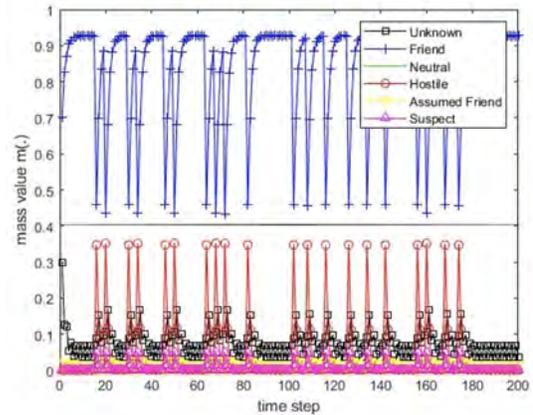


Figure 15. The values of the resulting belief mass for scenario 2 and the PCR1 rule.

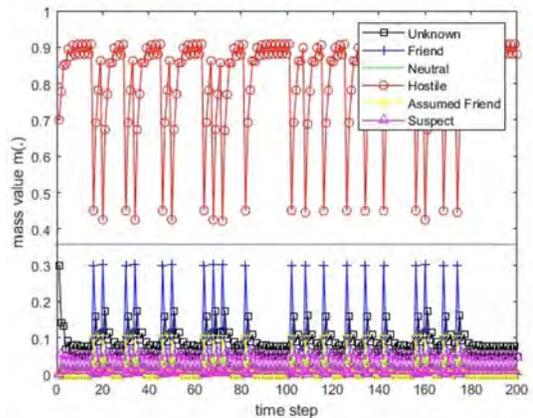


Figure 16. The values of the resulting belief mass for scenario 3 and the PCR1 rule.

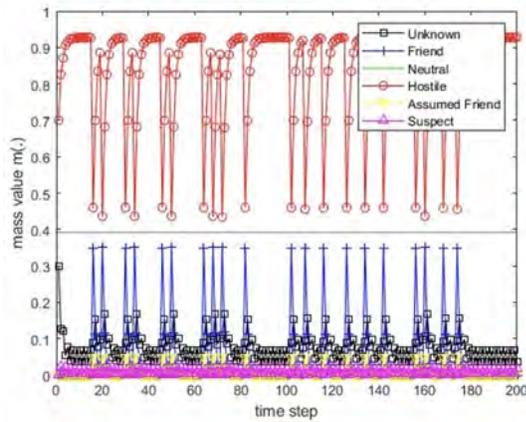


Figure 17. The values of the resulting belief mass for scenario 4 and the PCR1 rule.

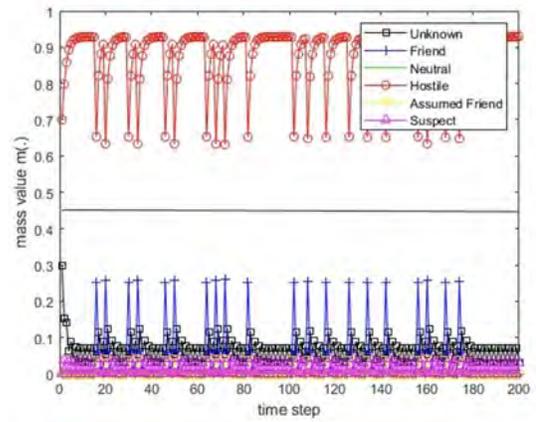


Figure 20. The values of the resulting belief mass for scenario 3 and the PCR3 rule.

### The PCR3 rule

The simulation results of the identification information fusion using the PCR3 rule for the deterministic scenarios 1, 2, 3 and 4 are presented in Figures 18–21.

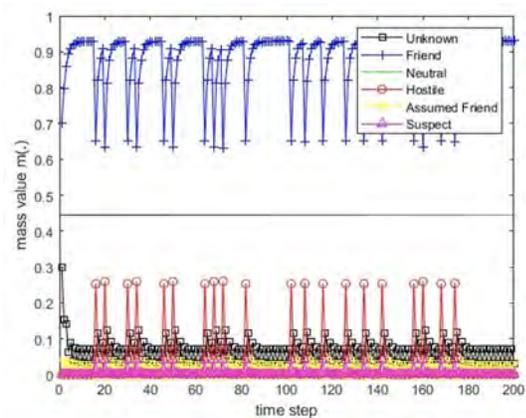


Figure 18. The values of the resulting belief mass for scenario 1 and the PCR3 rule.

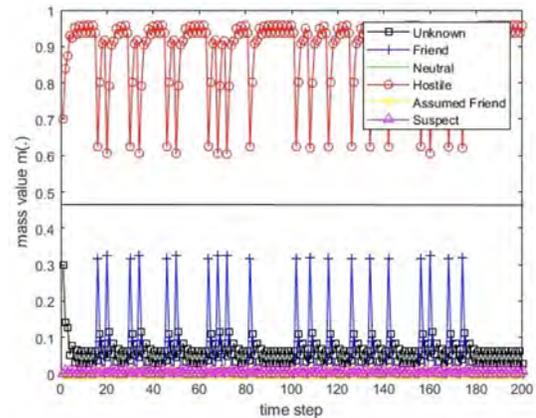


Figure 21. The values of the resulting belief mass for scenario 4 and the PCR3 rule.

### The PCR4 rule

The simulation results of the identification information fusion using the PCR4 rule for the deterministic scenarios 1, 2, 3 and 4 are presented in Figures 22–25.

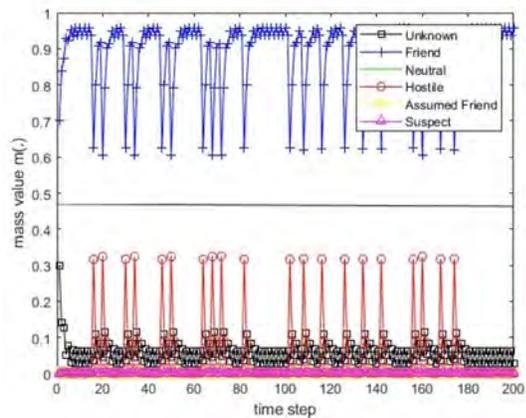


Figure 19. The values of the resulting belief mass for scenario 2 and the PCR3 rule.

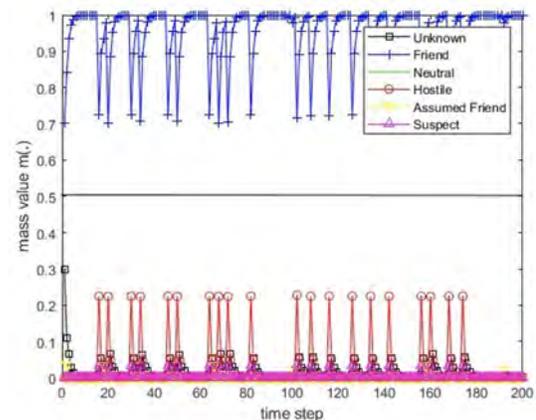


Figure 22. The values of the resulting belief mass for scenario 1 and the PCR4 rule.

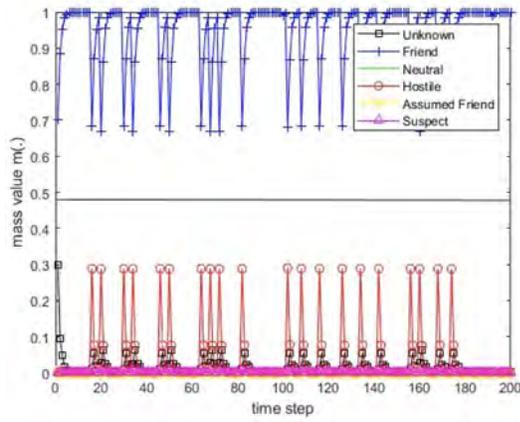


Figure 23. The values of the resulting belief mass for scenario 2 and the PCR4 rule.

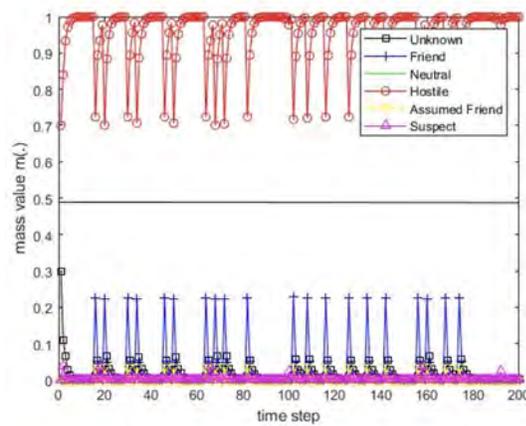


Figure 24. The values of the resulting belief mass for scenario 3 and the PCR4 rule.

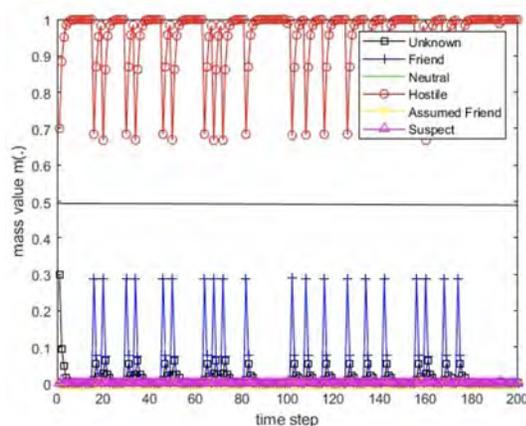


Figure 25. The values of the resulting belief mass for scenario 4 and the PCR4 rule.

**The PCR5 rule for 2 BBAs**

The simulation results of the identification information

fusion using the PCR5 rule for the deterministic scenarios 1, 2, 3 and 4 are presented in Figures 26–29.

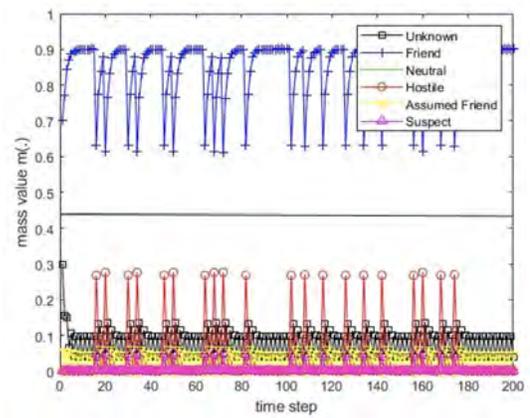


Figure 26. The values of the resulting belief mass for scenario 1 and the PCR5 rule for 2 BBAs.

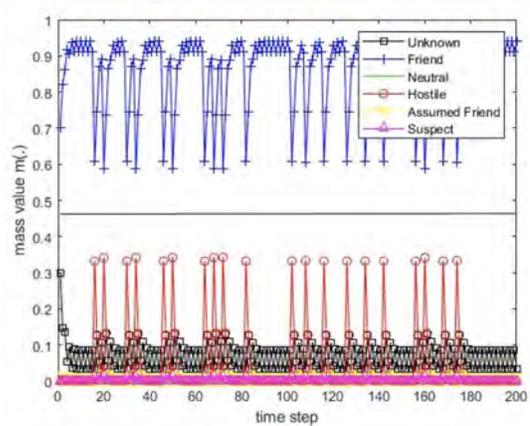


Figure 27. The values of the resulting belief mass for scenario 2 and the PCR5 rule for 2 BBAs.

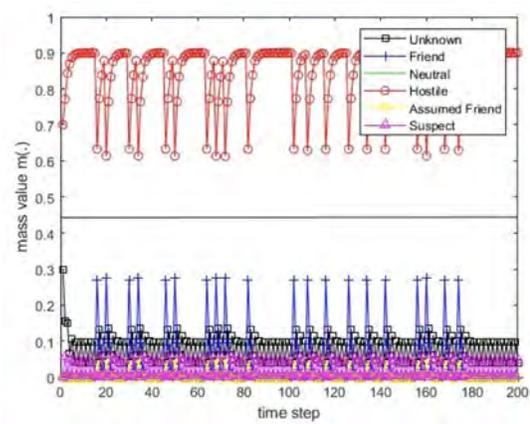


Figure 28. The values of the resulting belief mass for scenario 2a and the PCR5 rule for 2 BBAs.

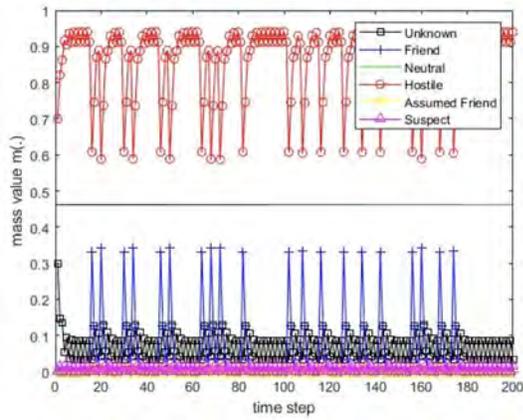


Figure 29. The values of the resulting belief mass for scenario 2b and the PCR5 rule for 2 BBAs.

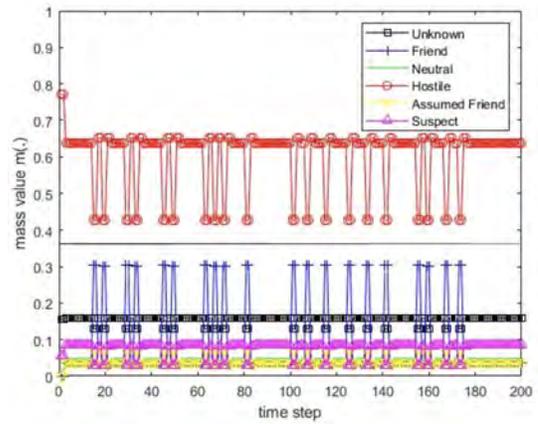


Figure 32. The values of the resulting belief mass for scenario 3 and the PCR5 rule for 3 BBAs.

**The PCR5 rule for 3 BBAs**

The simulation results of the identification information fusion using the PCR5 rule for 3 BBAs for the deterministic scenarios 1, 2, 3 and 4 are presented in Figures 30–33.

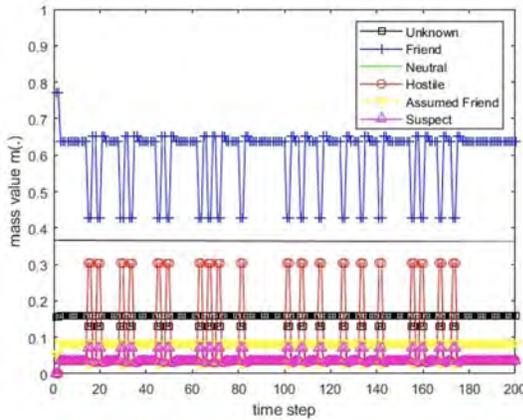


Figure 30. The values of the resulting belief mass for scenario 1 and the PCR5 rule for 3 BBAs.

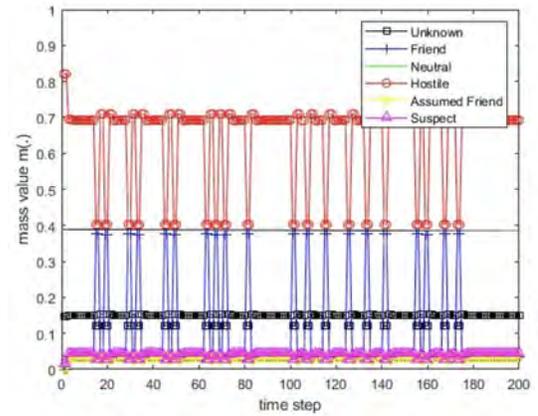


Figure 33. The values of the resulting belief mass for scenario 4 and the PCR5 rule for 3 BBAs.

**The PCR6 rule for 3 BBAs**

The simulation results of the identification information fusion using the PCR6 rule for the deterministic scenarios 1, 2, 3 and 4 are presented in Figures 34–37.

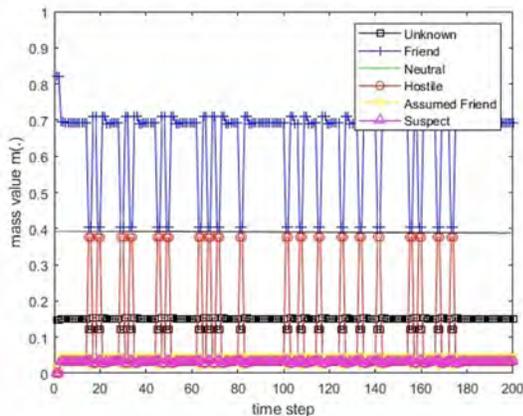


Figure 31. The values of the resulting belief mass for scenario 2 and the PCR5 rule for 3 BBAs.

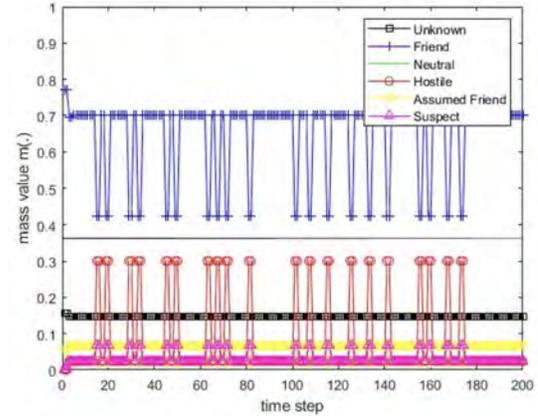


Figure 34. The values of the resulting belief mass for scenario 1 and the PCR6 rule for 3 BBAs.

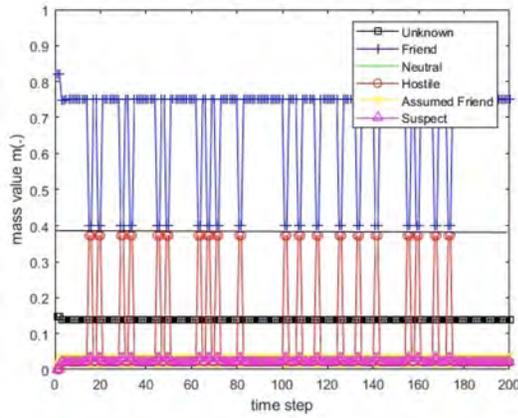


Figure 35. The values of the resulting belief mass for scenario 2 and the PCR6 rule for 3 BBAs.

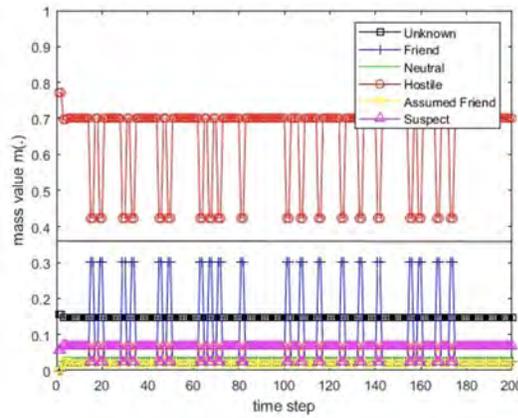


Figure 36. The values of the resulting belief mass for scenario 3 and the PCR6 rule for 3 BBAs.

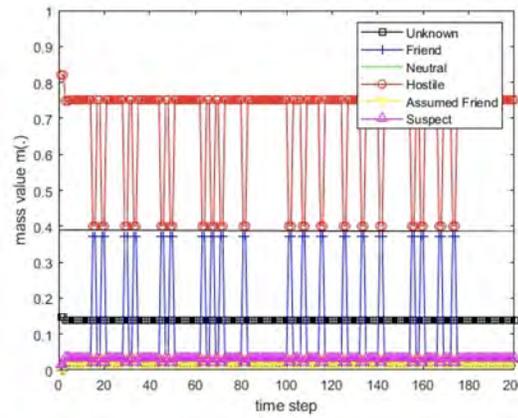


Figure 37. The values of the resulting belief mass for scenario 4 and the PCR6 rule for 3 BBAs.

The presented results (Figures 12-37) allow us to conclude that the applied methods of managing conflicts in the information fusion allow to draw correct conclusions about the real identification of the recognized object.

The application of the decision threshold for the belief mass at the level  $m_\alpha = 0.37$  for the PCR1 rule and  $m_\alpha = 0.45$  for the PCR3, PCR4 and PCR5 rules for scenarios 1 and 3 allows us to properly evaluate the identification of the recognized object: for scenario 1 – FRIEND and for scenario 3 – HOSTILE. For scenarios 2 and 4, the optimal thresholds are  $m_\alpha = 0.4$  for the PCR1 rule and  $m_\alpha = 0.45$  for the PCR3, PCR4 and PCR5 rules respectively. When assessing the interval between the minimum resultant mass for correct identification and the maximum resultant mass for misidentification, the worst results are reached by the PCR1 rule and the rules of PCR3, PCR4 and PCR5 behave similarly and are better than the rule PCR1.

The research carried out for deterministic scenarios shows that the PCR5 rule for 3 BBAs and the PCR6 rule for 3 BBAs behave very similarly. They restore the correct identification after the occurrence of temporary misidentification much faster than the rules PCR1 – PCR5 for 2 BBAs.

C. Calculation results for Monte Carlo scenarios

Dempster’s rule

In the Monte Carlo scenario, Dempster’s rule behaves similarly to a deterministic scenario. It is not resistant to a situation when the degree of conflict  $k_{Fi} = 1$ . This means the total conflict between the mass vector sent by the sensor and the mass vector of the information fusion center which occurs when each non-zero belief mass value sent by the sensor corresponds to zero belief mass value of the vector determined by the information fusion center and vice versa.

The simulation results of the identification information fusion using Dempster’s rule are presented for the Monte Carlo scenarios 5 and 6 in Figures 38 and 39.

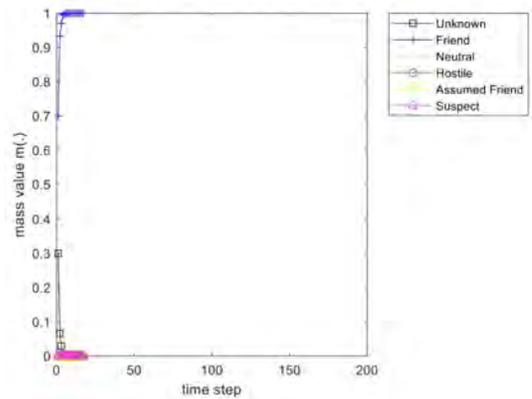


Figure 38. The values of the resulting belief mass for Monte Carlo scenario 5 and Dempster’s rule.

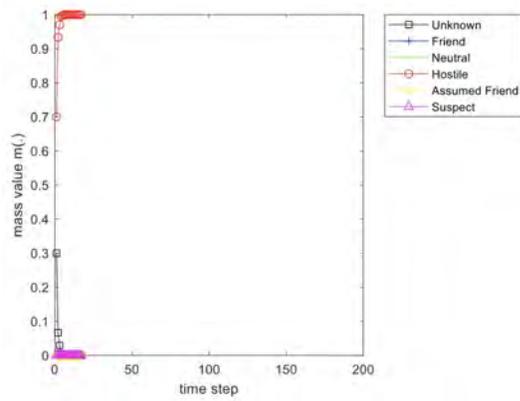


Figure 39. The values of the resulting belief mass for Monte Carlo scenario 6 and Dempster's rule.

### The PCR1 rule

The simulation results of the identification information fusion using the PCR1 rule for the Monte Carlo scenarios 5 and 6 are presented in Figures 40 and 41.

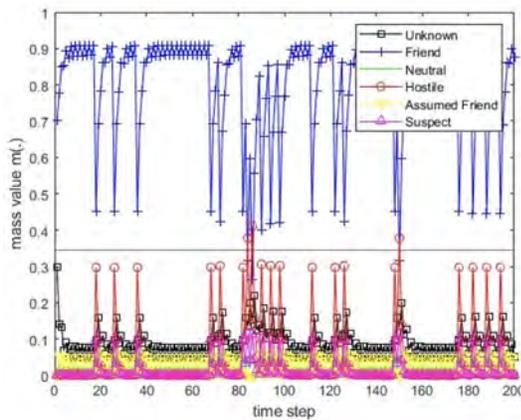


Figure 40. The values of the resulting belief mass for Monte Carlo scenario 5 and the PCR1 rule.

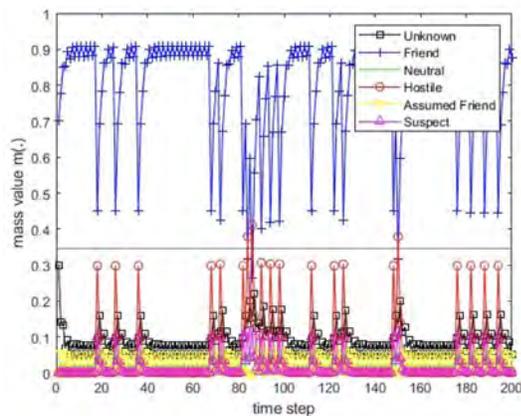


Figure 41. The values of the resulting belief mass for Monte Carlo scenario 6 and the PCR1 rule.

### The PCR3 rule

The simulation results of the identification information fusion using the PCR3 rule for the Monte Carlo scenarios 5 and 6 are presented in Figures 42 and 43.

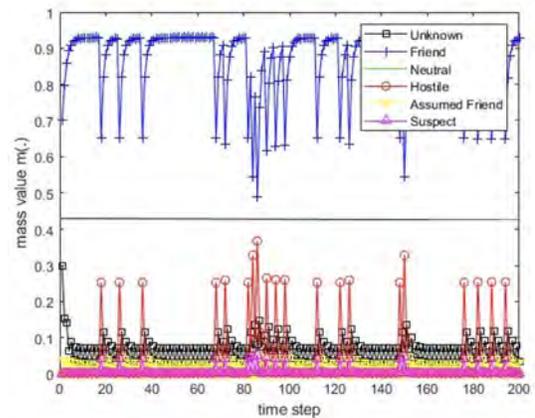


Figure 42. The values of the resulting belief mass for Monte Carlo scenario 5 and the PCR3 rule.

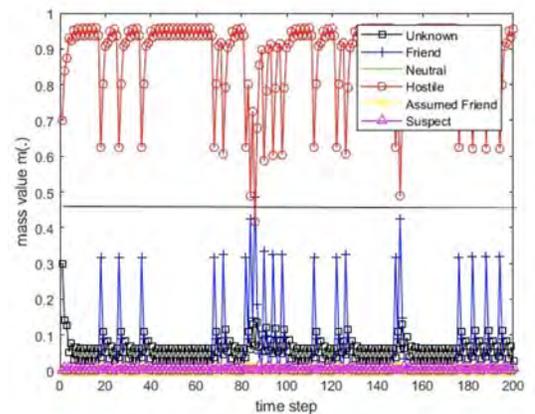


Figure 43. The values of the resulting belief mass for Monte Carlo scenario 6 and the PCR3 rule.

### The PCR4 rule

The simulation results of the identification information fusion using the PCR4 rule for the Monte Carlo scenarios 5 and 6 are presented in Figures 44 and 45.

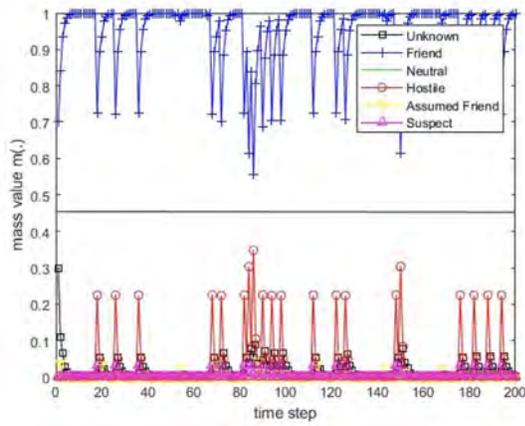


Figure 44. The values of the resulting belief mass for Monte Carlo scenario 5 and the PCR4 rule.

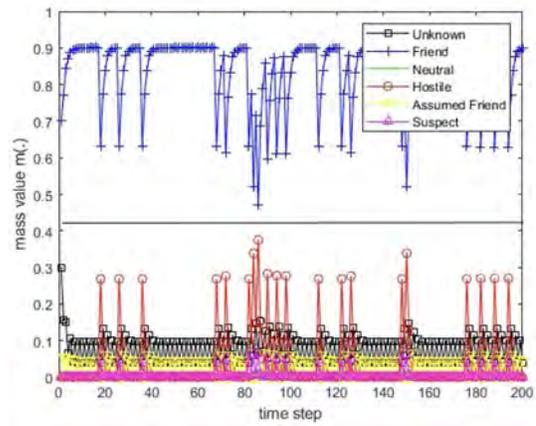


Figure 46. The values of the resulting belief mass for Monte Carlo scenario 5 and the PCR5 rule for 2 BBAs.

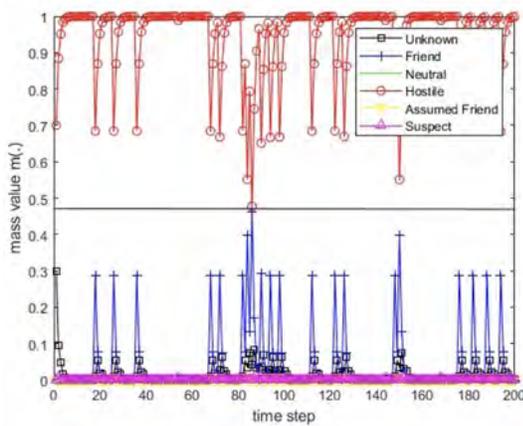


Figure 45. The values of the resulting belief mass for Monte Carlo scenario 6 and the PCR4 rule.

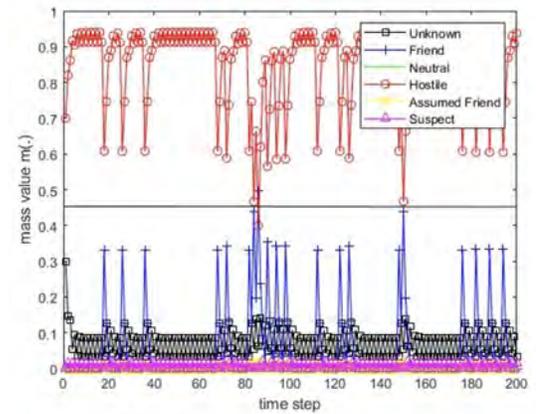


Figure 47. The values of the resulting belief mass for Monte Carlo scenario 6 and the PCR5 rule for 2 BBAs.

**The PCR5 rule for 2 BBAs**

The simulation results of the identification information fusion using the PCR5 rule for 2 BBAs for the Monte Carlo scenarios 5 and 6 are presented in Figures 46 and 47.

**The PCR5 rule for 3 BBAs**

The simulation results of the identification information fusion using the PCR5 rule for 3 BBAs for the Monte Carlo scenarios 5 and 6 are presented in Figures 48 and 49.

**The PCR6 rule for 3 BBAs**

The simulation results of the identification information fusion using the PCR6 rule for 3 BBAs for the Monte Carlo scenarios 5 and 6 are presented in Figures 50 and 51.

The presented results show that due to the high intensity of sending reports with incorrect identifications in the middle part of the scenarios, the information fusion rules (apart from

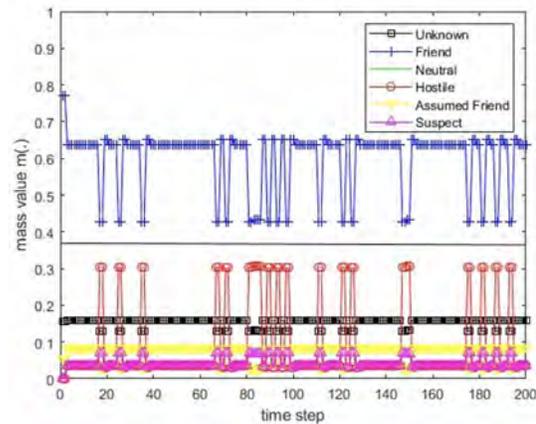


Figure 48. The values of the resulting belief mass for Monte Carlo scenario 5 and the PCR5 rule for 3 BBAs.

the PCR4, PCR 5 and PCR6 rules) determine the maximum resulting mass for incorrect identification. The PCR5 for 3 BBAs and PCR6 for 3 BBAs rules are the fastest to restore the correct identification after receiving several incorrect reports.

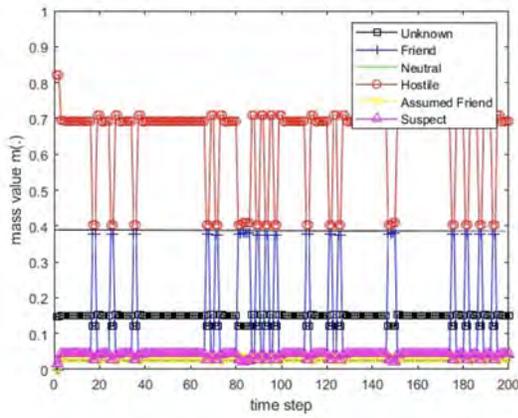


Figure 49. The values of the resulting belief mass for Monte Carlo scenario 6 and the PCR5 rule for 3 BBAs.

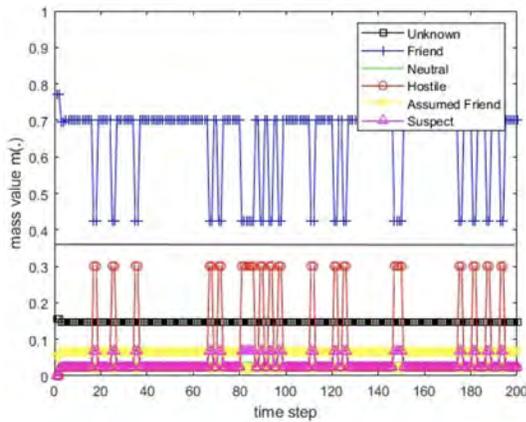


Figure 50. The values of the resulting belief mass for Monte Carlo scenario 5 and the PCR6 rule for 3 BBAs.

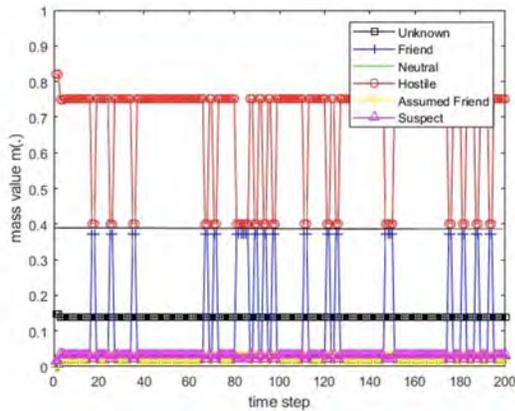


Figure 51. The values of the resulting belief mass for Monte Carlo scenario 6 and the PCR6 rule for 3 BBAs.

## VII. NUMERICAL EXPERIMENTS OF FUSION OF IDENTIFICATION IN INFORMATION FROM ESM SENSORS AND RADARS

### A. Numerical experiments scenarios

We assume that we will combine attribute information from 2 sensors: a combined primary and secondary surveillance

radar and an ESM sensor. These sensors work asynchronously. Upon receipt of the sensor's declaration in the form of a vector of masses, we fuse this vector with the vector of the actual values of the declaration masses for the fuser's frame of discernment. The frequency of transmission of sensor declarations depends on the rules of the data exchange network and on the technical characteristics of the sensors. Various combination methods are presented in [13], [14]. In this paper, two of the methods of proportional redistribution conflict (PRC5 and PCR6 [4], [11]) has been used. The numerical model of combined primary and secondary surveillance radars was taken from [5], [15].

Numerical experiments have been performed for the following data:

- for combined primary and secondary surveillance radars sensor:  $P_{fa} = 10^{-6}$ ,  $R_{max}^* = 100 \text{ km}$ ,  $P_d^* = 0.7$ ,  $\sigma_c^* = 2 \text{ m}^2$ ,  $P_{IFF} = 0.962$  and the following table of masses (compare Table I):

Table VI  
TRANSFORMATION OF THE BASE BELIEF ASSIGNMENT MASS INTO THE SECONDARY BELIEF ASSIGNMENT MASS FOR COMBINED PRIMARY AND SECONDARY SURVEILLANCE RADAR.

(Scenario #,Base identification)→	(1, $F_B$ )	(2, $N_B$ )	(3, $H_B$ )
$F_S$	0.8	0	0
$N_S$	0	0.5	0
$H_S$	0	0	0.7
$AF = N \cap F$	0.2	0.3	0
$S = N \cap H$	0	0.2	0.3

The flight path of air object was 30 km away from the sensor (in the horizontal plane), flight altitude was 1 km and radar cross-section, was 1 sq. m.

- for EMS sensor:
  - 1) the real value of identification is constant in each scenario and is equal to FRIEND ( $F$ ) – in the first scenario and HOSTILE ( $H$ ) – in the second scenario;
  - 2) the above declarations are transmitted by sensor number 1 with the real identification mass equal to 0.7 and the mass of complementary identification (UNKNOWN) equal to 0.3;
  - 3) the second sensor shall transmit its declarations in accordance with the tables I and II for the scenarios 1 and 2 and with the tables III and IV for the scenarios 3 and 4.

In the Tables VII–XII the mass values for all possible declarations for six scenarios for ESM sensor are presented

Table VII  
BELIEF MASSES FOR THE SENSOR 2 (ESM) FOR THE SCENARIO 1.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0.6	0.1	0	0.2	0	0.1
Incorrect identif. (20% of events)	0.7	0.1	0	0.1	0	0.1

Scenarios 1–6 for the sensor 1 have been presented in Figures 52, 53 and 54.

Table VIII  
BELIEF MASSES FOR THE SENSOR 2 (ESM) FOR THE SCENARIO 2.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0	0.5	0.3	0	0.2	0
Incorrect identif. (20% of events)	0	0.4	0.2	0	0.3	0.1

Table IX  
BELIEF MASSES FOR THE SENSOR 2 (ESM) FOR THE SCENARIO 3.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0	0.1	0.7	0	0.1	0.1
Incorrect identif. (20% of events)	0	0.1	0.6	0	0.2	0.1

Table X  
BELIEF MASSES FOR THE SENSOR 2 (ESM) FOR THE SCENARIO 4.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0.1	0.7	0.1	0	0	0.1
Incorrect identif. (20% of events)	0	0.1	0.6	0	0.2	0.1

Table XI  
BELIEF MASSES FOR THE SENSOR 2 (ESM) FOR THE SCENARIO 5.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0.6	0.1	0	0.2	0	0.1
Incorrect identif. (20% of events)	0	0.1	0.6	0	0.2	0.1

Table XII  
BELIEF MASSES FOR THE SENSOR 2 (ESM) FOR THE SCENARIO 6.

Type of identification	F	N	H	AF	S	U
Correct identif. (80% of events)	0.1	0.7	0.1	0	0	0.1
Incorrect identif. (20% of events)	0.6	0.1	0	0.2	0	0.1

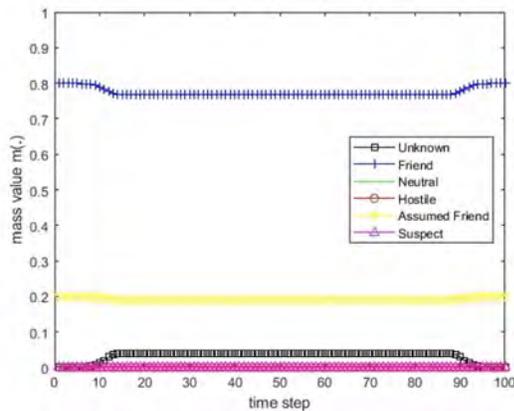


Figure 52. The course of scenarios number 1 and 4 for sensor 1.

All deterministic scenarios for the sensor 2 have been presented in Figures 55-60.

Scenarios 1–3 assume relatively small changes in the mass of all declarations. Scenarios 4–6 assume significant changes in the mass of all declarations (large errors).

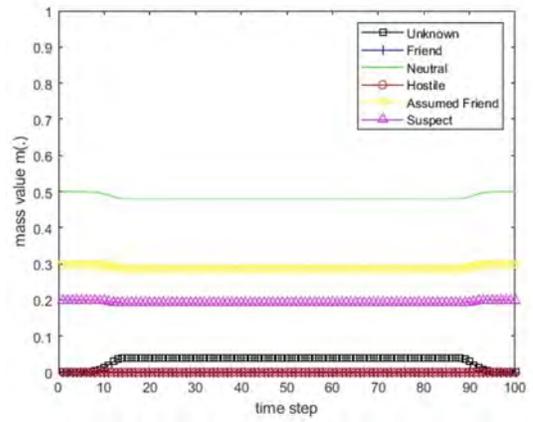


Figure 53. The course of scenarios number 2 and 5 for sensor 1.

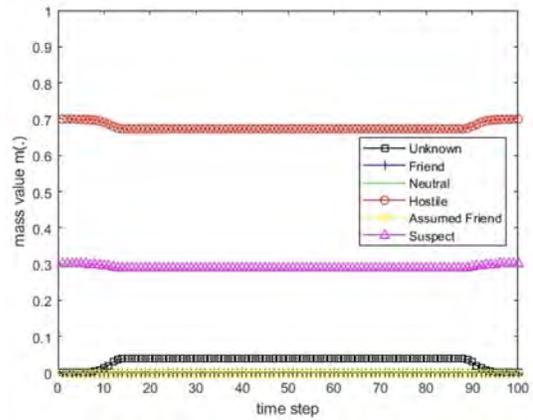


Figure 54. The course of scenarios number 3 and 6 for sensor 1.

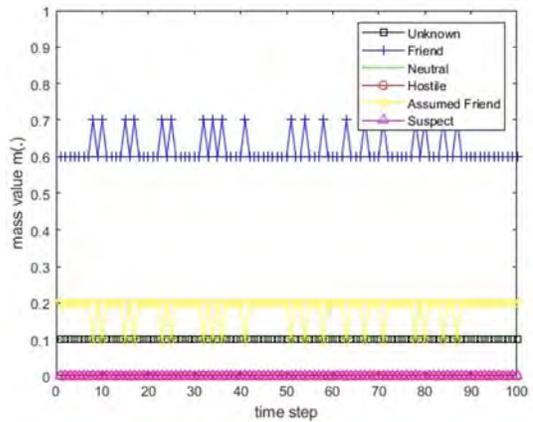


Figure 55. The course of scenarios number 1 for sensor 2.

Scenarios 4–6 assume significant changes in the mass of all declarations (large errors).

*B. Calculation results for four proportional conflict redistribution rules*

**The PCR5 rule for 2 BBAs**

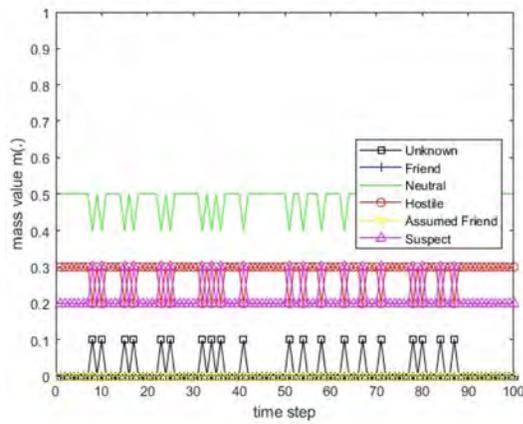


Figure 56. The course of scenarios number 2 for sensor 2.

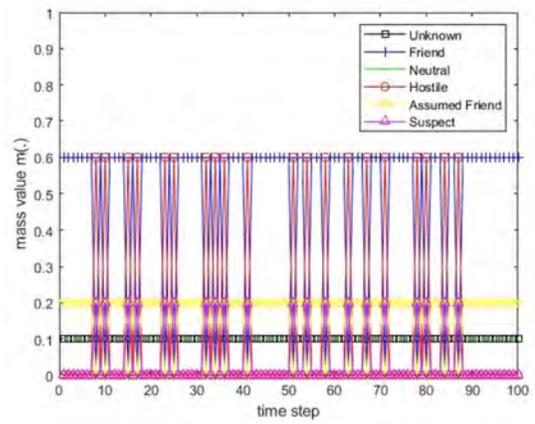


Figure 59. The course of scenarios number 5 for sensor 2.

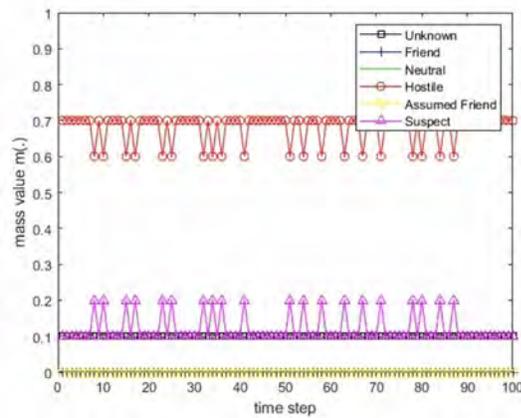


Figure 57. The course of scenarios number 3 for sensor 2.

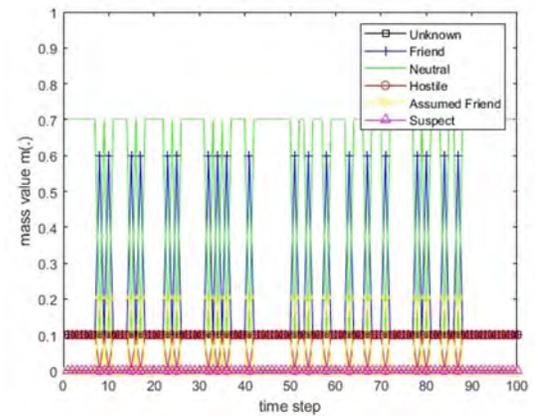


Figure 60. The course of scenarios number 6 for sensor 2.

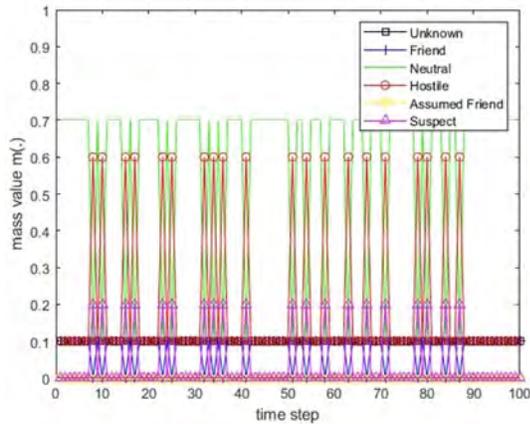


Figure 58. The course of scenarios number 4 for sensor 2.

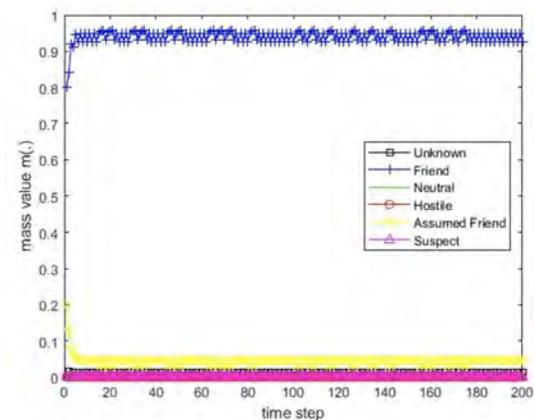


Figure 61. The values of the resulting belief mass for scenario 1 and the PCR5 rule for 2 BBAs.

The simulation results of the identification information fusion using the PCR5 rule for 2 BBAs for the deterministic scenarios 1-6 are presented in Figures 61–66.

For the PCR5 rule for 2 BBAs, the application of the decision thresholds at the belief mass level  $m_\alpha = 0.43$  for

scenario 4,  $m_\alpha = 0.35$  for scenario 5, and  $m_\alpha = 0.5$  for scenario 6 allows us to properly evaluate the identification of the recognized object for most time moments.

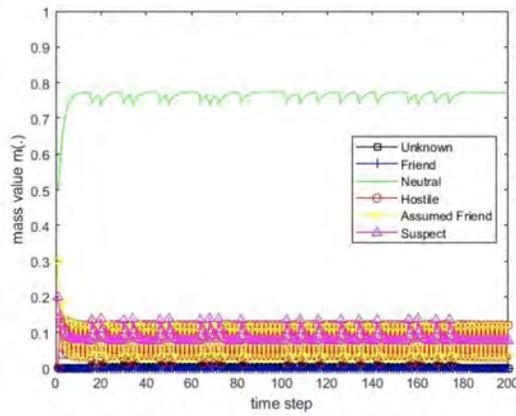


Figure 62. The values of the resulting belief mass for scenario 2 and the PCR5 rule for 2 BBAs.

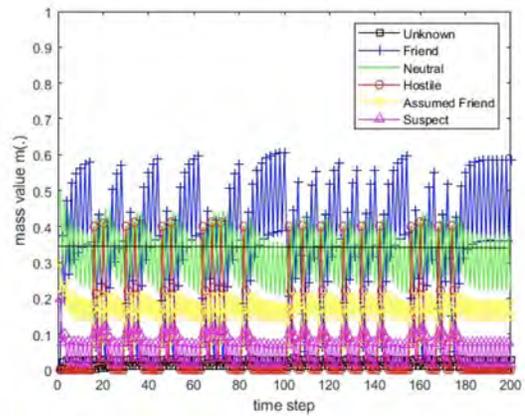


Figure 65. The values of the resulting belief mass for scenario 5 and the PCR5 rule for 2 BBAs.

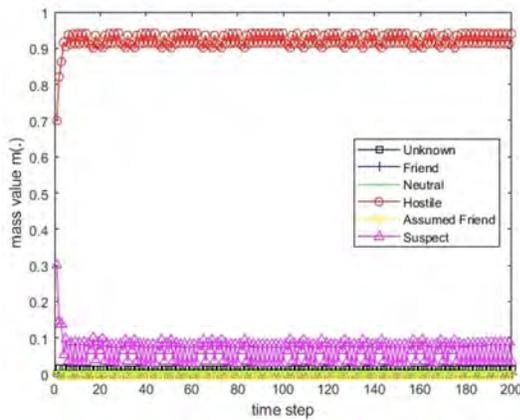


Figure 63. The values of the resulting belief mass for scenario 3 and the PCR5 rule for 2 BBAs.

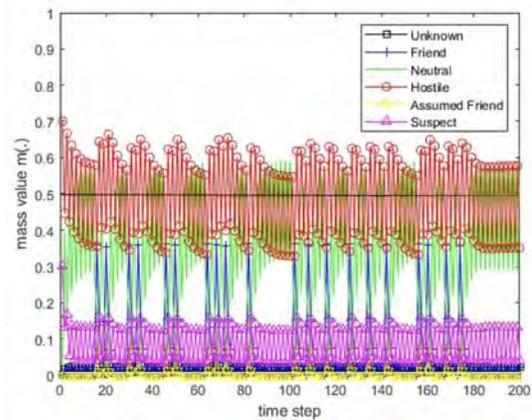


Figure 66. The values of the resulting belief mass for scenario 6 and the PCR5 rule for 2 BBAs.

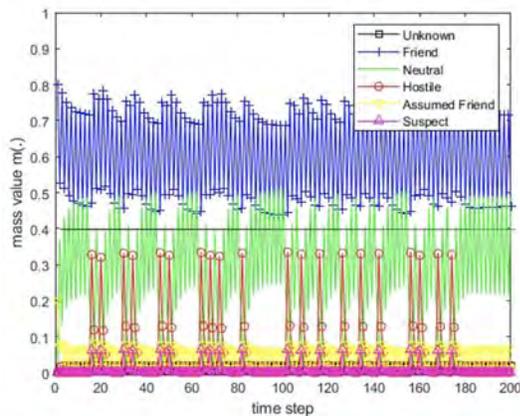


Figure 64. The values of the resulting belief mass for scenario 4 and the PCR5 rule for 2 BBAs.

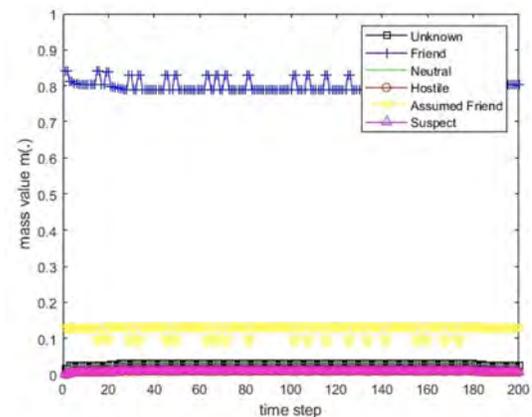


Figure 67. The values of the resulting belief mass for scenario 1 and the PCR5 rule for 3 BBAs.

### The PCR5 rule for 3 BBAs

The simulation results of the identification information fusion using the PCR5 rule for 3 BBAs for the deterministic scenarios 1–6 are presented in Figures 67–72.

For the PCR5 rule for 3 BBAs, the application of the decision thresholds at the belief mass level  $m_\alpha = 0.42$  for scenario 4 allows us to properly evaluate the identification

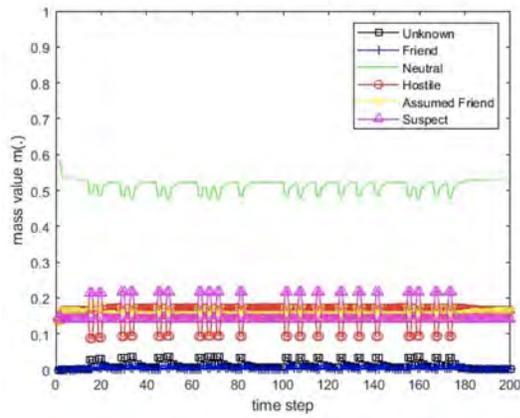


Figure 68. The values of the resulting belief mass for scenario 2 and the PCR5 rule for 3 BBAs.

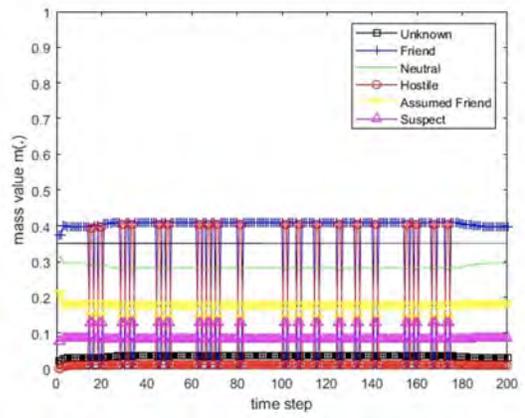


Figure 71. The values of the resulting belief mass for scenario 5 and the PCR5 rule for 3 BBAs.

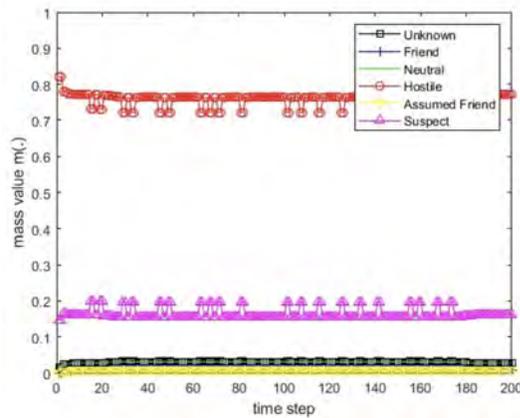


Figure 69. The values of the resulting belief mass for scenario 3 and the PCR5 rule for 3 BBAs.

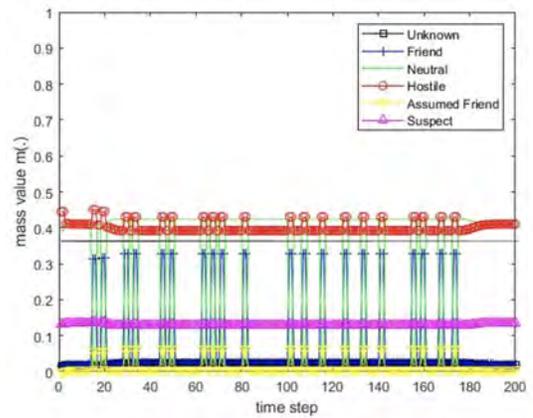


Figure 72. The values of the resulting belief mass for scenario 6 and the PCR5 rule for 3 BBAs.

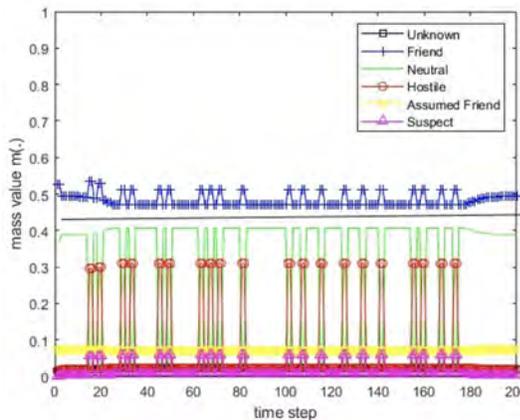


Figure 70. The values of the resulting belief mass for scenario 4 and the PCR5 rule for 3 BBAs.

time moments.

### The PCR6 rule for 3 BBAs

The simulation results of the identification information fusion using the PCR6 rule for 3 BBAs for the deterministic scenarios 1–6 are presented in Figures 73–78.

For the PCR6 rule for 3 BBAs, the application of the decision thresholds at the belief mass level  $m_\alpha = 0.45$  for scenario 4 allows us to properly evaluate the identification of the recognized object. For the PCR6 rule for 3 BBAs, the application of the decision thresholds at the belief mass level  $m_\alpha = 0.37$  for scenario 5 and  $m_\alpha = 0.4$  for scenario 6 allows us to properly evaluate the identification of the recognized object for most time moments. Comparing Figures 64–66 with Figures 70–72 and 76–78, one can conclude that the PCR5 for 3 BBAs and PCR6 for 3 BBAs rules provide more stable results of combined belief masses (smaller amplitude of changes). Due to the large dispersion of belief mass changes for scenarios 5 and 6, it is not possible to correctly evaluate the identification of the recognized object for all time moments.

of the recognized object. For the PCR5 rule for 3 BBAs, the application of the decision thresholds at the belief mass level  $m_\alpha = 0.37$  for scenarios 5 and 6 allows us to properly evaluate the identification of the recognized object for most

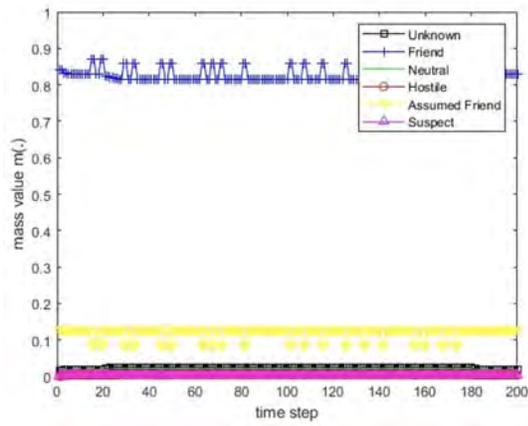


Figure 73. The values of the resulting belief mass for scenario 1 and the PCR6 rule for 3 BBAs.

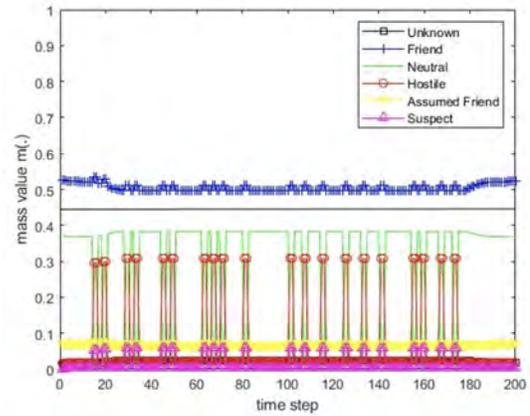


Figure 76. The values of the resulting belief mass for scenario 4 and the PCR6 rule for 3 BBAs.

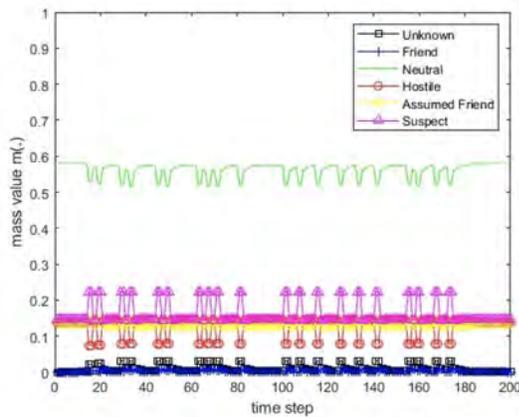


Figure 74. The values of the resulting belief mass for scenario 2 and the PCR6 rule for 3 BBAs.

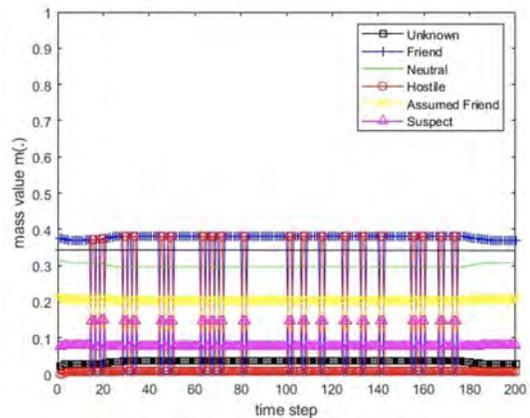


Figure 77. The values of the resulting belief mass for scenario 5 and the PCR6 rule for 3 BBAs.

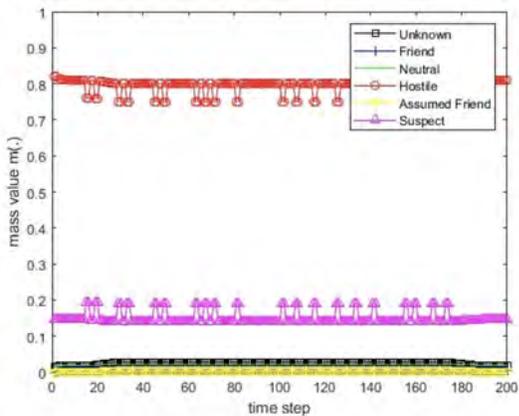


Figure 75. The values of the resulting belief mass for scenario 3 and the PCR6 rule for 3 BBAs.

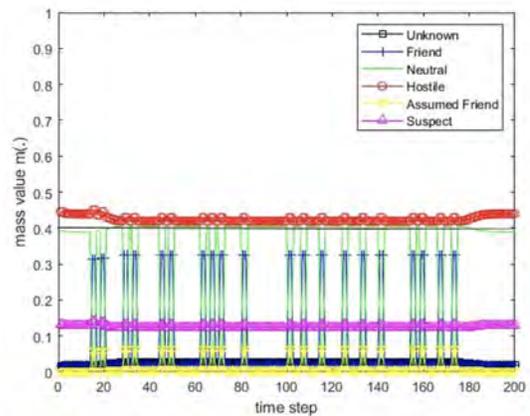


Figure 78. The values of the resulting belief mass for scenario 6 and the PCR6 rule for 3 BBAs.

The presented results (Figures 61-78) allow us to conclude that the applied methods of removing conflicts in the information fusion allow to draw correct conclusions about the real identification of the recognized object.

## VIII. CONCLUSION

The proposed basic belief assignment model for ELINT-ESM sensors and radars can be used to build identification information fusion systems. Practical meaning have, first of all, models conformable to STANAG 1241. Due to the assumption of conflicts between the ELINT-ESM sensor declarations in this work, Dezert-Smarandache theory is used to determine the basic belief assignment of declarations being the product of the process of fusion of identification information sent by these sensors. Supplementing standard reports on detected signals with random identification declarations allows the use of methods of identification information fusion in the information fusion center. The test results confirm the full usefulness of conflict redistribution rules in reports from ELINT-ESM sensors developed as a part of Dezert-Smarandache theory, with the best results presented in the PCR5 and PCR 6 rule.

The basic belief assignment model for ESM sensors and for combined primary and secondary (IFF) surveillance radars [15] can be applied to build models of different identification data fusion systems. The practical significance has first of all models compatible with STANAG 1241. It contains definitions corresponding to intersections of basic identification declarations. Therefore, the paper uses Dezert-Smarandache theory for calculation the basic belief assignment.

The conducted research showed that the best results obtained for the PCR6 rule when reports from three sources (from two sensors and the fusion system database) were processed simultaneously. This corresponds to synchronous processing of reports and involves delayed processing of a report from one of the sources. The research confirmed a slight advantage of the PCR6 rule over the PCR5 rule. This was mainly the case when the sensors sent information with a high degree of conflict.

## REFERENCES

- [1] NATO, *Standardization Agreement STANAG No. 1241*, NATO Standard Identity Description Structure for Tactical Use, Ed. 5, North Atlantic Treaty Organization, April 2005.
- [2] NATO, *Standardization Agreement STANAG No. 1241 AO*, NATO Joint Standard Identification Description, Ed. 6, Ratification Draft 1, North Atlantic Treaty Organization, 27 October 2009.
- [3] F. Smarandache, J. Dezert, *Applications and Advances of DSMT for Information Fusion (Collected works)*, Volume 2, American Research Press (ARP), Rehoboth, 2006.
- [4] T. Dezert, J. Dezert, F. Smarandache, *Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments*, Journal of Advances in Information Fusion, Volume 16, Issue 1, 48–73, June 2021.
- [5] T. Pietkiewicz, A. Kawalec A., *A method of determining the basic belief assignment for combined primary and secondary surveillance radars based on Dezert-Smarandache theory*, The 17th International Radar Symposium (IRS), Kraków, 1-6, 2016.
- [6] P. Djiknavorian, D. Grenier, P. Valin P., *DSm theory for fusing highly conflicting ESM reports*, 12th International Conference on Information Fusion, Seattle, WA, USA, pp. 1211-1217, July 6–9, 2009.
- [7] J. Matuszewski, A. Dikta, *Emitter location errors in electronic recognition system*, XI Conference on Reconnaissance and Electronic Warfare Systems, Proc. of SPIE - The International Society for Optical Engineering, Vol. 10418, No. 4/2017, art. no. 104180C, pp. C1–C8, 2017.
- [8] NATO, *Standardization Agreement STANAG No. 5516*, Tactical Data Exchange - Link 16, Ed. 3, North Atlantic Treaty Organization.
- [9] NATO, *Standardization Agreement STANAG No. 4162*, Identification Data Combining Process, Ed. 2, North Atlantic Treaty Organization.
- [10] P. Valin, E. Bossé, *Using A Priori Databases for Identity Estimation through Evidential Reasoning in Realistic Scenarios*, RTO IST Symposium on Military Data and Information Fusion, RTO-MP-IST-040, Prague, Czech Republic, 12-1–12-12, 20–22 October 2003.
- [11] J. Dezert, F. Smarandache, *Importance of sources using repeated fusion with the proportional conflict redistribution rules #5 and #6*, 2010, hal-00471839, <https://hal.science/hal-00471839>.
- [12] F. Smarandache, J. Dezert, *Applications and Advances of DSMT for Information Fusion (Collected works)*, Volume 1, American Research Press (ARP), Rehoboth, 2004.
- [13] P. Smets, *Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem*, International Journal of Approximate Reasoning, Vol. 9, 1–35, 1993.
- [14] P. Smets, R. Kennes R., *The transferable belief model*, Artificial Intelligence, Vol. 66, 191–234, 1994.
- [15] T. Pietkiewicz, A. Kawalec, B. Wajszczyk, *Analysis of Fusion Primary Radar, Secondary Surveillance Radar (IFF) and ESM Sensor Attribute Information under Dezert-Smarandache Theory*, The 18th International Radar Symposium (IRS), Prague, Czech Republic, pp. 1–10, June 28–30, 2017.
- [16] M.C. Stevens, *Secondary Surveillance Radar*, Artech House, 1988.
- [17] P. Djiknavorian, D. Grenier, P. Valin, *Analysis of information fusion combining rules under the DSm theory using ESM input*, 10th International Conference on Information Fusion, FUSION 2007, Québec, Canada, 9–12 July, 2007.

# Maritime Objects Upon FLIR Images Recognition by Fusion of Two SVM Classifiers Using Different Image Descriptors

Tadeusz Pietkiewicz, Alicja Wróbel

Institute of Radioelectronics, Military University of Technology

Gen. S. Kaliskiego str. 2, 00-908 Warszawa, Poland.

Emails: tadeusz.pietkiewicz@wat.edu.pl, alaala-rybak@wp.pl

**Abstract**—Infrared image recognition by means of FLIR cameras (forward-looking infrared) is one of the elements of the recognition of the maritime situation and it supports in many situations the creation of so-called maritime picture. This work presents results of two FLIR image classifiers research. The work presents the use of SVM (Support Vector Machine) to classify images of maritime objects. The SVM network uses to perform the multi-class classification the one-against-all method. Both classifiers use the pre-processed FLIR images as input data in the form of the brightness of all image pixels and a histogram of oriented gradients (for training and testing). All FLIR color images have been transformed into grayscale images, segmented using the Otsu algorithm with a possible manual correction, rescaled, centered and leveled. In the further part of the work a method of determining the basic belief assignment is proposed for SVM classifiers. In the final part of the work test results of the both classification methods and their fusion by Dezert-Smarandache PCR5 rule for a set of maritime objects FLIR images registered in the Baltic Sea are presented.

**Keywords:** FLIR images recognition, image classifiers, SVM networks, time series comparison, basic belief assignment, the Dezert-Smarandache rule of information fusion, proportional conflict redistribution.

## I. INTRODUCTION

FLIR (forwarded looking infra-red) passive infrared sensors are used for short- to medium-ranged recognition from 2 to 20 nautical miles depending on the size of the object being recognized and the conditions of observation. They are mounted on maritime and air platforms in the armed forces and border guards of many countries. FLIR cameras create a monochrome image in which the luminance of each pixel is proportional to the temperature of the observed point [1]. These cameras often artificially color the image to present it to the operator. The method of assigning the color to the temperature is usually shown on the image. The natural way is to assign higher temperatures to yellow colors, and the lowest temperatures to blue and purple colors. From the point of view of image recognition, these colors are artificial and should be removed from the image and converted into shades of gray.

Recognition of maritime objects based on FLIR images should first answer the question whether the registered object is a maritime object. If one gets a positive answer, one expects an answer to the next question, whether the object being

recognized belongs to one of the classes from the training set (previously recorded and classified images), or possibly state the inability to recognize the type.

FLIR images can be distorted due to specific atmospheric conditions (fog or rain) and solar lighting containing infrared radiation as well as due to physical processes taking place in the camera. The geometry of the silhouettes of maritime objects can be changed as a result of changing camera settings, different distance of the object from the camera and different object observation angles (so-called aspect angles). The above-mentioned factors make the process of recognizing maritime objects based on FLIR images a multi-stage process.

The process of comparing and recognizing objects takes into account their specific features called distinctive features. The choice of features is related to the specifics of the recognized objects, the method of recording images, and the methods and recognition algorithms used. In the case of recognition of maritime object images, the basic element of the image being analyzed is the silhouette of this object. It can be characterized by various sets of distinctive features. The principal component analysis (PCA) method or methods using deep neural networks take into account the luminance of all these silhouette pixels.

In this work, the descriptor of a grayscale image of a maritime object without compression is stored as a horizontal vector of original image rows concatenation. It is given the name of the linear image descriptor. The second type of descriptor is the histogram of oriented HOG gradients [2,3]. This descriptor compresses the image by determining the brightness gradients of the image in evenly spaced image sections (cells). Gradient vectors describe the shape of the objects contained in the image.

The work uses the method of classifying the silhouettes of marine objects based on multi-class classification using the SVM (Support Vector Machine) network [4,5,6,7,8]. This method is based on SVM classifiers, which in the literature are also called SVM networks because of their similarity to neural networks. These classifiers are, by their very nature, two-class classifiers, and thus they allow finding the answer to the question whether the recognized object belongs to one of two classes. In real problems, the set of patterns is usually multi-class, hence methods for using two-class classifiers for multi-class classification have been developed. One of such

methods called “one against all” is used in the work. It has been modified for the fusion of SVM networks in such a way that it is possible to determine the basic belief assignment over the set of pattern types.

Another problem to be addressed was the assessment of the linear separability of the training set histograms. It is related to the possible transformation of the original space of distinctive features into a space with a much larger number of dimensions and the use of the appropriate kernel function of the transformation. In this work, a hint contained in [6] is used. It allows the use of a linear kernel when the number of patterns in the training set is much smaller than the number of distinctive features of the patterns.

The SVM classifiers require FLIR images to be previously processed, preparing a histogram of the vertical brightness projection. The purpose of this process is, among other things, to eliminate unnecessary information about the background of the object and interference, as well as to normalize the silhouette of the object. The image pre-processing process may include segmentation, brightness normalization, silhouette scaling, silhouette centering, silhouette leveling and extraction of distinctive features. Some problems of information pre-processing have been presented in [4].

One of the important objectives of our work was to evaluate the effectiveness of information fusion methods [9,10] applied to the results of SVM classifiers. In papers [9,10,11], methods for determining the basic belief assignment on a set of possible decisions of SVM classifiers were proposed.

In the final part of the work, the results of tests of SVM classifiers with two different image descriptors and the results of the fusion of these classifiers using the Dezert-Smarandache PCR5 rule [9,11] are presented.

## II. HISTOGRAM OF ORIENTED GRADIENTS AS A VECTOR OF DISTINCTIVE FEATURES OF A MARITIME OBJECT IMAGE

As mentioned in section I of this work, it was assumed that the histogram of oriented gradients HOG is created on the basis of the image formed in the process of pre-processing, segmentation and secondary processing of the primary image.

Histograms of oriented gradients (HOG) are image descriptors that describe the shapes of objects within the image. The idea of their operation is based on counting gradients occurring in the same spatial orientation (at a specific angle), in a certain precisely defined fragment of the image. These gradients are counted in evenly distributed portions of the image. To improve the efficiency of object detection, local contrast normalization is applied in overlapping regions [2,3].

HOG descriptors are created by dividing the image into fragments, called cells. Gradients are then calculated for each fragment. The distribution of these gradients is represented by the so-called edge orientation histogram. Cells are grouped into blocks to perform contrast normalization. This operation is aimed at increasing resistance to shadows, differences in lighting [2,3]. The histograms computed in all blocks form a vector. This vector is called the HOG. An example of the HOG implementation is shown in Figure 1.

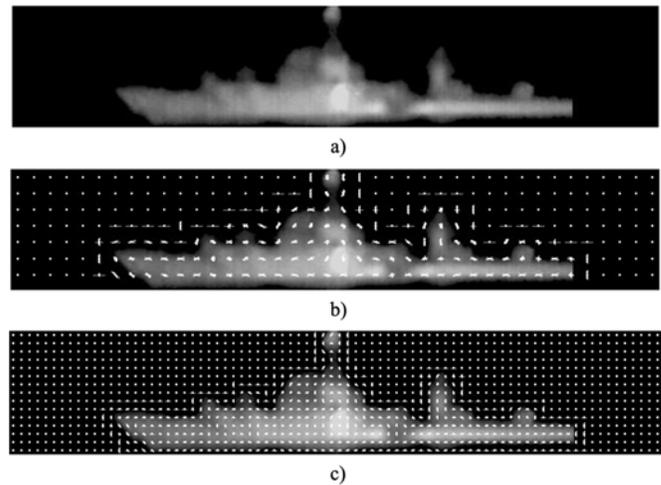


Figure 1. Example implementation of HOG: a) input image, b) gradients for the input image with cells  $16 \times 16$ , c) gradients for the input image with cells  $8 \times 8$ .

## III. LINEAR DESCRIPTOR AS A VECTOR OF DISTINCTIVE FEATURES OF A MARITIME OBJECT IMAGE

In this work, the descriptor of a grayscale image of a maritime object without compression is stored as a horizontal vector of original image rows concatenation. It is given the name of the linear image descriptor. If  $\mathbf{A}$  denotes an array containing a grayscale image of size  $m \times n$ , then the linear discriminant of this image can be represented as (in MATLAB notation):

$$[\mathbf{A}(1, :) \ \mathbf{A}(2, :) \ \dots \ \mathbf{A}(m, :)]. \quad (1)$$

## IV. CHARACTERISTICS OF SVM CLASSIFIER

### A. Introduction

Support Vector Machine SVM is a useful technique for data classification [5,6,7,8,12,13,14,15,16]. The SVM machine in the Polish literature on the subject is most often called the SVM network due to its simple interpretation using neural networks. First, basic information explaining the principles of data classification using SVM in the case of binary (two-class) classification will be presented. This technique will then be extended to the problem of multi-class classification.

### B. Binary classification

The SVM classifier belongs to the set of classifiers that maximize the separation margin [5,6,7,8]. The SVM classifier belongs to the set of classifiers that maximize the separation margin. These classifiers recognize patterns belonging to two classes by specifying a decision surface that provides maximum distance to the nearest points in the training set called support vectors.

The distance of any point  $\mathbf{x}$  from the hyperplane (3) is

$$d(\mathbf{x}) = \frac{|\mathbf{w}^T \mathbf{x} + b|}{\|\mathbf{w}\|} \quad (2)$$

Let us assume that a set of training pairs is given  $(\mathbf{x}_i, y_i)$  for  $i = 1, \dots, p$ , wherein each point  $\mathbf{x}_i \in \mathbb{R}^N$  belongs to one of two classes of patterns identified by labels  $y_i = +1$  (class 1) or  $y_i = -1$  (class 2). Assuming a linear separability of classes, the equation of the separating hyperplane can be written using the formula

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0, \quad (3)$$

where  $\mathbf{w} = [w_1, w_2, \dots, w_N]^T$  is an  $N$ -dimensional weight vector, and  $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$  is a vector of the distinctive features values of the object. The  $b$  value specifies the hyperplane offset relative to the origin of the coordinate system. Decision equations of classification take the following form:

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + b > 0, & \text{for } y_i = 1, \\ \mathbf{w}^T \mathbf{x}_i + b < 0, & \text{for } y_i = -1. \end{cases} \quad (4)$$

Intuitively, one can say that the greater the distance (2) of point  $\mathbf{x}$  from the hyperplane (3), the greater the reliability of the classification.

Because the assumption of the linear separability of the training data has been made, so no training data satisfies the equation  $\mathbf{w}^T \mathbf{x}_i + b = 0$ . It follows that the width of the separation margin is greater than zero, what means that inequalities (4) after their normalization can be written in the following form

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + b \geq 1, & \text{for } y_i = +1, \\ \mathbf{w}^T \mathbf{x}_i + b \leq -1, & \text{for } y_i = -1. \end{cases} \quad (5)$$

One can write these two inequalities in one formula

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1. \quad (6)$$

If a pair  $(\mathbf{x}_i, y_i)$  satisfies in (6) equality, then the vector  $\mathbf{x}_i$  is called a support vector SV.

Assuming a linear separability of training data, these vectors only decide on the location of the optimal separation hyperplane and the width of the separation margin, which has a value

$$d(\mathbf{x}) = \frac{2}{\|\mathbf{w}\|} \quad (7)$$

Optimal separating hyperplane (3) and separating margin in SVM in a two-dimensional space are presented in Figure 2.

The task of optimal separation margin design is to find such a margin, which has a maximum width. The problem of optimal selection of the separation hyperplane and the separation margin width comes down to solving the quadratic programming task in the following form

$$\min_{\mathbf{w}, b} \frac{1}{2} \mathbf{w}^T \mathbf{w}, \quad (8)$$

with constraints

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1. \quad (9)$$

This is a quadratic programming task with constraints that can be solved by the Lagrange multipliers method with  $\alpha = [\alpha_1 \alpha_2 \dots \alpha_N]^T$  multipliers using Karush-Kuhn-Tucker conditions [17]. The Lagrange function is as follows

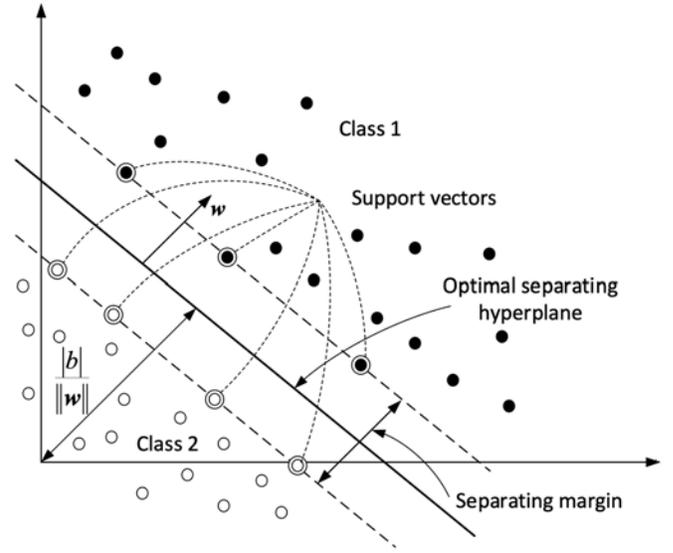


Figure 2. Optimal separating hyperplane and separating margin in SVM in a two-dimensional space.

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i y_i (\mathbf{w}^T \mathbf{x}_i + b) + \sum_{i=1}^N \alpha_i \quad (10)$$

The solution to this optimization task is as follows [2,3,4,5,6]

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i, \quad (11)$$

wherein non-zero Lagrange multipliers correspond only to support vectors.

To determine the constant value  $b$ , one can use the fact that, according to Karush-Kuhn-Tucker conditions at the saddle point of the Lagrange function, the product of the multiplier by the constraint associated with the support vector  $\mathbf{x}_{sv}$  is zero [17]  $\alpha_{sv}(\mathbf{w}^T \mathbf{x}_{sv} + b \pm 1) = 0$ ,  $\alpha_{sv} > 0$ . From here one can receive

$$b = -\mathbf{w}^T \mathbf{x}_{sv} \pm 1. \quad (12)$$

The equation of the optimal separation hyperplane is as follows

$$\sum_{i=1}^N \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b = 0, \quad (13)$$

while the decision function is as follows

$$\begin{cases} f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \geq 1, & \rightarrow y = 1, \\ f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \leq -1, & \rightarrow y = -1. \end{cases} \quad (14)$$

More complex models of SVM linear networks include the possibility of incompletely linearly separable training data. Suitable formulas can be found in [5,6,7,8].

In the above considerations, a linear separability of training data was assumed. The linear inseparability of training data does not mean a lack of their separability at all. A common

solution is the non-linear projection of original data into another functional space in which transformed patterns are linearly separable or the probability of their separability is very high. The condition is the use of non-linear transformation with a sufficiently high dimension  $K \gg N$  of the feature space. The above-mentioned construction of the separation hyperplane and decision rule can be applied in a new space that is specified by the projection function  $\Phi$ . The key property of the projection function  $\Phi$  is that the scalar product of vectors  $\Phi(\mathbf{x}_i)^T \Phi(\mathbf{x})$  in the result space can be represented as a certain kernel function  $K(\mathbf{x}_i, \mathbf{x})$ . The basic kernel functions are following:

- linear kernel:  $K(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i^T \mathbf{x}$ ,
- polynomial kernel:  $K(\mathbf{x}_i, \mathbf{x}) = (\alpha \mathbf{x}_i^T \mathbf{x} + r)^d$ ,  $\alpha > 0$ ,
- Gaussian kernel (RBF – radial basis function):  $K(\mathbf{x}_i, \mathbf{x}) = \exp(-\beta \|\mathbf{x}_i - \mathbf{x}\|^2)$ ,  $\beta > 0$ ,
- sigmoid kernel:  $K(\mathbf{x}_i, \mathbf{x}) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x} + r)$ .

The values of  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $r$  and  $d$  are the parameters of the kernels.

In [6] there are some guidelines regarding situations in which a linear kernel can be used and in which an RBF radial kernel can be used. If the number of patterns in the training set is much smaller than the number of distinctive features of the patterns, a linear kernel can be used. In this work, a linear kernel was used, because the number of training patterns was 340, and the length of the distinctive features vectors was 4800 or 1368.

The equation of the separating hyperplane after applying the transformation of the primary space of distinctive features by means of the kernel function is as follows

$$\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b = 0, \quad (15)$$

while the decision function is as follows while the decision function is as follows

$$\begin{cases} f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \geq 1, & \rightarrow y = 1, \\ f(\mathbf{x}) = \sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \leq -1, & \rightarrow y = -1. \end{cases} \quad (16)$$

### C. Multiclass classification

SVM networks divide data into two classes. Unlike classic neural networks, where we can have multiple outputs (each output is associated with one class), recognition of multiple classes requires the implementation of multiple classification tasks using multiple SVM networks. The best-known strategies for solving the problem of multi-class classification are “one-against-one” and “one-against-all” methods [5,15,18]. Suppose the training base has  $M$  types of patterns.

In the case of the “one-against-one” method,  $M(M-1)/2$  SVM classifiers are constructed. They distinguish sequentially two classes from the training set. One can receive a decision function for each pair of classes  $i$  and  $j$

$$f_{ij}(\mathbf{x}) = \mathbf{w}_{ij}^T \Phi(\mathbf{x}) + b_{ij}, \quad i, j \in \mathbf{M} = \{1, \dots, M\}, i \neq j. \quad (17)$$

After training all SVM networks, you can proceed to classify objects from the test set. If  $sgn(f_{ij}(\mathbf{x}))$  indicates the  $i$ -th class, one should increase by 1 the counter of this class indications, in the opposite case it should be increased by 1 the counter of the  $j$ -th class. Finally, we choose the class whose counter has reached the highest value.

In the case of the “one-against-all” method,  $M$  SVM classifiers are constructed, each network being trained on a different training set. Suppose we train the  $m$ -th SVM two-class network. Class 1 includes  $m$ -th type patterns, while class 2 includes other types. Finally, we receive a decision function for each network

$$f_m(\mathbf{x}) = \mathbf{w}_m^T \Phi(\mathbf{x}) + b_m, \quad m \in \mathbf{M} = \{1, \dots, M\}. \quad (18)$$

After training all SVM networks, one can proceed to classify objects from the test set. If  $sgn(f_{ij}(\mathbf{x}))$  indicates the  $m$ -th class, one should increase the number of indications in this class by 1, and in the opposite case one should increase the number of indications by 1 for all classes in the combined class. Finally, we choose the class whose counter achieved the highest value of wins. The authors [18] prefer the “one-against-all” method because of the linear dependence of the number of SVM networks on the number of pattern types in the training set.

The possibility of another extended interpretation of the results obtained by the “one-against-all” method is presented in [15]. According to [18], the higher the value of the function  $f_m(\mathbf{x})$  (18), the more reliable the classification result is. In the case of a linear kernel, such a criterion of reliability may be the distance of the recognized object from the separation plane, which is equal

$$f_m(\mathbf{x}) = \mathbf{w}_m^T \mathbf{x} + b_m, \quad m \in \mathbf{M} = \{1, \dots, M\}. \quad (19)$$

In point V-C of the work, the value of the  $f_m(\mathbf{x})$  decision function was used to construct the basic belief assignment on the results of the SVM multi-class classification.

In SVM multi-class classification, each  $m$ -th classifier determines the value of its decision function  $f_m(\mathbf{x})$ . Considering the classification results as a whole, one of three situations may occur:

- 1) Only one  $f_m(\mathbf{x})$  has a positive value, and all the others are negative. In this case, the number of positive classifier specifies the pattern type number.
- 2) More than one of the  $f_m(\mathbf{x})$  are positive. If we assume that the higher the value of the decision function, the more reliable the classification result is, then the number of the classifier corresponding to the highest value of the function  $f_m(\mathbf{x})$  determines the number of the pattern type, what can be written as follows

$$m^* = \arg \max_{m \in \mathbf{M} = \{1, \dots, M\}, f_m(\mathbf{x}) > 0} f_m(\mathbf{x}), \quad (20)$$

where  $id$  is the number of the recognized type.

- 3) None of the  $f_m(\mathbf{x})$  values is positive. That should be regarded as the new image belongs to a maritime object which type is not included in the training set (unknown object).

## V. FUSION OF INFORMATION FROM TWO SVM CLASSIFIERS

### A. The process of fusion of information from two classifiers

Each classifier used in the work transfers to the fusion module a vector of measures of basic belief assignment on a set of types of objects in the training set. In this work, it was assumed that both classifiers have the same training set. The set of possible hypotheses related to individual types of objects (the frame of discernment) is as follows

$$\Theta = \{\theta_i, i = 1, \dots, M\}, \quad (21)$$

wherein the index  $i$  numbers the type of the maritime object whose images are stored in a training set.  $M$  is the number of pattern types.

The hypotheses are exhaustive (21) and mutually exclusive, i.e.

$$\theta_i \cap \theta_j = \begin{cases} \theta_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \quad (22)$$

Each classifier sends its decisions in the form of a BBA measure vector (BBA - basic belief assignment).

$$\mathbf{m}_i = [m_i(\theta_1), \dots, m_i(\theta_M)], \quad (23)$$

wherein the index  $i = 1$  determines the BBA measure vector calculated by the SVM classifier of linear descriptor vectors and the index  $i = 2$  determines the BBA measure vector calculated by the SVM classifier of HOG feature vectors.

The information fusion procedure is described by the following formula:

$$\mathbf{m}_F = R_F(\mathbf{m}_1, \mathbf{m}_2), \quad (24)$$

wherein  $\mathbf{m}_F$  is a vector of the vector of BBA masses determined by the  $R_F$  information fusion rule based on the vectors  $\mathbf{m}_1$  and  $\mathbf{m}_2$  of BBA masses.

### B. The proportional conflict redistribution rule PCR5 for two bbas (two sources)

$$\begin{aligned} m'_F(\theta_i) &= m_{\text{PCR5}}(\theta_i) \\ &= m_{12}(\theta_i) + \sum_{\substack{j=1, \dots, M \\ \theta_i \cap \theta_j = \emptyset}} S_{12}^{\text{PCR5}}(\theta_i, \theta_j) \\ &= m_{12}(\theta_i) + \sum_{\substack{j=1, \dots, M \\ j \neq i}} S_{12}^{\text{PCR5}}(\theta_i, \theta_j), \end{aligned} \quad (25)$$

where

$$S_{12}^{\text{PCR5}}(\theta_i, \theta_j) = \begin{cases} 0, & \text{for } c_1 = 0 \text{ or } c_2 = 0, \\ \frac{m_1(\theta_i)^2 m_2(\theta_j)}{m_1(\theta_i) + m_2(\theta_j)} + \frac{m_2(\theta_i)^2 m_1(\theta_j)}{m_2(\theta_i) + m_1(\theta_j)}, & \\ \text{for } c_1 \neq 0 \wedge c_2 \neq 0, & \end{cases} \quad (26)$$

where  $c_1 \triangleq m_1(\theta_i) + m_2(\theta_j)$  and  $c_2 \triangleq m_2(\theta_i) + m_1(\theta_j)$ , and wherein

$$m_{12}(\theta_i) = m_1(\theta_i) m_2(\theta_i). \quad (27)$$

In the formula (25), the component  $S_{12}^{\text{PCR5}}(\theta_i, \theta_j)$  is equal to zero if both denominators are equal to zero. In the formula (26), if a denominator is zero, then component is discarded.

### C. The method of determining the BBA for the SVM classifiers

The procedure of image recognition by means of the SVM method “one-against-all” in accordance with the content of point IV-C and [15,18] allows to determine the basic belief assignment BBA on a set of pattern types. Each  $k$ -th pattern type is associated with one SVM and the identification process determines the value of the decision function

$$f_k(\mathbf{x}) = \sum_{i=1}^{N_k} \alpha_i^k y_i^k K(\mathbf{x}_i^k, \mathbf{x}) + b_k. \quad (28)$$

The value of this function can be used to determine the value of the degree of belief that the recognized (tested) object belongs to the class with the number  $k$  ( $k = 1, \dots, M$ ). In [8] it was proposed to use the logistic regression function in accordance with the following formula

$$m(\mathbf{x}, k) = \frac{e^{f_k(\mathbf{x})}}{1 + e^{f_k(\mathbf{x})}}. \quad (29)$$

In the formula (29)  $k$  numbers SVMs. As one can see  $0 < m(\mathbf{x}, k) < 1$ . The above measure is not normalized, therefore we normalize it by

$$\tilde{m}(\mathbf{x}, k) = \frac{m(\mathbf{x}, k)}{\sum_{i=1}^M m(\mathbf{x}, i)}. \quad (30)$$

One should note that the above method of mass determination is simplified, because it does not take into account the lack of the type of image pattern corresponding to the recognized (tested) image.

Each SVM network in the “one-against-all” method calculates the value of the decision function  $f_k(\mathbf{x})$  used to calculate the  $k$ -th component of the BBA vector using (29) and (30). The construction of the “one-against-all” method justifies the form of the frame of discernment (21). In this method, there are as many SVM networks as there are training image classes in the training set. The way of calculating the BBA vector (29) and (30) ensures that for each test image, its specific type will be determined. It follows from these considerations that the basic belief assignment is a Bayesian assignment regardless of how the vector of image features (linear descriptor or HOG) is determined. Formulas (21) and (22) are supplemented by the following formula:

$$\sum_{i=1}^M m_1(\theta_i) = \sum_{i=1}^M m_2(\theta_i) = 1. \quad (31)$$

## VI. RESULTS OF MARITIME OBJECT RECOGNITION USING SVM CLASSIFIER WITH LINEAR IMAGE DESCRIPTORS AND HOG IMAGE DESCRIPTORS AND THE FUSION OF THESE CLASSIFIERS

The study had a limited number of images of maritime objects of nine classes. The size of the training set is presented in the Table I.

Table I  
THE SIZE OF THE TRAINING SET.

Maritime object class	Size of the training set
1	56
2	38
3	54
4	22
5	56
6	38
7	22
8	30
9	24

Two training sets were created for each class. In one of the sets, the images were represented by linear descriptor vectors, while in the other, the images were represented by HOG feature vectors.

The next step was to train two types of SVM models. Thus, a total of 18 models were created (two models for each of the nine classes). The linear activation function of the kernel was used in this work according to the guidelines given in [6].

*A. Results of maritime objects recognition for SVM classifier with linear descriptor vectors*

Due to the small number of available original images from FLIR cameras, it was decided to extend the test sets with images on which Gaussian noise and salt and pepper noise were applied. Finally, the recognition of maritime objects was carried out on the basis of test sets, which included 130 images of objects belonging to each class. Nine training classes were used, so a total of 1170 images were tested. As part of the work, the achieved accuracy of classification was examined, and the results were presented in the form of tables in the following subsections.

The results of recognition of maritime object images using the SVM classifier with linear descriptor vectors are presented in the Table II. The right column of Table II corresponds to the mean value of correctly recognized maritime object (CRMO).

Table II  
THE RESULTS OF RECOGNITION OF MARITIME OBJECT IMAGES USING THE SVM CLASSIFIER WITH LINEAR DESCRIPTOR VECTORS.

Object type	# of tested images	# of correct recognitions	# of incorrect recognitions	Mean value of CRMO
1	130	130	0	100%
2	130	130	0	100%
3	130	130	0	100%
4	130	130	0	100%
5	130	130	0	100%
6	130	124	6	95.38%
7	130	130	0	100%
8	130	130	0	100%
9	130	130	0	100%
Total	1170	1164	6	<b>99.48%</b>

Only 6 out of 1170 images were not recognized correctly. The accuracy was 100% for all classes except class 6, despite the image noise. The average value of correctly

recognized objects in all classes was 99.48%. Such high accuracy may be due to the ability of the SVM network to generalize knowledge. This network is also characterized by low sensitivity to the number of used training data. Therefore, even achieving high accuracy is possible even with a small database [5]. In addition, the correctness of the classification can be positively influenced by the preprocessing of images, as well as a well-chosen kernel function. In this case, a linear function was used, which works well with input data of sufficiently many dimensions [8].

The Table III contains the measured execution times of the classification task by SVM models using linear image descriptors.

Table III  
RECOGNITION TIMES FROM THE TEST SET BY THE SVM CLASSIFIER USING LINEAR IMAGE DESCRIPTORS.

Maritime object class	Classification time of images from the test set (in s)
1	0.010306
2	0.011566
3	0.010264
4	0.011591
5	0.011203
6	0.010379
7	0.009611
8	0.011033
9	0.111119

*B. Results of maritime objects recognition for the SVM classifier using the histogram of oriented gradients*

The results obtained by testing an SVM network using a histogram of oriented gradients (HOG) are presented in the Table IV.

Table IV  
RESULTS OF MARITIME OBJECT IMAGE RECOGNITION WITH THE SVM CLASSIFIER USING THE HISTOGRAM OF ORIENTED GRADIENTS.

Object type	# of tested images	# of correct recognitions	# of incorrect recognitions	Mean value of CRMO
1	130	130	0	100%
2	130	130	0	100%
3	130	130	0	100%
4	130	130	0	100%
5	130	130	0	100%
6	130	129	1	99.23%
7	130	130	0	100%
8	130	130	0	100%
9	130	128	2	98.46%
Total	1170	1167	3	<b>99.74%</b>

The use of HOG allowed to achieve an even greater number of correctly recognized maritime objects. The average value of correctly recognized objects belonging to all classes was thus 99.74%. Therefore, it can be concluded that the use of SVM network learning data in the form of histograms of gradients allows to increase the accuracy of classification, compared to the use of full image information (brightness of all pixels).

The Table V contains the measured execution times of the classification task by SVM models using the histogram of oriented gradients.

Table V  
RECOGNITION TIMES FROM THE TEST SET BY THE SVM CLASSIFIER USING THE HISTOGRAM OF ORIENTED GRADIENTS.

Maritime object class	Classification time of images from the test set (in s)
1	0.002925
2	0.002263
3	0.002203
4	0.002268
5	0.002443
6	0.002189
7	0.002088
8	0.002155
9	0.002172

The time required to recognize objects from the test set by SVM classifiers whose learning data was in the form of HOG was an order of magnitude less than the time required by classifiers operating on the brightness of all image pixels. This is due to the reduced dimensionality of the problem, as the input vectors were characterized by a length of 4800 bytes for the whole image, while in the case of HOG their length was only 1368 bytes.

C. Information fusion results from SVM classifiers obtained using the PCR5 rule

The Table VI shows the classification results obtained by using a fusion of SVM classifiers based on training sets containing the brightness of all image pixels histograms of oriented gradients.

Table VI  
INFORMATION FUSION RESULTS FROM BOTH TYPES OF SVM CLASSIFIERS OBTAINED USING THE PCR5 RULE.

Object type	# of tested images	# of correct recognitions	# of incorrect recognitions	Mean value of CRMO
1	130	130	0	100%
2	130	130	0	100%
3	130	130	0	100%
4	130	130	0	100%
5	130	130	0	100%
6	130	129	1	99.23%
7	130	130	0	100%
8	130	130	0	100%
9	130	129	1	99.23%
Total	1170	1168	2	<b>99.83%</b>

The use of information fusion from both types of SVM classifiers allowed to obtain the largest number of correctly recognized objects. Only 2 objects out of 1170 were not assigned to the appropriate class. In this case, the average accuracy of classification of images belonging to all classes was 99.83%. This is the largest value compared to the previously described recognition methods.

It can therefore be concluded that fusion of information from different sources is reasonable and improves the quality of classification.

VII. CONCLUSIONS

The Table VII compares the effectiveness of maritime object recognition based on two SVM classifiers with linear image descriptor and HOG descriptor, and the classifier with information fusion according to the PCR5 rule. This comparison allows to positively assess the purposefulness of using the classifier fusion.

Table VII  
INFORMATION FUSION RESULTS FROM BOTH TYPES OF SVM CLASSIFIERS OBTAINED USING THE PCR5 RULE.

Object type	# of tested images	Effectiv. of SVM (linear descript.)	Effectiv. of SVM (HOG)	Effectiveness of PCR5 rule for SVM fusion
1	130	100%	100%	100%
2	130	100%	100%	100%
3	130	100%	100%	100%
4	130	100%	100%	100%
5	130	100%	100%	100%
6	130	95.38%	99.23%	99.23%
7	130	100%	100%	100%
8	130	100%	100%	100%
9	130	100%	98.46%	99.23%
Total	1170	<b>99.48%</b>	<b>99.74%</b>	<b>99.83%</b>

The results obtained are consistent with the information contained in publications on SVM networks. They allow to achieve very high efficiency, even when images are noisy. In addition, the support vector technique is not time-consuming, despite the use of data of significant volume, it shows low sensitivity to smaller training sets [5,13]. Both tested classifiers showed high efficiency in recognizing marine objects, however, the use of the histogram of oriented gradients allows for slightly higher accuracy than the use of training data in the form of a vector containing the brightness of all image pixels. The best results, however, were obtained using a fusion of both of these classifiers. The fusion was performed using one of the most accurate rules for proportional conflict redistribution - PCR5. The use of fusion helped to improve efficiency, so it can be considered reasonable to combine information from different classifiers when conflicts arise.

REFERENCES

- [1] [Ch. Dong, J. Liu, F. Xu, *Ship Detection in Optical Remote Sensing Images Based on Saliency and a Rotation-Invariant Descriptor*, Beijing, 2018.
- [2] T.R. Chandrashekar, A.K. Gautam, *Face Recognition based on Histogram of Oriented Gradients, Local Binary Pattern and SVM/HMM Classifiers*, IJESRT, International Journal of Engineering Sciences & Research Technology, Chandrashekar, 3(8), August, 2014.
- [3] N. Dalal, B. Triggs, *Histograms of oriented gradients for human detection*, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, pp. 886–893, San Diego, CA, USA, 2005.
- [4] T. Pietkiewicz, *Application of fusion of two classifiers based on principal component analysis method and time series comparison to recognize maritime objects upon FLIR images*, Proceedings of SPIE 11055, XII Conference on Reconnaissance and Electronic Warfare Systems, 110550Z, 2019.
- [5] S. Osowski, *Neural networks for information processing*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2020 (in Polish).

- [6] C.W. Hsu, C.C. Chang, C.J. Lin, *A practical guide to support vector classification*, 19 May 2016, <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> (10 December 2019).
- [7] V. Vapnik, *Statistical learning theory*, Wiley, New York, 1998.
- [8] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd Edition, Springer, New York, 2000.
- [9] F. Smarandache, J. Dezert, *Applications and Advances of DSMT for Information Fusion (Collected works)*, Volume 1, American Research Press (ARP), Rehoboth, 2004.
- [10] P. Djiknavorian, D. Grenier, P. Valin, *Analysis of information fusion combining rules under the DSMT theory using ESM input*, 10th International Conference on Information Fusion (FUSION 2007), Québec, Canada, 2007.
- [11] F. Smarandache, J. Dezert., *Proportional Conflict Redistribution Rules for Information Fusion*, pp. 3–68 in *Advances and Applications of DSMT for Information Fusion (Collected works)*, Vol. 2, American Research Press, Rehoboth, 2006.
- [12] K. Crammer, Y. Singer, *On the learnability and design of output codes for multiclass problems*, *Machine Learning*, Vol. 47, pp. 201–233, 2002.
- [13] B. Waske, J.A. Benediktsson, *Fusion of Support Vector Machines for Classification of Multisensor Data*, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 45, No. 12, pp. 3858–3866, Dec. 2007.
- [14] C.J.C. Burges, *A tutorial on support vector machines for pattern recognition*, *Data Mining and Knowledge Discovery*, *Data Mining and Knowledge Discovery*, Vol. 2, Issue 2, pp. 121–167, June 1998.
- [15] Y. Li, S. Gong, H. Liddel, *Support Vector Regression and Classification Based Multi-view Face Detection and Recognition*, *Proc. of 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pp. 300–305, Grenoble, France, 2000.
- [16] Y. Yang, J. Wang, Y. Yang, *Exploiting rotation invariance with SVM classifier for microcalcification detection*, in *Proc. of 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, Barcelona, 02–05 May 2012.
- [17] M. Chudy, *Selected optimization methods*, Bellona, Warszawa, 2001 (in Polish).
- [18] B. Heisele, P. Ho, T. Poggio, *Face recognition with support vector machines: global versus component-based approach*, in *Proc. of 8th IEEE International Conference on Computer Vision (ICCV 2001)*, pp. 688–694, Vancouver, BC, Canada, 2001.
- [19] T. Pietkiewicz, *Removal of conflicts in fusion of identification information from ELINT-ESM sensors*, *Proceedings of SPIE 11055, XII Conference on Reconnaissance and Electronic Warfare Systems, 110550S*, 2019.

# Multisource Weighted Domain Adaptation with Evidential Reasoning for Activity Recognition

Yilin Dong<sup>a</sup>, Xinde Li<sup>b</sup>, Jean Dezert<sup>c</sup>, Rigui Zhou<sup>a</sup>,  
Changming Zhu<sup>a</sup>, Lei Cao<sup>a</sup>, Mohammad Omar Khyam<sup>d</sup>, Shuzhi Sam Ge<sup>e</sup>

<sup>a</sup>College of Information Engineering, Shanghai Maritime University, Shanghai, China.

<sup>b</sup>Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing, China.

<sup>c</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>d</sup>School of Electrical Engineering, Central Queensland University, Melbourne, Australia.

<sup>e</sup>Social Robotics Laboratory, ECE Dept., Interactive Digital Media Institute, National University of Singapore, Singapore.

Emails: yldong@shmtu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr, rgzhou@shmtu.edu.cn,  
cmzhu@shmtu.edu.cn, lcao@shmtu.edu.cn, m.khyam@cqu.edu.au, samge@nus.edu.sg

**Abstract**—In recent years, wearable sensor-based human activity recognition (HAR) is becoming more and more attractive, especially in health monitoring and sports management. However, in order to obtain high-quality HAR, it is often necessary to get sufficient labeled activity data, which is very difficult, time-consuming and costly in a natural environment. To tackle this problem, multi-source domain adaptation is a promising method that aims to learn enough multi-source prior knowledge from labeled activity data, and then transfer this learned knowledge to the target unlabeled dataset. Thus, this paper presents a novel multi-source weighted domain adaptation with evidential reasoning (w-MSDAER) for HAR, which can effectively utilize complementary knowledge between multiple sources. Specifically, we first use the strategy of distribution alignment to learn local domain-invariant classifiers based on multi-source domains. And then, the reliabilities of these derived classifiers are comprehensively evaluated according to the belief function based the technique for order preference by similarity to ideal solution (BF-TOPSIS). Finally, the discounting fusion method is used to fuse the local classification results. Comprehensive experiments are conducted on two open-source datasets, and the results show that the proposed w-MSDAER significantly outperforms other state-of-art methods.

**Keywords:** human activity recognition, multi-source domain adaptation, evidential reasoning, reliability assessment.

## I. INTRODUCTION

### A. Background and Research Motivation

In recent years, wearable sensor-based human activity recognition (HAR) usually uses the raw signals collected by wearable sensors to identify human activities, and to help the patients to deal with chronic injuries or provide personalized medical advice. Given its good application prospect, sensor-based HAR has been comprehensively discussed in recent surveys [1], [2] and has also been widely used in many real-life scenarios such as health-care, ubiquitous computing, and human-computer interaction [3]–[8]. Generally speaking, high-precision activity recognition relies on good generalized recognition models, which means that sufficient and labeled data is always acquired to train reliable models in advance.

However, one typical and common scenario is that labeled data collected from specific positions or individuals are often limited. Still, we hope that the recognition models learned from such labeled data can identify the unknown activities of many other positions or people [9]. For example, the pre-trained HAR model built in the smartwatches needs to identify each user’s activities. However, it is impossible to label all consumers’ personal data in the manufacturing process of smart watches. Many recent cross-domain references [10]–[12] pointed out that the recognition model learned on some positions or specific people can not be well generalized to other positions and people. To directly demonstrate the negative effects of domain differences, the cross-location and cross-person activity recognition experiments were conducted on the mHealth<sup>1</sup> dataset based on previous works [3], [13]: 1) for the same subject, the Long-Short Term Memory (LSTM) was first trained by data collected from the accelerometer installed on the chest, and then this well-trained LSTM model was applied to identify the unknown activities based on the data collected by those sensors located on the chest, left-ankle and right-lower-arm, respectively; 2) the LSTM model trained for Subject1 (St1) was used to recognize the unknown activities of St1, Subject2 (St2), Subject3 (St3) and Subject4 (St4), respectively. As shown in Fig.1, the average accuracies of self-activity recognitions are 88.65% for *Chest* → *Chest* and 83.61% for *St1* → *St1*, while the accuracies of cross-position and cross-person recognition were only 14.84% (*Chest*→*Left Ankle*), 5.64% (*Chest*→*Right Lower Arm*) and 32.84% (*St1*→*St2*), 34.34% (*St1*→*St3*), 35.6% (*St1*→*St4*). This result fully shows that the domain difference affects the classifiers’ performances.

To reduce the negative impacts of domain differences, many recent works have successfully applied domain adaptation (DA) approaches to deal with cross-domain recognition problems. The classical single-domain adaptations often rely on the

<sup>1</sup><http://archive.ics.uci.edu/ml/datasets/mhealth+dataset>.

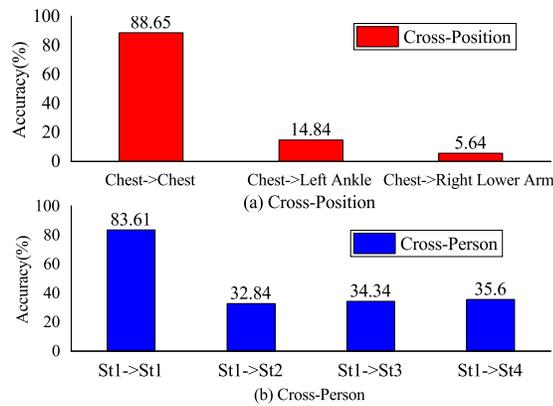


Figure 1. Classification Accuracy in Cross-Domain Activity Recognition.

domain-invariant feature extractions [11], or domain mapping strategies [12]. For example, the Joint Distribution Adaptation (JDA) method was proposed by Long et al. [14] to adapt both the marginal and conditional distributions simultaneously. Since deep neural networks have the capability of powerful feature extraction, many deep feature transformation methods have also been presented to adapt the distribution. Ghifary et al. [15] proposed a domain adaptive neural network (DaNN) to reduce the distribution mismatch between the source and target domains. However, those aforementioned methods only consider the situation of single source domain. In practical applications, there may be multiple source domains, and how to effectively utilize the complementary knowledge between different source domains has been a serious concern issue [16], [17]. Currently, many sophisticated techniques have been proposed to take advantage of differences among source domains or the relationship between sources and targets. Zhao et al. [18] proposed a separate domain classifier for each source domain and computed a loss based on the lowest domain error among these classifiers. Liu et al. [19] presented a novel evidential framework for combining multi-source domain adaptations. In [19], the reliabilities of source domains were directly evaluated and the multiple classification results derived from multi-source were fused with a novel decision-level cautious combination rule. Moreover, Liu et al. [20] also proposed a new method called distribution adaptation based on evidence theory to improve the classification accuracy by combining the complementary information derived from both the source and target domains.

As we discussed earlier in Fig.1, traditional sensor-based action recognition models have difficulties in solving cross-domain HAR problems. To tackle these problems, Hong et al. [21] proposed a novel single-source DA method based on semi-population strategy; Wang et al. [22] presented a novel deep network coupled with transfer learning for cross-position HAR (TNNAR); An extreme learning machine based kernel fusion was proposed by Wang et al. [23] to deal with

domain alignment in HAR. Besides, Garrett Wilson et al. [24] presented a convolutional deep domain adaptation model for time series data (CoDATS) from multiple source domains to improve the accuracy over prior single-source methods. However, the mentioned TNNAR or CoDATS did not fully consider the reliabilities/weights of source domains in their multi-source domain adaptations, which in some degree affects the final classification results. Inspired by the weighted combination strategy for multi-source domain adaptation proposed in [19], we propose a new multi-source weighted domain adaptation with evidential reasoning (w-MSDAER) for the human activity recognition problem. The difference in weight calculation between w-MSDAER and Liu's method in [19] is mainly reflected in the stage of reliability evaluation for source domains: Liu's method in [19] can be regarded as the pre-weight calculation strategy before basic belief assignments (BBAs) generation. In [19], each reliability of the source domain is directly evaluated by using the domain distance before and after distribution matching steps; In our proposed w-MSDAER, the reliabilities of source domains are indirectly evaluated based on the multi-criteria strategy after BBA generation. These BBAs are obtained from the outputs of local domain-invariant classifiers trained by the source domains and comprehensively evaluated from two aspects: the distance degrees between BBAs and the imprecision degree inside each BBA. Thus, w-MSDAER can be regarded as the post indirect weight evaluation method compared to Liu's method in [19].

### B. Challenges

It is essential to fuse the complementary multi-domain knowledge effectively. However, two main issues need to be addressed:

- How to comprehensively evaluate the reliability of different source domains? Since the classifiers using data from multi-source domains may have distinct abilities to classify activities in the unknown domain, it is necessary to analyze the reliabilities of multi-source domains in the process of domain adaptation;
- How to fuse the outputs of local domain-invariant classifiers learned from different source domains? Considering that conflicts may exist between the classification results from different source domains, our fusion rules are required to effectively solve the highly conflicting fusion problem.

### C. Main Contributions

To solve these two aforementioned problems, we first use the manifold embedded distribution alignment (MEDA) to learn local domain-invariant classifiers based on source domains. And then, the outputs of these classifiers are transformed into the BBAs and evaluated by the multi-criteria evaluation strategy: BF-TOPSIS<sup>2</sup>.

The main contributions of this work are summarized as follows:

<sup>2</sup>BF-TOPSIS is an extension of the technique for order preference by similarity to ideal solution (TOPSIS) [25] based on belief functions (BF).

- The novel fusion framework of weighted multi-source domain adaptation based on evidential reasoning (w-MSDAER) is presented. Here, w-MSDAER can combine the complementary knowledge among different source domains;
- In w-MSDAER, the reliabilities of local domain-invariant classifiers learned from multi-source domains are comprehensively assessed according to the BF-TOPSIS strategy. In BF-TOPSIS, two well-selected criteria are chosen: distances between BBAs and the self-imprecision degree inside the BBA;
- Through comprehensive experiments on two public activity recognition data sets, the superiority of w-MSDAER is shown.

The rest of this paper is organized as follows. In section II, the basics of evidential reasoning is presented. Section III describes the proposed w-MSDAER. The experimental results are given in details in Section IV. Finally, the article concludes and gives several future research directions in the last section.

## II. BASICS OF EVIDENTIAL REASONING

This section briefly introduces the basic knowledge of evidence reasoning, also known as Dempster-Shafer Theory (DST), which is necessary for presenting the proposed w-MSDAER.

In DST, the concept of Frame of Discernment (FoD) represents a set of exhaustive and exclusive elements which is denoted as  $\Theta \triangleq \{\theta_1, \dots, \theta_n\}$  ( $n \geq 2$ ). The power set of  $\Theta$ , which is the set of all subsets of  $\Theta$  (including the empty set  $\emptyset$ , and  $\Theta$  itself), is denoted  $2^\Theta$  because its cardinality is exactly equal to  $2^{|\Theta|}$ . A BBA  $m(\cdot)$  is defined by the mapping:  $2^\Theta \mapsto [0, 1]$ , verifying  $m(\emptyset) = 0$  and  $\sum_{\theta \in 2^\Theta} m(\theta) = 1$ .  $Bel(\theta) \triangleq \sum_{\theta' \in 2^\Theta | \theta' \subseteq \theta} m(\theta')$  and  $Pl(\theta) \triangleq \sum_{\theta' \in 2^\Theta | \theta' \cap \theta \neq \emptyset} m(\theta')$  define the belief and plausibility function, respectively. The interval  $BI(\theta) \triangleq [Bel(\theta), Pl(\theta)]$  is called the belief interval of  $\theta$ , which is usually interpreted as the interval where the value of unknown probability of  $\theta$  must belong.

In order to combine two distinct sources of evidence, the classical Dempster Shafer rule (DS) in [26] was proposed and defined by  $m_{DS}(\emptyset) = 0$  and  $\forall \theta \in 2^\Theta \setminus \{\emptyset\}$ :

$$m_{DS}(\theta) = \frac{\sum_{\theta', \theta'' \in 2^\Theta | \theta' \cap \theta'' = \theta} m_1(\theta') m_2(\theta'')}{1 - \sum_{\theta', \theta'' \in 2^\Theta | \theta' \cap \theta'' = \emptyset} m_1(\theta') m_2(\theta'')}. \quad (1)$$

To palliate DS rule drawbacks (see discussions in [27]), in Dezert-Smarandache Theory (DSMT) [28], the very interesting Proportional Conflict Redistribution-5 (PCR5) was defined by:  $m_{PCR5}(\emptyset) = 0$  and  $\forall \theta \in 2^\Theta \setminus \{\emptyset\}$ :

$$m_{PCR5}(\theta) = m_{12}(\theta) + \sum_{\theta' \in 2^\Theta \setminus \{\emptyset\} | \theta \cap \theta' = \emptyset} \left[ \frac{m_1(\theta)^2 m_2(\theta')}{m_1(\theta) + m_2(\theta')} + \frac{m_2(\theta)^2 m_1(\theta')}{m_2(\theta) + m_1(\theta')} \right], \quad (2)$$

where  $m_1(\cdot), m_2(\cdot)$  are two independent BBAs and  $m_{12}(\theta) = \sum_{\theta', \theta'' \in 2^\Theta | \theta' \cap \theta'' = \theta} m_1(\theta') m_2(\theta'')$ .

## III. MULTI-SOURCE WEIGHTED DOMAIN ADAPTATION WITH EVIDENTIAL REASONING FOR HAR

In this part, we first raise the issue of cross-domain HAR and then briefly present the specific steps of local domain-invariant classifier learned from MEDA and the multi-criteria assessment of these derived classifiers' reliabilities by using BF-TOPSIS [25]. At last, we present how to fuse all outputs of local classifiers based on the discounting combination rule and make the final decision.

### A. Problem Definition

Assuming that there exists  $M$  labeled source domains:  $\mathcal{D}_{s_1}, \mathcal{D}_{s_2}, \dots, \mathcal{D}_{s_M}$  and an unlabeled target domain  $\mathcal{D}_t$  in the multi-source DA problem. The source domains  $\mathcal{D}_s = \{(\mathbf{x}_u^{\mathcal{D}_s}, y_u^{\mathcal{D}_s})\}_{u=1}^{N_s}$  contains  $N_s$  labeled samples and  $\mathbf{x}_u^{\mathcal{D}_s}$  follows the domain distribution  $pro^{\mathcal{D}_s}(\mathbf{x}, y)$  and  $y_u^{\mathcal{D}_s} \in \{1, 2, \dots, n\}$  ( $n$  represents the number of categories) is its related label for samples in the source domain. Similarly,  $\mathcal{D}_t = \{\mathbf{x}_v^{\mathcal{D}_t}\}_{v=1}^{N_t}$  is the target domain, which includes  $N_t$  unlabeled samples, and  $\mathbf{x}_v^{\mathcal{D}_t}$  follows the target distribution  $pro^{\mathcal{D}_t}(\mathbf{x}, y)$ . In our problems discussed in this paper,  $\mathcal{D}_s$  and  $\mathcal{D}_t$  share the consistent feature spaces and the same label spaces:  $\mathcal{X}^{\mathcal{D}_s} = \mathcal{X}^{\mathcal{D}_t}$  and  $\mathcal{Y}^{\mathcal{D}_s} = \mathcal{Y}^{\mathcal{D}_t}$ . However,  $\mathcal{D}_s$  and  $\mathcal{D}_t$  belong to different distributions:  $pro^{\mathcal{D}_s}(\mathbf{x}, y) \neq pro^{\mathcal{D}_t}(\mathbf{x}, y)$ . With the help of the multi-source domains  $\mathcal{D}_{s_1}, \mathcal{D}_{s_2}, \dots, \mathcal{D}_{s_M}$ , the goal of our task is to obtain the label  $y^{\mathcal{D}_t}$  for the unlabeled samples in target domain. Here, we use a simple cross-position activity recognition problem to give more clear and specific descriptions of relevant terms discussed above. Assuming that there exists two source domains:  $\mathcal{D}_{s_1} = \{(\mathbf{x}^{\mathcal{D}_{s_1}}, y^{\mathcal{D}_{s_1}})\}$  and  $\mathcal{D}_{s_2} = \{(\mathbf{x}^{\mathcal{D}_{s_2}}, y^{\mathcal{D}_{s_2}})\}$ , where  $\mathcal{D}_{s_1}$  and  $\mathcal{D}_{s_2}$  represent the source domains derived from accelerators located on the chest and left arm, respectively. Besides,  $\mathbf{x}^{\mathcal{D}_{s_1}}$  and  $\mathbf{x}^{\mathcal{D}_{s_2}}$  are the raw signals collected by accelerators and the mathematical symbols  $y^{\mathcal{D}_{s_1}}$  and  $y^{\mathcal{D}_{s_2}}$  represent the categories of activities, such as standing, walking, running. Our main task is to classify those unlabeled samples  $\mathbf{x}^{\mathcal{D}_t}$  in the target domain  $\mathcal{D}_t = \{\mathbf{x}^{\mathcal{D}_t}\}$ , which are collected from the accelerator located on the right ankle.

### B. Local domain-invariant classifier based on manifold embedded distribution alignment

As mentioned in [19], our goal is to use the classifier  $f: \mathbf{x}^{\mathcal{D}_t} \mapsto y^{\mathcal{D}_t}$  learned from multi-source domains to realize the classification of those unlabeled samples in the target domain. Here, we use the MEDA strategy proposed by Wang et al. [29] to learn the local domain-invariant classifier before combination. More detailed discussions about MEDA are given in [29].

Specifically, according to the structural risk minimization (SRM) [30], the domain-invariant learned classifier  $f(\cdot)$  can be represented as

$$f = \arg \min_{f \in \mathcal{H}_K} \sum_{u=1}^{Ns} (y_u - f(\mathbf{z}_u))^2 + \eta \|f\|_K^2 + \lambda \overline{\mathcal{D}}_f(\mathcal{D}_s, \mathcal{D}_t) + \rho R_f(\mathcal{D}_s, \mathcal{D}_t), \quad (3)$$

where  $\mathbf{z}_u = g(\mathbf{x}_u)$  is the transformed manifold features; Considering the characteristics of computational efficiency, Geodesic Flow Kernel (GFK) [31] is applied to learn  $g(\cdot)$ ; The squared norm of  $f$  is denoted as  $\|f\|_K^2$ ; The dynamic distribution alignment is represented by  $\overline{\mathcal{D}}_f$ ; Besides,  $R_f$  is the Laplacian regularization [29]; Based on kernel function  $K(\cdot, \cdot)$ , the Hilbert space  $\mathcal{H}_K$  can be derived;  $\eta, \lambda, \rho$  are regularization parameters accordingly.

To obtain the local domain-invariant classifier  $f$ , the details of each term in (3) are reformulated as follows:

1) *Local classifier learned by SRM*: According to the representer theorem [32],  $f$  can be expressed by

$$f(\mathbf{z}) = \sum_{u=1}^{Ns+Nt} \beta_u K(\mathbf{z}_u, \mathbf{z}), \quad (4)$$

where the coefficient vector is denoted by  $\beta = (\beta_1, \beta_2, \dots, \beta_u, \dots, \beta_{(Ns+Nt)})^T \in \mathbb{R}^{(Ns+Nt) \times 1}$ ;  $K(\cdot, \cdot)$  is a kernel function. Afterwards, we can use the SRM strategy for  $\mathcal{D}_s$ :

$$\sum_{u=1}^{Ns} (y_u - f(\mathbf{z}_u))^2 + \eta \|f\|_K^2 = \sum_{u=1}^{Ns+Nt} \mathbf{A}_{uu} (y_u - f(\mathbf{z}_u))^2 + \eta \|f\|_K^2 = \|(\mathbf{Y} - \beta^T \mathbf{K}) \mathbf{A}\|_F^2 + \eta \text{tr}(\beta^T \mathbf{K} \beta), \quad (5)$$

where  $\mathbf{K} \in \mathbb{R}^{(Ns+Nt) \times (Ns+Nt)}$  represents the kernel matrix with  $\mathbf{K}_{uv} = K(\mathbf{z}_u, \mathbf{z}_v)$ ; The Frobenious norm and trace operators are denoted by  $\|\cdot\|_F$ ,  $\text{tr}(\cdot)$ , respectively;  $\mathbf{A} \in \mathbb{R}^{(Ns+Nt) \times (Ns+Nt)}$  is a binary diagonal domain indicator matrix with  $\mathbf{A}_{uu} = 1$  if  $u \in \mathcal{D}_s$ ,  $\mathbf{A}_{uu} = 0$  otherwise.  $\mathbf{Y} = [y_1, \dots, y_{Ns+Nt}]$  is the label matrix from source and the target domains.

2) *Distribution alignment*: Here, the distribution alignment is defined by:

$$\overline{\mathcal{D}}_f(\mathcal{D}_s, \mathcal{D}_t) = \text{tr}(\beta^T \mathbf{K} \mathbf{M} \mathbf{K} \beta), \quad (6)$$

where  $\mathbf{M} = (1 - \mu) \mathbf{M}_0 + \mu \sum_{h=1}^n \mathbf{M}_h$  represents the MMD matrix, and the elements inside can be computed [29] by

$$(\mathbf{M}_0)_{uv} = \begin{cases} \frac{1}{Ns^2}, z_u, z_v \in \mathcal{D}_s; \\ \frac{1}{Nt^2}, z_u, z_v \in \mathcal{D}_s; \\ -\frac{1}{Ns \cdot Nt}, \text{otherwise.} \end{cases} \quad (7)$$

$$(\mathbf{M}_h)_{uv} = \begin{cases} \frac{1}{Ns_h^2}, \mathbf{z}_u, \mathbf{z}_v \in \mathcal{D}_s^{(h)}; \\ \frac{1}{Nt_h^2}, \mathbf{z}_u, \mathbf{z}_v \in \mathcal{D}_t^{(h)}; \\ -\frac{1}{Nt_h \cdot Ns_h}, \begin{cases} \mathbf{z}_u \in \mathcal{D}_s^{(h)}, \mathbf{z}_v \in \mathcal{D}_t^{(h)}; \\ \mathbf{z}_u \in \mathcal{D}_t^{(h)}, \mathbf{z}_v \in \mathcal{D}_s^{(h)}; \end{cases} \\ 0, \text{otherwise.} \end{cases} \quad (8)$$

where  $Ns_h = |\mathcal{D}_s^{(h)}|$  and  $Nt_h = |\mathcal{D}_t^{(h)}|$ ;  $h \in \{1, 2, \dots, n\}$  is the number of categories.

3) *Laplacian regularization*: The regularization can be expressed [29] by

$$\begin{aligned} R_f(\mathcal{D}_s, \mathcal{D}_t) &= \sum_{u,v=1}^{Ns+Nt} \mathbf{W}_{uv} (f(\mathbf{z}_u) - f(\mathbf{z}_v))^2 \\ &= \sum_{u,v=1}^{Ns+Nt} f(\mathbf{z}_u) \mathbf{L}_{uv} f(\mathbf{z}_v) \\ &= \text{tr}(\beta^T \mathbf{K} \mathbf{L} \mathbf{K} \beta). \end{aligned} \quad (9)$$

where

$$\mathbf{W}_{uv} = \begin{cases} \text{sim}(\mathbf{z}_u, \mathbf{z}_v), \mathbf{z}_u \in \mathcal{N}_p(\mathbf{z}_u) \text{ or } \mathbf{z}_v \in \mathcal{N}_p(\mathbf{z}_v) \\ 0, \text{otherwise} \end{cases}$$

To measure the similarity between two points, we here use  $\text{sim}(\cdot, \cdot)$  (for example, cosine distance); The set of  $p$ -nearest neighbors to  $\mathbf{z}_u$  is denoted by  $\mathcal{N}_p(\mathbf{z}_u)$ ; Laplacian matrix  $\mathbf{L} = \mathbf{D} - \mathbf{W}$  with the diagonal matrix  $\mathbf{D}_{uu} = \sum_{v=1}^{Ns+Nt} \mathbf{W}_{uv}$ .

4) *Overall Reformulation*: Using the formulas (5), (6) and (9),  $f$  in (3) can be expressed as

$$f = \arg \min_{f \in \mathcal{H}_K} \|(\mathbf{Y} - \beta^T \mathbf{K}) \mathbf{A}\|_F^2 + \eta \text{tr}(\beta^T \mathbf{K} \beta) + \text{tr}(\beta^T \mathbf{K} (\lambda \mathbf{M} + \rho \mathbf{L}) \mathbf{K} \beta). \quad (10)$$

By setting the derivative  $\frac{\partial f}{\partial \beta} = 0$ , we get the corresponding solution:

$$\beta^* = ((\mathbf{A} + \lambda \mathbf{M} + \rho \mathbf{L}) \mathbf{K} + \eta \mathbf{I})^{-1} \mathbf{A} \mathbf{Y}^T. \quad (11)$$

C. *Reliability assessment of BBAs generated from local domain-invariant classifiers based on BF-TOPSIS*

As discussed earlier,  $M$  local domain-invariant classifiers learned from source domains  $(\mathcal{D}_{s_1}, \mathcal{D}_{s_2}, \dots, \mathcal{D}_{s_M})$  provide  $M$  BBAs  $m_i$  ( $i = 1, \dots, M$ ) (12):

$$\begin{matrix} & \theta_1 & \theta_2 & \theta_3 & \dots & \theta_{2^{|\Theta|}} \\ \mathcal{D}_{s_1} & m_1(\theta_1) & m_1(\theta_2) & m_1(\theta_3) & \dots & m_1(\theta_{2^{|\Theta|}}) \\ \mathcal{D}_{s_2} & m_2(\theta_1) & m_2(\theta_2) & m_2(\theta_3) & \dots & m_2(\theta_{2^{|\Theta|}}) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{D}_{s_M} & m_M(\theta_1) & m_M(\theta_2) & m_M(\theta_3) & \dots & m_M(\theta_{2^{|\Theta|}}) \end{matrix}, \quad (12)$$

where  $\theta \in 2^\Theta$ . Considering the possible conflicts and information redundancy among the evidence sources provided by multiple source domains, we need to evaluate the reliability of the evidence sources before processing them through a fusion step.

1) *Evaluation criteria*: Two widely used criteria are applied in our work which are described as follows:

a) *Distance degree*: The average distance between  $\mathcal{D}_{s_i}$  and other involved source domains  $\mathcal{D}_{s_j}$  are defined as follows:

$$d_{aver}(\mathcal{D}_{s_i}) = \frac{1}{M-1} \cdot \sum_{k=1}^{M-1} d_{BI}(m_{\mathcal{D}_{s_i}}(\cdot), m_{\mathcal{D}_{s_k}}(\cdot)), \quad (13)$$

where the interval distance  $d_{BI}$  was proposed by Han [33], and

- $d_{BI}(m_1, m_2) = \sqrt{N_f \cdot \sum_{ii=1}^{2^n-1} [d^I(BI_1(\theta_{ii}), BI_2(\theta_{ii}))]^2}$
- $N_f$  is the normalization factor:  $N_f = 1/2^{(n-1)}$ ,
- $BI_1(\theta_{ii}) : [Bel_1(\theta_{ii}), Pl_1(\theta_{ii})]$ ,
- $BI_2(\theta_{ii}) : [Bel_2(\theta_{ii}), Pl_2(\theta_{ii})]$ ,
- $d^I([a, b], [c, e]) = \sqrt{[\frac{a+b}{2} - \frac{c+e}{2}]^2 + \frac{1}{3}[\frac{b-a}{2} - \frac{e-c}{2}]^2}$ .

For the interval distance degree ( $d_{aver}(\mathcal{D}_{s_i})$ ), it mainly measures the difference between the specific domain  $\mathcal{D}_{s_i}$  itself and other involved domains  $\mathcal{D}_{s_{i'}}, i' \in \{1, \dots, M\}, i' \neq i$ . If this distance value is large, it means that this domain is quite different from other domains, so the reliability of this specific source domain is low and vice versa; Because we consider “the lower is better” preference ordering for  $d_{aver}(\mathcal{D}_{s_i})$ , we multiply its values by -1 and we take as first criterion  $Cr_1 = -d_{aver}(\mathcal{D}_{s_i})$  to apply BF-TOPSIS formulas of [25] that are established for “the greater is better” preference ordering.

b) *Imprecision degree*: Within evidential reasoning, the strife [34] is often used to determine the imperfection degree within a BBA. The measure of strife is defined as:

$$St(m) = - \sum_{\theta \in \mathcal{F}(m)} m(\theta) \log_2 \left[ \sum_{\theta' \in \mathcal{F}(m)} \frac{|\theta \cap \theta'|}{|\theta|} m(\theta') \right], \quad (14)$$

where  $|\theta \cap \theta'|$  and  $|\theta|$  refers to the cardinality of the subset  $\theta \cap \theta'$  and  $\theta$ .

As the formula (14) defines,  $St(m)$  does not describe the relationship between domains. It mainly measures the imprecision degree within the BBA and the value is determined by the belief masses of the focal elements in the BBA: if the output of local domain ( $\mathcal{D}_{s_1}$ ) classifier is denoted as  $m_1(\theta_1) = m_1(\theta_2) = \dots = m_1(\theta_{2|\Theta|}) = 1/(2^{|\Theta|})$ , the value of  $St(m_1)$  is largest which means that this BBA cannot give help to make the final decision; On the contrary,  $m_1(\theta_1) = 1, m_1(\theta_2) = m_1(\theta_3) = \dots = m_1(\theta_{2|\Theta|}) = 0$ , the value of  $St(m_1)$  is smallest because we can make the final decision ( $\theta_1$ ) easily according to the principle of maximum probability. Because we also consider “the lower is better” preference ordering for the strife measure, we multiply its values by -1 and consider as 2nd criterion  $Cr_2 = -St(m)$  in order to apply BF-TOPSIS formulas [25].

## 2) Evaluation of source domains' reliabilities:

a) *Scoring matrix*: We first compute the reliabilities of multi-source domains according to each criterion  $Cr_j, j =$

$1, \dots, N_c$  (in this paper,  $N_c = 2$ ) and then scoring matrix  $\mathbf{S}$  can be generated as follows:

$$\begin{matrix} & \mathcal{D}_{s_1} & \mathcal{D}_{s_2} & \mathcal{D}_{s_3} & \dots & \mathcal{D}_{s_M} \\ Cr_1 & S_1(\mathcal{D}_{s_1}) & S_1(\mathcal{D}_{s_2}) & S_1(\mathcal{D}_{s_3}) & \dots & S_1(\mathcal{D}_{s_M}) \\ Cr_2 & S_2(\mathcal{D}_{s_1}) & S_2(\mathcal{D}_{s_2}) & S_2(\mathcal{D}_{s_3}) & \dots & S_2(\mathcal{D}_{s_M}) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Cr_{N_c} & S_{N_c}(\mathcal{D}_{s_1}) & S_{N_c}(\mathcal{D}_{s_2}) & S_{N_c}(\mathcal{D}_{s_3}) & \dots & S_{N_c}(\mathcal{D}_{s_M}) \end{matrix} \quad (15)$$

b) *Construction of local BBA for source domain  $\mathcal{D}_{s_i}$* :

$$m_j(\mathcal{D}_{s_i}) \triangleq Bel_j(\mathcal{D}_{s_i}); \quad (16)$$

$$m_j(\bar{\mathcal{D}}_{s_i}) \triangleq Bel_j(\bar{\mathcal{D}}_{s_i}) = 1 - Pl_j(\bar{\mathcal{D}}_{s_i}); \quad (17)$$

$$m_j(\mathcal{D}_{s_i} \cup \bar{\mathcal{D}}_{s_i}) \triangleq Pl_j(\mathcal{D}_{s_i}) - Bel_j(\mathcal{D}_{s_i}). \quad (18)$$

Here, if we treat all involved source domains  $\mathcal{D}_{s_1}, \mathcal{D}_{s_2}, \dots, \mathcal{D}_{s_M}$  in (15) as the abstract focal elements in a special FoD:  $\Theta^{\mathcal{D}} = \{\mathcal{D}_{s_1}, \mathcal{D}_{s_2}, \dots, \mathcal{D}_{s_M}\}$ .  $\bar{\mathcal{D}}_{s_i}$  is defined as the complement of  $\mathcal{D}_{s_i}$  in  $\Theta^{\mathcal{D}}$ . Besides,  $Bel_j(\mathcal{D}_{s_i}), Pl_j(\mathcal{D}_{s_i})$  and  $Bel_j(\bar{\mathcal{D}}_{s_i})$  in (18) are defined as follows [25]:

$$Bel_j(\mathcal{D}_{s_i}) \triangleq \frac{Sup_j(\mathcal{D}_{s_i})}{\mathcal{D}_{jmax}^{max}} = \frac{\sum_{k \in \{1, \dots, M\} | S_j(\mathcal{D}_{s_k}) \leq S_j(\mathcal{D}_{s_i})} |S_j(\mathcal{D}_{s_i}) - S_j(\mathcal{D}_{s_k})|}{max_i Sup_j(\mathcal{D}_{s_i})}, \quad (19)$$

$$Bel_j(\bar{\mathcal{D}}_{s_i}) \triangleq \frac{Inf_j(\mathcal{D}_{s_i})}{\mathcal{D}_{jmin}^{min}} = \frac{-\sum_{k \in \{1, \dots, M\} | S_j(\mathcal{D}_{s_k}) \geq S_j(\mathcal{D}_{s_i})} |S_j(\mathcal{D}_{s_i}) - S_j(\mathcal{D}_{s_k})|}{min_i Inf_j(\mathcal{D}_{s_i})}, \quad (20)$$

$$Pl_j(\mathcal{D}_{s_i}) \triangleq 1 - Bel_j(\bar{\mathcal{D}}_{s_i}),$$

where the  $Bel_j(\mathcal{D}_{s_i})$  is the belief of  $\mathcal{D}_{s_i}$  as the evidential support of hypothesis: “ $\mathcal{D}_{s_i}$  is better than its competitors  $\bar{\mathcal{D}}_{s_i}$ ” and the definition of  $Bel_j(\bar{\mathcal{D}}_{s_i})$  is similar to  $Bel_j(\mathcal{D}_{s_i})$ .

c) *BF-TOPSIS method* [25]:

- Step 1: From the score matrix  $\mathbf{S}$ , compute BBAs  $m_{ij}(\mathcal{D}_{s_i}), m_{ij}(\bar{\mathcal{D}}_{s_i})$  and  $m_{ij}(\mathcal{D}_{s_i} \cup \bar{\mathcal{D}}_{s_i})$  using (18);
- Step 2: Calculate  $d_{BI}(m_{ij}, m_{ij}^{best})$  and  $d_{BI}(m_{ij}, m_{ij}^{worst})$ , where  $m_{ij}^{best}(\mathcal{D}_{s_i}) \triangleq 1$  and  $m_{ij}^{worst}(\bar{\mathcal{D}}_{s_i}) \triangleq 1$  represent the best and worst ideal BBAs, respectively;
- Step 3: Calculate  $d_{BI}(m_{ij}, m_{ij}^{best})$  and  $d_{BI}(m_{ij}, m_{ij}^{worst})$ :

$$d^{best}(\mathcal{D}_{s_i}) \triangleq \sum_{j=1}^{N_c} w_j \cdot d_{BI}(m_{ij}, m_{ij}^{best}); \quad (21)$$

$$d^{worst}(\mathcal{D}_{s_i}) \triangleq \sum_{j=1}^{N_c} w_j \cdot d_{BI}(m_{ij}, m_{ij}^{worst}). \quad (22)$$

- Step 4: Calculate the closeness degree:

$$Closeness(\mathcal{D}_{s_i}, \mathcal{D}_s^{best}) \triangleq \frac{d^{worst}(\mathcal{D}_{s_i})}{d^{worst}(\mathcal{D}_{s_i}) + d^{best}(\mathcal{D}_{s_i})}. \quad (23)$$

- Step 5: Compute the weights of each source domain based on  $Closeness(\mathcal{D}_{s_i}, \mathcal{D}_s^{best}) \in [0, 1]$  using (24):

$$w(\mathcal{D}_{s_i}) \triangleq \frac{Closeness(\mathcal{D}_{s_i}, \mathcal{D}_s^{best})}{\sum_{i=1}^M Closeness(\mathcal{D}_{s_i}, \mathcal{D}_s^{best})}. \quad (24)$$

#### D. Final combination with discounting and PCR5 rule

Once the weights of all local-invariant classifiers learned from source domains are derived, we just discount the final classification results which have been represented by BBAs. And the specific discounting rule is presented as follows for  $i = 1, \dots, M$ :

$$\begin{cases} \tilde{m}_i(\theta) = \omega_i \cdot m_{\mathcal{D}_{s_i}}(\theta), \theta \in 2^\Omega, & \theta \neq \Omega \\ \tilde{m}_i(\Omega) = 1 - \omega_i + \omega_i \cdot m_i(\Omega) \end{cases} \quad (25)$$

One can see that the mass assigned to each focal element is proportionally transferred to  $\Theta$  by the given weighting factor  $\omega_i$ . Thus, the small weighting factor will cause a big increase of mass of belief committed to ignorance. These  $M$  discounted classification results can be combined sequentially using the PCR5 fusion rule as (2) by

$$\mathbf{m}_{\text{fusion}} = ((\tilde{\mathbf{m}}_1 \otimes \tilde{\mathbf{m}}_2) \otimes \tilde{\mathbf{m}}_3) \cdots \otimes \tilde{\mathbf{m}}_M. \quad (26)$$

where  $\otimes$  symbol denotes the fusion rule.

To reduce the computational complexity of fusion rules, we consider the reduced power set (i.e.,  $2^{\Theta}_{\text{reduced}} = \{\theta_1, \theta_2, \dots, \theta_n, \Theta\}$ ), which only includes singletons and  $\Theta$  in the following activity recognition problem. Finally, according to the combination of multiple classification results provided by different source domains, a classification decision is made on the unlabeled samples in the target domain. In this paper, the final decision of the predicted class can be made as  $\theta^* = \arg \hat{A} \pm \max_{\theta} m(\theta)$  where  $\theta$  is a singleton of the  $2^{\Theta}$  based on the max of belief mass. And the brief framework of w-MSDAER is given in Fig.2, and the pseudo-code is described in Algorithm 1.

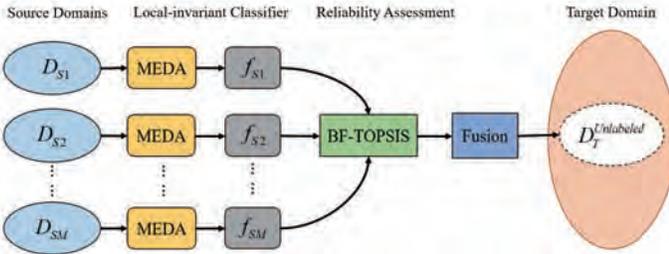


Figure 2. Framework of Our Proposed Method for Cross-Domain Activity Recognition.

#### IV. EXPERIMENTAL EVALUATION

In this part, our proposed w-MSDAER has been comprehensively evaluated on two open source datasets: the daily and physical activity dataset (DSADS)<sup>3</sup> [35] and the physical activity monitoring dataset (PAMAP2)<sup>4</sup> [36]. Our experiments mainly focus on two types: 1) cross-position HAR; 2) cross-person HAR. In the following subsections, we first introduce the experimental setting in detail. And then, the related recognition results and the analysis of parameter sensitivity are

<sup>3</sup><https://archive.ics.uci.edu/ml/datasets/daily+and+sports+activities>.

<sup>4</sup><https://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring>.

#### Algorithm 1: w-MSDAER for HAR

---

**Input:**  $M$  labeled source domains:  $\mathcal{D}_{s_1}, \mathcal{D}_{s_2}, \dots, \mathcal{D}_{s_M}$ ; Source domain  $\mathcal{D}_{s_i} = \{(\mathbf{x}_u^{\mathcal{D}_{s_i}}, y_u^{\mathcal{D}_{s_i}})\}_{u=1}^{N_{s_i}}$  and target domain  $\mathcal{D}_t = \{\mathbf{x}_v^{\mathcal{D}_t}\}_{v=1}^{N_t}$ .

**Output:** Prediction of the labels of target domain  $y^{\mathcal{D}_t}$ .

- 1 **Initialize:** manifold subspace dimension  $d$ ; regularization parameter  $\eta, \lambda, \rho$  and #neighbor  $p$ ; Iteration;
- 2 **Training of Local Domain-Invariant Classifiers:**
- 3 **for**  $i = 1, \dots, M$  **do**
- 4     Learn the transformed manifold feature using GFK:
- 5      $\mathbf{z}_u^i = g(\mathbf{x}_u^i)$ ;
- 6     Train a basic classifier using  $\mathcal{D}_{s_i}$  and predict on  $\mathcal{D}_t$  to get its labels  $\hat{y}^{\mathcal{D}_t}$ ;
- 7     Construct kernel  $\mathbf{K}^i$  using  $\mathbf{z}_u^i, \mathbf{z}_v^i$ ;
- 8     **Repeat:**
- 9     Compute  $(\mathbf{M}_0)_{uv}$  and  $(\mathbf{M}_h)_{uv}$  using (7) and (8);
- 10    Calculate  $\beta^*$  using (11) and obtain  $f_{\mathcal{D}_{s_i}}$  via the representer theorem in (4);
- 11    Update the soft label of  $\mathcal{D}_t$ ,  $\hat{y}^{\mathcal{D}_t} = f_{\mathcal{D}_{s_i}}(\mathbf{z}_u)$ ;
- 12    **Until Convergence;**
- 13    Output: classifier  $f_{\mathcal{D}_{s_i}}$ ;
- 14 **end**
- 15 **Calculation of Discounting Factors:**
- 16 Construct the BBAs matrix (15) and compute the scoring matrix  $\mathbf{S}$  using two criteria (13) and (14);
- 17 Calculate the weights of all involved BBAs using (24);
- 18 Update the BBAs based on the discounting rule (25);
- 19 **Fusion Step:**  $\mathbf{m}_{\text{fusion}} = ((\tilde{\mathbf{m}}_1 \otimes \tilde{\mathbf{m}}_2) \otimes \tilde{\mathbf{m}}_3) \cdots \otimes \tilde{\mathbf{m}}_M$ ;
- 20 **Decision Step:** Take as decision the maximum of belief mass of singleton focal elements  $\theta^* = \arg \hat{A} \pm \max_{\theta} m_{\text{fusion}}(\theta)$ ;
- 21 **final** ;
- 22 **return** Prediction of labels of target domain  $y^{\mathcal{D}_t}$ .

---

briefly presented. Similar to [9], [19], it is worth noting that different domains here refer to different sensor positions or different persons in cross-position HAR or cross-person HAR problems.

#### A. Data Set Description

In the HAR field, the mentioned two datasets: DSADS and PAMAP2 have been widely used. The DSADS dataset mainly includes nineteen activities, which were repeated by eight subjects. Those raw data collections mostly come from five IMUs located on the torso (T), right arm (RA), left arm (LA), right leg (RL), and left leg (LL). To facilitate the discussion, in this article, we only consider four subjects (Subject 1 to 4) and ten everyday activities in daily life (sitting, standing, lying on back and right side, ascending and descending stairs, standing in an elevator still, moving around in an elevator, walking in a parking lot and walking on a treadmill with a speed of 4 km/h). For the PAMAP2 dataset, due to lack of data, we only selected a subset of three people (Subject1-3) and four activities (lying, sitting, standing, and walking) for our following discussions about cross-domain HAR. And in PAMAP2, those involved wearable IMUs were installed on three different body positions: arm, chest, and leg.

#### B. Experimental Setup

Similar to [22], [37], we do not use the original time series data. Instead, we classify those unlabeled activity data based on the artificial features. Specifically, we use  $z$ -score [22], [37] to standardize the data, and combine the data from the three

axes of one sensor by using the simple averaging method given by [22], [37]. Then, we segment the data using a 5s-window-size and 3s-step-length moving window. Afterwards, 27 time and frequency features are extracted for a single sensor, and more details about feature representation can be referred to [37]. Finally, the values of extracted features are normalized into the interval  $[-1,1]$ . In this paper, we directly use these processed activity data after feature extraction and normalization, which can be downloaded from the link<sup>5</sup>. In the comparison experiments, we compare our proposed w-MSDAER with those baseline methods to illustrate the effectiveness of our proposed algorithm. Those discussed baseline methods include 1-nearest neighbor (1-NN), support vector machine (SVM), random forest (RF), extreme value learning machine (ELM); single-source domain adaptation methods, namely hierarchical transfer learning (STL), It performs marginal distribution within a group, joint distribution adaptation (JDA), balanced distribution adaptation (BDA), TNNAR and DaNN; multi-source domain adaptation methods, namely CoDATS, ELM+DS, ELM+PCR5, BDA+DS, BDA+PCR5, JDA+DS, JDA+PCR5, JDA+WDS. Here, these seven mentioned multi-source domain adaptations are manually generated according to the very simple principle: for each domain, we first use the baseline models such as ELM, BDA, JDA to learn the basic classifier, and then the outputs of these basic models are combined by using the classical DS rule (1), PCR5 rule (2) and weighted dempster rule (WDS) in the decision-level fusion. Besides, the hyperparameters in classical domain adaptation approaches were all determined in the same manner as previous references [11], [37]. For our proposed w-MSDAER, we set the main feature dimension  $d = 30$ , the number of iterations is set to 10, and the regularization parameters:  $\eta = 0.1, \lambda = 10, \rho = 1.0$  and  $p = 10$ .

### C. Measure of Performances

In this paper, we use *Accuracy* to measure the performance of w-MSDAER, which is defined by [38]

$$Accuracy = \frac{1}{n} \sum_{h=1}^n \frac{TP_h + TN_h}{TP_h + TN_h + FP_h + FN_h}, \quad (27)$$

where  $h$  denotes class index and  $n$  is the number of classes.  $TP_h, TN_h, FP_h$  and  $FN_h$  are respectively True Positives: TP, True Negatives: TN, False Positives: FP and False Negatives: FN.

### D. Results and discussions

For cross-position and cross-person HAR tasks, Tables I and II show the classification results of the proposed w-MSDAER compared to the classical single-source domain adaptation methods. In these two Tables, for convenience, we simply use “source domain→target domain” to name the specific cross-domain HAR tasks. For example, “RA→LA” means that the source domain is the data from the right arm and the

target domain is the data collected from the left arm in cross-position HAR; “St1→St2” means that the source domain is from subject1 and the target domain is from subject2 in cross-person HAR task. Table I shows that in the cross-position HAR experiments, w-MSDAER is superior to other methods in most cases. In particular, compared with traditional models (such as 1-NN, SVM, RF, and ELM), w-MSDAER can achieve higher classification accuracy in most cases, which indicates that w-MSDAER can guarantee a stable positive transfer in the cross-domain HAR task. In comparison with other transfer learning methods, especially JDA and TNNAR, the accuracy of w-MSDAER is improved by more than 10% and 2% respectively. This phenomenon further shows that the multi-source weighted domain adaptation method with evidential reasoning is an effective strategy to solve the cross-location HAR problem. Table II also shows a similar comparative phenomenon on the cross-person HAR. Another interesting observation is that the performances of most methods on cross-position HAR are relatively lower than those on the cross-person HAR. This is mainly because less involved sensor data are used in the cross-position HAR. However, these classical domain adaptations methods only consider one source domain to solve the transfer learning tasks. Differently, w-MSDAER performs multi-source fused domain adaptation. Compared with the single source domain, w-MSDAER reduces the uncertainty at the final decision layer and gives more accurate predictions. These characteristics lead to the best performance among compared single source domain-based DA methods.

In addition, we have also compared our method with the multi-source domain adaptive methods in multi-source cross-position and cross-person tasks, and the results are shown in Table III and Table IV. For convenience, we also simply use “multi-source domains→target domain” to name the specific multi-source cross-domain HAR tasks. For example, “T, RA, RL→LA” means that the source domains are the data from the right arm, right leg, torso and the target domain is the data collected from the left arm in cross-position HAR. “St1, St3→St2” means that the source domains are collected from subject1, subject3 and the target domain is from subject2 in cross-person HAR task. Overall, deep multi-source domain adaptation methods (CoDATS and CoDATS+WS) are better than the classical shallow multi-source domain adaptive methods. It is mainly because the deep multi-source domain adaptive methods can extract the depth features based on the depth network to complete the distribution matching of the target domain. However, our proposed w-MSDAER outperforms the CoDATS and CoDATS-WS. The main reason is that the traditional multi-source domain adaptation (both shallow and deep ones) does not evaluate the reliabilities of source domains from multiple perspectives. This is the main innovation of this paper, which has been clearly pointed out as our main contribution.

In order to clearly show the difference between w-MSDAER and Liu’s method in [19], we also use Liu’s method [19] to obtain the reliabilities of the involved domains for the final discounting steps in cross-position and cross-person tasks. In

<sup>5</sup><https://github.com/jindongwang/activityrecognition>.

Table I  
ACCURACY (%) COMPARISONS BETWEEN SINGLE-DOMAIN ADAPTATIONS AND W-MSDAER ON CROSS-POSITION HAR.

Method Dataset	Task	INN	SVM	RF	ELM	STL	JDA	BDA	TNNAR	DaNN	w-MSDAER
12*DSADS	T→LA	54.5	40.17	37	45.90	42.67	66.17	53.67	54.76	55.86	
	RA→LA	76.17	68.33	49.17	74.80	68.33	76	68.33	73.02	70.07	<b>80.67</b>
	RL→LA	42.33	30.17	37.5	44.82	38	65.5	48.67	48.84	49.82	
	RA→T	63	66.5	47.83	58.67	67.17	61.17	65	66.25	64.74	
	LA→T	57.17	50.17	38.83	49.17	51.33	59.83	58.17	66.34	54.54	<b>69.33</b>
	RL→T	40.17	42.83	41.83	40.17	50.67	45.67	47.67	53.79	55.53	
	T→RA	56	52.33	43.5	59.83	54.33	70.67	57.5	70.92	79.91	
	LA→RA	72.5	71.5	66.67	73.17	72.5	69.5	69.33	85.31	59.91	<b>91.67</b>
	RL→RA	47.17	38.33	47	45.83	53.67	66.83	49.5	62.24	60.48	
	T→RL	51.5	47.17	51.83	24.83	47.67	57.83	53.33	65.09	61.77	
RA→RL	56.83	59.33	46.17	51.33	56.5	<b>66</b>	59.17	63.67	62.24	62.33	
LA→RL	55.67	50.83	56.83	42.33	51.83	51.67	55.17	62.78	65.35		
6*PAMAP2	Chest→Arm	46.25	9.58	39.44	31.19	40.56	55.28	59.74	57.02	54.84	
	Leg→Arm	38.75	31.81	27.36	37.13	37.08	43.56	53.14	53.20	55.67	<b>61.94</b>
	Arm→Chest	34.03	26.53	29.44	39.27	36.11	51.16	51.82	54.13	54.43	
	Leg→Chest	28.89	21.67	32.5	32.84	30.83	32.84	26.07	57.16	57.71	<b>65.00</b>
	Arm→Leg	41.39	24.71	33.19	43.23	33.47	47.69	49.67	54.15	58.83	
Chest→Leg	35.42	22.92	35.83	34.82	35.28	47.19	39.6	51.28	50.74	<b>71.39</b>	
Average		49.87	41.93	42.32	46.07	48.22	57.47	53.64	61.10	59.64	<b>71.76</b>

Table II  
ACCURACY (%) COMPARISONS BETWEEN SINGLE-DOMAIN ADAPTATIONS AND W-MSDAER ON CROSS-PERSON HAR.

Method Dataset	Task	INN	SVM	RF	ELM	STL	JDA	BDA	TNNAR	DaNN	w-MSDAER
12*DSADS	St2→St1	57.17	48.5	54.33	66.33	52.83	57.33	74.17	76.23	69.02	
	St3→St1	51.67	50.67	38	40	50.17	47.33	51.83	65.21	60.48	78.50
	St4→St1	40.67	49.33	45.33	43.33	49	39.67	66.5	<b>82.35</b>	71.25	
	St1→St2	58	56.67	47.83	59.67	56.5	59.5	70.33	80.44	79.91	
	St3→St2	70.5	71.5	73.33	63.5	76.67	69.67	65.62	85.61	80.29	<b>88.67</b>
	St4→St2	57.5	53.83	53.83	58.5	61.17	57	67	68.79	63.23	
	St1→St3	50	47.67	43.33	48.5	48.83	41	54.17	60.37	55.42	
	St2→St3	67.33	71.33	<b>84.67</b>	69.67	83.33	69	74.5	82.47	80.58	80.00
	St4→St3	63.33	64.5	69.5	57	71	67.83	62.5	78.86	77.49	
	St1→St4	46.17	47.83	48.17	42.33	49.83	37.33	47.17	69.02	69.13	
	St2→St4	68.33	64.33	73.33	57.67	71	62	69.83	<b>73.45</b>	71.28	71.50
	St3→St4	63	61.17	57.83	61	65	66	63.67	68.74	68.13	
	6*PAMAP2	St2→St1	59.72	50.28	51.11	50.5	51.11	64.44	61.39	66.44	62.15
St3→St1		58.61	51.11	51.11	50	51.11	60.28	58.89	62.91	65.42	<b>65.28</b>
St1→St2		61.01	52.41	51.65	50.63	52.66	70.38	64.81	70.39	65.27	
St3→St2		51.9	82.03	<b>93.67</b>	73.92	83.04	66.33	65.57	85.49	80.24	78.99
St1→St3		57.18	50.65	50.13	39.43	50.65	59.27	58.49	66.34	62.17	
St2→St3	65.01	77.81	<b>79.63</b>	77.81	78.07	68.41	72.32	77.49	76.23	68.41	
Average		58.17	58.40	59.27	56.10	61.22	59.04	63.82	73.16	69.87	<b>75.91</b>

Table III  
ACCURACY (%) COMPARISONS BETWEEN MULTI-DOMAIN ADAPTATIONS AND W-MSDAER ON CROSS-POSITION HAR.

Method Dataset	Task	ELM+DSELM+PCR5BDA+DSBDA+PCR5JDA+DSJDA+PCR5JDA+WDS CoDATS CoDATS-WS	w-MSDAER								
4*DSADS	T, RA, RL→LA	73.33	65.33	50.83	65.50	70.33	74.67	77.83	75.27	<b>82.42</b>	80.67
	RA, LA, RL→T	60.00	58.50	36.67	56.67	65.17	67.54	66.89	64.58	62.01	<b>69.33</b>
	T, LA, RL→RA	82.67	79.83	49.67	68.17	65.18	72.67	73.41	85.55	84.38	<b>91.67</b>
	T, RA, LA→RL	50.17	47.00	48.83	<b>65.00</b>	52.83	59.50	62.82	56.12	62.00	62.33
3*PAMAP2	Chest, Leg→Arm	71.11	<b>72.22</b>	61.11	65.28	69.17	69.44	69.80	54.74	58.29	61.94
	Arm, Leg→Chest	53.89	61.94	53.96	56.94	55.56	55.83	60.78	60.31	64.93	<b>65.00</b>
	Arm, Chest→Leg	50.82	44.72	56.67	56.94	67.22	67.78	65.88	68.16	69.60	<b>71.39</b>
Average		63.14	61.36	51.11	62.07	63.64	66.78	68.20	66.39	69.09	<b>71.76</b>
Win/Total		0/7	1/7	0/7	1/7	0/7	0/7	0/7	0/7	1/7	4/7

Table IV  
ACCURACY (%) COMPARISONS BETWEEN MULTI-DOMAIN ADAPTATIONS AND W-MSDAER ON CROSS-PERSON HAR.

Method	Task	ELM+DS	ELM+PCR5	BDA+DS	BDA+PCR5	JDA+DS	JDA+PCR5	JDA+WDS	CoDATS	CoDATS-WS	w-MSDAER
4*DSADS	St2, St3, St4→St1	40.67	34.83	31.50	52.87	53.33	54.83	60.29	76.58	76.77	<b>78.50</b>
	St1, St3, St4→St2	65.67	59.00	49.00	73.83	52.50	79.00	79.86	82.24	87.11	<b>88.67</b>
	St1, St2, St4→St3	55.33	56.50	45.30	67.00	53.67	77.00	78.49	72.78	75.41	<b>80.00</b>
	St1, St2, St3→St4	58.50	48.67	42.00	64.83	37.83	63.67	66.28	64.75	64.49	<b>71.50</b>
3*PAMAP2	St2, St3→St1	47.50	45.83	56.67	59.44	61.67	63.33	62.30	66.43	<b>68.16</b>	65.28
	St1, St3→St2	54.18	58.73	65.06	70.38	72.15	72.66	75.39	71.63	74.18	78.99
	St1, St2→St3	49.87	49.35	57.44	66.32	68.15	71.02	71.48	60.39	<b>71.61</b>	68.41
Average		53.10	50.41	49.56	64.95	57.01	68.78	70.58	70.68	73.96	<b>75.91</b>
Win/Total		0/7	0/7	0/7	0/7	0/7	0/7	0/7	0/7	2/7	4/7

Table V  
COMPARISONS BETWEEN WEIGHT CALCULATION ON [19] AND W-MSDAER ON CROSS-POSITION (T,RA,RL→LA; RA,LA,RL→T) AND CROSS-PERSON (ST2,ST3,ST4→ST1; ST1,ST3,ST4→ST2) FOR DSADS DATASET.

Cross-Position Tasks	T,RA,RL→LA				RA,LA,RL→T			
	Weights of Source Domains			Accuracy(%)	Weights of Source Domains			Accuracy(%)
	T	RA	RL	LA	RA	LA	RL	T
Weight Calculation in [19]	0.3294	0.3407	0.3299	80.5	0.3333	0.3321	0.3346	66.83
Distance	0.2902±0.2883	0.3743±0.2602	0.3355±0.2773	79.67	0.3516±0.2698	0.3681±0.2860	0.2803±0.1815	68.5
Imprecision	0.2914±0.1811	0.5230±0.1584	0.1856±0.0413	77.67	0.4826±0.2617	0.3442±0.2755	0.1732±0.0469	68.33
Our Approach Multi-Criteria	0.2914±0.2045	0.4407±0.1207	0.2669±0.2018	<b>80.67</b>	0.4132±0.1485	0.3562±0.1842	0.2306±0.1978	<b>69.33</b>

Cross-Person Tasks	St2,St3,St4→St1				St1,St3,St4→St2			
	Weights of Source Domains			Accuracy(%)	Weights of Source Domains			Accuracy(%)
	Subject 2	Subject 3	Subject 4	Subject 1	Subject 1	Subject 3	Subject 4	Subject 2
Weight Calculation in [19]	0.3336	0.3332	0.3332	78	0.3316	0.334	0.3344	78.33
Distance	0.4635±0.2235	0.3017±0.2626	0.2349±0.1600	78.17	0.3674±0.2706	0.3951±0.2712	0.2376±0.1695	<b>90</b>
Imprecision	0.3121±0.2644	0.4786±0.3000	0.2093±0.1760	72.17	0.3136±0.2966	0.3732±0.2951	0.3131±0.3068	83.33
Our Approach Multi-Criteria	0.4012±0.1601	0.3753±0.1676	0.2234±0.1974	<b>78.5</b>	0.3407±0.1905	0.3861±0.1747	0.2732±0.2000	88.67

Table VI  
COMPARISONS BETWEEN WEIGHT CALCULATION IN [19] AND W-MSDAER ON CROSS-POSITION (T,LA,RL→RA; T,RA,LA→RL) AND CROSS-PERSON (ST1,ST2,ST4→ST3; ST1,ST2,ST3→ST4) FOR DSADS DATASET.

Cross-Position Tasks	T,LA,RL→RA				T,RA,LA→RL			
	Weights of Source Domains			Accuracy(%)	Weights of Source Domains			Accuracy(%)
	T	LA	RL	RA	T	RA	LA	RL
Weight Calculation in [19]	0.3298	0.3398	0.3304	<b>92.67</b>	0.3342	0.3336	0.3322	60.33
Distance	0.2950±0.2773	0.4171±0.2642	0.2879±0.2729	89	0.2462±0.1674	0.3399±0.2636	0.4138±0.2632	61.5
Imprecision	0.3552±0.2864	0.4566±0.2683	0.1882±0.0684	84	0.2693±0.1101	0.3918±0.1786	0.3390±0.1856	61
Our Approach Multi-Criteria	0.3261±0.1937	0.4317±0.1427	0.2422±0.1951	91.67	0.3261±0.1937	0.3261±0.1937	0.3261±0.1937	<b>62.33</b>

Cross-Person Tasks	St1,St2,St4→St3				St1,St2,St3→St4			
	Weights of Source Domains			Accuracy(%)	Weights of Source Domains			Accuracy(%)
	Subject 1	Subject 2	Subject 4	Subject 3	Subject 1	Subject 2	Subject 3	Subject 4
Weight Calculation in [19]	0.3314	0.3342	0.3344	73.83	0.3313	0.3345	0.3342	68.5
Distance	0.3034±0.2784	0.3608±0.2687	0.3358±0.2832	78.83	0.2032±0.1613	0.4576±0.2489	0.3391±0.2603	70.83
Imprecision	0.3034±0.2784	0.3608±0.2687	0.3358±0.2832	76.60	0.2439±0.0931	0.3777±0.2850	0.3784±0.2901	71
Our Approach Multi-Criteria	0.2924±0.2019	0.3879±0.1843	0.3197±0.1974	<b>80</b>	0.2298±0.1852	0.3637±0.2798	0.4065±0.2884	<b>71.5</b>

order to realize the weight calculation of Liu’s method in [19], we follow the specific steps given in [19] to use the  $\mathcal{A}$ -distance [39] to measure the distribution differences between source and target domains before and after distribution matching. Then, the derived distances are applied to calculate the weights of source domains for the final discounting steps. In addition, we also discuss the performance of w-MSDAER by using one single index and multi-criteria indexes. In Table V and VI, *Distance* means that we only use the single distance degree (13) to evaluate the weights of BBAs in w-MSDAER; *Imprecision* means that only the single imprecision degree (14) is applied in w-MSDAER; *Multi-Criteria* means that both the mentioned two indexes (13) and (14) are used in our method.

The results have been given in TableV and VI. In general, w-MSDAER with the multi-criteria strategy performs best except the two sub-tasks (St1,St3,St4→St2; T,LA,RL→RA). The main reason is that for each test sample, w-MSDAER aims to use the multi-criteria strategy (BF-TOPSIS) to evaluate the reliabilities/weights of the corresponding BBAs derived from classifiers trained by source domains.

To analyze the performance of w-MSDAER further, we present the confusion matrix in Fig.3. This figure shows the confusion matrix of RA, LA, RL→T tasks on the DSADS dataset (subject 1). We can find that w-MSDAER can achieve an accuracy more than 90% for those dynamic activities such as ascending and descending stairs, standing in an elevator still, moving around in an elevator, walking in a parking lot

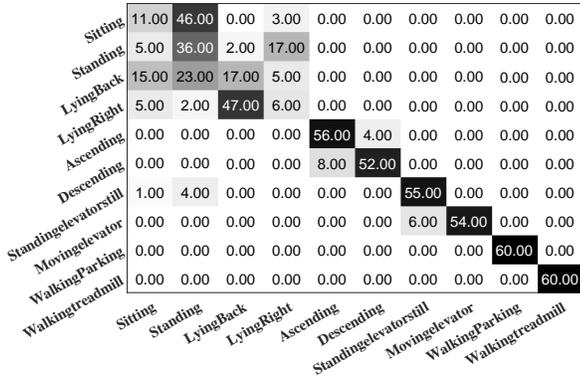


Figure 3. Confusion matrix of w-MSDAER on RA, LA, RL to T tasks on the Cross-Position DSADS dataset.

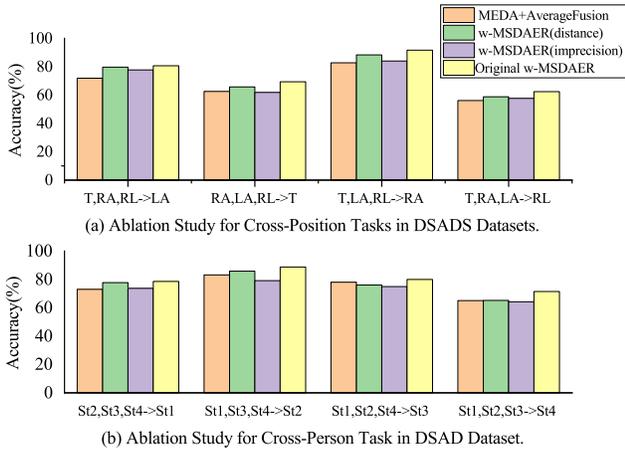


Figure 4. Ablation Study for Cross-Position and Cross-Person Tasks in DSADS Dataset.

and walking on a treadmill. However, there is some high misclassification between those static activities which includes sitting, standing, lying on back and right side. This can be explained by the fact that signals collected from sensors located on RA, LA, RL often have the similar features for the static activities.

E. Ablation Studies for Cross-Position and Cross-Person Tasks

Our proposed method w-MSDAER mainly includes two important parts: local domain-invariant learned classifier and decision-level weighted fusion strategy based on BF-TOPSIS. To further verify the effectiveness of our method, we design ablation experiments to evaluate the performance of various variant w-MSDAER from multiple perspectives. On the one hand, we ensure that the final fusion rules remain unchanged and adjust the local adaptive model. That is, we adjust MEDA in w-MSDAER to ELM, BDA, and JDA, respectively. From Table III and IV, original w-MSDAER performs better than other variants. On the other hand, we keep MEDA as the base model without adjustment. And we aim to adjust the fusion

strategy in the decision level and the evaluation criteria in BF-TOPSIS. In this case, four variants are considered: MEDA + AverageFusion, w-MSDAER (distance), w-MSDAER (imprecision), and original w-MSDAER. MEDA + AverageFusion means that we first obtain several classifiers from multi-source domains based on MEDA, and then the results of these classifiers are combined by the average fusion; w-MSDAER (distance) and w-MSDAER (imprecision) means that we just use one criterion ( $d_{aver}$  (13) or  $St(.)$  (14)) to evaluate the weights of source domains. The comparison results are shown in Fig.4. As we can see in this study, our proposed w-MSDAER outperforms all variants, which demonstrates the effectiveness of our method.

F. Parameter analysis

Similar to other state-of-the-art transfer learning methods [9], [22], we did also conduct a sensitivity analysis on the key parameters of w-MSDAER.

1) Subspace dimension  $d$  and neighbor  $p$ : In this part, the sensitivity of  $d$  and  $p$  have been investigated through experiments with a wide range of  $d \in \{5, 10, \dots, 45\}$  and  $p \in \{10, 20, \dots, 100\}$ . These related parameters were selected for two experiments: DSADS cross-position HAR (Fig.5 (a) and (b)) and PAMAP2 cross-person HAR (Fig.6 (a) and (b)). It can be observed that w-MSDAER was robust with regard to different values of  $d$  and  $p$ . Besides, for the cross-position HAR task, the accuracy of the sub-task: T, LA, RL to RA performed better than others, and for cross-person HAR, St1, St3 to S2 showed its best performance. The difference of transferring effect between different positions and persons further shows that it is difficult to achieve high-precision activity recognition by relying on a single source domain. And in the process of domain adaptation, it is necessary to evaluate the reliabilities of all involved source domains.

2) Regularization parameter  $\lambda$  and iteration: We also ran w-MSDAER, where the regularization parameters  $\lambda$  and iteration have a wide range of values, respectively. Specifically, the choices of these two important parameters are:  $\lambda \in [0.1, 0.5, 1, 10, 100, 1000]$  and  $Iteration \in [1, 5, 10, 15, 20, 25, 30]$ . Similarly, we observed that w-MSDAER could achieve robust performance over a wide range of parameter values in Fig.5 (c) and (d), Fig.6 (c) and (d). Besides, for the cross-position HAR task, the worst accuracy of the w-MSDAER algorithm is only 1.41% worse than its best accuracy. As the number of iterations increases, the accuracy does not decrease or increase. These results indicate that w-MSDAER is not very sensitive to the size of the iteration.

3) Time complexity: To evaluate the time complexity of w-MSDAER, on the DSADS and PAMAP2 datasets, we compared the running time of our method with classical domain adaptation methods on cross-position and cross-person HAR tasks. The running platform is Matlab<sup>TM</sup> on a personal laptop, using Intel Core i7-6810HQ CPU@3.40 GHz and 32.00 GB RAM. The running time is shown in Table VII. In comparison, the time complexity of BDA and w-MSDAER is significantly higher than other mentioned methods. This

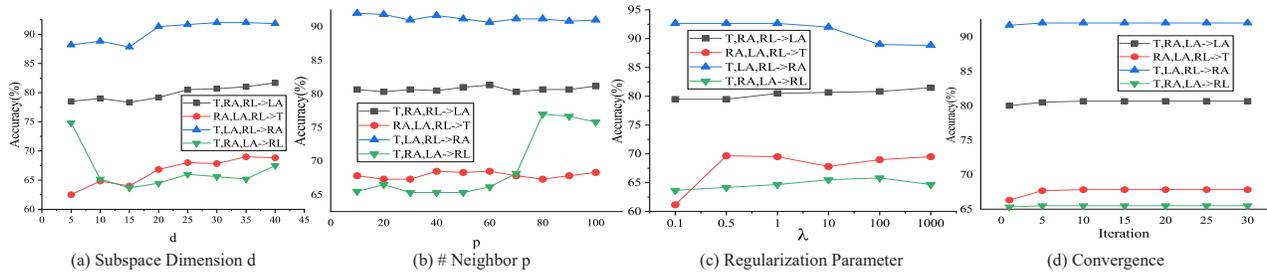


Figure 5. Classification Accuracy for DSADS Cross-Position HAR w.r.t.  $d$ ,  $p$ , and  $\lambda$ , respectively. (d) Convergence Analysis.

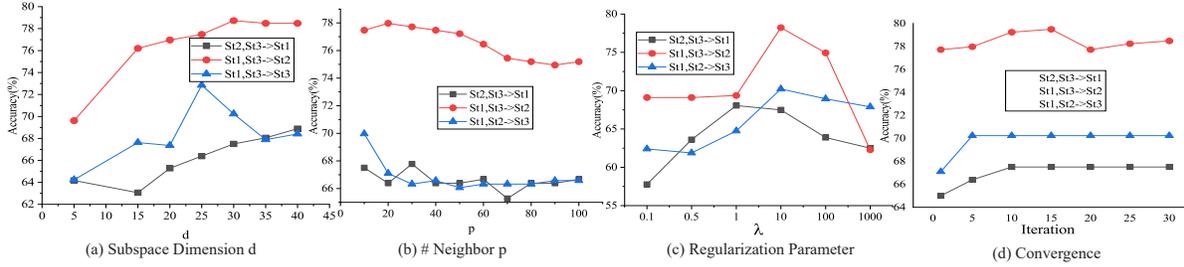


Figure 6. Classification Accuracy for PAMAP2 Cross-Person HAR w.r.t.  $d$ ,  $p$ , and  $\lambda$ , respectively. (d) Convergence Analysis.

Table VII  
TIME COMPLEXITY COMPARISON (S)

Method	Task	STL	JDA	BDA	w-MSDAER
2*DSADS	Cross Position	<b>6.0650</b>	12.3572	27.8737	57.3987
	Cross Person	<b>5.9608</b>	16.0091	30.1911	62.1064
2*PAMAP2	Cross Position	4.1261	<b>1.7503</b>	4.3576	5.5263
	Cross Person	4.7962	<b>2.5645</b>	5.4725	5.0670

is mainly because w-MSDAER involves reliability evaluation and integrated fusion process. To reduce the time complexity of w-MSDAER, our future work will focus on the use of faster fusion rules and simpler domain reliability evaluation strategies.

V. CONCLUSION AND FUTURE WORKS

In this paper, we propose a novel multi-source weighted domain adaptation with evidential reasoning (w-MSDAER) approach for cross-domain activity recognition. Compared to existing works, w-MSDAER first learns the domain-invariant classifiers from multi-source domains. Then, these classifiers have been evaluated by BF-TOPSIS and then fused with the discounting PCR5 rule. Finally, we conducted extensive experiments on two publicly available activity classification datasets. The results show that the w-MSDAER algorithm is superior to other advanced traditional and domain adaptation algorithms. In future works, we will focus on researching more robust evaluation methods for the reliability of source-domain. Besides, we will attempt more cautious decision-making strategies [19] based on belief functions to assign an unknown activity to the set of the classes, and if possible,

we will also evaluate the improved versions of PCR rules of combinations that are currently under development.

REFERENCES

- [1] H.F. Nweke, Y.W. Teh, M.A. Al-Garadi, U.R. Alo, *Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges*, Expert Systems with Applications, Vol. 105, pp. 233–261, 2018.
- [2] Q. Li, R. Gravina, Y. Li, S.H. Alsamhi, F. Sun, G. Fortino, *Multi-tuser activity recognition: Challenges and opportunities*, Information Fusion, Vol. 63, pp. 121–135, 2020.
- [3] Y. Dong, X. Li, J. Dezert, M.O. Khyam, M. Noor-A-Rahim, S.S. Ge, *Dezert-Smarandache theory-based fusion for human activity recognition in body sensor networks*, IEEE Trans. on Industrial Informatics, Vol. 16, No. 11, pp. 7138–7149, 2020.
- [4] Z. Ma, L.T. Yang, M. Lin, Q. Zhang, C. Dai, *Weighted support tensor machines for human activity recognition with smartphone sensors*, IEEE Trans. on Industrial Informatics, pp. 1–1, 2021 (early access)
- [5] T. Hasegawa, *Smartphone sensor-based human activity recognition robust to different sampling rates*, IEEE Sensors Journal, Vol. 21, No. 5, pp. 6930–6941, 2021.
- [6] A. Wang, S. Zhao, C. Zheng, H. Chen, L. Liu, G. Chen, *HierHAR: Sensor-based data-driven hierarchical human activity recognition*, IEEE Sensors Journal, Vol. 21, No. 3, pp. 3353–3365, 2021.
- [7] T. Huynh-The, C.-H. Hua, N. A. Tu, D.-S. Kim, *Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health*, IEEE Internet of Things Journal, Vol. 8, No. 3, pp. 1533–1543, 2020.
- [8] T. Huynh-The, C.-H. Hua, D.-S. Kim, *Encoding pose features to images with data augmentation for 3-d action recognition*, IEEE Trans. on Industrial Informatics, Vol. 16, No. 5, pp. 3100–3111, 2019.
- [9] J. Zhao, F. Deng, H. He, J. Chen, *Local domain adaptation for cross-domain activity recognition*, IEEE Trans. on Human- Machine Systems, Vol. 51, No. 1, pp. 12–21, 2021.
- [10] R. Fallahzadeh, Z.E. Ashari, P. Alinia, H. Ghasemzadeh, *Personalized activity recognition using partially available target data*, IEEE Trans. on Mobile Computing, Vol. 22, No.1, pp. 374–388, January 2023.

- [11] Y. Chen, J. Wang, M. Huang, H. Yu, *Cross-position activity recognition with stratified transfer learning*, Pervasive and Mobile Computing, Vol. 57, pp. 1–13, 2019.
- [12] W. Bian, D. Tao, Y. Rui, *Cross-domain human action recognition*, IEEE Trans on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 42, No. 2, pp. 298–307, 2012.
- [13] Y. Dong, X. Li, J. Dezert, R. Zhou, C. Zhu, S.S. Ge, *Multi-criteria analysis of sensor reliability for wearable human activity recognition*, IEEE Sensors Journal, Vol. 21, No. 17, pp. 19144–191561, September 2021.
- [14] M. Long, J. Wang, G. Ding, J. Sun, P. S. Yu, *Transfer feature learning with joint distribution adaptation*, in Proceedings of the IEEE Int. Conf. on Computer Vision, pp. 2200–2207, 2013.
- [15] M. Ghifary, W.B. Kleijn, M. Zhang, *Domain adaptive neural networks for object recognition*, in Pacific Rim Int. Conf. on Artificial Intelligence. Springer, pp. 898–904, 2014.
- [16] L. Duan, D. Xu, I.W.-H. Tsang, *Domain adaptation from multiple sources: A domain-dependent regularization approach*, IEEE Trans on Neural Networks and Learning Systems, Vol. 23, No. 3, pp. 504–518, 2012.
- [17] S. Sun, H. Shi, Y. Wu, *A survey of multi-source domain adaptation*, Information Fusion, Vol. 24, pp. 84–92, 2015.
- [18] H. Zhao, S. Zhang, G. Wu, J.M. Moura, J.P. Costeira, G.J. Gordon, *Adversarial multiple source domain adaptation*, Advances in Neural Information Processing Systems, Vol. 31, pp. 8559–8570, 2018.
- [19] Z.-G. Liu, L.-Q. Huang, K. Zhou, T. Denœux, *Combination of transferable classification with multisource domain adaptation based on evidential reasoning*, IEEE Trans. on Neural Networks and Learning Systems, Vol. 32, No. 5, pp. 2015–2029, 2021.
- [20] L.-Q. Huang, Z.-G. Liu, J. Dezert, *Cross-domain pattern classification with distribution adaptation based on evidence theory*, IEEE Trans.on Cybernetics, Vol. 53, No.2, pp. 718–731, February 2023.
- [21] J.-H. Hong, J. Ramos, A.K. Dey, *Toward personalized activity recognition systems with a semipopulation approach*, IEEE Trans. on Human-Machine Systems, Vol. 46, No. 1, pp. 101–112, 2016.
- [22] J. Wang, V.W. Zheng, Y. Chen, M. Huang, *Deep transfer learning for cross-domain activity recognition*, in Proc. of the 3rd Int. Conf. on Crowd Science and Engineering, ser. ICCSE'18. New York, NY, USA: Association for Computing Machinery, 2018.
- [23] Z. Wang, D. Wu, R. Gravina, G. Fortino, Y. Jiang, K. Tang, *Kernel fusion based extreme learning machine for cross-location activity recognition*, Information Fusion, Vol. 37, pp. 1–9, 2017.
- [24] G. Wilson, J.R. Doppa, D.J. Cook, *Multi-source deep domain adaptation with weak supervision for time-series sensor data*, in Proc. of the 26th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining, pp. 1768–1778, 2020.
- [25] J. Dezert, D. Han, H. Yin, *A new belief function based approach for multi-criteria decision-making support*, in Proc. of 2016 19th Int. Conf. on Information Fusion (Fusion 2016), pp. 782–789, Heidelberg, Germany, July 2016
- [26] G. Shafer, *A mathematical theory of evidence*, Princeton university press, 1976.
- [27] F. Xiao, Z. Cao, A. Jolfaei, *A novel conflict measurement in decision-making and its application in fault diagnosis*, IEEE Trans. on Fuzzy Systems, Vol. 29, No. 1, pp. 186–197, 2021.
- [28] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion*, Vol. 1–4, American Research Press, 2004–2015.
- [29] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, P.S. Yu, *Visual domain adaptation with manifold embedded distribution alignment*, in Proc. of the 26th ACM Int. Conf. on Multimedia, pp. 402–410, 2018.
- [30] V.N. Vapnik, *An overview of statistical learning theory*, IEEE Trans. on Neural Networks, Vol. 10, No. 5, pp. 988–999, 1999.
- [31] B. Gong, Y. Shi, F. Sha, K. Grauman, *Geodesic flow kernel for unsupervised domain adaptation*, in Proc. of 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2066–2073, Providence, RI, USA, July 2012.
- [32] M. Belkin, P. Niyogi, V. Sindhwani, *Manifold regularization: A geometric framework for learning from labeled and unlabeled examples*, Journal of Machine Learning Research, Vol. 7, No. 11, 2006.
- [33] D. Han, J. Dezert, Y. Yang, *Belief interval-based distance measures in the theory of Belief Functions*, IEEE Trans. on Systems Man and Cybernetics Systems, Vol. 48, No. 6, pp. 833–850, 2018.
- [34] Y. Dong, X. Li, J. Dezert, S.S. Ge, *A novel multi-criteria discounting combination approach for multi-sensor fusion*, IEEE Sensors Journal, Vol. 19, No. 20, pp. 9411–9421, 2019.
- [35] K. Altun, B. Barshan, O. Tunel, *Comparative study on classifying human activities with miniature inertial and magnetic sensors*, Pattern Recognition, Vol. 43, No. 10, pp. 3605–3620, October 2010.
- [36] A. Reiss, D. Stricker, *Creating and benchmarking a new dataset for physical activity monitoring*, in Proc. of the 5th Int. Conf. on Pervasive Technologies Related to Assistive Environments, pp. 1–8, Heraklion, Crete, June 2012.
- [37] J. Wang, Y. Chen, L. Hu, X. Peng, S.Y. Philip, *Stratified transfer learning for cross-domain activity recognition*, in Proc. of 2018 IEEE Int. Conf. on Pervasive Computing and Communications (PerCom), pp. 1–10, 2018.
- [38] M. Sokolova, G. Lalpalme, *A systematic analysis of performance measures for classification tasks*, Information Processing & Management, Vol. 45, No. 4, pp. 427–437, 2009.
- [39] S. Ben-David, J. Blitzer, K. Crammer, F. Pereira, *Analysis of representations for domain adaptation*, Advances in Neural Information Processing Systems 19 (NIPS 2006) Edited by: B. Schölkopf and J. Platt and T. Hoffman, Vol. 19, 2006.

# Flexible User-Oriented Rust Toolbox for Information Fusion

Frédéric Dambreville, Jean Dezert

The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

Emails: frederic.dambreville@onera.fr, jean.dezert@onera.fr

**Abstract**—This paper presents a generic a flexible user-oriented rusted toolbox for information fusion. This framework implements common belief or belief-derived functions. It implements combination rules in a generic way by means of rule definitions based on referee functions. It implements two logical frameworks, powersets and taxonomies, to define beliefs. The framework is based on a multithreaded and asynchronous architecture, which also makes it possible to deploy the software on multiple machines. The architecture is highly modular and adaptive, allowing simple definition and layout of the software using editable configuration files. One of the objectives is to modularize and compare both computation engines and user interfaces. The rust programming language is used to implement the framework because of its speed, security and efficiency in maintaining the library code.

**Keywords:** rust, toolbox, belief functions, decision-making, distance between BBAs.

## I. INTRODUCTION

The page of the Belief Functions and Applications Society<sup>1</sup>, lists some software and toolboxes dedicated to the computation of belief functions and combination of belief functions. These toolboxes are generally implemented in R, python or Matlab languages. Toolbox *ibelief* [1], implemented in the R language, especially offers a variety of possible rules and decision criteria. Considering these already existing libraries, we need to clarify our motivation for offering this new FURTIF<sup>2</sup> toolbox. Our motivation is fourfold.

- (a) *Use the full power of a system programming language.* The performance of languages such as MatLab, Python or R is significantly lower than that of a system language such as C, Rust (both of which are official Linux kernel languages [2]) or C++. The reasons for this are manifold and depend on the languages involved (typing strategy, use of garbage collection, interpreted versus compiled code). Often, optimized language components are based on fast system language implementations. Typically, vectorization can offer better performance with languages like R, MatLab and Python, but this limits programming to vector data modeling. For our part, we would like to have libraries that allow full freedom in the implementation of data and methods. For this reason, we need high-performance general-purpose languages such

as system programming languages. In this environment, we propose two types of lattice on which to represent belief information. The first is obviously the powerset, which provides a general Boolean algebra representation based on bit vectors. The second is the taxonomy, which is a lattice without negation operators or operator distributivity properties. The results of the combination rules are significantly different for these two types of lattice. From a propositional point of view, Boolean algebra allows a finer representation of information than taxonomy, which constrains knowledge to taxa. However, the use of a taxonomy allows real control of the combinatorial explosion, and makes more sense from the point of view of the human operator. A vector representation of the taxonomy, although we are approaching it, is not necessarily the most efficient. For this reason, we need to be able to implement dedicated structures while maintaining a very high level of performance.

- (b) *Easy-to-implement parallelization capabilities.* The concern for parallelism is, of course, a consequence of the need for performance. There are many aspects to managing parallelism, not least synchronism or asynchronism, and secure data sharing. We need a programming language that facilitates this. In addition, we are looking for an agnostic approach to setting up exchange channels between processes. In particular, this agnosticism presents a certain difficulty in reconciling single-machine and multi-machine modes of operation.
- (c) *Easily expand the toolbox.* One of our objectives is to enable *users* to easily adapt and extend the scope of the toolbox. In this perspective, the possibility of creating and testing new rules guides our approach. We should mention our work presented in [3], in which we described referee functions, a generic and semantically meaningful way of defining rules for combining belief functions; based on this work, we proposed a (now obsolete) java library enabling these functionalities. This kind of definition indeed can be combined with various generic engines for effective fusion computation, enabling the rule designer to limit his development efforts to rule definition alone. Furthermore, as part of future perspectives, we aim to make the software framework easily extensible, typically with the possibility of including new processing codes, new human-machine interfaces, and the

<sup>1</sup><https://bfasociety.org>

<sup>2</sup>FURTIF being the acronym of Flexible User-oriented Rusted Toolbox for Information Fusion.

possibility of interacting with other information processing paradigms. All of this, again, with the intention of being user-oriented. Here again, asynchronous parallelism capabilities are an interesting prospect for this objective.

- (d) *Maintainability*. From our point of view, it is essential that our toolbox is easy to maintain. With this in mind, we need the tools of a modern programming language, and its environment, to facilitate and automate this task. These tools can be grammatical, based on high-level language expressivity, or managerial, based on the well organization of internal or imported packages.

In fact, there are very few system programming languages that have the qualities of a high-level language, with mechanisms for securing access to data, particularly with a view to parallelism or asynchronism. Rust language<sup>3</sup>, is among them. For this reason, we have chosen the Rust language for our toolbox project.

In the following sections, we proceed to:

- a brief overview of belief functions and referee fusion approaches,
- a brief introduction to the Rust language, mentioning some of its qualities and components,
- details of the framework we have developed. In particular, we will detail our approach to implementing taxonomies as lattices. We will present the FURTIF toolbox in its global context, then focus on the particular aspects related to belief functions, and finally present future developments of our work.

## II. BELIEF FUNCTIONS

We begin this section of the paper with a quick introduction to belief functions [4]–[6]. To begin with, we introduce the notion of bounded lattice, which is a fairly general logical support for the description of belief functions. Next, we introduce the notion of belief mass, and then some belief and other functions that can be derived from it. We then present various rules for combining beliefs. Finally, we introduce the notion of referee functions, which is a general formalism that allows combination rules to be defined independently of the implementation of rule computation processes.

### A. Bounded lattices

A bounded lattice is a partially ordered set  $(L, \prec, \perp, \top)$  with lower and upper bounds  $\perp$  and  $\top$  in which any pair of elements  $X, Y \in L$  has a greater minorant  $X \wedge Y$  and a lesser majorant  $X \vee Y$ . It therefore follows that the lattice induces two operators,  $\wedge$  and  $\vee$ , which have interesting logical properties for representing the conjunction or disjunction of two propositions respectively. In a general bounded lattice, the operators  $\wedge$  and  $\vee$  are associative and commutative. Bounded lattices verify the absorption property  $X \wedge (X \vee Y) = X = X \vee (X \wedge Y)$  and it is possible to retrieve the order relation  $X \prec Y$  from operators using the properties  $X \prec Y \iff X \wedge Y = X$  or  $X \prec Y \iff X \vee Y = Y$ . On the other hand, the operators

$\wedge$  and  $\vee$  are not necessarily distributive, and a bounded lattice does not have a negation operator in all generality.

1) *Powerset*: Powersets are well-known examples of bounded lattices, and are used in the vast majority of belief function literature, as well as in the libraries mentioned in the introduction. We define a frame of discernment (FoD)  $\Theta$ , which is the (finite) set of elementary propositions characterizing the knowledge frame. The powerset  $2^\Theta \triangleq \{X \subset \Theta\}$  is the set of subsets of  $\Theta$ . The powerset is a bounded lattice whose order relation is the ensemblistic inclusion  $\subset$  and whose smallest and largest elements are respectively the empty set  $\emptyset$  and the total set  $\Theta$ . The conjunction and disjunction operators are the set intersection and union  $\cap$  and  $\cup$  respectively. A powerset is also a Boolean algebra, i.e. it is a distributive bounded lattice (the operators  $\cap$  and  $\cup$  are distributive) which has a negation operator  $\neg X \triangleq \Theta \setminus X$ , i.e. verifying  $X \cap \neg X = \emptyset$  and  $X \cup \neg X = \Theta$ .

2) *Taxonomy*: In our (simplified) definition, a taxonomy is a tree in which each node or leaf is characterized by a symbol representing a proposition (a taxon). To this tree, we add as a descendant of each leaf the impossible taxon  $\perp$  representing a contradiction.

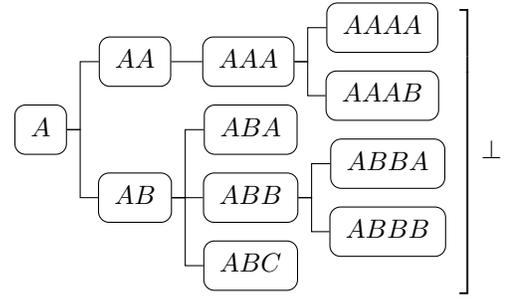


Figure 1. Example of taxonomy (taxon names are unimportant, except for the sake of presentation).

Figure 1 shows an example of taxonomy, where the set of taxa is:

$$L = \{A, AA, AAA, AAAA, AAAB, AB, ABA, ABB, ABBA, ABBB, ABC, \perp\}$$

The greatest taxon of this taxonomy is  $\top = A$  while the smallest taxon is  $\perp$ . The disjunction between two taxa is the closest common ancestor of both (or one of the taxa, if it is an ancestor of the other). For example, we have the following results:

$$ABA \vee ABBA = AB; AB \vee \perp = AB; AB \vee ABBA = AB$$

The conjunction between two taxa is the closest descendant common to both (or one of the taxa, if it is a descendant of the other). For example, we have the following results:

$$ABA \wedge ABBA = \perp; AB \wedge \perp = \perp; AB \wedge ABBA = ABBA$$

In practice, the conjunction of two non-impossible taxa is  $\perp$  unless one of the taxa is a descendant of the other.

Considering these elements, it follows that a taxonomy is a

<sup>3</sup>www.rust-lang.org

bounded lattice. On the other hand, taxonomy has no negation operator and is not associative, since:

$$(ABA \wedge ABC) \vee (ABB \wedge ABC) = \perp \vee \perp = \perp,$$

while:

$$(ABA \vee ABB) \wedge ABC = AB \wedge ABC = ABC.$$

### 3) Motivation for using taxonomies instead of powersets:

The size of a powerset is exponential with respect to the size of the discernment frame. In contrast, the size of a taxonomy remains reasonable and controlled. This is particularly useful for avoiding the combinatorial explosion in the calculation of combination rules presented below. Another motivation is less technical: a taxonomy has an immediate operational meaning that a powerset does not.

### B. Belief mass

A belief mass on the bounded lattice  $(L, \prec, \perp, \top)$  is defined as a function:

$$m : L \rightarrow \mathbb{R}_+,$$

which verifies:

$$\sum_{X \in L} m(X) = 1. \quad (1)$$

We will not discuss here the possibility and interpretation of having a non-zero mass put on the contradiction, that is  $m(\perp) > 0$ . Our library allows for this situation, but does not interpret it. But it is the user's usage that will give  $m(\perp)$  its semantics. In addition, certain belief functions derived from masses and combination rules can assign a specific role to  $\perp$  or  $\top$ .

### C. Belief functions

Certain belief functions of particular interest can be derived from belief masses, and this is implemented in the FURTIF toolbox we propose. We briefly present their definitions without going into further detail.

1) *Implicability, credibility, commonality, plausibility*: Let a belief mass  $m$  be defined on the bounded lattice  $(L, \prec, \perp, \top)$ . Implicability of  $X \in L$  is defined as follows [7]:

$$b(X) = \sum_{Y: Y \prec X} m(Y).$$

Credibility of  $X \in L$  is defined as follows:

$$Bel(X) = \sum_{Y: \perp \neq Y \prec X} m(Y).$$

The transformation of implicability into credibility, and vice versa, is immediate.

Commonality of  $X \in L$  is defined as follows:

$$Q(X) = \sum_{Y: X \prec Y} m(Y).$$

Plausibility of  $X \in L$  is defined as follows:

$$Pl(X) = \sum_{Y: Y \wedge X \neq \perp} m(Y).$$

If the bounded lattice is a Boolean algebra of negation operator  $\neg$ , it is well known that plausibility and implicability are dual by relation:

$$\text{If } L \text{ is a Boolean algebra, then } Pl(X) = 1 - b(\neg X). \quad (2)$$

Under the assumption  $m(\perp) = 0$ , we obtain that  $b = Bel$  and we find the well-known property,  $Pl(X) = 1 - Bel(\neg X)$ , valid under this hypothesis.

The property (2) is irrelevant for a general bounded lattice.

2) *Equivalence*: On the lattice  $(L, \prec, \perp, \top)$ , it is possible to compute any of the functions  $m, b, Bel, Q, Pl$  from any of the functions  $m, b, Bel, Q$ , see [7] for details.

If the bounded lattice is a Boolean algebra, then it is possible to compute any of the functions  $m, b, Bel, Q$  from function  $Pl$  thanks to duality (2). This property is false for general bounded lattice.

3) *Pignistic probability*: Given a mass function  $m$  defined on a powerset  $2^\Theta$ , the pignistic probability on  $\Theta$  is computed from  $m$  by a cardinality ratio:

$$BetP(\{\theta\}) = \sum_{X: \theta \in X} \frac{m(X)}{|X|}.$$

In the FURTIF toolbox, a similar definition of pignistic probability is proposed for taxonomies, based on a predefined weight on each leaf. Note, however, that this kind of generalization is not suitable for all bounded lattices [9].

### D. Combination rules

Let  $(L, \prec, \perp, \top)$  be a bounded lattice. We assume several belief masses,  $m_1, \dots, m_N$ , defined on  $L$  and obtained from several sources of information. A key problem is to deduce a fused belief mass from the various sources. The reference combination rule is the Dempster-Shafer rule:

$$m_1 \oplus \dots \oplus m_N(X) = \frac{1}{1 - Z} \sum_{\substack{(Y_1: N) \in L^N \\ Y_1 \wedge \dots \wedge Y_N \neq \perp}} \prod_{k=1: N} m_k(Y_k).$$

The normalization term  $1 - Z$  fulfills condition (1). The value  $Z$  is a measure of conflict between sources.

The conflict resulting from the combination of belief masses is a problem that has led to a multiplication of viewpoints. As a result, numerous rules have been proposed to deal with the conflict differently. We will not go into them all, but we will mention the following rule in particular, since we've implemented them in our toolbox: conjunctive rule, disjunctive rule, Dubois & Prade rule [11], PCR6 rule [10], which is an example of proportional conflict redistribution rule [12].

### E. Referee functions and combination rules

The Referee function was introduced in the DSMT book 3 in 2009 [12], but it followed some work on a probabilistic version of PCR5 [13]. A presentation of the notion can also be found in [3], as well as some more advanced thoughts in [8].

1) *General*: Let's assume a bounded lattice  $(L, \prec, \perp, \top)$  and a sequence of belief masses  $m_{1:N}$  defined on this lattice. Let  $Y_{1:N}$  be a sequence of propositions of  $L$ . Then the conditional function  $X \mapsto F(X|Y_{1:N}, m_{1:N})$  defined for any proposition  $X \in L$  is a referee function if it verifies:

$$F(\cdot|Y_{1:N}, m_{1:N}) \geq 0 \quad \text{and} \quad \sum_{X \in L} F(X|Y_{1:N}, m_{1:N}) \in [0, 1]. \quad (3)$$

Given a referee function, it is then possible to define a combination rule, where the merged belief mass takes the form:

$$\oplus [m_{1:N}|F](X) = \frac{1}{1-Z} \sum_{Y_{1:N} \in L^N} F(X|Y_{1:N}, m_{1:N}) \prod_{k=1}^N m_k(Y_k). \quad (4)$$

The normalization term  $1-Z$  fulfills condition (1). The value  $Z$  is a measure of conflict between sources for rule  $\oplus[\cdot|F]$ .

We showed in [3], [8] that referee functions can be used to define a large number of common rules and facilitate the design of new custom rules. From an implementation point of view, this formulation also makes it possible to distinguish:

- rule definition, which, for example, is left to the user if he wishes to test customized rules. This is formalized by  $F(X|Y_{1:N}, m_{1:N})$  in (4),
- from the generic implementation of the computation of this rule, which is formalized by  $\sum_{Y_{1:N} \in L^N} \cdots \prod_{k=1}^N m_k(Y_k)$  in (4).

Of course, the computation of the combination rule is the generic part that must be handled by the toolbox.

2) *Rules definitions*: Given the logical proposition  $P$ , we define  $[P]$ , the Iverson bracket [14] on  $P$ , by:

$$[P] = \begin{cases} 0 & \text{if } P \text{ is false,} \\ 1 & \text{if } P \text{ is true.} \end{cases}$$

See also the appendix A for a note on this notation.

The toolbox implements the referee functions of the following rules as standard:

a) *Conjunctive rule*:

$$F_{\wedge}(X|Y_{1:N}, m_{1:N}) = \left[ X = \bigwedge_{k=1:N} Y_k \right].$$

The rule produces no conflict, i.e.  $Z = 0$ , but can produce non-zero mass on  $\perp$ .

b) *Disjunctive rule*:

$$F_{\vee}(X|Y_{1:N}, m_{1:N}) = \left[ X = \bigvee_{k=1:N} Y_k \right].$$

The rule produces no conflict.

c) *Dempster Shafer rule*:

$$F_{DS}(X|Y_{1:N}, m_{1:N}) = \left[ \bigwedge_{k=1:N} Y_k \neq \perp \right] \left[ X = \bigwedge_{k=1:N} Y_k \right].$$

The rule produces conflict.

d) *Dubois & Prade rule*:

$$F_{DP}(X|Y_{1:2}, m_{1:2}) = [Y_1 \wedge Y_2 \neq \perp] [X = Y_1 \wedge Y_2] + [Y_1 \wedge Y_2 = \perp] [X = Y_1 \vee Y_2].$$

The rule produces no conflict.

e) *PCR6 rule*:

$$F_{PCR6}(X|Y_{1:N}, m_{1:N}) = \left[ \bigwedge_{k=1:N} Y_k \neq \perp \right] \left[ X = \bigwedge_{k=1:N} Y_k \right] + \left[ \bigwedge_{k=1:N} Y_k = \perp \right] \frac{\sum_{k=1:N} m_k(Y_k) [X = Y_k]}{\sum_{k=1:N} m_k(Y_k)}.$$

The rule produces no conflict.

f) *PCR# rule*: Referee function of PCR# rule is algorithmically defined in [3], [8]. This rule proportionally redistributes the conflict over the largest possible consensus (a  $k$ -consensus is the non-empty intersection of  $k$  propositions  $Y_i$ ). In comparison, PCR6 only redistributes on consensus of size 1 (single  $Y_i$ ). The rule produces no conflict.

### III. RUST PROGRAMMING LANGUAGE

We feel it is important to outline the features of the Rust language that can help us achieve the four objectives we set out in the introduction. The aim here is not to give a formal, detailed description of the language, and we shall restrict ourselves to providing qualitative insights. Language learning resources are available at [17].

#### A. System programming

To begin with, Rust is indeed a system language, insofar as Rust is one of the 2 official languages, along with C, of the Linux kernel [2]. Microsoft has decided to use Rust code to secure certain parts of the windows operating system (OS) [15], and there is an OS, Redox, in the advanced development phase, essentially based on the Rust language [16]. The Rust language has thus won the support of major players in the field of critical code development. From a practical point of view, the motivations reported by these language users are of several kinds. In particular, the language is recognized as fast and optimized: it can be compiled as native code or as web assembly code; it is based on direct memory management and not on a garbage collector; it is based on a strong but simple typing system; certain pointer characteristics enable specific compiler optimizations. The language is recognized as safe: pointers are characterized by pointer borrowing semantics that greatly consolidate memory management and prevent memory usage violations. This memory security also makes this language ideal for parallel and asynchronous applications.

These features are of interest to us, as they offer the prospect of fast, massive computational tools. The possibility of parallelization also allows us to consider an increase in computing power. But it also makes it easier to build a set of modular tools that can be used to create complex applications that can be deployed on several processes or machines.

## B. Secure memory and resources management with Rust

Rust's secure use of memory is not based on a garbage collector, but on direct access to memory through pointers, constrained by borrowing semantics and data lifetime semantics. This enables the compiler to ensure that memory and resources are used legally and securely at compile time.

Borrowing semantics allow us to distinguish three states of data:

- the process may have ownership of the data,
- the process may have a non-mutable pointer to the data,
- the process may have a mutable pointer to the data.

The semantics then guarantee either that there can only be a single mutable access to the data, or that there are one or more non-mutable accesses to the data. This ensures that processes do not use memory in a conflicting way.

Data lifetime semantics enable the compiler to trace throughout the code that the use of a data item is valid given its lifetime. Of course, this contributes to the automatic allocation and deallocation of memory, but it also makes it possible to condition memory usage in parallelized contexts.

## C. Type and trait

The multiparadigm of Rust and all its libraries, including standard ones, is based on two hierarchies: a data hierarchy (Type) and a functional hierarchy (Trait).

1) *Type hierarchy*: The type hierarchy is the result of the type construction mechanism:

- primitive types such as integers, reals, characters, ...
- structures made up of several sub-type fields and pointers,

```
1 struct MyStructure { // type contains three subfields:
2   field_1: u128, // field is a 128-bits unsigned int
3   field_2: Type2, // field is of type Type2
4   field_3: Type3, // field is of type Type3
5 }
```

- labeled unions (Rust's enum), which enable several data types to be used in the same memory field,
- other constructors such as arrays and tuples.

There are no classes in the object programming sense, so data structures remain simple. In contrast, Rust features an elaborate pattern matching system for data types.

It is in data types that we concretely implement the components of our toolbox. For example: the lattices, Powerset and Taxonomy; the referee functions, Disjunctive, Conjunctive, DempsterShafer, DuboisPrade, Pcr6, PcrSharp; the engine for computing combination rules.

2) *Traits hierarchy*: A trait in the Rust programming language is a collection of methods corresponding to a group of functionalities and properties. Some of the methods in a trait can be defined by default, while others are simply declared and must then be defined by any type implementing that trait. A trait is not a class or an abstract class in that it is not associated with any data. A trait can inherit another trait.

As an example, in the toolbox we define the `Lattice`

and `ComplementedLattice` traits, whose incomplete and simplified definitions are shown below:

```
1 trait Lattice {
2   // encoding type for the elements of the lattice
3   type Item;
4   // hash code of the lattice
5   fn lattice_hash(&self) -> u128;
6   // least element of the lattice
7   fn bottom(&self) -> Self::Item;
8   // greatest element of the lattice
9   fn top(&self) -> Self::Item;
10  // greatest lower bound
11  fn meet(&self, left: Self::Item, right: Self::Item)
12     -> Self::Item;
13  // least upper bound
14  fn join(&self, left: Self::Item, right: Self::Item)
15     -> Self::Item;
16  // test if left and right are disjoint
17  fn disjoint(&self, left: Self::Item, right: Self::Item)
18     -> bool {
19      self.meet(left, right) == self.bottom()
20  }
21  // test if left implies (i.e. is contained by) right
22  fn implies(&self, left: Self::Item, right: Self::Item)
23     -> bool {
24      self.join(left, right) == right
25  }
26  ...
27 }
```

```
1 trait ComplementedLattice: Lattice {
2   // Negation operator
3   fn neg(&self, element: Self::Item) -> Self::Item;
4   ...
5 }
```

The `Lattice` trait describes the features and qualities of a bounded lattice. This feature declares the `meet` and `join` methods, which correspond to the  $\wedge$  and  $\vee$  operators. It also declares the `bottom` and `top` methods, which respectively produce the minimum  $\perp$  and maximum  $\top$  elements of the lattice. In addition, this trait provides a default implementation of the `implies` and `disjoint` methods, which correspond respectively to the lattice's order relation and to a disjunction test of two lattice elements.

The `ComplementedLattice` trait describes the features and qualities of a Boolean algebra (distributive bounded lattice with negation operator). We see that `ComplementedLattice` inherits `Lattice`'s features, through the declaration `ComplementedLattice: Lattice`, to which it adds the declaration of the `neg` method.

As such, a trait represents only a set of features, but remains an abstraction that conveys no intrinsic data. A trait is useful for implementing certain functionalities in a generic way. The resulting code can be reused by future developers, for example, by implementing the trait on their own data types: only declarations not implemented by the trait need be implemented by the data type. In fact, to actually use a trait, you need to do so through a data type that implements the trait.

Let us go on with this toolbox-inspired example. A data type for powersets can be defined by (the actual definition is a little more complex):

```
1 struct Powerset {
2   hash: u128, // a hash code for this powerset
3   card_theta: u8, // size of the FoD (8-bit integer)
4 }
```

This simplified definition of Powerset contains two fields. The `hash` field is a quasi-unique code that almost certainly certifies the identity of the lattice. This field is used by the toolbox to check data consistency. The `card_data` field gives the value of  $|\Theta|$ , the cardinality of  $\Theta$ . The `Lattice` and `ComplementedLattice` traits are then implemented by the `Powerset` type as follows:

```

1 impl Lattice for Powerset {
2     type Item = u128;
3     fn lattice_hash(&self) -> u128 { self.hash }
4     fn bottom(&self) -> Self::Item { 0 }
5     fn top(&self) -> Self::Item {
6         1 << self.card_theta - 1
7     }
8     fn meet(&self, left: Self::Item, right: Self::Item)
9         -> Self::Item {
10        left & right
11    }
12    fn join(&self, left: Self::Item, right: Self::Item)
13        -> Self::Item {
14        left | right
15    }
16 }

1 impl ComplementedLattice for Powerset {
2     fn neg(&self, element: Self::Item) -> Self::Item {
3         self.top() & !element
4     }
5 }

```

In this implementation, powerset elements are encoded by the type `Item = u128`, i.e. an unsigned 128-bit integer. In this implementation, each bit of the integer represents an element of  $\Theta$ . It is therefore possible to manage a frame of discernment containing up to 128 elements.

The operators  $\wedge$  and  $\vee$  derive directly from the bitwise Boolean operators `&` and `|` on integers. As expected,  $\perp$  and  $\top$  are respectively defined as 0 and  $2^{|\Theta|} - 1$  in this binary representation. The negation operator is defined by `self.top() & !element`, i.e. a bitwise-not of the 128-bit integer combined with a conjunction with  $\top$ .

The definition of type `Taxonomy` is more complex, and we will not go into it in this section. However, we can say that taxonomy elements are also encoded by a primitive type `u128`, but the encoding of both taxa and operators is different. The `Taxonomy` type implements `Lattice`, but not the `ComplementedLattice`.

#### D. Rust and code maintainability

Previously, we presented the basic and fundamental concepts of traits and types in the Rust language. These concepts are combined with other language features and tools to facilitate library development and maintenance.

1) *Conditional implementation*: The grammar of the Rust language allows sophisticated designs in terms of generic programming. We will not go into this in detail, but we will nevertheless present an important consequence: Rust allows conditional implementations. More precisely, *the designer of a library can have a set of traits automatically implemented by a user-defined type if certain conditions are met.*

Let us illustrate this with an example.

We have defined two further traits, which are `IterableLattice` and `BeliefTransform`. Trait

`IterableLattice` takes into account the lattice's graph structure when exploring its elements. For example, this trait provides methods for iterating lattice elements with monotonicity properties with respect to the partial lattice order. Its incomplete and simplified definition takes the form:

```

1 trait IterableLattice: Lattice {
2     type IntoIterUp: Iterator<Item = Self::Item>;
3     type IntoIterDown: Iterator<Item = Self::Item>;
4     // lattice iterator, non decreasing with inference
5     fn bottom_to_top(&self) -> Self::IntoIterUp;
6     // lattice iterator, non increasing with inference
7     fn top_to_bottom(&self) -> Self::IntoIterDown;
8     ...
9 }

```

Trait `IterableLattice` contains type and method declarations that must be defined by the types implementing the trait. Types `Powerset` and `Taxonomy` thus implement `IterableLattice`.

Trait `BeliefTransform` provides methods for the transformation and inverse transformation of belief masses or belief functions. Mass and belief function data are carried by the generic type `Assignment<Self::Item>`, which depends on the type parameter `Self::Item`. The incomplete and simplified definition of `BeliefTransform` takes the form:

```

1 trait BeliefTransform where Self: IterableLattice {
2     // mass to credibility transform
3     fn mass_to_credibility(
4         &self, mass: Assignment<Self::Item>
5     ) -> Assignment<Self::Item> {
6         // some codes
7         ...
8     }
9     // credibility to mass transform
10    fn mass_from_credibility(
11        &self, credibility: Assignment<Self::Item>
12    ) -> Assignment<Self::Item> {
13        // some codes
14        ...
15    }
16    ...
17 }

```

Methods of `BeliefTransform` are all implemented, and it should be noted that this trait requires that the types implementing it (symbolized by `Self`) also implement `IterableLattice`. Actually, `BeliefTransform` uses methods of `IterableLattice` to explore the lattice.

The definition of `BeliefTransform` is followed by the following piece of code:

```

1 impl<L> BeliefTransform for L where L: IterableLattice { }

```

This code means that any type `L`, which implements the trait `IterableLattice`, will automatically implement the trait `BeliefTransform`. This is an example of conditional implementation.

2) *Macros*: Basically, macros are used to modify program code according to different contexts during the compilation process. Rust has a double macro system [18]. In both cases, Rust macros work at the level of the program's *abstract syntax tree*. This is done in an advanced phase of code text analysis, but before the compiled code is generated (in comparison, `C/C++` macros work at an early stage). In this respect, Rust macros are a powerful metaprogramming tool, enabling elaborate manipulation of the generated code. The

two Rust macro systems are:

- declarative macros, simpler to use but less powerful. The metacode of these macros is integrated directly into the program's source code. It is mostly based on pattern matching for its expansion.
- procedural macros, more complex but very powerful, require the development and execution of a Rust sub-program to analyze and modify the abstract syntax tree of the main program.

While the design of macros is highly technical, the use of a macro by a third-party developer is relatively straightforward, making it a powerful tool for defining and maintaining libraries.

To illustrate this, we give an example of a macro from our toolbox. As our toolbox can be implemented on several processes or machines, we need to control the consistency of data types exchanged between several processes. To this end, we have defined the `HashedTypeDef` trait, which must be implemented by all data types that can be exchanged between processes. The simplified definition of this trait is as follows:

```
1 trait HashedTypeDef {
2     // hash code for type Self
3     const TYPE_HASH: u128;
4 }
```

We can see here that any type implementing `HashedTypeDef` will be associated with the hash code `TYPE_HASH`. `TYPE_HASH` must be almost unique for all types defined with this trait. For this reason, this feature must be implemented by a dedicated macro, which builds the hash code at compile time in such a way as to ensure this quasi-uniqueness property.

There is little point in presenting the implementation details of the macro in itself. However, it is interesting to see how it can be deployed by the designer of a data type. Let us consider type `Assignment` mentioned earlier. This data type must be exchanged between processes, and must therefore implement `HashedTypeDef`. To do this, all we need to do is mention the use of the macro when defining the structure:

```
1 #[derive(HashedTypeDef)]
2 // simplified definition of assignment
3 pub struct Assignment<X> {
4     assignment: HashMap<X, f64>,
5     lattice_hash: u128,
6 }
```

It follows that `Assignment::<X>::TYPE_HASH` will then contain the constant value of the hash code of type `Assignment<X>`. On the other hand, the fields `assignment` and `lattice_hash` are not constant and respectively represent the content of a belief mass or function and the hash code of the lattice on which this mass or function is defined.

3) *Package management*: `cargo` is Rust's package manager. It tracks all the packages needed to compile an executable or library, along with their version constraints and options. It uses the `rustc` compiler for the various compilation phases and interacts with the `crates.io` package repository. In fact,

the majority of Rust libraries of interest are available from `crates.io`. The `crates.io` repository also offers rather efficient package search facilities.

To use `cargo`, the project and its dependencies must be described in `cargo.toml` files. These descriptive files are relatively user-friendly.

### E. Paradigms

Rust combines several programming paradigms:

- functional programming, with mutability control, rich pattern matching and iterators,
- imperative and structured programming,
- object-oriented programming using traits; there is no notion of classes in Rust,
- generic programming. Rust's generic programming and metaprogramming capabilities are rich,
- parallel and concurrent programming, whose implementation is greatly secured by the language's memory security capabilities. In particular, Rust handles asynchronous programming very well.

### F. Multithread and asynchronism

We need to be able to operate concurrently to give the toolbox real flexibility in its interaction with external software and environments. Fortunately, Rust supports different parallelism paradigms.

The standard library offers tools to facilitate multithreaded programming, including different communication channels between threads, and precise management of thread-safe types and data sharing (thread-safe reference counting pointer, mutex and atomic types). In addition, a number of leading libraries are available on `crates.io` to facilitate the implementation of multithreaded parallelism.

Rust supports asynchronous programming using futures. The language itself incorporates the keywords `async` and `await`, which facilitate asynchronous programming using futures. The `tokio` library<sup>4</sup> [19], among others, offers various extensions and the availability of an asynchronous runtime environment.

## IV. DESCRIPTION OF THE TOOLBOX AND PERSPECTIVES

In this section, we present the architecture and components already available for the toolbox, the components in the near future, and the prospects that are foreseeable but not yet planned. In addition, we present a simple example of implementation in the distributed and concurrent framework enabled by the tools.

### A. Global context of the toolbox

The toolbox is implemented around the `Silx` middleware we developed. The belief function components can nevertheless be used independently of `Silx` as libraries used by other programs. However, the role of the `Silx` middleware is to link the belief function components to other components, enabling the belief

<sup>4</sup><https://crates.io/crates/tokio>

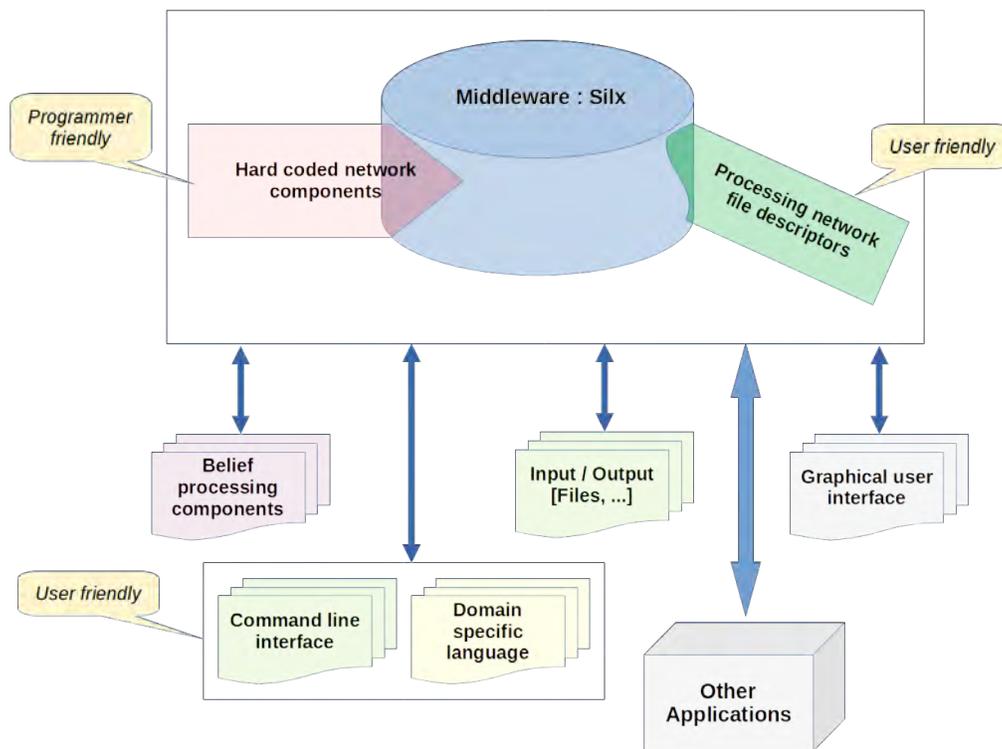


Figure 2. FURTIF toolbox: global context.

function tools to be used without compilation, or to potentially interact with other tools and software.

Figure 2 shows the overall architecture of the toolbox. We can see that the Silx middleware (blue cylinder) integrates hard-coded components, which comprise all the functionalities embraced by the considered distribution of the toolbox (light red arrow shape on the left). These components alone do not constitute an application as such, but need to be arranged in a computing network. To this end, Silx takes configuration files as parameters, which are used to define the computation network (green rectangle on the right). With Silx, the component codes and the configuration files, the toolbox enables or will enable the implementation and interaction of the following functional components:

- belief function computation components,
- input/output components,
- command-line interface components,
- components defining domain-specific languages for belief functions (currently being defined),
- future unplanned components.

We will now describe some of the middleware's features. To begin with, Silx is based on the tokio library [19], a reference Rust library for asynchronous concurrent programming offering facilities for networked application design. Based on the tokio infrastructure, we have built an environment which, based on user-friendly and editable descriptive parameter files, launches asynchronous components and establishes communication channels between these components. The communica-

tion channels are also asynchronous and of different types. In particular, they can be used to broadcast messages to a set of recipients, or to obtain from a recipient a reply to a request. These channels can operate on a single machine or be established on several machines using network sockets. The interface of these channels remains essentially the same whether the context is single-machine or multi-machine. Last but not least, Silx uses a common binary serialization format for both communication contexts.

The serialization library we use is called rkyv [20], and is a zero-copy deserialization framework. This means that it is possible to directly access the fields of a serialized structure without having to deserialize it: the result is an almost negligible performance penalty, even when running on a single machine. In addition, we have opted for endian agnostic digital formats (i.e. the order in which the bytes of a number are encoded remains the same, whatever the type of microprocessor).

### B. Belief functions components

Figure 3 describes all the components implemented or planned for the manipulation and fusion of belief functions.

- Transformations between belief masses, implicabilities, credibilities, plausibilities and commonalities are implemented if lattice properties allow (dashed arrows if implementation is constrained by lattice type).
- Several combination rules are implemented by means of referee functions (Conjunctive, Disjunctive, Dempster-Shafer, Dubois & Prade, PCR6, PCR#) and an exact but

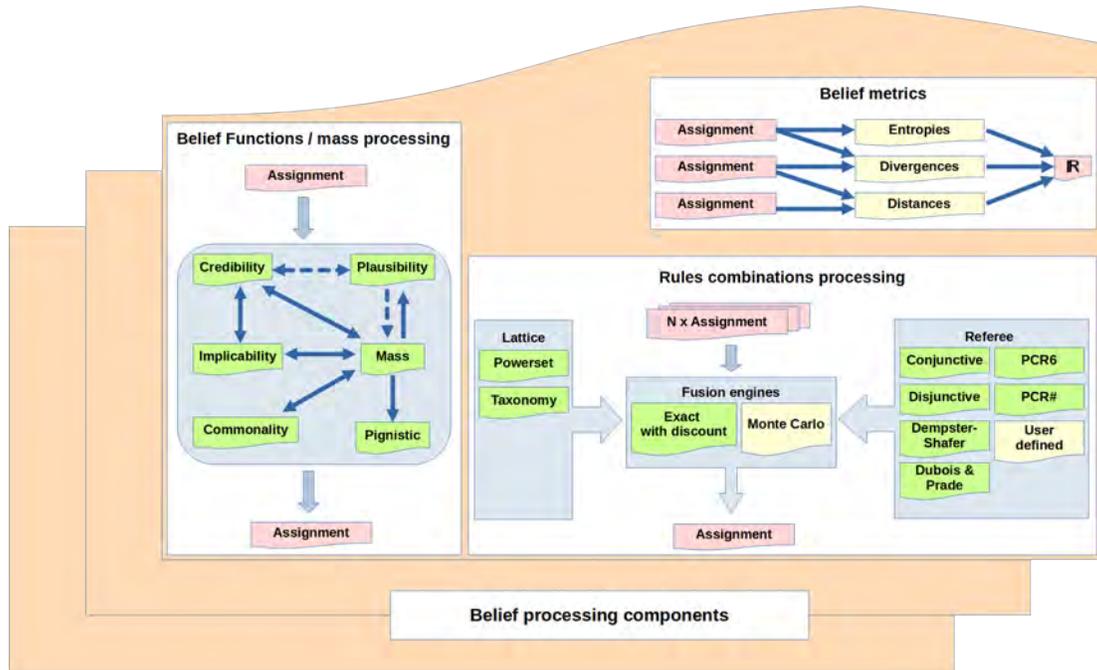


Figure 3. FURTIF toolbox: belief functions components.

discounted combination engine (the number of propositions with a non-zero mass is limited). A Monte Carlo combination rule is planned, as well as a generic arbitration function that can be designed by the user. For the user-designed referee function, reflection on domain-specific languages is planned.

- Implementations of Dezert entropy, divergence and distance [21], [22] are also planned.

### C. Taxonomy and Lattice implementation

There are several possible approaches to encoding a taxonomy element. One approach is to build two dictionaries that map each pair  $(X, Y)$  of elements of the taxonomy to  $X \wedge Y$  and  $X \vee Y$  respectively. The use of a dictionary based on a hash function (HashMap) makes it possible to compute the operators in constant time. However, the quadratic size of the structures may be prohibitive for a large taxonomy.

Figures 4 and 5 describe a more efficient approach to encoding the taxonomy and calculating the  $\wedge$  and  $\vee$  operators. In fact, there is a direct encoding based on binary calculus of taxa and operators when the taxonomy is in the form of a binary tree. An example of this encoding is shown in figure 5. The situation is a little more complex when the taxonomy is in non-binary form, and requires a prior transformation of the taxonomy tree into a binary tree (figure taxo:to:binary). We will see that this transformation subsequently induces the necessary use of a dictionary of the same order of magnitude as the taxonomy, in order to finalize the calculation of the  $\vee$  operator. This is perfectly fine in practice. Now let us get down to the details:

- As shown in Figure 4, the taxonomy is rendered binary by introducing fake taxa (taxa named  $*$ ) at nodes containing more than two children.
- Based on resulting binary taxonomy tree, the encoding is done by means of a pair of values  $(d, c)$ , respectively depth and code, which are constructed as follows (see figure taxo:coding):
  - $d$  is simply the depth of the node within the tree. For example, 0 for root  $\top$ , 1 for its children, and so on.
  - $c$  is set to maximal possible value for root  $\top$ .
  - $c$  is obtained from the code of the parent node by changing the  $d$ -th bit from the left (so there are only 2 possibilities, 0 and 1).
  - $d$  is set to maximal possible value for  $\perp$  and  $c$  is set to 0.

In practice,  $(d, c)$  is encoded on the same integer (a 128-bit integer in our implementation, a 8-bits integer in figure 5). If the integer is encoded on  $2^\lambda$  bits, then we encode  $d$  on  $\lambda$  bits and  $c$  on  $2^\lambda - \lambda$  bits. Thus, the maximum value for  $d$  is  $d_\perp = 2^\lambda - 1$  and the maximum value for  $c$  is  $c_\top = 2^{2^\lambda - \lambda} - 1$ . In figure 5, this gives  $\lambda = 3$ ,  $d_\perp = 7$  and  $c_\top = 0b11111$ .

The operators are then computed by means of simple binary operations (match these results with figure 5):

**Computation of  $\wedge$ :** The operator is symmetric and:

- If  $d = d'$  and  $c \neq c'$ , then  $(d, c) \wedge (d', c') = \perp$ .
- If  $d < d'$  and the first  $d$  bits of  $c$  and  $c'$  are identical, then  $(d, c) \wedge (d', c') = (d', c')$ .
- If  $d < d'$  and the first  $d$  bits of  $c$  and  $c'$  are not identical, then  $(d, c) \wedge (d', c') = \perp$ .

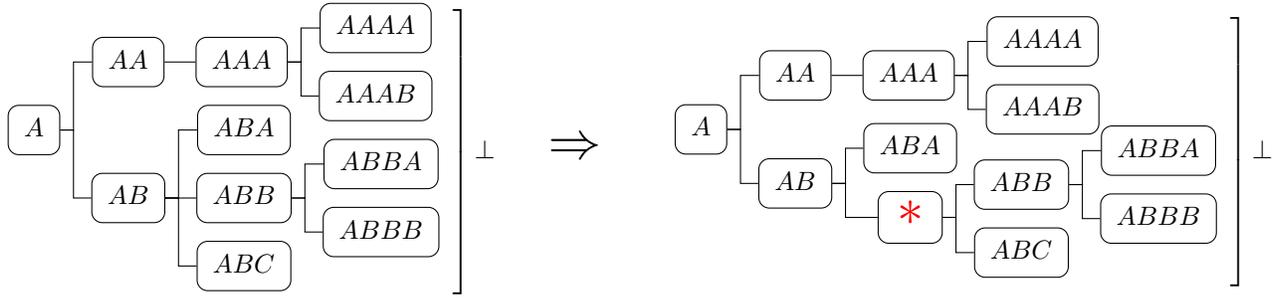
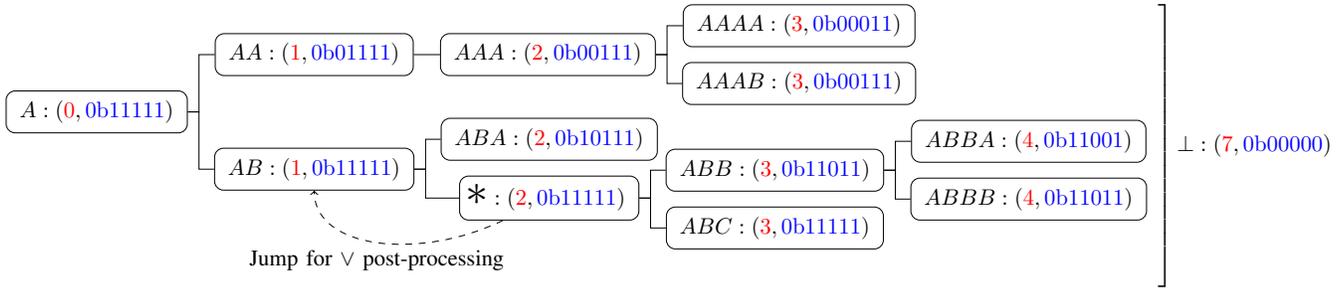


Figure 4. Turning taxonomy into a binary taxonomy: add fake nodes.


 Figure 5. Encoding binary taxonomy on 8 bits;  $3 = \log_2(8)$  bits for depth;  $5 = 8 - 3$  bits for code.

**Computation of  $\vee$ :** Consider the prefix common to  $c$  and  $c'$ . Note  $d''$  the length of this prefix and  $c''$  this prefix completed with 1's. Then  $(d, c) \vee (d', c') = (d'', c'')$ .

**Case of  $\perp$ :** Taxon  $\perp$  is managed specifically.

At this stage, there is still the problem of fake taxa in regards to operator  $\vee$ . The  $\vee$  operation could very well result in a code corresponding to a fake taxon. In Figure 5, for example, we see that  $(4, 0b11011) \vee (3, 0b11111) = (2, 0b11111)$ , which is equivalent to writing  $ABBB \vee ABC = *$ . Of course, this result is inadequate, and you have to go up the tree until you get a real taxon; in our example, this gives  $(1, 0b11111)$ , i.e.  $AB$ . To this end, our library uses a hash-code dictionary.

#### D. A simple distributed and concurrent example

Our long-term goal is to use the Silx middleware to make the library of belief functions interact with a command-line console or graphical user interfaces in association with domain-specific languages. Only then will our toolbox be truly user-oriented. We still have some work to do to achieve this goal. In particular, we need to develop a large part of the user interaction components.

Although we cannot demonstrate the functionality of the toolbox in its entirety, we can illustrate its flexibility by showing how the components integrate with the Silx middleware through a toy example.

1) *Toy example:* We propose a multi-machine processing architecture composed of the following processing nodes:

- a single-machine processing node, machine 0, consisting of the following elementary processes:
  - the `lattice` component that defines the lattice being worked on. They transmit the lattice definition

through dedicated channels to the `fuser`, `writer` and `reader_i` components.

- the `fuser` component that performs the fusion computations. This component receives the lattice definition from component `lattice`, and vectors of belief masses from sources `reader_1`, `reader_2` and `reader_3`. The component produces a result vector, which it sends to the `writer` component.
- the `writer` component, which receives the fusion results from the `fuser` component and saves them, serialized, on the local disk,
- the `shutdown_0` component, which manages the shutdown of the computing node,
- three single-machine processing nodes, running on machines  $i \in \{1, 2, 3\}$ , each with the following elementary processes respectively:
  - the `reader_i` component, which reads and deserializes a vector of belief functions from the local disk of machine  $i$  and sends it to the `fuser` component. This component obtains the lattice definition from component `lattice`.
  - the `shutdown_i` component, which manages the shutdown of the processing node.

The multi-machine processing architecture is illustrated in figure 6. This example thus includes 10 asynchronous processes and 12 communication channels between processes. Of course, this is an excessively rich implementation for such a simple problem, but the aim here is to be illustrative. These various communication channels can be intra-machine or cross-machine (via network sockets). In this figure, these

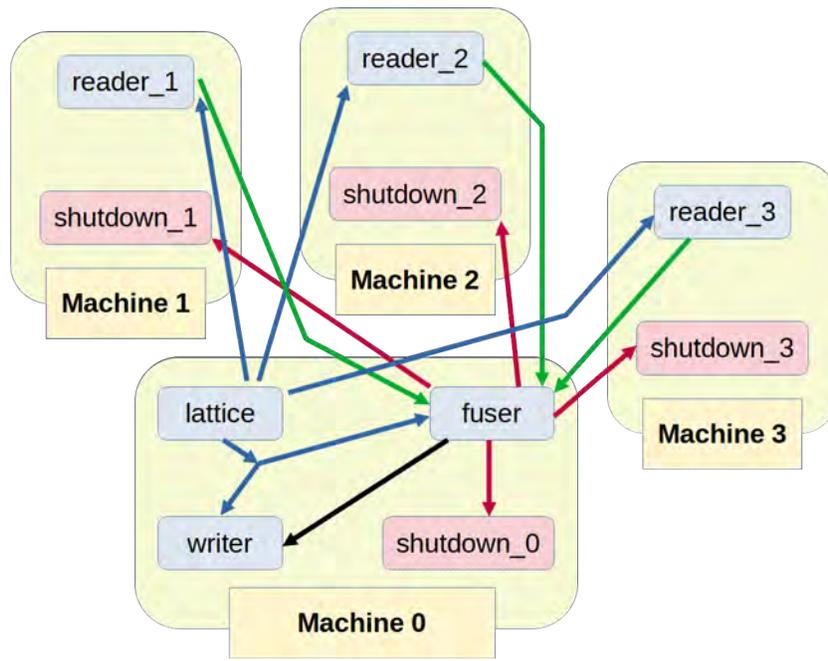


Figure 6. Example: processing graph.

communication channels are grouped by color: in blue, the 4 channels carrying the lattice definition; in green, the 3 channels carrying the data to be fused; in black, the channel carrying the fusion result; in red, the 4 channels triggering the shutdown of the processing networks.

In the medium term, computing networks will be set up using graphical user interfaces. At present, the network is entirely defined using yaml configuration files [23]. There are four sets of configuration files:

- startup files: these files are defined for each machine (or IP address) implemented.
- local processing network builders: these builders are defined for each machine (or IP address) used.
- builders of processing components: these constructors define all component parameters, including input/output channels.
- communication channel definitions.

All these files are located in appropriate, well-organized directories.

*a) Startup files:* Startup files are defined differently depending on whether you are on the master machine or one of the slave machines. For a slave machine, the startup file indicates that the starter must listen to the master machine. The master's IP address and port are labelled `main`, while those of the slave machine are labelled `this`, as shown in figure 7. On the other hand, the master machine's start-up file contains all the information needed to set up the processing networks of the master and slave machines, i.e.: the list of configuration files of the local network builders, and the list of communication channels (their names and configuration files). In this way, the master machine's starter builds a complete plan

```
!Listener
main: 127.0.0.1:8180
this: 127.0.0.1:8181
```

Figure 7. Slave 1 startup file: `slave_reader1.yaml`.

of the processing networks, which it transmits to the slave machines, so that each machine then deploys the processing networks according to the plans. Figure 8 shows one such master configuration file.

*b) Local processing network builders:* The parameter files of local processing network builders simply list all the processing components that make them up, i.e. the list of their names associated with their respective configuration files. Figure 9 shows the configuration file for the master processing network. Figure 10 shows the configuration file for machine 2's slave processing network: for this application, the slave configuration files are essentially identical, apart from the indices.

The communication channels do not appear in this list, but the way in which they connect to the processing components is specified in the configuration files for these components. Please note that, for reasons of simplicity, we have sometimes given the same name to the component and to one of its incoming or outgoing channels. Nevertheless, channel names are capitalized – e.g. `Writer`, `Reader_i`, `Shutdown_i` – whereas component names are lower-case – e.g. `writer`, `reader_i`, `shutdown_i`.

*c) Component builders:* Processing components are parameterizable structures that implement computations specific to their nature, and a glue that enables them to interact with

```
!Main
builders:
  127.0.0.1:8180: !unloaded
    path: dsmtbook/builders/main_builder.yaml
  127.0.0.1:8181: !unloaded
    path: dsmtbook/builders/slave_reader1_builder.yaml
  127.0.0.1:8182: !unloaded
    path: dsmtbook/builders/slave_reader2_builder.yaml
  127.0.0.1:8183: !unloaded
    path: dsmtbook/builders/slave_reader3_builder.yaml
flow:
  Lattice_0: !unloaded
    path: dsmtbook/channels/channel_lattice0.yaml
  Lattice_1: !unloaded
    path: dsmtbook/channels/channel_lattice1.yaml
  Lattice_2: !unloaded
    path: dsmtbook/channels/channel_lattice2.yaml
  Lattice_3: !unloaded
    path: dsmtbook/channels/channel_lattice3.yaml
  Reader_1: !unloaded
    path: dsmtbook/channels/channel_reader1.yaml
  Reader_2: !unloaded
    path: dsmtbook/channels/channel_reader2.yaml
  Reader_3: !unloaded
    path: dsmtbook/channels/channel_reader3.yaml
  Shutdown_0: !unloaded
    path: dsmtbook/channels/channel_shutdown0.yaml
  Shutdown_1: !unloaded
    path: dsmtbook/channels/channel_shutdown1.yaml
  Shutdown_2: !unloaded
    path: dsmtbook/channels/channel_shutdown2.yaml
  Shutdown_3: !unloaded
    path: dsmtbook/channels/channel_shutdown3.yaml
  Writer: !unloaded
    path: dsmtbook/channels/channel_writer.yaml
main: 127.0.0.1:8180
```

Figure 8. Master startup file: main.yaml.

```
net_size: 16
named_servant:
  fuser: !unloaded
    path: dsmtbook/servants/servant_fuser.yaml
  lattice: !unloaded
    path: dsmtbook/servants/servant_lattice.yaml
  shutdown_0: !unloaded
    path: dsmtbook/servants/servant_shutdown0.yaml
  writer: !unloaded
    path: dsmtbook/servants/servant_writer.yaml
ctrl_ch_capacity: 16
```

Figure 9. Master network builder: main\_builder.yaml.

```
net_size: 16
named_servant:
  reader_2: !unloaded
    path: dsmtbook/servants/servant_reader2.yaml
  shutdown_2: !unloaded
    path: dsmtbook/servants/servant_shutdown2.yaml
ctrl_ch_capacity: 16
```

Figure 10. Slave 2 network builder: slave\_reader2\_builder.yaml.

the Silx middleware. Implementing this glue is facilitated by Silx macros, which will not be detailed here. The design of such components is the responsibility of the programmer using our library. However, the parameterization of these components is the responsibility of the library user. In practice, the component parameter file is derived from the serialization (in yaml language) of its respective parameterizable structure.

Let us now take a closer look at the features of each of the components presented here:

- Figure 11 describes the setup file, which determines the computation of the fusing component, `fuser`. The internal structure providing this definition is `DsmtbookFuserBuilder`.

```
servant: DsmtbookFuserBuilder
referee: Pcr6
channel_lattice: Lattice_0
channels_reader:
- Reader_1
- Reader_2
- Reader_3
channel_writer: Writer
channels_shutdown:
- Shutdown_0
- Shutdown_1
- Shutdown_2
- Shutdown_3
```

Figure 11. Component builder: servant\_fuser.yaml.

The `referee: Pcr6` field specifies that the fusion will use the PCR6 rule.

The `channel_lattice: Lattice_0` field specifies that the `Lattice_0` channel is used to receive the lattice definition (from the `lattice` component).

The `channel_writer: Writer` field specifies that the `Writer` channel will be used to transmit the result of the calculation (to the `writer` component).

The field prefixed by `channels_reader:` lists the channels – i.e. `Reader_1`, `Reader_2` and `Reader_3` – that the component listens to in order to receive input data to be fused from components `reader_1`, `reader_2` and `reader_3` respectively.

The field prefixed with `channels_shutdown:` lists the channels to which to send the network shutdown signal, once the fusion calculation is complete.

- Figure 12 shows the configuration of the `lattice` component.

This component contains the taxonomy definition shown in figure 13.

This taxonomy is sent to components `fuser`, `writer`, `reader_1`, `reader_2` and `reader_3` through channels `Lattice_0`, `Lattice_1`, `Lattice_2` and `Lattice_3`.

- Figures 14 and 15 show the parameters of components `reader_1` and `reader_3` respectively.

The `reader_1` component reads input data to be fused from the `input_1.json` file, deserializing it from the `json` format. The `reader_3` component reads input data to be fused from the `input_3.yaml` file, deserializing it from the `yaml` format.

Components `reader_1` and `reader_3` receive taxonomy information through channels `Lattice_1` and `Lattice_3` respectively. The `reader_1` and `reader_3` components send their input data to

```

servant: DsmtbookLatticeBuilder
channels_lattice:
- Lattice_0
- Lattice_1
- Lattice_2
- Lattice_3
lattice: !Taxonomy
  taxonomy: !Node
    name: Object
    children:
    - !Node
      name: Air
      children:
      - !Leaf
        name: Airplane
        weight: 0.1
      - !Leaf
        name: UAV
        weight: 0.1
    - !Node
      name: Amphibian
      children:
      - !Leaf
        name: Hovercraft
        weight: 0.05
    - !Node
      name: Ground
      children:
      - !Leaf
        name: Bike
        weight: 0.15
      - !Leaf
        name: Car
        weight: 0.2
      - !Leaf
        name: Truck
        weight: 0.15
    - !Node
      name: Water
      children:
      - !Leaf
        name: Boat
        weight: 0.15
      - !Leaf
        name: Ship
        weight: 0.1
  
```

Figure 12. Component builder: `servant_lattice.yaml`.

the fuser component through the `Reader_1` and `Reader_3` channels respectively.

The `reader_2` component, not shown, is parameterized in a similar way to the `reader_3` component.

- Figure 16 shows the parameters of the writer component.

The writer component receives taxonomy information through the `Lattice_0` channel, and receives fused data from the fuser component through the `Writer` channel.

The writer component writes the fused data to the `output.yaml` file, serializing it in `yaml` format.

d) *Channel definitions:* Figures 17, 18 and 19 show the configuration files for channels `Lattice_0`, `Lattice_1` and `Shutdown_3` respectively. The various configuration files have a common structure. First, the file begins with a label that defines the channel type: here `!Broadcast` and `!NetBroadcast`, which characterize a data broadcast within a single machine or between two machines (or IP

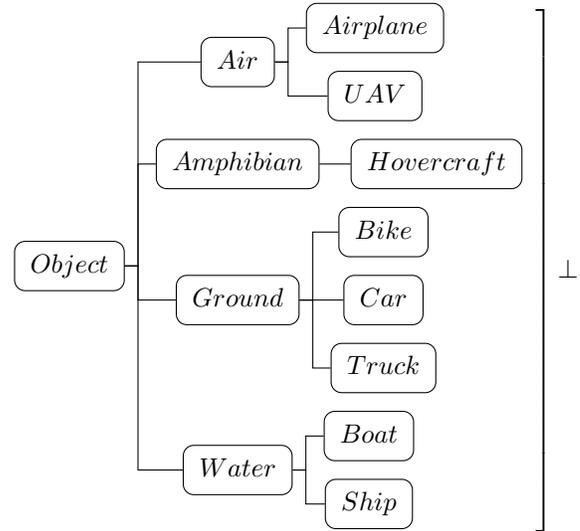


Figure 13. Taxonomy encoded in `servant_lattice.yaml`.

```

servant: DsmtbookReaderBuilder
channel_reader: Reader_1
channel_lattice: Lattice_1
serializer: Json
file: dsmtbook/data/input_1.json
  
```

Figure 14. Component builder: `servant_reader1.yaml`.

```

servant: DsmtbookReaderBuilder
channel_reader: Reader_3
channel_lattice: Lattice_3
serializer: Yaml
file: dsmtbook/data/input_3.yaml
  
```

Figure 15. Component builder: `servant_reader3.yaml`.

```

servant: DsmtbookWriterBuilder
channel_writer: Writer
channel_lattice: Lattice_0
serializer: Yaml
file: dsmtbook/data/output.yaml
  
```

Figure 16. Component builder: `servant_writer.yaml`.

```

!Broadcast
cluster: 127.0.0.1:8180
max_ping:
  secs: 0
  nanos: 50000000
data_type: 59f5bfc8-6116-0b9c-149b-50c0e4f645cb
size: 16
input:
- lattice
output:
- fuser
- writer
  
```

Figure 17. Channel: `channel_lattice0.yaml`.

addresses). Next, the field preceded by `data_type:` contains a quasi-unique code for the type of data carried by the channel. This code is used as a security check when generating the processing networks. Typically, this code is the same for

```

!NetBroadcast
max_ping:
  secs: 0
  nanos: 50000000
data_type: 59f5bfc8-6116-0b9c-149b-50c0e4f645cb
size: 16
input:
- 127.0.0.1:8180
- - lattice
output:
- 127.0.0.1:8181
- - reader_1
    
```

Figure 18. Channel: channel\_lattice1.yaml.

```

!NetBroadcast
max_ping:
  secs: 0
  nanos: 50000000
data_type: 223a70f4-30a0-ee42-21a3-c2365df732c8
size: 16
input:
- 127.0.0.1:8180
- - fuser
output:
- 127.0.0.1:8183
- - shutdown_3
    
```

Figure 19. Channel: channel\_shutdown3.yaml.

channels `Lattice_0` and `Lattice_1`, as they transmit the same type of data. However, the code is different for channels `Lattice_1` and `Shutdown_3`, as the shutdown signal is of a different type than the lattice definition. The fields preceded by `input:` and `output:` respectively list the components transmitting or listening on the channel. Note the difference between `!Broadcast` and `!NetBroadcast` channels. For the `!Broadcast` channel, the IP address within which the channel is transmitting is indicated in the field preceded by `cluster:.` For the `!NetBroadcast` channel, there is a transmit IP address placed after `input:` and a receive address placed after `output:.`

2) *Processing results:* In this presentation, we have limited the input and output data to sequences of two belief masses. We denote  $[m_i, m'_i]$  the input data produced by the `reader_i` component and  $[m_{\oplus}, m'_{\oplus}]$  the fused data produced (row by row) by the `fuser` component. In our example, the input data is defined by:

$$\begin{array}{lll}
 m_1(Object) = 0.2 & m_1(Air) = 0.3 & m_1(Truck) = 0.5 \\
 m'_1(Object) = 0.3 & m'_1(Ground) = 0.4 & m'_1(Hovercraft) = 0.3 \\
 m_2(Object) = 0.1 & m_2(UAV) = 0.5 & m_2(Amphibian) = 0.4 \\
 m'_2(Object) = 0.4 & m'_2(Car) = 0.2 & m'_2(Water) = 0.3 \\
 m_3(Object) = 0.2 & m_3(Ground) = 0.2 & m_3(Bike) = 0.6 \\
 m'_3(Object) = 0.4 & m'_3(Air) = 0.4 & m'_3(Ship) = 0.2
 \end{array}$$

We obtain the following fused data, by PCR6 rule as indicated in figure 11:

$$\begin{array}{lll}
 m_{\oplus}(Object) = 0.06 & m_{\oplus}(Air) = 0.07 & m_{\oplus}(Ground) = 0.04 \\
 m_{\oplus}(Truck) = 0.20 & m_{\oplus}(UAV) = 0.22 & m_{\oplus}(Amphibian) = 0.14 \\
 m_{\oplus}(Bike) = 0.27 & & \\
 m'_{\oplus}(Object) = 0.20 & m'_{\oplus}(Air) = 0.19 & m'_{\oplus}(Ground) = 0.18 \\
 m'_{\oplus}(Water) = 0.12 & m'_{\oplus}(Car) = 0.10 & m'_{\oplus}(Hovercraft) = 0.13 \\
 m'_{\oplus}(Ship) = 0.08 & &
 \end{array}$$

Figures 20 and 21 illustrate the contents of the serialized data of  $[m_1, m'_1]$  and  $[m_2, m'_2]$  as stored in the files `input_1.json` and `input_2.yaml` with json and yaml formats respectively. Files `input_3.yaml` and `output.yaml` are formed in a similar way to `input_2.yaml`.

```

[
  [
    [ "Object", 0.2 ],
    [ "Air", 0.3 ],
    [ "Truck", 0.5 ]
  ],
  [
    [ "Object", 0.3 ],
    [ "Ground", 0.4 ],
    [ "Hovercraft", 0.3 ]
  ]
]
    
```

 Figure 20.  $[m_1, m'_1]$ : input\_1.json.

```

- - - Object
- - 0.1
- - Amphibian
- - 0.4
- - UAV
- - 0.5
- - - Object
- - 0.4
- - Car
- - 0.2
- - Water
- - 0.3
    
```

 Figure 21.  $[m_2, m'_2]$ : input\_2.yaml.

## E. Perspectives

Our toolbox is designed to be extended. Of course, global extensions can be envisaged, such as the addition of a graphical user interface, or interaction with other applications connected to the middleware. On the other hand, the belief function components are by no means exhaustive, and numerous completions are foreseeable, though not yet planned.

## V. CONCLUSION

In this work, we presented the *FURTIF* toolbox for belief function fusion. This toolbox is implemented in Rust, both for performance reasons and to open up the possibility of integrating it into a distributed, multithreaded and concurrent environment. To this end, we have developed the middleware *Silx* in which *FURTIF* is integrated. It should be noted,

however, that the library's components can be used independently of the middleware. At present, the toolbox is almost complete, but lacks a few components that were planned but are still under development (e.g. domain-specific language; generic referee function; Monte Carlo for computing combination rules; entropy, divergence and distance). Although not currently planned, the implementation of a graphical user interface, both for using the fusion library as well as for building and deploying computational networks, is a future goal of interest. More generally, we intend to use Silx middleware to support mixed implementations combining *FURTIF* with other information processing paradigms.

The *FURTIF* toolbox will be available in open source at [24] and [25]. Also, updated information about this toolbox will be regularly posted on [26] for convenience.

It is worth mentioning that it is possible to compile and link one or more rust source files written with the Matlab™ Data API for RUST into a binary MEX<sup>5</sup> file if necessary<sup>6</sup>. This possibility has not yet been tested for interfacing the *FURTIF* toolbox with Matlab™.

## APPENDIX

### A. Iverson bracket

Although the Iverson bracket [14] is an elegant and concise notation, it is not ambiguous. The point to remember is that the context of Iverson's bracket is propositional on the inside and numerical on the outside, as in  $[\text{false}] + 1$ , which has the value  $0 + 1$ . Take the following example:

$$[[1 + 3] \times 5, [[x = 3] \wedge [x < 0]] + 1]$$

We understand unambiguously that we have defined the row vector  $[20, 1]$ . Indeed:

**Purely numerical context:**  $[1 + 3] \times 5$  is another way of writing  $(1 + 3) \times 5 = 20$ ,

**Purely propositional context:**  $[x = 3] \wedge [x < 0]$  is another way of writing  $(x = 3) \wedge (x < 0)$ , which is **false**,

**Propositional context inside & numerical context outside:**  $[[x = 3] \wedge [x < 0]] + 1$  is  $[\text{false}] + 1$ , that is 1,

**Matrix context:**  $[[1 + 3] \times 5, [[x = 3] \wedge [x < 0]] + 1]$  is  $[20, 1]$ .

## REFERENCES

- [1] K. Zhou, A. Martin, *ibelief: Belief Function Implementation*, Version 1.3.1, 2021.  
<https://cran.rstudio.com/web/packages/ibelief/>
- [2] S. Vaughan-Nichols, *Linus Torvalds: Rust will go into Linux 6.1*, 2022.  
<https://www.zdnet.com/article/linus-torvalds-rust-will-go-into-linux-6-1/>
- [3] F. Dambreville, *Generic implementation of fusion rules based on Referee function*, in Proc. of Workshop on the Theory of Belief Functions, BELIEF 2010.
- [4] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [5] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. Math. Statist., Vol. 38, pp. 325–339, 1967.
- [6] P. Smets, *Practical uses of belief functions*, In: K. B. Laskey and H. Prade, Editors, *Uncertainty in Artificial Intelligence*, Vol. 15, pp. 612–621, Stockholm, Sweden, 1999.
- [7] P. Smets, *The application of the matrix calculus to belief functions*, Int. J. of Approximate Reasoning, Vol. 31, pp. 1–30, 2002.
- [8] F. Dambreville, *Generic and Concurrent Computation of Belief Combination Rules*, in: C. Francalanci, M. Helfert, (Eds), *Data Management Technologies and Applications*, DATA 2016, Communications in Computer and Information Science, Vol 737, Springer, Cham, 2016.  
[https://doi.org/10.1007/978-3-319-62911-7\\_7](https://doi.org/10.1007/978-3-319-62911-7_7)
- [9] J. Dezert, F. Smarandache, M. Daniel, *A Generalized Pignistic Transformation*, in: F. Smarandache & J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion (Collected works)*, Vol. 1, American Research Press, Rehoboth, 2004.
- [10] A. Martin, C. Osswald, *Toward a combination rule to deal with partial conflict and specificity in belief functions theory*, in Proc. of the 10th Int. Conf. on Information Fusion (Fusion 2007), Québec, QC, Canada, July 2007.
- [11] D. Dubois, H. Prade, *On the unicity of Dempster rule of combination*, International Journal of Intelligent Systems, Vol. 1(2), pp. 133–142, 1986.
- [12] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion (Collected Works)*, Volumes 1–4, American Research Press, USA 2004–2015.  
<https://www.onera.fr/fr/staff/jean-dezert/references>
- [13] A. Kirchner, F. Dambreville, F. Celeste, J. Dezert, F. Smarandache, *Application of probabilistic PCR5 fusion rule for multisensor target tracking*, in Proc. of the 10th Int. Conf. on Information Fusion (Fusion 2007), Québec, QC, Canada, July 2007.
- [14] Wikipedia, *Iverson bracket*,  
[https://en.wikipedia.org/wiki/Iverson\\_bracket](https://en.wikipedia.org/wiki/Iverson_bracket)
- [15] T. Claburn, *Microsoft is busy rewriting core Windows code in memory-safe Rust*, The Register, April 2023.  
[https://www.theregister.com/2023/04/27/microsoft\\_windows\\_rust/](https://www.theregister.com/2023/04/27/microsoft_windows_rust/)
- [16] *Redox, a Unix-like microkernel operating system written in the programming language Rust*.  
<https://www.redox-os.org/>
- [17] *Learn Rust*.  
<https://www.rust-lang.org/learn>
- [18] D. Keep, L. Wirth, *The Little Book of Rust Macros*.  
<https://veykril.github.io/tlbrm/>
- [19] Tokio Community, *Build reliable network applications without compromising speed*.  
<https://tokio.rs/>
- [20] D. Koloski, *Rkyv: Zero-copy deserialization framework for Rust*.  
<https://crates.io/crates/rkyv>
- [21] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, in Proc. of the 25th Int. Conf. on Information Fusion (Fusion 2022), Linköping, Sweden, July 2022.
- [22] J. Dezert, F. Dambreville, *Cross-Entropy and Relative Entropy of Basic Belief Assignments*, in Proc. of the 26th Int. Conf. on Information Fusion (Fusion 2023), Charleston, SC, USA, June 2023.
- [23] Yaml Community, *YAML Ain't Markup Language*.  
<https://yaml.org/>
- [24] <https://crates.io/crates/furtif> and <https://crates.io/crates/furtif-core>
- [25] <https://github.com/onera/silx-furtif>
- [26] <https://www.onera.fr/fr/staff/jean-dezert>

<sup>5</sup><https://www.mathworks.com/help/matlab/call-mex-functions.html>

<sup>6</sup><https://crates.io/crates/rustmex>



# E-FPN: Evidential Feature Pyramid Network for Ship Classification

Yilin Dong<sup>a</sup>, Kunhai Xu<sup>a</sup>, Changming Zhu<sup>a</sup>, Enguang Guan<sup>b</sup>, Yihai Liu<sup>c</sup>,

<sup>a</sup>College of Information Engineering, Shanghai Maritime University, Shanghai, China.

<sup>b</sup>College of Logistics Engineering, Shanghai Maritime University, Shanghai, China.

<sup>c</sup>Jiangsu Automation Research Institute, Lianyungang, China.

Emails: yldong@shmtu.edu.cn, 202230310302@stu.shmtu.edu.cn, cmzhu@shmtu.edu.cn, egguan@shmtu.edu.cn, liuyihai@126.com

Originally published as: Y. Dong, K. Xu, C. Zhu, E. Guan, Y. Liu, *E-FPN: Evidential Feature Pyramid Network for Ship Classification*, *Remote Sensing* 2023, 15, 3916, 2023, and reprinted with permission..

**Abstract**—Ship classification, as an important problem in the field of computer vision, has been the focus of research for various algorithms over the past few decades. In particular, convolutional neural networks (CNN) have become one of the most popular models for ship classification tasks, especially in deep learning methods. Currently, several classical methods have used single-scale features to tackle ship classification, without paying much attentions to the impact of multi-scale feature. Therefore, this paper proposes a multi-scale feature fusion ship classification method based on evidence theory. In this method, multiple scales of features are utilized to fuse feature maps of three different sizes (40x40x256, 20x20x512, 10x10x1024), which are used to perform ship classification tasks separately. Finally, the multi-scales-based classification results are treated as pieces of evidence and fused at the decision level using evidence theory to obtain the final classification result. Experimental results demonstrate that compared to classical classification networks, this method can effectively improve classification accuracy.

**Keywords:** ship classification, multi-scale, evidence theory, feature fusion, deep learning.

## I. INTRODUCTION

Image classification, as an important problem in the field of computer vision, aims to assign input images to predefined categories. Over the past few decades, significant progress has been made in image classification, especially with deep learning-based methods. CNN can automatically extract rich feature representations from input images and perform classification using fully connected layers. Compared to traditional machine learning methods, deep learning approaches can learn more discriminative features automatically from data, leading to higher classification accuracy. Practical applications of image classification techniques have become relatively mature and have been widely used in various domains, such as visual recognition [1], medical image analysis [2], industrial quality inspection [3], agriculture [4]–[6], surveillance [7], and autonomous driving [8].

However, due to the complex and diverse characteristics of image data and the variety of practical application scenarios, improving the accuracy of image classification further remains a challenging task. For instance, challenges persist in satellite remote sensing image classification [9]–[12] and fine-grained image classification [13], [14].

For example, ship satellite remote sensing images present specific challenges compared to traditional natural images in the image classification task [15]–[17]:

- 1) Variations in ship size and shape: The appearance and shape of ships in satellite remote sensing images can be influenced by various factors such as distance, lighting conditions, and viewing angles. Therefore, ships of the same type may exhibit different sizes and shapes in different satellite remote-sensing images, making image classification difficult.
- 2) Complexity of the background: Ship satellite remote sensing images often include complex backgrounds such as waves, clouds, and ports. These backgrounds can introduce interference in the classification of ships.
- 3) Similarity: ship satellite remote sensing images encompass various types of ships, including different ship types and purposes such as cargo ships, passenger ships, and fishing boats. However, apart from some specific ship types, most ship outlines exhibit an elongated shape with axis symmetry and a pointed bow, which can pose challenges for classification algorithms.
- 4) Resolution: ship satellite remote sensing images typically have lower resolution compared to traditional natural images. This can impact the extraction of fine-grained ship details and features, thus affecting the performance of classification algorithms.
- 5) Data quality: Ship satellite remote sensing images are susceptible to natural factors such as lighting, weather conditions, and cloud cover, which can result in lower image quality. Issues like blurring, distortion, and occlusion can arise, affecting the accuracy of ship classification.

Currently, most existing improvement methods for ship classification, which rely on CNN to automatically extract abstract features, mainly focus on modifying network structures, optimizing training strategies, or redesigning loss functions in an iterative manner. However, they overlook further processing of the classification results.

In the case of fine-grained image classification, which is different from general ship classification tasks, the main challenge lies in categorizing objects from closely related subcategories. These objects often exhibit subtle category differences, and the crucial information containing these differences is typically localized in small regions of the image. When extracting features using deep neural networks, smaller-sized features in the images may get diluted as the network deepens, thereby affecting the classification results [18]. Utilizing multi-scale feature fusion methods allows deep networks to learn small-sized features that may have been diluted due to network depth, thereby enhancing the accuracy of classification. Therefore, solely focusing on network structure or loss function improvements may pose challenges in further enhancing classification performance.

In the CNN-based methods, initially, researchers focused on deepening the network structure to improve classification performance and address issues arising from deeper networks in order to enhance the classification network. Later, attention shifted towards better feature propagation or utilizing detailed features to strengthen classification performance. For example, attention mechanisms were introduced to emphasize more discriminative features [19], or multiple feature extraction networks were used in combination with extracted feature maps to complement missing features [20]. Knowledge distillation was also employed to transfer detailed image features to smaller primary networks, resulting in improved performance for the classification network [21]. However, the approaches above added additional complexity to the network structures in order to better extract features.

This paper proposes a multi-scale ship classification network that applies evidence theory to decision-level fusion to break free from the improvement loop mentioned earlier and enhance classification accuracy from a different perspective. Three main modules are utilized in this method to ensure better classification accuracy: (1) Multi-scale output module of the feature extraction network; (2) Pyramid feature fusion module; (3) Decision-level fusion module based on evidence theory. The first two parts focus on improving accuracy using network structures, while the final part emphasizes optimizing classification performance using the final probability distribution information.

To validate the feasibility of this method, experiments were conducted on a traditional natural image dataset and a remote-sensing image dataset for fine-grained ship classification. Several comparisons were made with classical classification methods. The experimental results demonstrate that the proposed method: E-FPN achieves better classification accuracy and consistency compared to classical classification methods. The main contributions of this paper are as follows:

- 1) To address the issue of information loss during the feature extraction process, feature-level fusion is performed by selecting feature maps of different depths from the backbone feature extraction network. This fusion aims to supplement the lost information.

- 2) The classification results from multiple scales are further fused at the decision level using fusion rules based on evidence theory. The different classification results are treated as pieces of evidence, and the differences in probability distributions are utilized to optimize the classification results.

The remaining sections of this paper are composed as follows. Section II provides a review of related works. Section III introduces the relevant background knowledge. Section IV presents the overall network structure of the E-FPN. Section V provides detailed explanations of the experimental setup, including parameter settings, experimental procedures, and parameter discussions. Finally, in Section VI, the paper concludes with a summary and discusses future research directions.

## II. RELATED WORK

At the algorithmic level, deep learning-based image classification methods can be divided into two categories based on different feature extractors. The first category is CNN-based image classification methods, which have achieved remarkable breakthroughs in the past decade based on modern deep learning techniques. Krizhevsky et al. introduced rectified linear units (ReLU) in convolutional neural networks to achieve nonlinearity and used the Dropout technique to mitigate overfitting and learn more complex objects [22]. Karen Simonyan and Andrew Zisserman improved upon AlexNet by stacking  $3 \times 3$  convolutions and deepening the network structure to enhance classification accuracy [23]. However, as the networks became deeper, issues such as network degradation, vanishing gradients, and exploding gradients emerged. To address these problems, Kaiming He et al. introduced Batch Normalization (BN) to replace Dropout and solve the issues of vanishing and exploding gradients. They also introduced residual connections to address network degradation [24]. Saining Xie et al. introduced Inception on top of ResNet, transforming single-path convolutions into multi-path convolutions with multiple branches [25]. Gao Huang et al. proposed DenseNet in 2017, which connects each layer with all previous layers in a feed-forward fashion to alleviate the vanishing gradient problem and enhance feature propagation [26]. Tsung-Yu Lin et al. used two feature extractors to extract features from input images and combined them using a bilinear pooling function before performing classification to compensate for the features lost by a single feature extractor (B-CNN) [27]. To fully exploit the small features that can differentiate different categories, Jianlong Fu et al. proposed RA-CNN, which focuses the classification operation on regions with differentiating features using a recurrent attention projection mechanism [28].

To enhance the classification accuracy of CNN-based classification networks on satellite remote sensing images, Linqing Huang et al. proposed a classification method that converts images in the dataset into different color spaces and trains separate CNNs on each color space. Finally, the output results of each classifier are fused using evidence theory [29]. Yue

Chen et al. presented a method called Destruction and Construction Learning (DCL), which disrupts and shuffles input images to emphasize local detailed features. They employ a region alignment network to restore the image layout and learn semantic information from local regions, thereby strengthening the connections between neighboring regions [30]. Heliang Zheng et al. introduced a technique that extracts precise attention maps to highlight target regions with rich detailed features at a high resolution. They also employ knowledge distillation to transfer image detail features to the main network for image classification [31].

The second approach is based on the visual transformer method [32]. Similar to CNN, transformers have dominated the field of Natural Language Processing (NLP) in the past decade. Initially, when transformers were introduced to computer vision, they were primarily used to extract global contextual information from images, but their performance was not satisfactory. In the past two years, there have been breakthroughs in using large-scale pretraining on transformer-based CNN classification networks, which have surpassed the dominance of CNNs in traditional image domains. Examples of such networks include Vision Transformer (ViT) [33] and Shifted Window Transformer (SWIN-Transformer) [34].

In recent years, advancements in ship classification algorithms have involved various improvement approaches in academic research. For instance, Chen et al. employed a contrastive learning method to replace classical classification techniques. They designed a loss function to separate different categories and bring together similar ones [35]. Zhang et al. adopted a combination of traditional feature extraction methods and modern abstract feature extraction methods to enhance the representation capability of ship features [18]. Guo et al. utilized shape-aware feature extraction techniques, allowing the feature extraction process to better align with the distinctive spindle-shaped appearance of ships [36]. Building upon the bilinear pooling method, Zhang et al. made improvements to make it more suitable for ship classification tasks [37]. Additionally, Jahan et al. employed knowledge distillation and class balancing methods to achieve ship classification in SAR ship image [38].

### III. PRELIMINARIES

#### A. Cross Stage Partial Darknet(CSPDarkNet)

CSPDarkNet [39] can be divided into five main parts: Focus, Dark2, Dark3, Dark4, and Dark5, in sequential order. The Focus module focuses on aggregating the width and height information of the image into the channel information by subsampling the image's pixel values. The structures of Dark2 to Dark4 are well demonstrated in Figure 1, where each Dark part consists of a BaseConv layer and a CSPLayer. Each BaseConv layer consists of a convolutional layer, a BatchNorm2d layer, and an activation function. The entire CSPLayer can be viewed as a residual module, where one side of the residual branch passes through the BaseConv layer once, while the other side goes through  $n$  Bottleneck units after the BaseConv layer. The two parts are then concatenated

and subjected to another BaseConv operation. The structure of the Bottleneck unit, as shown in Figure 1, involves a  $1 \times 1$  and a  $3 \times 3$  convolution for the main branch, while the residual branch remains unchanged, and the two parts are finally added together. The Dark5 part is slightly different from the previous three parts. It introduces a Spatial Pyramid Pooling (SPP-bottleneck) module between the BaseConv and CSPLayer, which utilizes max pooling with three different kernel sizes to extract features and then combines them to increase the network's receptive field. Its structure is depicted in Figure 2.

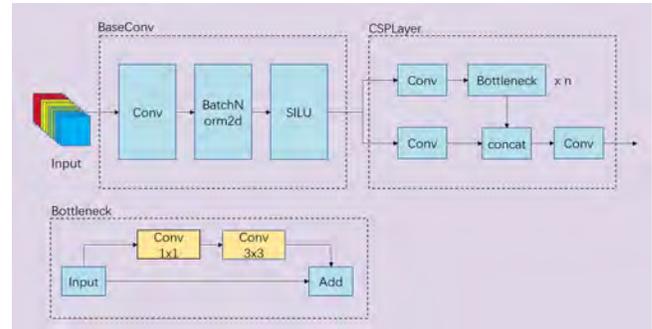


Figure 1. CSPDarkNet network structure.

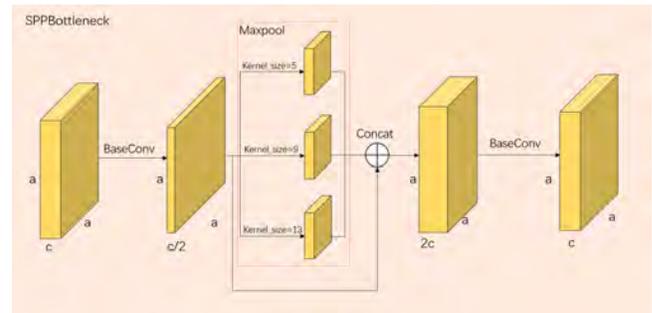


Figure 2. SPP-Bottleneck network structure.

#### B. Feature Pyramid Networks (FPN)

In convolutional networks, deep layers are more responsive to semantic features, while shallow layers are more responsive to image details. In image classification tasks, it has been validated by Karen Simonyan and others that deeper networks have a positive impact on image classification. However, deep convolutional layers tend to lose fine-grained details. Therefore, the FPN [40] model can be used to fuse features from shallow and deep layers, allowing the deep layers to complement the information lost during multiple convolutional operations, which is beneficial for subsequent classification tasks. The FPN structure is illustrated in Figure 3.

#### C. Evidence Theory

The evidence theory, established by Dempster and Shafer, represents propositions using mathematical sets [41]. Unlike

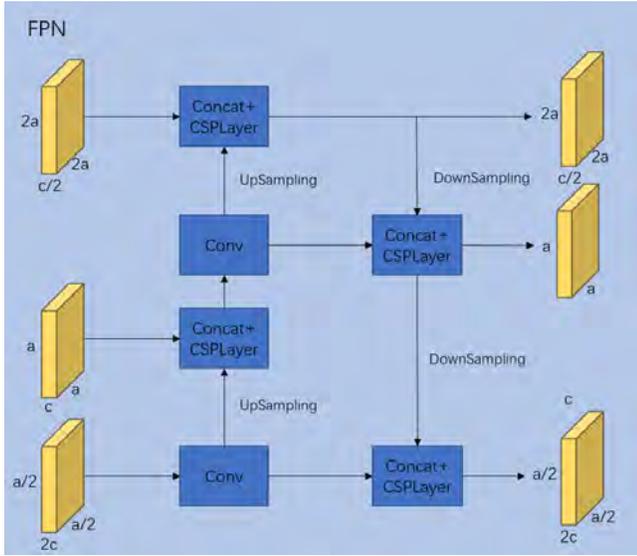


Figure 3. FPN network structure.

probability theory, which considers only single elements, evidence theory allows for multiple elements within a set. This theory is characterized by its ambiguity and the ability to perform imprecise reasoning at different levels of abstraction. It can differentiate between ignorance and equiprobability, enabling better representation of uncertain propositions. Evidence theory simulates the normal human thinking process, where one observes and collects information before synthesizing it from various aspects to make judgments and obtain results for a given problem.

In the Dempster-Shafer (DS) evidence theory, the sample space composed of all propositions is defined as a discernment framework, denoted as  $\Theta$ . It is a set comprising a group of mutually exclusive and collectively exhaustive propositions representing all possible answers to a given question. Let's assume another discernment framework defined as  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , where  $\theta_1, \theta_2, \dots, \theta_n$  represents a set of basic hypotheses, and  $\theta_i \cap \theta_j = \emptyset, i \neq j, i, j = 1, 2, \dots, n$  are subsets of it. The power set of  $\Theta$  is the set of all its subsets and is denoted as  $2^\Theta$ .

Basic Probability Assignment (BPA) refers to the process of calculating the basic probabilities for each piece of evidence in the discernment framework  $\Theta$ . This process is accomplished using the basic probability assignment function, denoted as the mass function  $m(x)$ , which reflects the degree of belief or confidence in a proposition. The mass function satisfies the following properties:

$$m : 2^\Theta \rightarrow [0, 1], \quad (1)$$

$$m(\emptyset) = 0, \sum_{A \subseteq \Theta} m(A) = 1. \quad (2)$$

In evidence theory, the uncertainty of evidence can be quantified using the belief function  $Bel(A)$  and the plausibility

function  $Pl(A)$ . The definitions and the relationship between the belief and plausibility functions are as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B), \quad (3)$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad (4)$$

$$= 1 - Bel(\bar{A}). \quad (5)$$

Dempster-Shafer fusion rule of two BPAs is given by

$$\begin{aligned} m(A) &= [m_1 \oplus m_2](A) \\ &= \begin{cases} 0, & A = \emptyset, \\ \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1-K}, & A \neq \emptyset \end{cases}, \end{aligned} \quad (6)$$

with

$$K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C) < 1. \quad (7)$$

$K$  represents the conflict coefficient, which can describe the magnitude of conflict between items of evidence, a higher value of  $K$  indicates a greater degree of conflict between the evidence.  $\frac{1}{1-K}$  serves as a normalization factor. For the combination of multiple items of evidence, the calculation follows a similar approach. Multiple belief functions can be combined using orthogonal sum to generate a new mass function, denoted as  $m_1 \oplus m_2 \oplus m_3 \oplus \dots \oplus m_n$ . If this combination exists, the order of calculation does not affect the result, satisfying the commutative and associative properties.

Suppose there are  $n$  sets of evidence  $E_1, E_2, \dots, E_n$ , with their corresponding basic belief assignment functions  $m_1, m_2, \dots, m_n$ , and focal elements  $A_1, A_2, \dots, A_n$ , within the given recognition framework. The classical Dempster's combination rule for these sets can be defined as follows:

$$m(A) = \begin{cases} \frac{\sum_{\cap A_i = A} \prod_{1 \leq i \leq n} m_i(A_i)}{1-K}, & A \neq \emptyset \\ 0, & A = \emptyset \end{cases}, \quad (8)$$

$$K = \sum_{\cap A_i = \emptyset} \prod_{1 \leq i \leq n} m_i(A_i). \quad (9)$$

The classical Dempster's combination rule is susceptible to paradoxes [42], and there are several classic paradoxical situations:

- 1) Conflict of evidence: When the basic belief assignment functions of multiple evidence sources exhibit strong conflicts, the fusion process may lead to highly unreasonable results and even fail to generate a consistent synthesis (complete conflict, i.e.,  $K = 1$ ).
- 2) One-vote veto: If there is a piece of evidence for which the basic belief assignment function for a specific proposition is 0, the fusion result will be 0 regardless of the values of other evidence's belief assignment functions. This reflects the limitation of the DS fusion rule in allocating conflict properly. For example, assuming there is evidence E1:  $m_1(a) = 0.999, m_1(b) = 0.001, m_1(c) = 0$ ; E2:  $m_2(a) = 0, m_2(b) = 0.001, m_2(c) = 0.999$ . Using the

formula, calculate the results as  $m(a) = m(c) = 0$ , and  $m(b) = 1$ . Clearly, the results are unreasonable.

- 3) Poor Robustness: Although the changes in the basic belief assignment values of the focal elements in the evidence are minimal, the synthesized results can be completely different. For example, modifying the evidence E1 in the previous example to:  $m_1(a) = 0.998$ ,  $m_1(b) = 0.001$ ,  $m_1(c) = 0.001$ , the synthesized result shows  $m(b) = 0.001$ , contrary to the previous result.

The Dezert-Smarandache Theory (DSMT) has made improvements to address the aforementioned issues. One of these improvements is the Proportional Conflict Redistribution rules no.5 (PCR5) [43], which reduces the generation of unreasonable results caused by significant conflicts between items of evidence compared to the DS fusion method. Additionally, weights can be assigned to the outputs of the FPN before performing the fusion operation to mitigate conflicts. In the PCR5 fusion rule, the conflicting degrees are proportionally allocated to each focal element, enabling a more reasonable fusion of two sources of evidence with high conflicts. The PCR5 fusion method for two BPAs is described as follows:

$$m_{1,2}^{PCR5}(A) = m_{1,2}^{Conj}(A) + \sum_{\substack{X \in 2^\Theta \\ X \cap A = \emptyset}} \left[ \frac{m_1(A)^2 m_2(X)}{m_1(A) + m_2(X)} + \frac{m_2(A)^2 m_1(X)}{m_2(A) + m_1(X)} \right], \quad (10)$$

where

$$m_{1,2}^{Conj}(A) = \sum_{A \cap B = A} m_1(A) m_2(B). \quad (11)$$

Among them,  $m_1$ , and  $m_2$  represent the two items of evidence;  $A$  and  $B$  denote the focal elements contained in the evidence;  $m_{1,2}^{Conj}(A)$  represents the non-conflicting product, and the latter part of the sum represents the allocation of all conflicting products containing  $A$  on  $A$ .

The weighting calculation method used in the experiment referred to the approach proposed by Zhunga Liu et al. [44], which adds a weight to the prefused data by calculating the difference between two BPA. The mass values corresponding to the two classifiers indicate the likelihood of the corresponding class being true, with higher values suggesting a higher probability. The collection of all classes judged as true can be represented as follows:

$$\Phi_i = \left\{ A \mid \frac{m_i(A)}{\max_{B \in \Omega} m_i(B)} > \lambda \right\}. \quad (12)$$

Among them,  $\Phi_i$  represents the set of true classes.  $\lambda$  denotes a threshold set between 0 and 1. When the ratio of the mass value corresponding to a class to the maximum mass value in that BPA exceeds the threshold, it is considered that the class may also be true. This approach increases tolerance for differences between the two classification results while retaining information that is beneficial for the final

classification result. The calculation method for the difference between the two BPAs is as follows:

$$K = \begin{cases} 0, & \text{if } \Phi_1 \cap \Phi_2 \neq \emptyset, \\ \sqrt{\left( \max_{A \in \Omega} m_1(A) \right) \left( \max_{B \in \Omega} m_2(B) \right)}, & \text{if } \Phi_1 \cap \Phi_2 = \emptyset. \end{cases} \quad (13)$$

The weight can be represented as follows:

$$\omega = 1 - K. \quad (14)$$

The weights are used to discount the two BPAs using Shafer's discounting operation, aiming to reduce conflicts between the two classifiers:

$$\begin{cases} m_i(A) = \omega \cdot m_i(A), \forall A \in \Omega, \\ m_i(\Omega) = 1 - \omega. \end{cases} \quad (15)$$

By employing the operation of adding weights, it is possible to reduce the negative impact of the classifier with lower classification ability on the final fusion results when there is significant conflict between the two classifiers. This, in turn, enhances the accuracy of the ultimate fusion outcome.

#### IV. METHODOLOGY

Most existing CNN-based models only utilize features or scales from the final stage as the ultimate classification features, making them single-scale classification models. However, shallow-level features of the network contain more detailed information. Neglecting shallow-level features without considering them can lead to decreased classification accuracy for similar or small objects during the classification process. Particularly when the image resolution is low, shallow-level features can retain more information and reduce the risk of feature loss. To better utilize the features of shallow-level networks, this paper proposes a method that uses multi-scale features and employs the Feature Pyramid Network (FPN) to fuse features from different scales. The fusion of multiple classification results is achieved using the fusion rules of evidence theory. This approach enables the model to learn abstract features at different levels of abstraction on different scales, thereby improving the model's classification accuracy and enhancing decision-making capabilities. Consequently, the E-FPN consists of three main components: the feature extraction network, the FPN feature fusion part, and the decision-level fusion based on evidence theory. Specifically, the feature extraction part is responsible for extracting abstract features from images, the FPN feature fusion part combines features from different scales, and the decision-level fusion part, based on evidence theory, integrates the classification results from multiple scales into the final classification result. Figure 4 illustrates the overall network structure of the proposed method in this paper. The feature extraction part utilizes the backbone network structure of YOLOX, using Darknet53 as the main network for extracting image features. Darknet53 combines the characteristics of ResNet and uses a residual network to ensure that the gradient problem caused by excessively deep networks is avoided during feature representation. From Figure 4, it can be observed that the Dark3, Dark4, and Dark5 parts

of the Darknet53 feature extraction network output feature maps of three different dimensionalities. These feature maps contain features of the objects to be classified at three different scales, and all three scales of feature maps are involved in the final classification decision step. In other words, by utilizing feature maps from different depths of the network for multi-scale feature fusion, better feature representation capability is ensured. For this feature extraction network, the chosen image input size is set to  $320 \times 320$ . As the network layers increase, the input image dimension transitions from  $320 \times 320$  to  $40 \times 40$  for Dark3,  $20 \times 20$  for Dark4, and  $10 \times 10$  for Dark5. Considering the depth of the feature extraction network, choosing larger or smaller image input sizes would result in insufficient feature extraction or feature loss, which is not conducive to classification decision-making. In fact, selecting an appropriate input size is also consistent with mature CNN models, such as VGG and ResNet.

By extracting features at different stages of the feature extraction network, different scales of feature maps are obtained, capturing information at three different scales. However, directly performing classification operations on these feature maps is not sufficient. Although the shallow-level feature information can propagate to deeper layers in the network, it may get diluted during convolutional operations, leading to the neglect of detailed information in the resulting deep-level features and a decrease in classification accuracy. From Figure 4, it can be observed that feature maps from different stages or scales participate in the object classification task. Therefore, the classification results obtained from feature maps at different scales will affect the final classification accuracy. It is necessary to enhance the classification accuracy of feature maps at different scales involved in the classification task as much as possible. To address this issue, this paper introduces a multi-scale feature fusion method. This method allows the deep-level network to learn detailed feature information from the shallow-level network, while the shallow-level network can learn abstract feature information from the deep-level network, thus improving the feature representation capability. With this approach, each scale of the feature map can learn richer information, leading to better classification accuracy in the subsequent classification process and ultimately improving the final classification accuracy. Subsequent experiments demonstrate that using the multi-scale feature fusion method can improve the accuracy of object classification. It achieves better classification results compared to using single-scale feature-based classification methods.

In light of the above, this paper employs Feature Pyramid Network (FPN) to perform feature-level fusion of the three feature maps obtained from the backbone network, aiming to complement the diluted detailed features during the feature extraction process. By obtaining three feature maps with the same input dimension, classification operations are separately performed on two of the feature maps, resulting in two sets of classification results. In this paper, the evidence theory is used to fuse the classification results from different scales. The evidence theory can handle uncertainty and incomplete infor-

mation by combining multiple pieces of evidence to improve classification accuracy. The multi-scale output classification results are treated as distinct sources of evidence, which are fused at the decision level using evidence theory to obtain the final classification result. Specifically, the classification results obtained from feature maps of different scales can be regarded as different sources of evidence, and the obtained classification results can be seen as probability distributions where each element represents the probability value of a corresponding class. Therefore, the maximum probability value in the obtained probability distribution cannot solely represent the current target class, as other higher probability values may correspond to the correct class as well. Hence, the obtained multiple probability distributions can serve as references from different aspects, rather than being definitive classification results. The use of the evidence theory enables the integration of the probability distributions obtained from different scales as different pieces of evidence, and through analyzing the differences between these pieces of evidence, a new probability distribution is derived as the classification result. This classification method resembles the decision-making process of human experts, who analyze and study information from multiple sources to make an informed judgment, resulting in a relatively accurate answer. Subsequent experiments have demonstrated that fusing the multi-scale classification results using the evidence theory can further improve the accuracy of ship classification, validating the effectiveness and applicability of the evidence theory in ship classification.

In this paper, the input images to the network are set to a size of  $320 \times 320$  in order to retain detailed features in the images. Various image augmentation techniques, such as random horizontal flipping, occlusion, and cropping, are applied to augment the dataset and enhance the network's performance. The input network used is CSPDarkNet, where the images are processed through the Focus module to extract a value for every other pixel, resulting in four feature maps that are then combined together. This process reduces the width and height information of the image while increasing the number of channels. This reduces the number of parameters and improves the network's performance while minimizing the loss of original information.

$$I = \text{concat}(X[\dots, :: 2, :: 2], X[\dots, 1 :: 2, :: 2], X[\dots, :: 2, 1 :: 2], X[\dots, 1 :: 2, 1 :: 2]) \quad (16)$$

Among them, the input image  $X$  undergoes a slicing operation denoted as  $X[\cdot]$ , where every pixel value is extracted to obtain four feature maps. The concatenation operation  $\text{concat}()$  is then applied to combine these four feature maps. After the Focus operation, the size of the resulting feature maps becomes  $160 \times 160 \times 12$ .

After the Focus module, the feature extraction stage follows, consisting of Dark2, Dark3, Dark4, and Dark5. The Dark5 part includes the SPPbottleneck module, which applies pooling layers with different kernel sizes to the image to increase the network's receptive field and extract more features. In

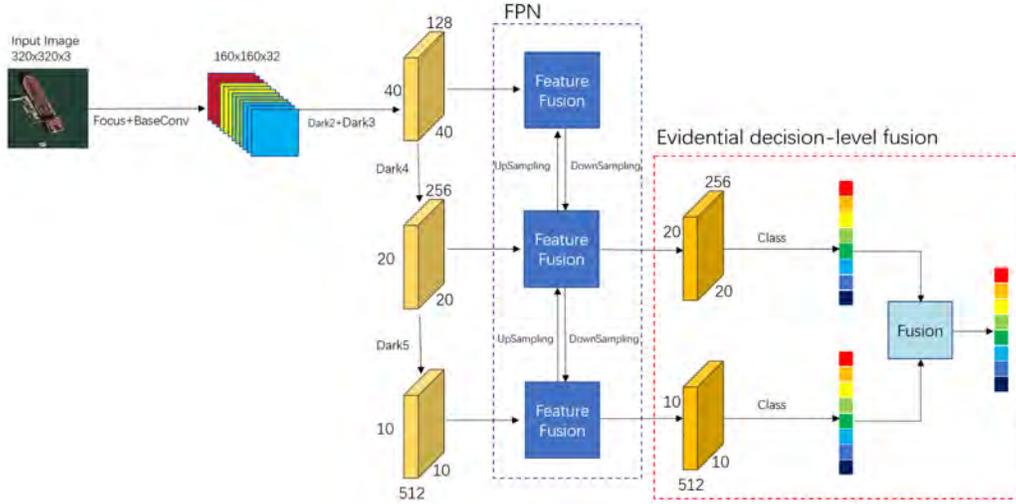


Figure 4. E-FPN network structure.

this study, the SPP-Bottleneck module utilizes pooling kernels of sizes  $5 \times 5$ ,  $9 \times 9$ , and  $13 \times 13$ . The feature maps obtained from Dark3, Dark4, and Dark5, denoted as  $I_3$ ,  $I_4$ , and  $I_5$ , are chosen as the outputs of the feature extraction network. The sizes of these feature maps are  $40 \times 40 \times 256$ ,  $20 \times 20 \times 512$ , and  $(10 \times 10 \times 1024)$ , respectively. Subsequently, these three feature maps are fed into the FPN network for feature-level fusion. In the fusion stage, the FPN layer takes the three feature maps with different dimensions and performs upsampling and downsampling operations to integrate the features from multiple scales, enriching the information within the feature maps at different scales.

$$I'_j = \text{concat}(f(I_j), g(I_j)), \quad (17)$$

$$f(I_j) = WI_j, \quad (18)$$

$$g(I_j) = \text{DownSampling}(\text{UpSampling}(f(I_j))). \quad (19)$$

In the provided formula,  $f(I_j)$  represents a convolutional operation applied to the feature map, while  $g(I_j)$  indicates the process of upsampling the feature map, followed by fusion with a shallow-level feature map, and then downsampling. Finally, the resulting feature map is concatenated with the feature map processed through the  $f(I_j)$  operations to obtain the final feature map used for classification. During the upsampling and downsampling process, the combined feature map is further integrated using the CSPLayer. This results in three feature maps ( $I'_3, I'_4, I'_5$ ) with the same dimensions as the input. Among these, the feature maps corresponding to the Dark4 and Dark5 dimensions ( $I'_4, I'_5$ ) are selected for the classification process. The classification component consists of a BaseConv, two convolutional layers, and three linear layers. In the linear layers, the flattened feature maps are sequentially reduced to dimensions of 256, 64, and 10, where the parameter 10 represents the number of classes for classification. The  $\text{softmax}()$  activation function is applied

to obtain the probability distributions ( $m_1$  and  $m_2$ ) for the output feature maps corresponding to the Dark4 and Dark5 scales. These probability distributions from the two scales are considered as evidence sources for decision-level fusion.

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^C e^{z_c}}, \quad (20)$$

where  $e^{z_i}$  represents the  $i$ -th value, and  $C$  represents the number of outputs, which is the number of classes.

Although the feature maps between different dimensions complement each other with feature information through the FPN operation, the probability distribution results obtained from different-sized feature maps still exhibit variations after classification. Hence, the differences in information between these two probability distributions can be utilized to optimize the classification results. Treating these two probability distributions as two sources of evidence, denoted as  $m_1$  and  $m_2$ , the DS fusion rule is employed to merge them. Initially, the conflict coefficient  $K$ , which represents the degree of dissimilarity between the two pieces of evidence, is computed using Equation (7) based on  $m_1$  and  $m_2$ . Subsequently, Equation (10) is applied to fuse the probability values of each class in  $m_1$  and  $m_2$ , resulting in a unique classification result. During the fusion process, the probability values corresponding to classes with relatively higher degrees of credibility in the probability distribution are accentuated, while the probability values corresponding to other classes are attenuated. If a scenario arises where two probability values in the distribution are similar, indicating hesitation between two classes, this method can leverage the differential information from other probability distributions to make decisions, thereby enhancing the reliability of the final classification result. Consequently, the final classification result is obtained. This approach provides a more reliable classification outcome compared to the individual fused results. The pseudocode for the E-FPN is outlined in Algorithm 1.

Algorithm 1: The method processing of an image.

---

**Input:** A ship image  $X$   
**begin**  
**Do** abstract feature extraction  
 $I_1 = X \otimes Focus$   
 $I_2 = I_1 \otimes Darknet2$   
 $I_3 = I_2 \otimes Darknet3$   
 $I_4 = I_3 \otimes Darknet4$   
 $I_5 = I_4 \otimes Darknet5$   
**End**  
**Do** FPN feature fusion  
 $I'_4, I'_5 = f_{FPN}(I_3, I_4, I_5)$   
**End**  
**Do** Classification  
 $I''_4 = flatten(I'_4)$   
 $I''_5 = flatten(I'_5)$   
 $m_1 = softmax\{FC(I''_4)\}$   
 $m_2 = softmax\{FC(I''_5)\}$   
**End**  
**Do** Decision fusion  
 $result = PCR5(m_1, m_2)$   
**End**  
**Output:** Classification tensor result.

---

In this case,  $I_1, I_2, I_3, I_4, I_5$  represent the outputs of each stage in the backbone network,  $DarknetN$ , and  $N \in 2, 3, 4, 5$  represents different parts of the backbone network,  $f_{FPN}(\cdot)$  refers to the feature fusion operation,  $K$  represents the conflict coefficient between evidence.  $flatten(\cdot)$  denotes the operation of flattening the feature map,  $FC(\cdot)$  represents the classification operation, and  $softmax\{\cdot\}$  maps the obtained classification results to the range  $[0,1]$ .

## V. EXPERIMENTAL

### A. Dataset

In this section, to validate the effectiveness of the E-FPN and compare it with other image classification algorithms, two datasets, CIFAR-10 and FGSCR10, were used.

CIFAR-10 is a small-scale dataset used for general object recognition. It consists of 10 classes of RGB images, with 6,000 images per class. The dataset is divided into a training set of 50,000 images and a test set of 10,000 images. The images have a size of  $32 \times 32$  pixels. This dataset is used to evaluate the classification performance of traditional natural images.

The FGSCR-42 dataset is a publicly available dataset for fine-grained ship classification in remote sensing images. It contains 42 classes with a total of 9,320 images, and the images have varying resolutions. For the experiments in this section, we selected 10 classes with a larger number of image samples, resulting in a total of 5,220 images. This dataset is used to evaluate the classification performance in the context of remote sensing images and fine-grained object classification. The composition and sample images are shown in Table I and Figure 5, respectively.

### B. Experimental Parameter Settings

In this study, we will compare classic classification algorithms, namely ResNet50, ResNeXt50, VGG19, and VGG16,

Table I  
SHIP IMAGE CATEGORY.

Category	Train	Test
Arleigh_Burke-class_destroyer	290	290
Cargo_ship	189	189
Civil_yacht	389	388
Container_ship	228	227
Medical_ship	161	161
Nimitz-class_aircraft_carrier	277	276
San_Antonio-class_transport_dock	160	159
Ticonderoga-class_cruiser	304	303
Towing_vessel	389	389
Wasp-class_assault_ship	227	226

along with the fine-grained image classification algorithms B-CNN and DCL, against E-FPN to evaluate its effectiveness. For the classic classification algorithms, the image size was uniformly adjusted to  $224 \times 224$ . Data augmentation techniques, including random horizontal flipping, random occlusion, and random cropping, were applied to the dataset images. The initial learning rate was set to 0.0001, and the training batch size, weight decay, and decay epoch were set to 64, 0.1, and 50, respectively. The Adam optimizer was selected, and the cross entropy loss function was employed for calculating the loss. In the proposed method, to preserve more image feature information, the dataset images were uniformly resized to  $320 \times 320$  while keeping the remaining parameters consistent with the aforementioned settings. This was done to evaluate the effectiveness of E-FPN in terms of classification performance, by comparing it with the baseline models. Further details regarding the metrics and evaluation will be presented in the following sections. The experiments were conducted using the GPU resource A5000-24G.

### C. Evaluation Indices

In this experiment, overall accuracy (OA) and the Kappa statistic were employed as evaluation metrics to assess the classification performance of the models. The details are as follows:

- 1) OA: Overall Accuracy is defined as the ratio of correctly classified samples to the total number of samples. The calculation method is as follows:

$$OA = \frac{1}{N} \sum_i^N f(i), \quad (21)$$

where  $N$  represents the total number of image samples in the dataset.  $f(i)$  represents whether the classification of the  $i$ -th sample is correct. If the classification is correct, the value of  $f(i)$  is 1; otherwise, it is 0.

- 2) The Kappa coefficient is used for consistency testing and can also be used to measure classification accuracy. Its calculation is based on the confusion matrix. The calculation method is as follows:

$$k = \frac{p_0 - p_e}{1 - p_e}, \quad (22)$$

where  $p_0$  represents the ratio of the sum of correctly classified samples in each class to the total number

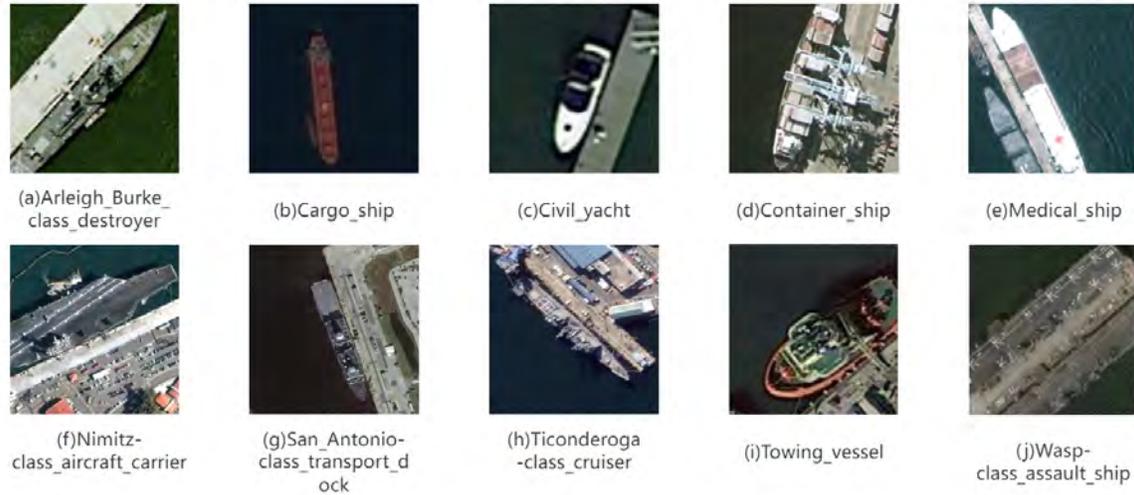


Figure 5. FGSCR-10 Image Examples.

of samples, which corresponds to the overall accuracy. Assuming the true number of samples in each class is denoted as  $a_1, a_2, \dots, a_c$ , the predicted number of samples in each class is denoted as  $b_1, b_2, \dots, b_c$ , and the total number of samples is  $n$ , then the equation can be expressed as follows:

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_c \times b_c}{n \times n}. \quad (23)$$

The calculation result of Kappa falls between  $[-1, 1]$ , but it typically ranges  $[0, 1]$ . It can be categorized into five levels to represent different levels of agreement:  $[0.0, 0.20]$  slight agreement,  $[0.21, 0.40]$  fair agreement,  $[0.41, 0.60]$  moderate agreement,  $[0.61, 0.80]$  substantial agreement, and  $[0.81, 1]$  almost perfect agreement.

#### D. Performance Evaluation

In this section, the effectiveness of the proposed method is evaluated by comparing it with classical image classification networks on the CIFAR-10 and FGSCR-10 datasets. The validation results are shown in Table II and Table III. Where bold typeface represents the best while underlining represents the second-best.

Table II  
COMPARED WITH CLASSICAL NETWORK OA.

Method	FGSCR-10	CIFAR-10
Resnet50	0.9677	0.9320
Resnext50	0.9631	0.9319
VGG16	0.9685	0.9330
VGG19	0.9405	0.9451
B-CNN	0.9663	0.9242
DCL	0.9731	<b>0.9504</b>
E-FPN	<b>0.9804</b>	<u>0.9478</u>

In Tables II and III, two metrics are used to evaluate the classification performance, compare four classical classification

Table III  
COMPARED WITH CLASSICAL NETWORK KAPPA.

Method	FGSCR-10	CIFAR-10
Resnet50	0.9638	0.9220
Resnext50	0.9681	0.9327
VGG16	0.9573	0.9424
VGG19	0.9336	0.9390
B-CNN	0.9621	0.9157
DCL	0.9693	0.9449
E-FPN	<b>0.9776</b>	<b>0.9450</b>

networks and two fine-grained image classification networks with E-FPN. The proposed method is evaluated on the CIFAR-10 dataset using two metrics, OA and Kappa. The results indicate that E-FPN achieved excellent performance in both metrics, with an OA of 94.78% and a Kappa value of 0.945, obtaining the second-best and best scores, respectively. This demonstrates the effectiveness of the E-FPN in the traditional natural image dataset.

In the FGSCR-10 dataset, the proposed method achieved an OA of 98.04% and a Kappa value of 0.9776. Compared to the other four classical methods, the E-FPN showed an improvement in OA ranging from 1.15% to 3.95% and an improvement in the Kappa metric ranging from 0.0095 to 0.044. When compared with the other two fine-grained image classification algorithms, E-FPN also achieved excellent results, with the highest OA and Kappa values.

Through the experiments on the two datasets, it can be observed that all algorithms showed similar performance on the CIFAR-10 dataset, and in some cases, B-CNN even exhibited lower accuracy compared to the baseline model. This could be attributed to the low resolution of the images in this dataset, as certain algorithmic improvements may not perform as effectively under such conditions.

In the FGSCR-10 dataset, the performance of the proposed

method surpassed that of the other four baseline models. This may be due to the fact that the FGSCR-10 dataset involves fine-grained classification targets. After feature extraction by the backbone network, the E-FPN utilizes the FPN method to fuse features at different scales, which allows for complementary details among the three-dimensional feature maps. Finally, the classification results of different feature maps are fused using the evidence theory-based decision-level fusion method, further correcting the classification results. For example, when an image is misclassified, its correct classification has a probability value that is close to the probability value of the current misclassification. When another set of probability distributions is fused, the probability value corresponding to the correct classification is also large. After fusion, the probability value of the correct classification may become the largest, resulting in the final correct result. As a result, the proposed method demonstrates an advantage over the other methods in the FGSCR-10 dataset. Compared to the other two fine-grained image classification algorithms, our proposed E-FPN outperforms B-CNN and DCL. This may be attributed to the effective extraction of object's fine-grained features using our multi-scale approach, and the decision-level fusion enables comprehensive analysis of classification results from different perspectives.

In terms of the Kappa metric, all classification methods achieved a performance exceeding 90% on both datasets, indicating a level of consistency considered "almost perfect." Compared to the other four baseline models, E-FPN exhibited further improvement in this metric, signifying enhanced classification accuracy for each class and its general applicability. Additionally, when compared to the fine-grained image recognition algorithms (B-CNN and DCL), E-FPN also shows improvement in terms of Kappa. Furthermore, Figure 6 provides a detailed visualization of the classification results for each class, demonstrating the proposed method's performance in terms of confusion matrices on both the CIFAR-10 and FGSCR-10 datasets. There are very few dark areas outside the diagonal, indicating a reduced number of misclassifications. This visual representation intuitively demonstrates the effectiveness of E-FPN.

Table IV presents the number of parameters, FLOPs (floating-point operations), and inference time for the seven models.

Table IV  
COMPARISON OF THE NUMBER OF PARAMETERS AND FLOPS.

Method	Params(M)	FLOPs(G)	Inference time(ms)
Resnet50	23.53	4.13	35.854
Resnext50	23	3.82	35.418
VGG16	134.33	15.52	<b>33.658</b>
VGG19	139.62	19.96	34.574
B-CNN	<b>17.34</b>	61.93	49.243
DCL	23.57	16.53	48.165
E-FPN	79.22	<b>3.58</b>	45.224

It can be observed that VGG16 and VGG19 have significantly higher numbers of parameters and FLOPs compared

to the other baseline models. This is likely due to their deeper network architectures and the utilization of numerous convolutional layers. On the other hand, ResNet50 and ResNeXt50 have smaller numbers of parameters and FLOPs. This reduction can be attributed to the utilization of residual structures, which help reduce network depth and complexity. Among the five methods, E-FPN has a higher number of parameters compared to ResNet50 and ResNeXt50 but lower than VGG16 and VGG19. However, its FLOPs are the lowest among the five methods, indicating relatively low computational cost when performing the classification task. This is because the proposed method introduces an additional FPN network while the backbone network adopts the residual approach to reduce its depth. Comparing the fine-grained image classification models, E-FPN has the highest number of parameters, suggesting higher storage requirements. However, its FLOPs remain the lowest, indicating that, compared to the other six models, E-FPN requires fewer computational resources during the inference phase, making it suitable for deployment on mobile and edge devices. This observation is evident from the inference speed, where all three fine-grained image recognition models, including E-FPN, require higher inference time than the four baseline models. However, in fine-grained image recognition models, the inference time of the E-FPN model is lower than the other two (B-CNN and DCL). This demonstrates the advantage of E-FPN in terms of inference speed.

Additionally, the DS fusion method used in the E-FPN incurs minimal additional computational cost for the network. As a result, the increase in network parameters is relatively small, and the FLOPs are the lowest among the all models.

By comparing the experimental results from the two aforementioned tables, it can be concluded that the E-FPN is effective on both traditional natural image datasets and fine-grained remote sensing image datasets. In the description of the FPN network structure, it was mentioned that three feature maps of different dimensions were utilized, but during the final decision-level fusion, only the results from the deeper two scales of feature maps were selected for fusion. In the following, we will discuss the impact of choosing different dimension feature maps for decision-level fusion on the final results. The results of these experiments are presented<sup>1</sup> in Table V and Table VI.

Table V  
FUSION OF RESULTS FROM DIFFERENT SCALES IN CIFAR-10 DATASET.

Dataset	Dark3	Dark4	Dark5	OA	Kappa
CIFAR-10	✓	×	×	0.9374	0.9304
	×	✓	×	0.9438	0.9375
	×	×	✓	<b>0.9516</b>	0.9462
	✓	✓	×	0.9431	0.9367
	×	✓	✓	0.9492	<b>0.945</b>
	✓	×	✓	0.948	0.9422
	✓	✓	✓	0.9478	0.942

<sup>1</sup>The symbol "✓" indicates the usage of feature maps at that scale, while "×" indicates their exclusion.

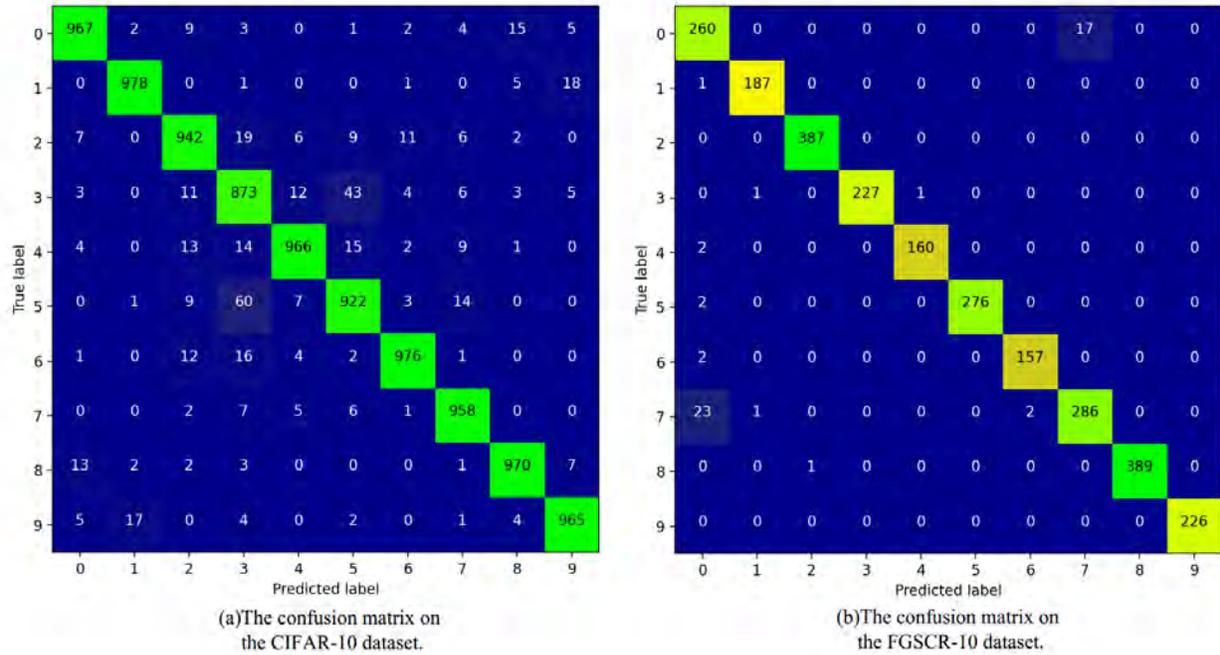


Figure 6. The confusion matrices of E-FPN on CIFAR-10 and FGSCR-10 are presented. (a) shows the confusion matrix obtained on the CIFAR-10 dataset, while (b) shows the confusion matrix obtained on the FGSCR-10 dataset.

Table VI  
FUSION OF RESULTS FROM DIFFERENT SCALES IN FGSCR-10 DATASET.

Dataset	Dark3	Dark4	Dark5	OA	Kappa
FGSCR-10	✓	×	×	0.9773	0.9746
	×	✓	×	0.9773	0.9746
	×	×	✓	0.9773	0.9746
	✓	✓	×	0.978	0.975
	×	✓	✓	<b>0.9804</b>	<b>0.978</b>
	✓	×	✓	<b>0.9804</b>	0.9478
	✓	✓	✓	<b>0.9804</b>	<b>0.978</b>

In this experiment, different combinations of feature maps were fused for each dataset, and the impact of pairwise fusion of different feature maps on the final results was compared. The last line represents the results obtained by fusing all three feature maps together. Dark3, Dark4, and Dark5 represent the probability distributions of the classification results from the FPN fused outputs of the backbone network. In Table V, the CIFAR-10 dataset was used. It can be observed that before decision-level fusion, the OA gradually improves as the network layers deepen. However, after fusion, the OA is lower than the OA of the Dark5 output result. Among the fused results, the fusion of Dark4 and Dark5 achieves the highest OA of 94.92%. Furthermore, its Kappa value is superior to the other three results, being 0.945. Preliminary analysis suggests that this may be due to significant conflicts in the probability values among certain categories before fusion, resulting in an unreasonable probability distribution after fusion, thus leading to incorrect fusion results. Further investigation of this issue will be discussed in subsequent sections.

Table VI displays the results obtained from the FGSCR-10 dataset, Dark3, Dark4, and Dark5 have the same classification OA of 97.73%. However, the OA improves after fusion. The fusion of Dark3 and Dark4 achieves an OA of 97.8%, while the fusions of Dark4 and Dark5, Dark3 and Dark5, and all three (Dark3, Dark4, and Dark5) have an OA of 98.04%. The performance of the Kappa index is consistent with the OA, with the fusion of Dark3 and Dark4 resulting in a Kappa value of 0.975, while the other three fusions all have a Kappa value of 0.978. By comparing the results before and after fusion, it can be observed that the samples correctly classified by Dark3, Dark4, and Dark5 are not entirely the same, and in the probability distributions of misclassified samples, the probability values for the correct class are close to those of the misclassified class. Therefore, after fusion, some misclassified samples are corrected, resulting in an improvement in the final OA of the results.

According to Table VI, it can be observed that the highest OA results after fusion are obtained by combining Dark5 with other parts, and these results are superior to the results obtained by fusing Dark3 and Dark4. It can be seen that the results obtained from the deeper parts of the network have a more reliable probability distribution. However, the results obtained by fusing all three parts together show a slight decrease compared to the fusion of Dark4 and Dark5. This may be due to the fact that during fusion, the probability values of the correct class and the misclassified class for all three inputs are very close, and since Dark3 has classification errors, the final result was not corrected to the correct class during fusion, resulting in a decrease in OA. Therefore, in the

experiment, this study chooses to fuse Dark4 and Dark5 for the fusion process.

The E-FPN in this paper consists of three parts: the feature extraction network, the FPN network, and the decision fusion part. During the training process, the crossentropy loss values of the three outputs from FPN are summed to calculate the overall loss value. Specifically, the obtained loss values in the network are referred to as loss\_0, loss\_1, and loss\_2. However, for the final decision, only the output results from Dark4 and Dark5 are selected for fusion. Therefore, the next step is to explore the impact of the loss\_0 value obtained from Dark2 on the classification performance and the effect of using FPN for fusion at the feature level.

According to the experimental results in Table VII, on the CIFAR-10 dataset, the removal of loss\_0 slightly improves the OA to 95.13%. This may be because FPN has multiple output classification results, and adding loss\_0 during the training process may lead to oscillation and decision risk. Additionally, the OA gap between Dark4 and Dark5 is very small, and their Kappa values are similar. Without using FPN for feature-level fusion, the OA is lower in both cases, and the fused OA and Kappa values are also lower compared to the two cases with FPN. This indicates that the classification results of the shallow layers may have a negative impact on decision-making and fusion in datasets with clear image features.

Table VII

THE ABLATION EXPERIMENTS OF E-FPN ON THE CIFAR-10 DATASET.(DARK3, DARK4, AND DARK5 REPRESENT THE CLASSIFICATION OA ON THREE DIFFERENT SCALES.)

CIFAR-10	DarkNet +FPN+loss_0	DarkNet +FPN	DarkNet +loss_0
Dark3	0.9374		0.8865
Dark4	0.9438	0.9401	0.9344
Dark5	<b>0.9516</b>	0.9511	0.9428
E-FPN OA	0.9492	<b>0.9513</b>	<b>0.944</b>
E-FPN Kappa	0.945	0.945	0.9377

However, the positive impact of the shallow layers in the feature-level fusion should not be ignored, as shown in Table VIII.

Table VIII

THE ABLATION EXPERIMENTS OF E-FPN ON THE FGSCR-10 DATASET.(DARK3, DARK4, AND DARK5 REPRESENT THE CLASSIFICATION OA ON THREE DIFFERENT SCALES.)

FGSCR-10	DarkNet +FPN+loss_0	DarkNet +FPN	DarkNet +loss_0
Dark3	0.9773		0.9605
Dark4	0.9773	0.9743	0.972
Dark5	0.9773	<b>0.9754</b>	0.9735
E-FPN OA	<b>0.9804</b>	0.975	<b>0.9781</b>
E-FPN Kappa	0.978	0.972	0.9754

When conducting experiments on the FGSCR-10 dataset, it was found that adding loss\_0 and using FPN for feature-level fusion resulted in higher OA and Kappa values compared to not using FPN or not adding loss\_0, achieving 98.04%

and 0.978, respectively. This indicates that the classification performance for each class object in the dataset is excellent. Under the conditions of removing FPN and removing loss\_0, the OA gap between Dark4, Dark5, and the fused result is small. However, it can be observed that the OA of the fused result is better than the individual results. As mentioned earlier, although the outputs of the shallow network can have a negative impact on the final decision-level fusion, the features learned by the shallow network still have a positive influence on the classification results in the feature-level fusion process.

Based on the experiments and discussions, it can be concluded that using the FPN structure and training the shallow network for classification improves the classification performance on the fine-grained remote sensing image dataset. The FPN structure complements the detailed features lost in the deep network. Since the FPN structure used in the paper involves the fusion of information from three layers, adding loss\_0 for classification training in the top layer of the network can facilitate the learning of more useful feature information, further enhancing the feature fusion effect. The results in Table VIII indicate that employing the feature-level fusion method helps improve the classification performance of fine-grained remote sensing image classification and further enhances the classification performance after decision-level fusion.

The experimental parameter section in this paper mentions that, unlike the four other classification methods used in the comparative experiments, the image input size for the E-FPN in this paper is  $320 \times 320$ , while the four classical classification methods use an image input size of  $224 \times 224$ . The purpose of this choice is to preserve more image feature information. However, it should be noted that a larger input image size does not necessarily guarantee better performance. Tables IX and X present a comparison of the impact of different input image sizes on the classification performance.

Table IX

INPUT IMAGES OF DIFFERENT SIZES IN CIFAR-10.(DARK3, DARK4, AND DARK5 REPRESENT THE CLASSIFICATION OA ON THREE DIFFERENT SCALES.)

CIFAR-10	640×640	320×320	224×224	160×160
Dark3	0.9161	0.9374	0.9171	0.9012
Dark4	0.9164	0.9438	0.9277	0.9123
Dark5	0.9242	<b>0.9516</b>	0.9341	0.9169
E-FPN OA	<b>0.9248</b>	0.9492	<b>0.9348</b>	<b>0.9174</b>
E-FPN Kappa	0.9164	0.945	0.9275	0.9082

In the experiments comparing the impact of different input image sizes on classification performance, the image size of  $224 \times 224$ , which is the same as the other four classical algorithms, was selected. Additionally, scaled versions of  $640 \times 640$  and  $160 \times 160$  were used. Based on the data in Tables IX and X, it was found that the input image size of  $320 \times 320$  achieved the best performance in terms of classification OA and Kappa value. Furthermore, in both tables, as the input image size decreased from large to small, the classification OA initially increased and then decreased. Therefore, it is not necessarily

Table X  
INPUT IMAGES OF DIFFERENT SIZES IN FGSCR-10.(DARK3, DARK4, AND DARK5 REPRESENT THE CLASSIFICATION OA ON THREE DIFFERENT SCALES.)

FGSCR-10	640×640	320×320	224×224	160×160
Dark3	<b>0.9758</b>	0.9773	0.9551	<b>0.9677</b>
Dark4	0.9746	0.9773	0.9605	0.9674
Dark5	0.9654	0.9773	<b>0.9635</b>	0.9616
E-FPN OA	0.9693	<b>0.9804</b>	0.9628	0.9658
E-FPN Kappa	0.9655	0.978	0.9582	0.9616

true that a larger input image size leads to better performance, and the appropriate size should be chosen based on the specific circumstances.

Table IX presents the influence of different input image sizes on classification performance, revealing that the classification OA for each size increases with the depth of the network. Except for the 320×320 size group, all other size groups exhibit an increase in classification OA after decision-level fusion. However, the final accuracy remains lower than that of the 320×320 size group. Regarding the Kappa index, the 320×320 size group still performs the best. These data indicate that for traditional natural image datasets, which have easily discernible image features, adequate feature extraction enables effective classification, necessitating only the selection of an appropriate input image size.

Table X demonstrates the impact of different input image sizes on classification performance in the FGSCR-10 dataset. In contrast to Table IX, Table X does not observe an increase in classification OA with network depth. In the 640×640 and 160×160 size groups, a decline in classification OA is observed as network depth increases. This may be due to excessively large or small feature maps that fail to effectively propagate relevant features in FPN feature fusion. For the 160×160 size group, the small image size may lead to the loss of crucial detail features, resulting in reduced classification OA. This could also result in significant conflicts between the generated probability distributions, making it difficult to correct misclassifications during decision-level fusion and ultimately decreasing the OA of the fused results. In the 320×320 size group, Dark3, Dark4, and Dark5 exhibit higher classification OA than the other groups. Although these three groups have the same classification accuracy, decision-level fusion further enhances their OA. These data demonstrate that the proposed classification method, when applied to fine-grained remote sensing image datasets, benefits from using appropriately sized input images. This enables the extraction of abstract features while retaining some detailed features, facilitating subsequent image classification operations.

In the previous sections, we discussed the network architecture and input image data. In Section II, the limitations of the DS fusion method were mentioned, specifically, the issue of unreasonable fusion results when significant conflicts exist between two input evidence. To overcome this problem, this paper adopts the PCR5 fusion method and utilizes the Shafer discounting method to weigh the evidence, reducing conflicts

between input evidence. The obtained results are compared with those of the DS fusion method.

Table XI and Table XII present the OA and Kappa values obtained using three different fusion rules on the CIFAR-10 dataset. The DS fusion rule is the fusion rule adopted in this paper, PCR5 is the proportional conflict redistribution method mentioned in Section II of this paper, and wPCR5 refers to the addition of weights to the probability distributions before using PCR5 fusion rule, applying the Shafer discounting method to discount the evidence and reduce conflicts between input data. From Table XI, it can be observed that the OA of Dark3×Dark4, Dark3×Dark5, and Dark4×Dark5 combinations under the DS fusion rule and PCR5 fusion rule is almost indistinguishable. However, for the Dark4×Dark5 combination, the OA decreases when using the PCR5 rule compared to the DS rule. After applying the wPCR5 fusion rule, the OA improves compared to both the DS rule and the PCR5 rule for all three combinations. This improvement may be attributed to the already high classification OA before fusion, indicating a relatively small conflict between the probability distributions of the two input data. The PCR5 fusion rule primarily aims to mitigate the impact of conflicts on fusion results and prevent the generation of unreasonable output values. By adding weights and employing the PCR5 rule, conflicts between the two inputs can be further effectively reduced, leading to better results. The Kappa values generally exhibit a similar pattern to the OA results. The wPCR5 rule yields slightly better results compared to the DS and PCR5 rules, but the improvement is marginal, while there is little difference between the DS rule and the PCR5 rule.

Table XI  
PRECISION COMPARISON OF DIFFERENT DECISION - LEVEL FUSION METHODS IN CIFAR-10.

CIFAR-10 OA	Dark3×Dark4	Dark3×Dark5	Dark4×Dark5
E-FPN with DS	0.9431	0.948	0.9492
E-FPN with PCR5	0.9432	0.9479	0.9489
E-FPN with wPCR5	<b>0.9442</b>	<b>0.9508</b>	<b>0.9509</b>

Table XII  
KAPPA COMPARISON OF DIFFERENT DECISION - LEVEL FUSION METHODS IN CIFAR-10.

CIFAR-10 Kappa	Dark3×Dark4	Dark3×Dark5	Dark4×Dark5
E-FPN with DS	0.9367	0.9422	<b>0.945</b>
E-FPN with PCR5	0.9368	0.9421	0.9432
E-FPN with wPCR5	<b>0.938</b>	<b>0.945</b>	<b>0.945</b>

Table XIII and Table XIV compare the OA and Kappa values on the FGSCR-10 dataset. Similar to the results obtained on the CIFAR-10 dataset, the DS fusion rule and the PCR5 fusion rule yield nearly identical results. However, for the Dark3×Dark5 combination with higher conflicts, the PCR5 rule slightly outperforms the DS rule. When using the wPCR5 rule, the performance is slightly worse than the previous two rules. The same trend is observed in the Kappa values. However, in the case of fine-grained remote

sensing image datasets, the probability values of each class in the classification distributions are close, making it difficult to compute favorable weights, as mentioned in Section II. Consequently, the weighting approach weakens the confidence of certain correctly classified classes during discounting operations, resulting in suboptimal final results. Regarding the Kappa values, there is little difference among the three fusion methods.

Table XIII  
PRECISION COMPARISON OF DIFFERENT DECISION - LEVEL FUSION METHODS IN FGSCR-10.

FGSCR-10 OA	Dark3×Dark4	Dark3×Dark5	Dark4×Dark5
E-FPN with DS	<b>0.978</b>	0.9804	<b>0.9804</b>
E-FPN with PCR5	<b>0.978</b>	<b>0.9812</b>	<b>0.9804</b>
E-FPN with wPCR5	0.977	0.9796	0.98

Table XIV  
KAPPA COMPARISON OF DIFFERENT DECISION - LEVEL FUSION METHODS IN FGSCR-10.

FGSCR-10 Kappa	Dark3×Dark4	Dark3×Dark5	Dark4×Dark5
E-FPN with DS	0.975	<b>0.978</b>	<b>0.978</b>
E-FPN with PCR5	<b>0.976</b>	<b>0.978</b>	<b>0.978</b>
E-FPN with wPCR5	0.975	0.977	0.9776

Based on the analysis above, it can be observed that the DS fusion rule and the PCR5 fusion rule yield almost identical results on both datasets. The wPCR5 method performs slightly better than the previous two methods on traditional natural image datasets but slightly worse on fine-grained remote sensing image datasets. Additionally, the computation complexity of PCR5 and wPCR5 is higher than that of the DS rule, and the complexity increases more noticeably with a larger number of classes to be classified. Therefore, when there is no significant conflict between the two probability distributions, the DS fusion rule is chosen in this paper.

In the previous experiments, it was mentioned that in the fine-grained remote sensing image dataset, the method of adding weights to reduce conflicts between evidence actually weakened the credibility of some correctly classified results. In the process of calculating the weights, a threshold was set for the ratio between the mass values of each class and the maximum mass value to preserve the differences between the two classification results. In the previous experiments, a threshold of 0.5 was set. The impact of the threshold value on the OA and Kappa value after fusion can be seen in Table XV and Table XVI, where the value  $\lambda$  represents the threshold chosen for calculating weights.

The Tables XV and XVI demonstrates the influence of threshold values ranging from 0.1 to 0.9 on the classification OA and consistency in two datasets. In the CIFAR-10 dataset, as the threshold value increases from 0.1 to 0.9, the classification OA gradually rises to 95.17%. Compared to the threshold value of 0.1, there is an improvement of 0.24%. The Kappa value increases from 0.9436 to 0.9463. In this dataset, when the threshold value increases, it filters out categories

Table XV  
COMPARISON OF DIFFERENT THRESHOLDS UNDER CIFAR-10.

$\lambda$	OA	Kappa
0.1	0.9493	0.9436
0.2	0.9497	0.9441
0.3	0.95	0.9444
0.4	0.9501	0.9445
0.5	0.9509	0.945
0.6	0.9516	0.9462
0.7	0.9516	0.9462
0.8	<b>0.9517</b>	<b>0.9463</b>
0.9	<b>0.9517</b>	<b>0.9463</b>

Table XVI  
COMPARISON OF DIFFERENT THRESHOLDS UNDER FGSCR-10.

$\lambda$	OA	Kappa
0.1	<b>0.98</b>	<b>0.9776</b>
0.2	0.9796	0.9771
0.3	0.9796	0.9771
0.4	0.9796	0.9771
0.5	<b>0.98</b>	<b>0.9776</b>
0.6	<b>0.98</b>	<b>0.9776</b>
0.7	<b>0.98</b>	<b>0.9776</b>
0.8	<b>0.98</b>	<b>0.9776</b>
0.9	0.9796	0.9771

with lower probability values in the probability distributions, retaining other potential options for correct classification. This preserves some differences between classifiers as complementary information, which benefits subsequent fusion operations.

In the FGSCR-10 dataset, changing the threshold value from 0.1 to 0.9 has almost no impact on the classification OA and Kappa values. This indicates that the threshold value has little effect on the fusion results in this dataset. Table XVII displays partial probability distributions generated by Dark5. It can be observed that the reason for this phenomenon is that one class in the probability distribution before fusion has a significantly high probability value, and the ratios of other probabilities to it might be lower than 0.1. Consequently, the variation of the threshold value does not affect the final result.

Based on the experiments, it can be concluded that the threshold value has almost no impact on the classification of OA in the FGSCR-10 dataset. In the CIFAR-10 dataset used in this experiment, setting a higher threshold value allows for the rational utilization of differences between different classifiers, obtaining complementary information and thereby improving the OA of the classification results.

## VI. CONCLUSIONS

This study proposes a feature fusion and decision fusion method that combines FPN with evidence theory to improve classification accuracy. The effectiveness of this method is validated on both traditional natural image datasets and fine-grained remote sensing image classification datasets. For the fine-grained remote sensing image dataset, FPN is utilized for feature-level fusion to capture the lost detailed features in shallow networks. Simultaneously, evidence theory is applied to modify the generated probability distributions. In the

Table XVII  
PARTIAL PROBABILITY DISTRIBUTION UNDER THE FGSCR-10  
WITH E-FPN.

Category	$m_1$	$m_2$	$m_3$
Arleigh_Burke-class_destroyer	0.98235	1	0
Cargo_ship	0.00001	0	0
Civil_yacht	0	0	0
Container_ship	0	0	0
Medical_ship	0	0	0
Nimitz-class_aircraft_carrier	0	0	0
San_Antonio-class_transport_dock	0	0	0
Ticonderoga-class_cruiser	0.01764	0	1
Towing_vessel	0	0	0
Wasp-class_assault_ship	0	0	0

experimental section, the network architecture and parameters of this method are discussed, and the impact of different fusion rules on the final classification accuracy is compared. The experimental results demonstrate that selecting appropriate sizes of input images and using both feature-level fusion and decision-level fusion can effectively improve classification accuracy. Additionally, reducing conflicts between different classifier results through the addition of weights contributes to the enhancement of classification results in certain cases.

The proposed E-FPN method still has some issues that need to be optimized. For instance, as demonstrated in Tables II, III, and IV in Section V-D of the paper, currently, E-FPN did not achieve significant improvement compared to the other three fine-grained image classification algorithms in the ship fine-grained classification task. Furthermore, when compared to the baseline models on the CIFAR-10 dataset, the improvement of our proposed method was not significant. We believe this is due to the small image resolution in this dataset, where the utilization of multi-scale features might not effectively extract and fuse large-scale and small-scale features, leading to the incomplete exploitation of the advantages of multi-scale features. Additionally, E-FPN has a higher number of parameters than other algorithms, which demands significant storage resources when deployed, and this limitation requires optimization in future work.

Moreover, the current usage of E-FPN involves the classification of single, complete images, which poses significant challenges when encountering scenarios with multiple objects or complex background environments in the image.

Future work should focus on applying this method to different feature extraction networks and exploring its generalizability. Additionally, further research should explore detail-oriented feature extraction and fusion methods to replace the fusion of entire feature maps, aiming to reduce the complexity and number of parameters of the method. Simultaneously, it is important to explore methods that prioritize the object's location in the image to mitigate interference caused by background objects in the classification process.

## REFERENCES

- [1] B. Mahaur, K. Mishra, N. Singh, *Improved Residual Network based on norm-preservation for visual recognition*, Neural Networks, Vol. 157, pp. 305–322, 2023.
- [2] M.H. Hesamian, W. Jia, X. He, P. Kennedy, *Deep learning techniques for medical image segmentation: achievements and challenges*, Journal of digital imaging, Vol. 32, pp. 582–596, 2019.
- [3] S. Sundaram, A. Zeid, *Artificial intelligence-based smart quality inspection for manufacturing*, Micromachines, Vol. 14, pp. 570–589, 2023.
- [4] L.M. Azizah, S.F. Umayah, S. Riyadi, C. Damarjati, N.A. Utama, *Deep learning implementation using convolutional neural network in mangosteen surface defect detection*, in Proc. of the 7th IEEE Int. Conf. on Control System, Computing and Engineering (ICCSCE 2017), pp. 242–246, 2017.
- [5] J. Kurihara, T. Nagata, H. Tomiyama, *Rice Yield Prediction in Different Growth Environments Using Unmanned Aerial Vehicle-Based Hyperspectral Imaging*, Remote Sensing, Vol. 15, pp. 2004–2022, 2023.
- [6] M. Liu, W.H. Su, X.Q. Wang, *Quantitative Evaluation of Maize Emergence Using UAV Imagery and Deep Learning*, Remote Sensing, Vol. 15, pp. 1979–1994, 2023.
- [7] S. Akcay, T. Breckon, *Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging*, Pattern Recognition, Vol. 122, pp. 108245–108266, 2022.
- [8] Y. Xu, H. Wang, X. Liu, H.R. He, Q. Gu, W. Sun, *Learning to see the hidden part of the vehicle in the autopilot scene*, Electronics, Vol. 8, pp.331–347, 2019.
- [9] G. Wang, H. Chen, L. Chen, Y. Zhuang, S. Zhang, T. Zhang, H. Dong, P. Gao, *P<sup>2</sup>FEViT: Plug-and-Play CNN Feature Embedded Hybrid Vision Transformer for Remote Sensing Image Classification*, Remote Sensing, Vol. 15, pp. 1773–1799, 2023.
- [10] W. Li, H. Chen, Q. Liu, H. Liu, Y. Wang, G. Gui, *Attention mechanism and depthwise separable convolution aided 3DCNN for hyperspectral remote sensing image classification*, Remote Sensing, Vol. 14, pp. 2215–2241, 2022.
- [11] L. Liang, S. Zhang, J. Li, A. Plaza, Z. Cui, *Multi-Scale Spectral-Spatial Attention Network for Hyperspectral Image Classification Combining 2D Octave and 3D Convolutional Neural Networks*, Remote Sensing, Vol. 15, pp. 1758–1782, 2023.
- [12] C. Shi, X. Zhang, J. Sun, L. Wang, *Remote sensing scene image classification based on self-compensating convolution neural network*, Remote Sensing, Vol. 14, pp. 545–573, 2022.
- [13] X. Ke, Y. Cai, B. Chen, H. Liu, W. Guo, *Granularity-Aware Distillation and Structure Modeling Region Proposal Network for Fine-Grained Image Classification*, Pattern Recognition, pp. 109305–109319, 2023.
- [14] P. Zhao, Y. Li, B. Tang, H. Liu, S. Yao, *Feature relocation network for fine-grained image classification*, Neural Networks, Vol. 161, pp. 306–317, 2023.
- [15] L. Chen, W. Shi, D. Deng, *Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images*, Remote Sensing, Vol. 13, p. 660, 2021.
- [16] L. Bo, X. Xiaoyang, W. Xingxing, T. Wenting, *Ship detection and classification from optical remote sensing images: A survey*, Chinese Journal of Aeronautics, Vol. 34, pp. 145–163, 2021.
- [17] Y. Dong, F. Chen, S. Han, H. Liu, *Ship object detection of remote sensing image based on visual attention*, Remote Sensing, Vol. 13, p. 3192, 2021.
- [18] T. Zhang, X. Zhang, X. Ke, C. Liu, X. Xu, X. Zhan, C. Wang, I. Ahmad, Y. Zhou, D. Pan, et al., *HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 60, pp. 1–22, 2021.
- [19] W. Xiong, Z. Xiong, Y. Cui, *An explainable attention network for fine-grained ship classification using remote-sensing images*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 60, pp. 1–14, 2022.
- [20] L. Ouyang, L. Fang, X. Ji, *Multigranularity Self-Attention Network for Fine-Grained Ship Detection in Remote Sensing Images*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 25, pp. 9722–9732, 2022.
- [21] C.S. Jahan, A. Savakis, E. Blasch, *Cross-modal knowledge distillation in deep networks for SAR image classification*, in Proc. of the Geospatial Informatics XII, SPIE, Vol. 12099, pp. 20–27, 2022.
- [22] A. Krizhevsky, I. Sutskever, G.E. Hinton, *Imagenet classification with deep convolutional neural networks*, Advances in Neural Information Processing Systems, 2012.
- [23] K. Simonyan, A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, arXiv preprint arXiv:1409.1556, 2014.
- [24] K. He, X. Zhang, S. Ren, J. Sun, *Deep residual learning for image recognition*, in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

- [25] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, *Aggregated residual transformations for deep neural networks*, in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1492–1500, 2017.
- [26] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, *Densely connected convolutional networks*, in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 4700–4708, 2017.
- [27] T.Y. Lin, A. RoyChowdhury, S. Maji, *Bilinear CNN models for fine-grained visual recognition*, in Proc. of the IEEE Int. Conf. on Computer Vision, pp. 1449–1457, 2015.
- [28] J. Fu, H. Zheng, T. Mei, *Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition*, in Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVFPR 2017), pp. 4438–4446, 2017.
- [29] L. Huang, W. Zhao, A.W.C. Liew, Y. You, *An evidential combination method with multi-color spaces for remote sensing image scene classification*, Information Fusion, 2023.
- [30] Y. Chen, Y. Bai, W. Zhang, T. Mei, *Destruction and construction learning for fine-grained image recognition*, in Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVFPR 2019), pp. 5157–5166, 2019.
- [31] H. Zheng, J. Fu, Z.J. Zha, J. Luo, *Looking for the devil in the details: Learning trilinear attention sampling network for fine-grained image recognition*, in Proc. of the IEEE Int. Conf. on Computer Vision (CVF 2019), pp. 5012–5021, 2019.
- [32] C.F.R. Chen, Q. Fan, R. Panda, *Crossvit: Cross-attention multi-scale vision transformer for image classification*, in Proc. of the IEEE Int. Conf. on Computer Vision (CVF 2021), pp. 357–366, 2021.
- [33] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. *An image is worth  $16 \times 16$  words: Transformers for image recognition at scale*, arXiv preprint arXiv:2010.11929, 2020.
- [34] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, *Swin transformer: Hierarchical vision transformer using shifted windows*, in Proc. of the IEEE Int. Conf. on Computer Vision (CVF 2021), pp. 10012–10022, 2021.
- [35] J. Chen, K. Chen, H. Chen, W. Li, Z. Zou, Z. Shi, *Contrastive learning for fine-grained ship classification in remote sensing images*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 60, pp. 1–16, 2022.
- [36] B. Guo, R. Zhang, H. Guo, W. Yang, H. Yu, P. Zhang, T. Zou, *Fine-Grained Ship Detection in High-Resolution Satellite Images With Shape-Aware Feature Learning*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Vol. 16, pp. 1914–1926, 2023.
- [37] Z. Zhang, T. Zhang, Z. Liu, P. Zhang, S. Tu, Y. Li, M. Waqas, *Fine-grained ship image recognition based on BCNN with inception and AM-Softmax*, 2022.
- [38] C.S. Jahan, A. Savakis, E. Blasch, *Sar image classification with knowledge distillation and class balancing for long-tailed distributions*, in Proc. of the 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), pp. 1–5, 2022.
- [39] J. Redmon, A. Farhadi, *Yolov3: An incremental improvement*, arXiv preprint arXiv:1804.02767, 2018.
- [40] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, *Feature pyramid networks for object detection*, in Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2117–2125, 2017.
- [41] G. Shafer, *Dempster-Shafer theory*, Encyclopedia of artificial intelligence, Vol. 1, pp. 330–331, 1992.
- [42] Y. Lin, Y. Li, X. Yin, Z. Dou, *Multisensor fault diagnosis modeling based on the evidence theory*, IEEE Trans. on Reliability, Vol. 67, pp. 513–521, 2018.
- [43] T. Dezert, J. Dezert, F. Smarandache, *Improvement of proportional conflict redistribution rules of combination of basic belief assignments*, Journal of Advances in Information Fusion (JAIF), Vol. 16, pp. 48–74, 2021.
- [44] Z. Liu, Q. Pan, J. Dezert, J.W. Han, Y. He, *Classifier fusion with contextual reliability evaluation*, IEEE Trans. on Cybernetics, Vol. 48, pp. 1605–1618, 2017.

Part 3:

**Contributions to Belief  
Functions**



# A Real Z-box Experiment for Testing Zadeh's Example

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, A. Tchamova, D. Han, *A Real Z-box Experiment for Testing Zadeh's Example*, in Proc. of Fusion 2015 Int. Conf. on Information Fusion, Washington D.C, USA, July 6–9, 2015, and reprinted with permission.

**Abstract**—In this paper, we propose a real experiment for building and realizing the physical combination of basic belief assignments associated with two independent, informative, and equi-reliable sources of information, according to the famous Zadeh's example. This experiment is based on a particular electronic circuit box, called Z-box, enabling to observe and to check the fusion result experimentally. Our experimental results clearly invalidate the fusion result obtained by Dempster-Shafer's rule of combination and show that it is physically possible to consider in a natural fusion process two independent and equi-reliable sources of evidences at same time, even if they appear as highly conflicting in Shafer's sense.

**Keywords:** belief function, Zadeh's example, Z-box experiment, information fusion, Dempster's rule.

## I. INTRODUCTION

Dempster-Shafer Theory (DST), introduced by Shafer in 1976 [1] offers an elegant theoretical framework for modeling epistemic uncertainty and for combining distinct bodies of evidence collected from different sources. In DST, the combination (fusion) of several distinct sources of evidences is done with Dempster-Shafer (DS)<sup>1</sup> rule of combination, which corresponds to the normalized conjunctive consensus operator [1], assuming that the sources are not in total conflict<sup>2</sup>. Since 1976, DST has been used in many fields of applications, including information fusion, pattern recognition, decision making, etc, but it also has been seriously criticized by some authors [2]–[12].

In spite of it, starting from Zadeh's criticism [2]–[4], many questions have arisen about the validity and the consistency of this theory when combining uncertain and conflicting evidences expressed as basic belief assignments (BBAs). Zadeh's "paradox" [2] is the first example where DS rule gives an apparently counter-intuitive result in highly conflicting case. Another very interesting example showing the counter-intuitive behavior of DS rule in some very low conflicting cases has been discovered recently and discussed by the authors in [11].

<sup>1</sup>Although the rule has been proposed originally by Dempster, we call it Dempster-Shafer rule because it has been widely promoted by Shafer in DST.

<sup>2</sup>otherwise DS rule is mathematically not defined because of 0/0 indeterminacy.

Since the publication of Zadeh's example, many researchers and engineers [5]–[9], [14] working in applications with belief functions have observed and admitted that DS rule is problematic for evidence combination, especially when the sources of evidence are highly conflicting. A most recent detailed discussion on the validity of DS rule can be found in [10]–[12]. It is worth noting that the discussion of the choice of semantics for the justification of a rule of combination is not the purpose of this paper. We just want to revisit and discuss here the most well-known Zadeh's emblematic example only from a physical-based standpoint because we are very concerned with fusion in real applications, especially for defense and security.

This paper was inspired by our curiosity to revisit Zadeh's example on the base of a real experiment, in order to become aware of the authentic physical fusion process (validated by the Nature's physical laws) and to understand the way how this emblematic example is "resolved" in actual fact by the Nature. In this paper, we propose a real experiment for generating BBAs from physical quantities that are consistent with BBAs inputs given in Zadeh's example, and that can be fused automatically by a pure natural phenomenon. Our paper shows that in this Z-box experiment, Dempster's rule of combination is inconsistent with physical (fusion) law of Nature and thus it cannot be used to predict the experimental results. Our experiment can be reproduced and verified by any reader who wants to check by him/herself the validity of our results. In this experiment, we have considered and generated two independent Bayesian BBAs that are equi-reliable and fit with Zadeh's BBAs inputs and let the Nature combine them physically, and we just observe what happens. Even if the two Zadeh's Bayesian BBAs appear as highly conflicting (in Shafer's sense), we have shown that it is however possible to make a physical experiment in which each source provides a BBA as chosen by Zadeh. This is possible because each source has only a partial knowledge of the state of the world.

In this work, we have just designed a simple physical experiment in which the fusion procedure is just governed by the physical law of Nature. All the fusion rules aim to obtain good and reasonable fusion results. We do think that to use such a physical experiment for testing DS rule (a type of fusion rule) makes sense and is rational, and our results indicate that DS rule does not agree with the physical (natural)

fusion process. To certify that DS fusion rule is undoubtedly valid and really useful in practical applications, it should be proved valid through an undisputed experimental protocol and tested on real experiments, and not claimed valid from specific justifications conditioned by particular choices of semantics that have been disputed since more than four decades in the scientific community. The choice of a semantic interpretation of fusion, although interesting, is not our major concern here. So far, and to authors knowledge, there is no undisputed experimental protocol proving that Dempster's rule is valid, even if Shafer proposed an interpretation based on a *random-code* interpretation of belief functions (BF) in [13]. It is also worth recalling that DS rule is not a generalization of Bayesian inference because even when BBAs are Bayesian, DS and Bayes rules become incompatible as soon as the a priori is truly informative (i.e. it is not vacuous, nor uniform) – as it is in the vast majority of practical cases in fact, see [12] and references inside for justifications with examples. That is why, it is vain (in our opinion) to search for a real valid and general physical experiment validating DS rule in the general context of belief functions.

After a brief recall of the basics of DST and Zadeh's example, we will present in details our Z-box experiment and discuss its results in the next sections.

## II. BASICS OF DST

Let  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  be a frame of discernment of a problem under consideration containing  $n$  distinct exclusive and exhaustive elements  $\theta_i$ ,  $i = 1, \dots, n$ . A basic belief assignment<sup>3</sup> (BBA),  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  is a mapping from the power set of  $\Theta$  (i.e. the set of subsets of  $\Theta$ ), denoted  $2^\Theta$ , to  $[0, 1]$ , that must satisfy the following conditions: 1)  $m(\emptyset) = 0$ , i.e. the mass of empty set (impossible event) is zero; 2)  $\sum_{X \in 2^\Theta} m(X) = 1$ , i.e. the mass of belief is normalized to one. The quantity  $m(X)$  represents the mass of belief exactly committed to  $X$ . An element  $X \in 2^\Theta$  is called a focal element if and only if  $m(X) > 0$ . The set  $\mathcal{F}(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$  of all focal elements of a BBA  $m(\cdot)$  is called the core of the BBA. The vacuous BBA characterizing the full ignorance is defined by  $m_v(\cdot) : 2^\Theta \rightarrow [0, 1]$  such that  $m_v(X) = 0$  if  $X \neq \Theta$ , and  $m_v(\Theta) = 1$ .

From any BBA  $m(\cdot)$ , the belief function  $Bel(\cdot)$  and the plausibility function  $Pl(\cdot)$  are defined for  $\forall X \in 2^\Theta$  as:  $Bel(X) = \sum_{Y|Y \subseteq X} m(Y)$  and  $Pl(X) = \sum_{Y|X \cap Y \neq \emptyset} m(Y)$ .  $Bel(X)$  and  $Pl(\bar{X})$  are classically interpreted as lower and upper bounds of an unknown subjective probability  $P(\cdot)$  and one has the following inequality satisfied  $Bel(X) \leq P(X) \leq Pl(X)$ ,  $\forall X \in 2^\Theta$ . In DST, the combination (fusion) of several distinct sources of evidences is done with DS rule of combination, which corresponds to the normalized conjunctive consensus operator [1], assuming that the sources are not in total conflict<sup>4</sup>. DS combination of two independent BBAs

$m_\Theta^1(\cdot)$  and  $m_\Theta^2(\cdot)$  is defined by  $m_\Theta(\emptyset) = 0$ , and for all  $X \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_\Theta(X) = \frac{1}{1 - K_{12}} \times \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_\Theta^1(X_1) m_\Theta^2(X_2), \quad (1)$$

where

$$K_{12} \triangleq \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = \emptyset}} m_\Theta^1(X_1) m_\Theta^2(X_2), \quad (2)$$

defines the so-called *conflict* between the two sources of evidence characterized by the BBAs  $m_\Theta^1(\cdot)$  and  $m_\Theta^2(\cdot)$ .

## III. ZADEH'S EXAMPLE

The famous Zadeh's example considers two doctors examining a patient who suffers from either meningitis (A), concussion (B) or brain tumor (C). The frame of discernment is chosen as  $\Theta = \{A, B, C\}$  and it is assumed as exhaustive and exclusive. Both doctors agree in their low expectation of a tumor, but disagree in likely cause and provide the following diagnosis, described by the following BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  satisfying

$$m_1(A) = 0.90, \quad m_1(B) = 0.00, \quad m_1(C) = 0.10, \quad (3)$$

$$m_2(A) = 0.00, \quad m_2(B) = 0.90, \quad m_2(C) = 0.10. \quad (4)$$

If one combines the two BBAs using DS rule of combination, the following counter-intuitive final conclusion is obtained

$$m_{DS}(A) = 0.0, \quad m_{DS}(B) = 0.0, \quad m_{DS}(C) = 1.0. \quad (5)$$

The conclusion made on the base of DS rule is that the patient has for sure a brain tumor because it is the only diagnose that both doctors agree on even if the two experts (doctors) agree that tumor is unlikely but are in almost full contradiction for the other causes of the disease. What is even more questionable is that the same conclusion (the brain tumor is unlikely) would be obtained regardless of the probabilities associated with the other possible diagnoses. This very simple but interesting example shows the limitations of practical use of the DST for automated reasoning and has widely been discussed in the literature [2]–[12].

A more emblematic and interesting example, involving possibly low conflicting sources, has been discovered recently and discussed in [10]–[12]. It corresponds to the case where the two equi-reliable doctors' reports concern the following BBAs satisfying  $m_1(A) = a$ ,  $m_1(A \cup B) = 1 - a$  and  $m_2(A \cup B) = b_1$ ,  $m_2(C) = 1 - b_1 - b_2$ ,  $m_2(A \cup B \cup C) = b_2$ , with parameters  $0 < a, b_1, b_2, < 1$ . It is easy to verify that the conflict given by (2) is equal to  $K_{12} = m_1(A)m_2(C) + m_1(A \cup B)m_2(C) = 1 - b_1 - b_2$ . Surprisingly, this conflict does not impact (it can be very high, or very low) the DS fusion result because one always has in this new example  $m_{DS}(\cdot) = m_1(\cdot)$ . This result is also abnormal and counter-intuitive because the second source  $m_2(\cdot)$  (the 2nd doctor diagnosis) does not count at all in DS fusion process, even if  $m_2(\cdot)$  is not vacuous (it is informative) and truly conflicting with the first doctor's diagnosis  $m_1(\cdot)$ .

<sup>3</sup>also called a belief mass function (BMF) by some authors, or a basic probability assignment (BPA) by Shafer.

<sup>4</sup>otherwise DS rule is mathematically not defined because of 0/0 indeterminacy.

## IV. A REAL Z-BOX EXPERIMENT

In this section, we propose an electronic circuit (called Z-box scheme) as shown in Fig. 1 to generate BBAs according to Zadeh's example and to test experimentally the physical fusion of these BBAs.

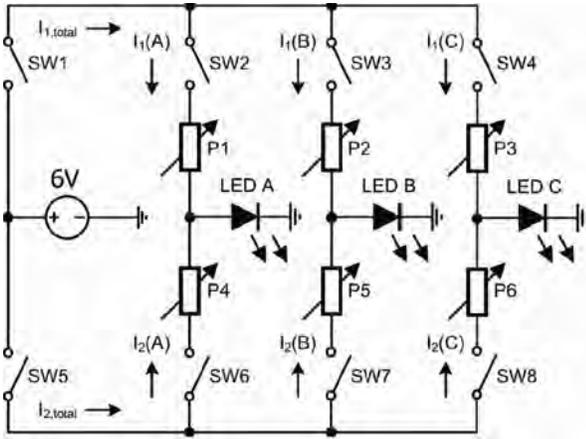


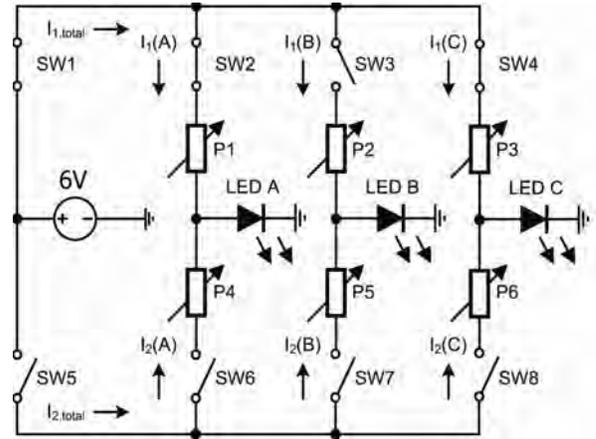
Figure 1. Z-box Scheme.

It is clear that this scheme can be easily extended to build and combine more than three Bayesian sources of evidence as well, which is out of the scope of this paper. This scheme utilizes a simple battery of 6 Volts as an only circuit's power supply. The switches SW1 and SW5 are used to obtain two independent sub-circuits, in order to realize two independent sources of information for the purpose of our task. Three simple linear potentiometers ( $P_1, P_2, P_3$ ) and three switches (SW2, SW3, and SW4) are used to establish the first source (sub-circuit 1), respectively three potentiometers ( $P_4, P_5, P_6$ ) and three switches (SW6, SW7, and SW8) for the second source (sub-circuit 2). Each of these two sources of information provides its relative truth, established on its own knowledge only, by setting the special tuning of corresponding sets of potentiometers. Three white Light Emitting Diodes (LED's -  $LED_A, LED_B,$  and  $LED_C$ ) are put to be utilized as light indicators. The light intensity is proportional to the current values through the LED's. We are concerned with the answer of the question: which LED emits the light with strongest intensity? Our frame is  $\Theta = \{A \triangleq LED_A, B \triangleq LED_B, C \triangleq LED_C\}$ . The Z-box experiment consists in three main steps: 1) tuning the source no. 1 (Sub-circuit 1) to generate BBA  $m_1(\cdot)$ ; 2) tuning the source no. 2 (Sub-circuit 2) to generate BBA  $m_2(\cdot)$ ; and 3) the physical fusion of the two BBAs. The descriptions of these steps are given in the sequel and are illustrated in the figures 2-4.

**Step 1:** Tuning the first source (Sub-circuit 1) according to Fig.2. Only the upper branch of the circuit is active with the following settings:

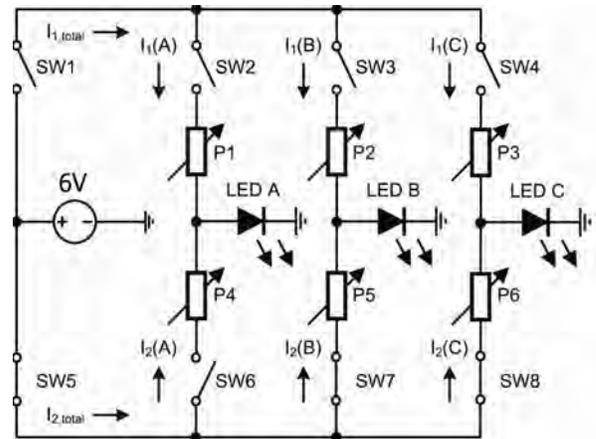
- Switch SW1 is closed and switch SW5 is open.

- Switches SW2 and SW4 are closed. Switch SW3 is left open, providing a zero-current through  $LED_B$ :  $I_1(LED_B) = 0.0$  mA.
- The potentiometers ( $P_1, P_3$ ) are tuned to provide the following current values through the LED's:  $I_1(LED_A) \approx 32.5$  mA,  $I_1(LED_B) = 0.0$  mA and  $I_1(LED_C) \approx 3.6$  mA, where the index  $\{1\}$  is used to denote the 1st source of information.


 Figure 2. Step 1 of the experiment : setting the BBA  $m_1(\cdot)$ .

**Step 2:** Tuning the second source (Sub-circuit 2) according to Fig. 3. Only the lower branch of the circuit is active with the following settings:

- Switch SW1 is open and switch SW5 is closed.
- Switches SW7 and SW8 are closed. Switch SW6 is left open, providing a zero-current through  $LED_A$  as  $I_2(LED_A) = 0.0$  mA.
- The potentiometers ( $P_5, P_6$ ) are tuned to provide the following current values through the LED's:  $I_2(\cdot) = \{I_2(LED_A) = 0.0$  mA,  $I_2(LED_B) \approx 32.5$  mA, and  $I_2(LED_C) \approx 3.6$  mA, where the index  $\{2\}$  is used to denote the 2nd source of information.


 Figure 3. Step 2 of the experiment : setting the BBA  $m_2(\cdot)$ .

**Step 3:** Both branches of the circuit are active at the same time for making the physical fusion. More precisely, we set the switches SW2, SW3 and SW4 and tune the potentiometers  $P_1$ ,  $P_2$  and  $P_3$  according to Step 1, and we set the switches SW6, SW7 and SW8 and tune the potentiometers  $P_4$ ,  $P_5$  and  $P_6$  according to Step 2. The switches SW1 and SW5 are closed to implement the fusion of the sources as shown in Fig. 4.

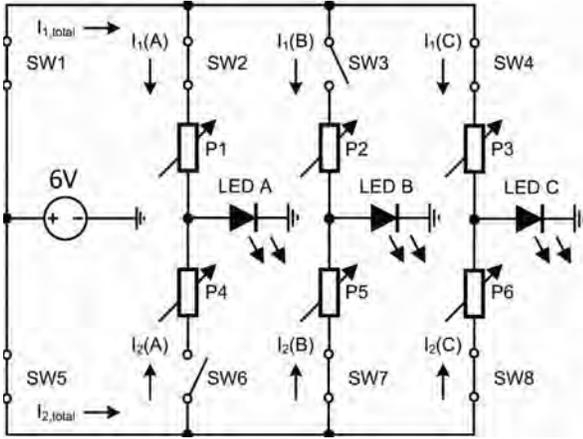


Figure 4. Step 3 of the experiment : the (physical) fusion of BBAs.

At this step, one gets:

$$\begin{cases} I_1(LED_A) \approx 32.5 \text{ mA}, \\ I_1(LED_B) = 0.00 \text{ mA}, \\ I_1(LED_C) \approx 3.6 \text{ mA}. \end{cases} \quad (6)$$

and

$$\begin{cases} I_2(LED_A) = 0.00 \text{ mA}, \\ I_2(LED_B) \approx 32.5 \text{ mA}, \\ I_2(LED_C) \approx 3.6 \text{ mA}. \end{cases} \quad (7)$$

The total current intensities are respectively equal to

$$\begin{cases} I_{1,total} = \sum_{i \in \{A,B,C\}} I_1(LED_i) \approx 36.1 \text{ mA}, \\ I_{2,total} = \sum_{i \in \{A,B,C\}} I_2(LED_i) \approx 36.1 \text{ mA}. \end{cases}$$

Fig. 5 shows the different LED's current values obtained in each step during the experiment's time duration of 5 sec. In the left subplots of Fig. 5 (result of step 1), one sees that the current through  $LED_A$  is 9 times higher than the current through  $LED_C$ , while the current through  $LED_B$  is almost zero, whereas in the middle subplots of Fig. 5 (result of step 2), one sees that the current through  $LED_B$  is 9 times higher than the current through  $LED_C$ , while the current through  $LED_A$  is almost zero. The observed results make perfect sense. Because the light intensity is proportional to current values through the LEDs, the same proportions are valid for the intensity of the light emitted from the LEDs. One sees that these settings fit with the input BBAs of Zadeh's example because after the normalization of current values one has the

following masses of belief in the origin of the strongest light emission:

$$\begin{cases} m_1(A) \triangleq \frac{I_1(LED_A)}{I_{1,total}} \approx 0.9, \\ m_1(B) \triangleq \frac{I_1(LED_B)}{I_{1,total}} = 0.0, \\ m_1(C) \triangleq \frac{I_1(LED_C)}{I_{1,total}} \approx 0.1, \end{cases} \quad (8)$$

and

$$\begin{cases} m_2(A) \triangleq \frac{I_2(LED_A)}{I_{2,total}} = 0, \\ m_2(B) \triangleq \frac{I_2(LED_B)}{I_{2,total}} \approx 0.9, \\ m_2(C) \triangleq \frac{I_2(LED_C)}{I_{2,total}} \approx 0.1. \end{cases} \quad (9)$$

The results of steps 1 and 2 show that both of the sources (corresponding to 1st and 2nd sub-circuits), taken independently, are able to make a correct physical assessment of the real physical situation. The right subplots of Fig. 5 (result of step 3) show the real physical fusion results simulated from MicroSim DesignLab 8 [18], as shown through the screen copy given in Fig. 6. Here we use the index  $\{12\}$  to denote that both sources (sub-circuits) are active. The observed current intensities are  $I_{12}(LED_A) \approx 32.5$  mA,  $I_{12}(LED_B) \approx 32.5$  mA, and  $I_{12}(LED_C) \approx 6.9$  mA. After the normalization of  $I_{12}(\cdot)$ , we get finally the combined BBA  $m_{12}(\cdot)$  over the frame of discernment  $\Theta \triangleq \{A, B, C\}$  that is given by  $m_{12}(A) \triangleq I_{12}(LED_A)/I_{12,total} \approx 0.45$ ,  $m_{12}(B) \triangleq I_{12}(LED_B)/I_{12,total} \approx 0.45$ , and  $m_{12}(C) \triangleq I_{12}(LED_C)/I_{12,total} \approx 0.10$ , where  $I_{12,total} = I_{12}(LED_A) + I_{12}(LED_B) + I_{12}(LED_C) \approx 71.9$  mA.

Clearly, the observed fact is that after the real physical fusion, the current through  $LED_A$  is just equal to the current through  $LED_B$ , and both are approximately 5 times higher than the current through  $LED_C$ . The experimental fusion result does not fit with the predicted result based on DS rule (5), nevertheless in this experiment both BBA inputs match the medical experts' opinions as in Zadeh's example, and they are considered to be in high "conflict" according to the classical interpretation in DST. This result brings to light the fact that DS rule result (5) is not consistent in this experiment with what the physical fusion system provides. This real Z-box experiment supports Zadeh's intuition about the non-adequate behavior of DS rule, and the counter-intuitive decisions that can be drawn from it. Stated otherwise, the natural physical fusion does not follow DS rule of combination. In fact, the notion of "conflict", which plays an important role when manipulating belief functions, is questionable, since it appears quite artificial in physics (in natural phenomenon). The conflict plays however a main role in decision-making in human reasoning. The way in which the total or partial conflicts are managed by Shafer's evidential reasoning is incompatible with this simple physical experiment.

It is worth noting that the physical fusion of sources of Zadeh's example is consistent with the simple averaging rule, and (relatively) consistent with PCR6 fusion rule [17] (Vol. 2) which will provide in this example  $m_{PCR6}(A) = 0.486$ ,  $m_{PCR6}(B) = 0.486$ , and  $m_{PCR6}(C) = 0.028$ . Contrarily to DS rule, PCR6 is fully consistent with the averaging

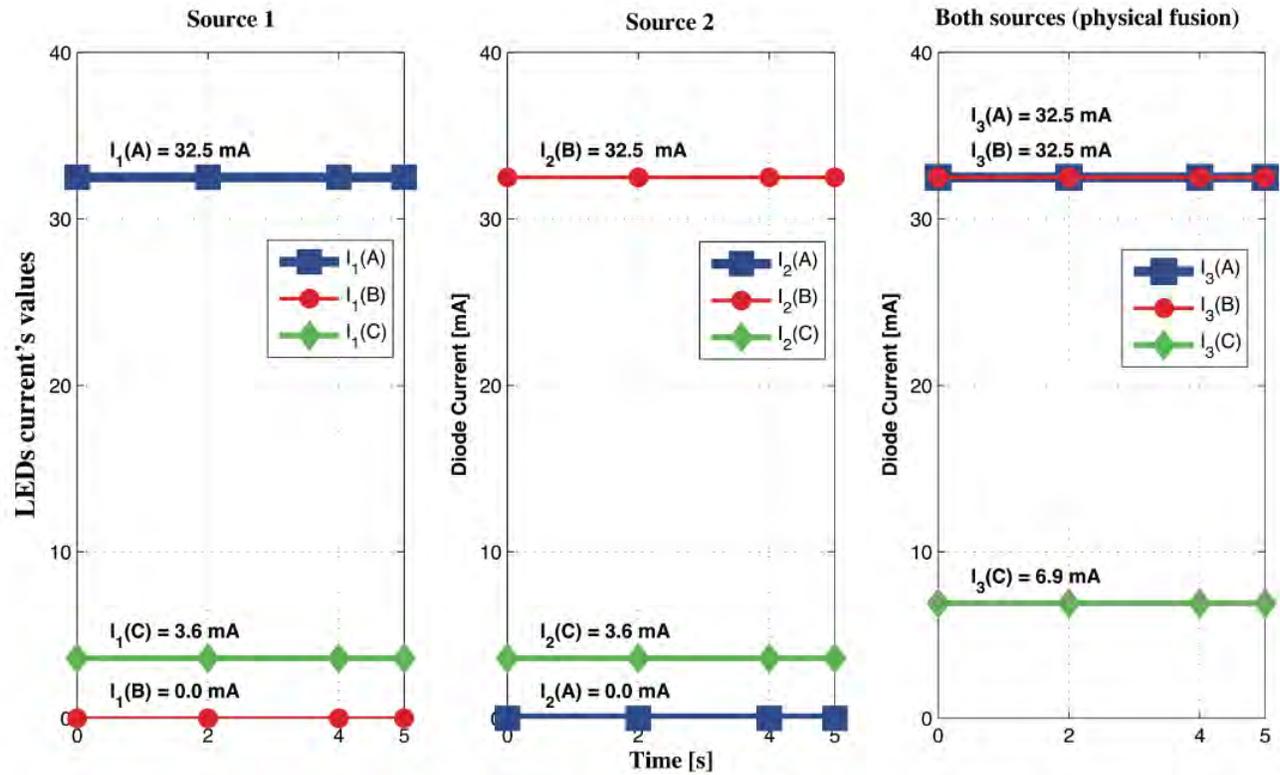


Figure 5. LEDs current values for source 1, source 2, both sources (by physical fusion).

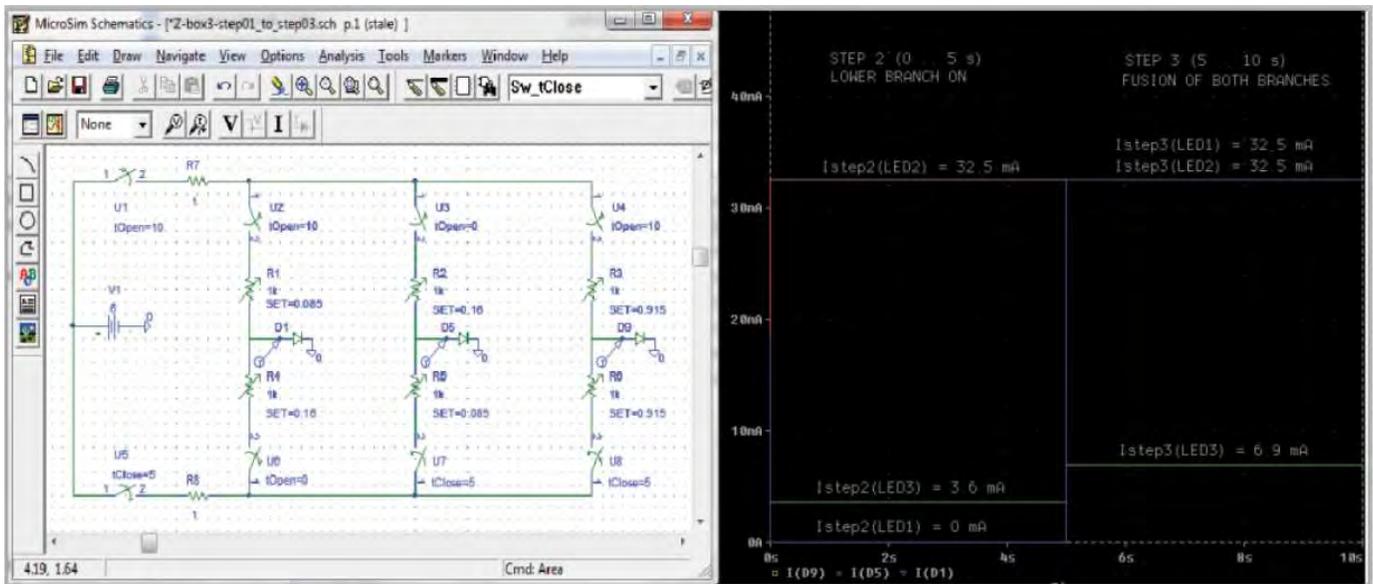


Figure 6. Screen copies of MicroSim schematics and its physical fusion result.

rule for estimating frequentist probabilities in binary random experiments, see [19] for details with examples.

## V. CONCLUSIONS

In this paper a real experimental method for building basic belief assignments associated with two independent, informative, and equireliable sources of information, following the emblematic Zadeh's example has been presented. It is based on a particular electronic circuit box (called Z-box), enabling to observe and to check the fusion result experimentally. Zadeh's intuition about the non-adequate behavior of DS rule and the counter-intuitive decisions obtained on its base is perfectly defended by Nature through this experiment. A similar experiment, called Z-aquarium experiment can also be done with fluids (with a container filled of water) instead of an electronic circuit, but it is more complex to set up and it has not been reported in this paper. Our conclusion is that Dempster-Shafer Theory does not agree with the physical fusion process at least for a situation that fits with Zadeh's example. The more general question on the validity of DST (especially, when subjective beliefs are considered) was not the purpose of this paper because this question has already been addressed in details in our previous research works put in references.

## ACKNOWLEDGEMENTS

The authors thank Dr. Nikolay N. Tchamov for his help to test the Z-Box with the MicroSim software. This work was supported by Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National Natural Science Foundation of China (No.61104214, No. 61203222), the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20120201120036) and by project AComIn, grant 316087, funded by the FP7 Capacity Programme.

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ, U.S.A.: Princeton University Press, 1976.
- [2] L.A. Zadeh, *On the validity of Dempster's Rule of Combination*, Memo M79/24, Berkeley, U.S.A., Univ. of California, 1979.
- [3] L.A. Zadeh, *Book review: A mathematical Theory of Evidence*, The AI Magazine, vol. 5 (3), pp. 81–83, 1984.
- [4] L.A. Zadeh, *A simple view of the Dempster-Shafer Theory of Evidence and its Implication for the Rule of Combination*, The AI Magazine, vol. 7 (2), pp. 85–90, 1986.
- [5] J. Lemmer, *Confidence Factors, Empiricism and the Dempster-Shafer Theory of Evidence*, Proc. of 1st Conference on Uncertainty in AI, pp. 160–176, 1985.
- [6] F. Voorbraak, *On the Justification of Dempster's Rule of Combination*, Dept. of Philosophy, Univ. of Utrecht, The Netherlands, Logic Group Preprint Series, no. 42, 1988.
- [7] P. Wang, *A Defect in Dempster-Shafer Theory*, Proc. of 10th Conference on Uncertainty in AI, pp. 560–566, 1994.
- [8] J. Pearl, *Reasoning with Belief Functions: An Analysis of Compatibility*, Int. Journal of Approximate Reasoning, vol. 4, pp. 363–389, 1990.
- [9] A. Gelman, *The boxer, the Wrestler, and the Coin Flip: a Paradox of Robust Bayesian Inference and Belief Functions*, American Statistician, vol. 60 (2), pp. 146–150, 2006.
- [10] J. Dezert, P. Wang and A. Tchamova, *On the Validity of Dempster-Shafer Theory*, Proc. of International Fusion Conference, Singapore, 2012.
- [11] A. Tchamova and J. Dezert, *On the Behavior of Dempster's rule of Combination and the Foundations of Dempster-Shafer Theory*, 6th IEEE Int. Conference Intelligent Systems, Sofia, Bulgaria pp. 108–113, 2012.
- [12] J. Dezert and A. Tchamova, *On the Validity of Dempster's Fusion Rule and its Interpretation as a Generalization of Bayesian Fusion Rule*, Int. Journal of Intelligent Systems, vol. 29, pp. 223–252, 2014 (with erratum in this DSMT Book - Vol. 5)
- [13] G. Shafer, *Constructive Probability*, Synthese, Vol. 48, pp. 1–60, 1981.
- [14] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, London, Chapman and Hall, pp. 278–281, 1991.
- [15] P. Smets and R. Kennes, *The Transferable Belief Model*, Artif. Int., vol. 66, pp. 191–234, 1994.
- [16] P. Smets, *Practical Uses of Belief Functions*, in K. B. Laskey and H. Prade, Editors, UAI, Stockholm, Sweden, pp. 612–621, 1999.
- [17] F. Smarandache and J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, Volumes 1, 2, 3 & 4, ARP, 2004–2015. <http://www.gallup.unm.edu/~smarandache/DSMT.htm>
- [18] [http://kahuna.sdsu.edu/engineering/electrical/faculty/szeto\\_software.html](http://kahuna.sdsu.edu/engineering/electrical/faculty/szeto_software.html)
- [19] F. Smarandache and J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, Proc. of Fusion 2013 Int. Conference on Information Fusion, Istanbul, Turkey, July 9-12, 2013.

# Two Novel Methods for BBA Approximation Based on Focal Element Redundancy

Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>

<sup>a</sup>Center for Information Engineering Science Research, Xi'an Jiaotong University, Xi'an, China.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>c</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an, China.

Emails: deqhan@xjtu.edu.cn, jean.dezert@onera.fr, jiafeiyy@mail.xjtu.edu.cn

Originally published as: D. Han, J. Dezert, Y. Yang, *Two Novel Methods for BBA Approximation Based on Focal Element Redundancy*, in Proc. of Fusion 2015 Int. Conf. on Information Fusion, Washington D.C, USA, July 6–9, 2015, and reprinted with permission.

**Abstract**—The theory of belief functions is a very appealing theory for uncertainty modeling and reasoning which has been widely used in information fusion. However, when the cardinality of the frame of discernment and the number of the focal elements are large the fusion of belief functions requires in general a high computational complexity. To circumvent this difficulty, many methods were proposed to implement more efficiently the combination rules and to approximate basic belief assignments (BBA's) into simplest ones to reduce the number of focal elements involved in the fusion process. In this paper, we present a novel principle for approximating a BBA by withdrawing more redundant focal elements of the original BBA. Two methods based on this principle are presented (using batch and recursive implementations). Numerical examples, simulations and related analyses are provided to illustrate and evaluate the performances of this new BBA approximation method.

**Index Terms**—Evidence theory; belief functions; basic belief assignment; approximation.

## I. INTRODUCTION

The original theory of belief functions, also known as Dempster-Shafer Theory (DST) [1] has been widely used in information fusion, pattern recognition and decision making due to its advantages in representing uncertain information and partial knowledge. However, the computational complexity is one of its drawbacks [2], specially for combining sources of evidences expressing their BBA's with respect to large frames of discernment (FoD). The computational complexity of the evidence combination is strongly affected by the cardinality of the FoD and the number of focal elements of the BBA of the sources to combine.

To reduce the computational complexity of evidence combination, various approaches have been proposed, which generally fit within the following two categories:

- Efficient implementation for performing exact computations of the chosen rule of combination. For example, an optimal algorithm for Dempster's rule of combination was proposed by Kennes [3]. Barnett [4], Shafer and Logan's [5] works are also representatives of this aspect.
- Approximation of simplification of BBA's. For example,  $k-l-x$  approach [6], summarization approach [7], the D1 approximation [8], inner and outer approximations [9], Monte-Carlo based approximation [10], etc., remove

focal elements and redistribute the corresponding mass assignments. In our previous works, we also had proposed hierarchical proportional redistribution approach [11], and the optimization-based BBA approximations [12].

The work presented in this paper focuses on the reduction of evidence combination's computational cost thanks to BBA approximations. In the aforementioned works of category b), the different methods propose to remove some focal elements according to some criteria, typically based either on their mass values or on their cardinalities. We think that only mass values or focal element cardinality are not enough for selecting the focal elements to remove for making good BBA approximation. We propose a novel approach using the notion of focal element redundancy. Those relatively redundant focal elements should be removed and those relatively non-redundant ones should be remained. To quantify this notion of redundancy, we use the average distance between a given focal element and all the other focal elements. Smaller average distance means that the given focal element carries similar information when compared with others, i.e., it is more redundant and should be removed at first. User can preset the desired number of remaining focal elements (also the number of removed focal elements). Two removing procedures (including a batch mode and a iterative mode) are proposed in the sequel, followed by the re-normalization or redistribution. Numerical examples, simulations and related analyses are provided to show the rationality and interest of these novel BBA approximation approaches.

## II. BASICS OF BELIEF FUNCTIONS

The theory of belief functions has been developed by Shafer [1] in 1976 from early works of Dempster. In DST, the elements in frame of discernment (FoD)  $\Theta$  are mutually exclusive and exhaustive. A basic belief assignment (BBA), also called a mass function, is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and

$$\sum_{A \in 2^\Theta} m(A) = 1 \quad (1)$$

If  $m(A) > 0$ ,  $A$  is called a focal element of the BBA  $m(\cdot)$ . In DST, the combination of two distinct bodies of evidence

(BOEs)  $m_1(\cdot)$  and  $m_2(\cdot)$  is done using Dempster's rule as follows.  $\forall A \in 2^\Theta$  :

$$m(A) = \begin{cases} 0, & \text{if } A = \emptyset \\ \frac{1}{1-K} \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j), & \text{if } A \neq \emptyset \end{cases} \quad (2)$$

where  $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)$  is the total conflicting mass assignments, which is discarded by normalization in Dempster's rule. It can be found from Eq. (2) that Dempster's rule is both commutative and associative. Dempster's rule has been seriously criticized for its counter-intuitive behaviors both in high conflicting and low conflicting situations [13], and other rules of combination have been developed in the literature – see [14] for details. These modified or refined combination rules focus on suppressing the counter-intuitive behaviors of Dempster's rule. However, like Dempster's rule, they all have to face the problem of high computational complexity with the increase of the FoD's cardinality and the quantity of the focal elements.

To reduce the computational cost of combination of BBA's and make the fusion process tractable, we can as a first strategy switch to more simple rules of combination or try to develop efficient implementations of sophisticate rules, or as a second strategy simplify (approximate) original BBA to combine by simplest BBA with less focal elements of smaller cardinalities, or we can mix both strategies as well. In this paper, we focus on the second strategy devoted to BBA approximation, which is more intuitive for human to catch the meaning [15].

### III. EXISTING BBA APPROXIMATION APPROACHES

Some existing BBA approximation approaches are briefly recalled in this section for the purpose of comparisons with the novel methods proposed in this paper.

1) *k-l-x method* [6]: This approach has been proposed by Tessem in 1993. The simplified BBA is obtained by

- keeping no less than  $k$  focal elements;
- keeping no more than  $l$  focal elements;
- by deleting the masses which are no greater than  $x$ .

In  $k-l-x$ , all original focal elements are sorted according to their mass values in a decreasing order. Then, the first  $p$  focal elements are chosen such that  $k \leq p \leq l$  and such that the sum of the mass assignments of these first  $p$  focal elements is no less than  $1-x$ . The removed mass values are redistributed to remaining focal elements by a classical normalization procedure.

2) *Summarization method* [7]: This method is similar to the  $k-l-x$  and it also keeps focal elements having highest mass values. The mass values of focal elements to remove are accumulated and assigned to their union set. Suppose  $k$  is the desired number of focal elements in the approximated BBA  $m_S(\cdot)$  of a given BBA  $m(\cdot)$ . Let  $M$  be the set of  $k-1$  focal elements with the highest mass values in  $m(\cdot)$ . Then  $m_S(\cdot)$  is obtained from  $m(\cdot)$  by

$$m_S(A) = \begin{cases} m(A), & \text{if } A \in M \\ \sum_{A' \subseteq A, A' \notin M} m(A'), & \text{if } A = A_0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where  $A_0$  is determined by

$$A_0 \triangleq \bigcup_{A' \notin M, m(A') > 0} A' \quad (4)$$

3) *D1 method* [8]: Let  $m(\cdot)$  be the original BBA to approximate.  $m_S(\cdot)$  denotes the approximated BBA and the desired number of focal elements is  $k$ . Let  $M$  be the set of  $k-1$  focal elements with the highest mass values in  $m(\cdot)$  and  $M^-$  be the set including all the other focal elements of  $m(\cdot)$ . The basic idea of the D1 method is to keep all the members of  $M$  as the focal elements of  $m_S(\cdot)$  and to assign the mass values of the focal elements in  $M^-$  among the focal elements in  $M$  according to the following procedure.

Given a focal element  $A \in M^-$ , in  $M$ , find all the supersets of  $A$  to form the collection  $M_A$ . If  $M_A$  is not empty, the mass value of  $A$  is uniformly assigned among the focal elements with smallest cardinality in  $M_A$ . When  $M_A$  is empty, then construct  $M'_A$  as

$$M'_A = \{B \in M \mid |B| \geq |A|, B \cap A \neq \emptyset\} \quad (5)$$

Then, if  $M'_A$  is not empty,  $m(A)$  is assigned among the focal elements with smallest cardinality in  $M'_A$ . The value assigned to a focal element  $B$  depends on the value of  $|B \cap A|$ . Such a procedure is executed iteratively until all  $m(A)$  have been assigned to the focal elements in  $M$ .

If  $M'_A$  is empty, there are two possible cases:

- 1) If the total set  $\Theta \in M$ , the sum of mass values of the focal elements in  $M^-$  will be added to  $\Theta$ ;
- 2) If  $\Theta \notin M$ , then set  $\Theta$  as a focal element of  $m_S(\cdot)$  and assign the sum of mass values of the focal elements in  $M^-$  to  $m_S(\Theta)$ .

More details on D1 method with examples can be found in [8].

The basic principle of these three previous approaches of BBA approximation is to remove the focal elements having smaller mass values because they are deemed as unimportant. Besides these methods, there exist other works on BBA approximations. For example, Denœux inner and outer approximations [9], Grabisch's  $k$ -additive BBA approximation [16], and our previous works based on hierarchical proportional distribution (HPR) [11] and optimization-based BBA approximations [12]. In these methods, the aim is to remove the focal elements with larger cardinalities because they bring more computational cost in the fusion process in general (see related references for details).

### IV. NEW BBA APPROXIMATIONS USING THE PRINCIPLE OF FOCAL ELEMENT REDUNDANCY

As briefly shown in the previous section, the existing BBA approximation approaches propose to remove some focal elements by eliminating those with smaller mass values, or with larger cardinalities. Although these methods have some rational justification, only mass values or cardinalities are not enough in our opinion for judging which focal elements should be removed for making BBA approximation. We consider that

it is quite hazardous (risky) to deem focal elements having small mass values as unimportant. It may also be dangerous to remove the focal elements with large cardinality justified only by the possible high computational cost they may cause in the fusion process. So, we should be cautious when adopting a BBA approximation technique. We agree with the fact that focal elements that are considered unimportant must be removed at first in an approximation method. However, focal elements' mass values are not enough for judging their importance. A more solid index (criterion) should be found to estimate the importance of a focal element to keep. Because the very redundant focal elements can reasonably be considered as unimportant and the relatively non-redundant focal elements can reasonably be considered as important, we define the degree of non-redundancy for a focal element at first. From this degree of non-redundancy, we can then develop new BBA approximation methods as it will be shown.

#### A. Degree of non-redundancy of focal elements

Suppose a BBA  $m(\cdot)$  has  $l$  focal elements. A distance between focal elements  $A_i$  and  $A_j$  proposed by Denœux [9] is defined as

$$\delta_{\cap}(A_i, A_j) = m(A_i) \cdot |A_i| + m(A_j) \cdot |A_j| - [m(A_i) + m(A_j)] \cdot |A_i \cap A_j| \quad (6)$$

If a focal element  $A_i$  has the smallest average distance with other focal elements  $A_j \subseteq \Theta, j \neq i$ , then  $A_i$  shares most common information with other focal elements, i.e.,  $A_i$  is the most redundant. Therefore, we can define the degree of non-redundancy based on the average distance between a focal element and others. First, we calculate the distance matrix for all the focal elements of  $m(\cdot)$  as

$$Mat_{FE} \triangleq \begin{bmatrix} \delta_{\cap}(A_1, A_1) & \delta_{\cap}(A_1, A_2) & \cdots & \delta_{\cap}(A_1, A_l) \\ \delta_{\cap}(A_2, A_1) & \delta_{\cap}(A_2, A_2) & \cdots & \delta_{\cap}(A_2, A_l) \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{\cap}(A_l, A_1) & \delta_{\cap}(A_l, A_2) & \cdots & \delta_{\cap}(A_l, A_l) \end{bmatrix}$$

It should be noted that  $\delta_{\cap}(A_i, A_i) = 0$  and  $\delta_{\cap}(A_i, A_j) = \delta_{\cap}(A_j, A_i)$  where  $i = 1, \dots, l$ . Hence, it is not necessary to calculate all the elements in  $Mat_{FE}$  because the matrix is symmetric.

We define the degree of non-redundancy of the focal element  $A_i$  by

$$\text{nRd}(A_i) \triangleq \frac{1}{l-1} \sum_{j=1}^{l-1} \delta_{\cap}(A_i, A_j) \quad (7)$$

The larger  $\text{nRd}(A_i)$  value, the larger non-redundancy (less redundancy) for  $A_i$ . The less  $\text{nRd}(A_i)$  value, the less non-redundancy (larger redundancy) for  $A_i$ .

Based on the focal element redundancy, i.e., to use the degree of non-redundancy in (7), we propose two new BBA approximation methods described in the next subsections, where the more non-redundant focal elements will be remained and the more redundant ones will be removed.

#### B. Batch approximation method

Let  $m(\cdot)$  denote the original BBA to approximate with  $l$  focal elements. In the approximation, we want to keep  $k < l$  focal elements. First, we propose a BBA approximation with a batch processing, which means that the number of focal elements is reduced from  $l$  to  $k$  in one processing cycle as follows.

- **Step 1:** Calculate  $Mat_{FE}$  at first, and for each  $A_i, i = 1, \dots, l$  compute its non-redundancy value  $\text{nRd}(A_i)$ ;
- **Step 2:** Sort all the elements in descending order according to the values of  $\text{nRd}(A_i)$ ;
- **Step 3:** Remove the  $l - k$  bottom focal elements;
- **Step 4:** Normalize the mass values of the remaining  $k$  focal elements and output the approximated BBA  $m_S^{BRd}(\cdot)$ .

#### C. Iterative approximation method

In this method, we remove iteratively one most redundant focal element (with the least nRd value) in each cycle until  $k$  focal elements are remained. This method consists of the following steps:

- **Step 1:** Calculate  $Mat_{FE}$  and nRd for each  $A_i, i = 1, \dots, l$ ;
- **Step 2:** Sort all the elements in descending order according to their values of  $\text{nRd}(A_i)$ ;
- **Step 3:** Remove the bottom focal element  $A_r$ ;
- **Step 4:** If the number of remaining focal element is larger than  $k$ , recalculate  $\text{nRd}(A_i)$  for  $i = 1, \dots, l, i \neq r$  and go to Step 3. Otherwise, go to Step 5 ;
- **Step 5:** Normalize the mass values of the remaining  $k$  focal elements and output the approximated BBA  $m_S^{IRd}(\cdot)$ .

For this iterative method, the degrees of non-redundancy are recalculated in each cycle after removing a focal element in the previous cycle. That is to say, in each cycle, only the non-redundancy of the current remaining focal elements are concerned.

#### D. Illustrative examples

Here we provide a simple numerical example to illustrate the implementation procedures of some available BBA approximation approaches with respect to our two new methods.

**Example 1:** Let consider the BBA  $m(\cdot)$  defined over the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$  listed in Table I.

TABLE I  
FOCAL ELEMENTS AND MASS VALUES OF  $m(\cdot)$

Focal Elements	Mass values
$A_1 = \{\theta_1, \theta_2\}$	0.50
$A_2 = \{\theta_1, \theta_3, \theta_4\}$	0.30
$A_3 = \{\theta_3\}$	0.10
$A_4 = \{\theta_3, \theta_4\}$	0.05
$A_5 = \{\theta_4, \theta_5\}$	0.05

1) *Using  $k-l-x$  method [6]:* Here  $k$  and  $l$  are set to 3.  $x$  is set to 0.1. The focal elements  $A_4 = \{\theta_3, \theta_4\}$  and  $A_5 = \{\theta_4, \theta_5\}$  are removed without violating the constraints in  $k-l-x$ . The remaining total mass value is  $1 - 0.05 - 0.05 = 0.9$ . Then, all the remaining focal elements' mass values are divided by 0.9 to accomplish the normalization. The approximated BBA  $m_S^{klx}(\cdot)$  obtained by  $k-l-x$  method is listed in Table II, where  $A'_i, i = 1, 2, 3$  are the focal elements of  $m_S^{klx}(\cdot)$ .

TABLE II  
 $m_S^{klx}(\cdot)$  OBTAINED USING  $k-l-x$

Focal Elements	Mass values
$A'_1 = \{\theta_1, \theta_2\}$	0.5556
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.3333
$A'_3 = \{\theta_3\}$	0.1111

2) *Using summarization method [7]:* Here  $k$  is set to 3. According to the summarization method, the focal elements  $A_3 = \{\theta_3\}$ ,  $A_4 = \{\theta_3, \theta_4\}$  and  $A_5 = \{\theta_4, \theta_5\}$  are removed, and their union  $\{\theta_3, \theta_4, \theta_5\}$  is generated as a new focal element with mass value  $m(\{\theta_3\}) + m(\{\theta_3, \theta_4\}) + m(\{\theta_4, \theta_5\}) = 0.2$ . The approximated BBA  $m_S^{Sum}$  is listed in Table III below.

TABLE III  
 $m_S^{Sum}(\cdot)$  OBTAINED USING SUMMARIZATION

Focal Elements	Mass values
$A'_1 = \{\theta_1, \theta_2\}$	0.50
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.30
$A'_3 = \{\theta_3, \theta_4, \theta_5\}$	0.20

3) *Using DI method [8]:* Here  $k$  is still 3. It can be obtained that  $A_1, A_2$  belong to  $M$ , and  $A_3, A_4, A_5$  belong to  $M^-$ . The focal element  $A_1 = \{\theta_1, \theta_2\}$  has empty intersection with the focal elements in  $M^-$ , therefore its value will be unchanged. In  $M$ ,  $A_2$  is the unique superset of  $A_3$  and  $A_4$ , therefore,  $m(A_3) + m(A_4) = 0.10 + 0.05$  is added to its original mass value.  $A_2$  also covers half of  $A_5$ , therefore,  $m(A_5)/2 = 0.025$  is further added to the mass of  $A_2$ . Finally, the rest mass value is assigned to the total set  $\Theta$ . The approximated BBA  $m_S^{D1}$  is listed in Table IV.

TABLE IV  
 $m_S^{D1}(\cdot)$  OBTAINED USING SUMMARIZATION

Focal Elements	Mass values
$A'_1 = \{\theta_1, \theta_2\}$	0.50
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.475
$A'_3 = \Theta$	0.025

4) *Using Denœux inner approximation [9]:* Because this method uses the focal element distance in Eq. (6), we also apply it in this example for comparison. With the inner approximation method, the focal elements pair with smallest distance are removed, and then their intersection is set as

the supplemented focal element whose mass value is the sum of the removed two focal elements' mass values. Such a procedure is repeated until the desired number of focal elements is reached. The results at each step are listed in Table V.

TABLE V  
BBA'S OBTAINED USING INNER APPROXIMATION

Step 1		Step 2	
Focal elements	Mass values	Focal elements	Mass values
$A'_1 = \{\theta_1, \theta_2\}$	0.5	$A'_1 = \{\theta_1, \theta_2\}$	0.5
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.3	$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.3
$A'_3 = \{\theta_3\}$	0.15	$A'_3 = \emptyset$	0.2
$A'_4 = \{\theta_4, \theta_5\}$	0.05		

As we can see in Table V, it generates the empty set as a focal element, which is not allowed in the classical Dempster-Shafer evidence theory under close-world assumption.

5) *Using the redundancy-based batch approximation method:* The desired remaining focal element is set to  $k = 3$ . We first calculate the distance matrix  $Mat_{FE}$  and we get

$$Mat_{FE} = \begin{matrix} & A_1 & A_2 & A_3 & A_4 & A_5 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{matrix} & \begin{bmatrix} 0 & 1.10 & 1.10 & 1.10 & 1.10 \\ 1.10 & 0 & 0.60 & 0.30 & 0.65 \\ 1.10 & 0.60 & 0 & 0.05 & 0.20 \\ 1.10 & 0.30 & 0.05 & 0 & 0.10 \\ 1.10 & 0.65 & 0.20 & 0.10 & 0 \end{bmatrix} \end{matrix}$$

Based on this matrix, the degree of non-redundancy for each focal elements of  $m(\cdot)$  can be obtained. It is listed in Table VI.

TABLE VI  
NON-REDUNDANCY FOR DIFFERENT FOCAL ELEMENTS

Focal Elements	Mass values	nRd( $A_i$ )
$A_1 = \{\theta_1, \theta_2\}$	0.50	1.10
$A_2 = \{\theta_1, \theta_3, \theta_4\}$	0.30	0.6625
$A_3 = \{\theta_3\}$	0.10	0.4875
$A_4 = \{\theta_3, \theta_4\}$	0.05	0.3875
$A_5 = \{\theta_4, \theta_5\}$	0.05	0.5125

Since  $A_3$  and  $A_4$  at the bottom have the two least nRd values, they correspond the two focal elements with the lowest non-redundancy, i.e., the highest redundancy. Therefore, they are removed and their mass values are redistributed thanks to the classical normalization step. The approximated BBA  $m_S^{BRd}$  is listed in Table VII.

TABLE VII  
 $m_S^{BRd}(\cdot)$  OBTAINED USING THE BATCH APPROXIMATION BASED ON REDUNDANCY

Focal Elements	Mass values
$A'_1 = \{\theta_1, \theta_2\}$	0.5882
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.3530
$A'_3 = \{\theta_4, \theta_5\}$	0.0588

6) *Using the redundancy-based iterative approximation method:* The number of remaining focal elements is still set to  $k = 3$ , so that two focal elements have to be removed. In the iterative mode, only one focal element is removed in each cycle, thus two cycles are needed.

In cycle I, the degree of non-redundancy is the same as listed in Table V. Then, the focal element  $A_4$  is removed in first cycle.

In cycle II, recalculate nRd for  $A_i, i = 1, \dots, 5, i \neq 4$  according to  $nRd(A_i) = \sum_{j=1, j \neq 4, j \neq i}^5 \delta(A_i, A_j)$ . The results are

$$\begin{aligned} nRd(A_1) &= 1.1000, nRd(A_2) = 0.7833, \\ nRd(A_3) &= 0.6333, nRd(A_5) = 0.6500 \end{aligned}$$

Then,  $A_3$  is removed in this cycle due to its the lowest nRd value (the highest redundancy among the remaining focal elements). The approximated BBA obtained using iterative way is the same as the one listed in Table VII. It should be noted that the batch approximation and the iterative approximation will not always output the same results as shown in the next example.

**Example 2:** Suppose that FoD is  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The BBA to approximate is listed in Table VIII, and the desired number of remaining focal elements is  $k = 3$ .

TABLE VIII  
FOCAL ELEMENTS AND MASS VALUES OF  $m(\cdot)$

Focal Elements	Mass values
$A_1 = \{\theta_1, \theta_2\}$	0.1780
$A_2 = \{\theta_2, \theta_3\}$	0.2477
$A_3 = \{\theta_2\}$	0.2322
$A_4 = \{\theta_3\}$	0.1758
$A_5 = \Theta$	0.1662

The distance matrix  $Mat_{FE}$  is

$$\begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{matrix} \begin{bmatrix} 0 & 0.4258 & 0.1780 & 0.5319 & 0.1662 \\ 0.4258 & 0 & 0.2477 & 0.2477 & 0.1662 \\ 0.1780 & 0.2477 & 0 & 0.4080 & 0.3325 \\ 0.5319 & 0.2477 & 0.4080 & 0 & 0.3325 \\ 0.1662 & 0.1662 & 0.3325 & 0.3325 & 0 \end{bmatrix} \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{matrix}$$

The degree of non-redundancy of focal elements are

$$\begin{aligned} nRd(A_1) &= 0.3255, nRd(A_2) = 0.2719, \\ nRd(A_3) &= 0.2916, nRd(A_4) = 0.3800, nRd(A_5) = 0.2494 \end{aligned}$$

With the batch approximation, the focal elements  $A_2$  and  $A_5$  are removed. After normalization, we get the approximated BBA listed in Table IX.

With the iterative approximation method, the degree of non-redundancy obtained at Cycle I are also

$$\begin{aligned} nRd^I(A_1) &= 0.3255, nRd^I(A_2) = 0.2719, \\ nRd^I(A_3) &= 0.2916, nRd^I(A_4) = 0.3800, \\ nRd^I(A_5) &= 0.2494 \end{aligned}$$

TABLE IX  
 $m_S^{BRd}(\cdot)$  OBTAINED USING THE BATCH APPROXIMATION BASED ON REDUNDANCY

Focal Elements	Mass values
$A'_1 = \{\theta_1, \theta_2\}$	0.3038
$A'_2 = \{\theta_2\}$	0.3962
$A'_3 = \{\theta_3\}$	0.3000

The iterative approximation first removes the focal element  $A_5$  because it has the least nRd value. Then we recalculate the nRd values for  $A_1, A_2, A_3$ , and  $A_4$  which gives us

$$\begin{aligned} nRd^{II}(A_1) &= 0.3786, nRd^{II}(A_2) = 0.3071, \\ nRd^{II}(A_3) &= 0.2779, nRd^{II}(A_4) = 0.3959 \end{aligned}$$

At Cycle II, the focal element  $A_3$  having the least nRd value is removed. After normalization, we get the approximated BBA  $m_S^{SRd}(\cdot)$  using iterative approximation as listed in Table X.

TABLE X  
 $m_S^{SRd}(\cdot)$  OBTAINED USING THE BATCH APPROXIMATION BASED ON REDUNDANCY

Focal Elements	Mass values
$A'_1 = \{\theta_1, \theta_2\}$	0.2960
$A'_2 = \{\theta_2, \theta_3\}$	0.4118
$A'_3 = \{\theta_3\}$	0.2922

which is different of the result of Table IX using the batch approximation.

## V. COMPARATIVE ANALYSIS

In this section, we present simulation results to compare the different BBA approximation approaches in terms of the computational cost and the closeness to the original one in average meaning. A BBA transformation with less computational cost and more closeness is preferred. To measure the closeness or the dissimilarity between different BBAs, a distance measure between BBA is used. In this work, we use Jousselme's distance [17] because it remains one of the most widely used distance of evidence. This distance is defined as

$$d_J(m_1, m_2) \triangleq \sqrt{\frac{1}{2} \cdot (m_1 - m_2)^T \mathbf{Jac} (m_1 - m_2)} \quad (8)$$

where  $\mathbf{Jac}$  is the so-called Jaccard's weighting matrix whose elements  $J_{ij} = \mathbf{Jac}(A_i, B_j)$  are defined by

$$\mathbf{Jac}(A_i, B_j) = \frac{|A_i \cap B_j|}{|A_i \cup B_j|} \quad (9)$$

A BBA  $m(\cdot)$  here can be considered as a column vector according to the geometric interpretation of the theory of belief functions [18]. There are also other types of distance of evidence [18]. We choose to use Jousselme's distance of evidence in this paper, because it has been proved to be a strict distance metric [19].

Our comparative analysis is based on a Monte Carlo simulation using  $M = 200$  random runs. In  $j$ -th simulation

run, the BBA to approximate  $m^j(\cdot)$  is randomly generated and the different approximation results  $\{m_{S_i}^j(\cdot)\}$  are obtained using the different approximation approaches, where  $i$  denotes the  $i$ -th BBA approximation approach. We calculate the computational time of the original evidence combination of  $m^j(\cdot) \oplus m^j(\cdot)$  with Dempster’s rule, and the computation time of Dempster’s combination of each approximated BBA  $m_{S_i}^j(\cdot) \oplus m_{S_i}^j(\cdot)$ . As stated before, there are many available BBA approximation approaches. Here we only compare our proposed approaches with  $k-l-x$  method, D1 method, Summarization method because with these methods the number of the remaining focal elements and the empty set is never considered as a valid focal element (contrarily to inner approximation method which will bring troubles for making the comparisons because Jousselme distance cannot be computed if one allows to put mass on empty set because  $|\emptyset| = 0$ ).

In our simulations, the cardinality of the FoD $\Theta$  is chosen to 3. In each random generation, there are 7 focal elements in the original BBA to approximate. The remaining number of focal elements for all the approaches used here are set to 6, 5, 4, 3, and 2. Random generation of BBA is based on Algorithm 1 [18] below.

TABLE XI  
ALGORITHM 1: RANDOM GENERATION OF BBA.

---

**Input:**  $\Theta$ : Frame of discernment;  
 $N_{max}$ : Maximum number of focal elements  
**Output:**  $m$ : BBA  
 Generate  $\mathcal{P}(\Theta)$ , which is the power set of  $\Theta$ ;  
 Generate a random permutation of  $\mathcal{P}(\Theta) \rightarrow \mathcal{R}(\Theta)$ ;  
 Generate an integer between 1 and  $N_{max} \rightarrow l$ ;  
**FOReach** First  $k$  elements of  $\mathcal{R}(\Theta)$  do  
 Generate a value within  $[0, 1] \rightarrow m_i, i = 1, \dots, l$ ;  
**END**  
 Normalize the vector  $m = [m_1, \dots, m_l] \rightarrow m'$ ;  
 $m(A_i) = m'_i$ ;

---

The average distance values over 200 runs between the original BBA and the approximated BBA’s obtained using different approaches given different remaining focal elements’ numbers are shown in Fig. 1. The average (over all runs and all numbers of remaining focal elements) computation time and distance are shown in Table XII.

TABLE XII  
COMPARISONS BETWEEN DIFFERENT BBA APPROXIMATIONS IN TERMS OF TIME AND CLOSENESS

Approaches	Distance	Time (ms)
Batch-redundancy	0.1162	0.1026
Iterative-redundancy	0.1147	0.1059
$k-l-x$	0.1181	0.1073
D1	0.1718	0.1039
Summarization	0.1624	0.1034

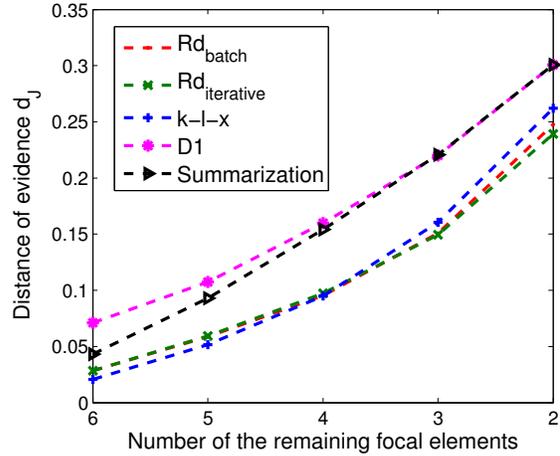


Fig. 1. Comparisons between different approximations in terms of the distance of evidence.

As we can see in Fig. 1 and in Table XII, all the method have the average computation time around 0.1 ms, which is reduced when compared with the original average computation time which is 0.2011 ms. It means that all the methods can well reduce the computational cost. Our new BBA approximation approaches based on focal element redundancy outputs BBA’s which are closer to the original one when compared with other approaches. This means that our proposed approximation approaches output BBA’s which are most faithful and with the least loss of information when compared with other approaches. So based on this comprehensive evaluation using two criteria including computation time and the closeness to the original BBA, our comparative analysis shows that our new methods perform better. The iterative version (having the smallest average distance) performs better than the batch version.

## VI. CONCLUSION

The degree of non-redundancy of focal elements is defined, based on which, two novel BBA approximation methods have been proposed in this paper including a batch version and an iterative version. Our Monte Carlo simulation results show that these new methods can well reduce the computational cost when compared with other available approaches; at the same time, the approximated BBA’s obtained using our new approaches are closer the original BBA in average, which represents the less loss of information in the approximation procedure.

In our future work, further theoretical analyses on the definition of the focal element non-redundancy or redundancy are needed, based on which, we will also attempt to design some new types of the focal element redundancy and to make additional comparison with the one used in this paper. Besides the computation time and the distance of evidence used in this paper, we will explore more comprehensive evaluation criteria of the BBA approximation approaches, and test other distance

measures of evidence [20] in our proposed approaches. This is crucial for the design of more effective approximations.

#### ACKNOWLEDGMENT

This work was supported by the Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National Natural Science Foundation (No. 61203222, No. 61104214), Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 61221063), Science and technology project of Shaanxi Province (No. 2013KJXX-46), Specialized Research Fund for the Doctoral Program of Higher Education (20120201120036), and Fundamental Research Funds for the Central Universities (No. xjj2012104, No. xjj2014122).

#### REFERENCES

- [1] G. Shafer, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.
- [2] P. Smets, "Practical uses of belief functions," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, vol. 15. Stockholm, Sweden: Morgan Kaufmann Publishers Inc., 1999, pp. 612–621.
- [3] R. Kennes, "Computational aspects of the mobius transformation of graphs," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 22, no. 2, pp. 201–223, 1992.
- [4] J. A. Barnett, "Computational methods for a mathematical theory of evidence," in *Proceedings of IJCAI-81*. Vancouver, Canada: Springer, 1981, pp. 868–875.
- [5] G. Shafer and R. Logan, "Implementing dempster's rule for hierarchical evidence," *Artificial Intelligence*, vol. 33, no. 3, pp. 271–298, 1987.
- [6] B. Tessem, "Approximations for efficient computation in the theory of evidence," *Artificial Intelligence*, vol. 61, no. 2, pp. 315–329, 1993.
- [7] J. D. Lowrance, T. D. Garvey, and T. M. Strat, "A framework for evidential-reasoning systems," in *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Philadelphia, Pennsylvania, USA: Springer, August 11-15 1986, pp. 419–434.
- [8] M. Bauer, "Approximations for decision making in the dempster-shafer theory of evidence," in *Proceedings of the 12th international conference on Uncertainty in artificial intelligence*. Portland, Oregon, USA: Morgan Kaufmann Publishers Inc., August 1-3 1996, pp. 73–80.
- [9] T. Dencœur, "Inner and outer approximation of belief structures using a hierarchical clustering approach," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 04, pp. 437–460, 2001.
- [10] S. Moral and A. Salmerón, "A monte carlo algorithm for combining dempster-shafer belief based on approximate pre-computation," in *Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU)*, A. Hunter and S. Pearsons, Eds. Springer, 1999, pp. 305–315.
- [11] J. Dezert, D. Han, Z. Liu, and J.-M. Tacnet, "Hierarchical proportional redistribution for bba approximation," in *Belief functions: theory and applications*. Springer, 2012, pp. 275–283.
- [12] D. Han, J. Dezert, and C. Han, "New basic belief assignment approximations based on optimization," in *Proceedings of the 15th International Conference on Information Fusion, ISIF*. Singapore: IEEE, July 9-12 2012, pp. 286–293.
- [13] F. Smarandache and J. Dezert, Eds., *Applications and Advances of DSMT for Information Fusion (Vol IV)*. Rehoboth, DE, USA: Amer. Res. Press, 2015. [Online]. Available: <http://www.gallup.unm.edu/~smarandache/DSMT-book4.pdf>
- [14] —, *Applications and Advances of DSMT for Information Fusion (Vol III)*. Rehoboth, DE, USA: Amer. Res. Press, 2009. [Online]. Available: <http://www.gallup.unm.edu/~smarandache/DSMT-book3.pdf>
- [15] T. Burger, "Defining new approximations of belief functions by means of dempsters combination," in *Proc. of the 1st International Workshop on the Theories of Belief Functions (BELIEF 2010), Brest, France, April 1st-April 2nd, 2010*.
- [16] M. Grabisch, "Upper approximation of non-additive measures by k-additive measures—the case of belief functions," in *Proceedings of the 1st International Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Gent, Belgium, June 1999, pp. 158–164.
- [17] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Information fusion*, vol. 2, no. 2, pp. 91–101, 2001.
- [18] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, no. 2, pp. 118–145, 2012.
- [19] M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the jaccard index matrix," *International Journal of Approximate Reasoning*, vol. 54, no. 5, pp. 615–626, 2013.
- [20] D. Han, J. Dezert, and Y. Yang, "New distance measures of evidence based on belief intervals," in *Belief Functions: Theory and Applications*. Springer, 2014, pp. 432–441.



# Evaluation of Probability Transformations of Belief Functions for Decision-Making

Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Zhansheng Duan<sup>c</sup>

<sup>a</sup>School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049 China.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>c</sup>CIESR, Xi'an Jiaotong University, Xi'an, Shaanxi 710049 China.

Emails: deqhan@gmail.com, jean.dezert@onera.fr, zsduan@mail.xjtu.edu.cn

Originally published as: D. Han, J. Dezert, Z. Duan, *Evaluation of Probability Transformations of Belief Functions for Decision-Making*, IEEE Trans. on SMC, Vol. 46(1), pp. 93–108, January 2016, and reprinted with permission.

**Abstract**—The transformation of belief function into probability is one of the most important and common ways for decision-making under the framework of evidence theory. In this paper, we focus on the evaluation of such probability transformations, which are crucial for their proper applications and the design of new ones. Shannon entropy or probabilistic information content (PIC) measure is traditionally used in evaluating probability transformations. The transformation having the lowest entropy or highest PIC is considered as the best one. This standpoint is questioned in this paper by comparing a probability transformation based on uncertainty minimization with other available probability transformations. It shows experimentally that entropy or PIC is not comprehensive to evaluate a probability transformation. To make a comprehensive evaluation, some new approaches are proposed by the joint use of PIC and the distance of evidence according to the value based and the rank based fusion. A pattern classification application oriented evaluation approach for probability transformations is also proposed. Some desired properties for probability transformations are also discussed. Experimental results and related analysis are provided to show the rationality of the new evaluation approaches.

**Index Terms**—Evidence theory, Probability transformation, Probabilistic information content, Entropy, Decision-making.

## I. INTRODUCTION

Dempster-Shafer theory (DST), also known as the theory of belief functions [1], [2], provides a way to reason with imprecise, uncertain and incomplete information. DST can distinguish “unknown” and “imprecision” and provides a method to fuse different evidences by using the commutative and associative Dempster’s rule of combination. That is why the DST is widely used in information fusion. There are, however, some drawbacks [3], [4] of the DST, e.g., counter-intuitive combination results, high computational cost, and lack of evaluation criteria. So some modified models were proposed, e.g., the transferable belief model (TBM) [3] and Dezert-Smarandache theory (DSMT) [4].

The final goal of uncertainty reasoning is usually decision-making. To make decision easier, the mass assignment for a compound focal element is usually assigned to each singleton by a probability transformation. The probability transformation aims to approximate a basic belief assignment (BBA) by a probabilistic measure. The pursuit of efficient probability transformations has attracted great attention in recent years

and many probability transformations have been proposed [4]–[13].

The most well-known probability transformation is the pignistic probability transformation (PPT) [4] in TBM. PPT maps a belief defined on subsets to a probability measure defined on singletons, based on which a classical decision under probabilistic framework can be readily applied. PPT uses equal weights when splitting mass assignments of the compound focal elements and redistributing them to singletons included in them. Other modified probability transformations were also proposed [5]–[13], which assign the mass assignments of compound focal elements to singletons according to some ratio constructed from the available information (e.g. the belief and the plausibility). Typical examples include Sudano’s probabilities [8] and Cuzzolin’s intersection probability [12], etc. Under the DSMT framework, other probability transformations called DSMP [9] and HDSMP [13] were proposed. DSMP takes into account both the masses and the cardinality of focal elements in the proportional redistribution process and HDSMP is a hierarchical version of DSMP. They can also be used in the DST framework.

To compare all the available transformations for the purpose of appropriate application and design of new transformations, evaluation is required. In almost all the existing works on probability transformations, Shannon entropy or its dual, Probabilistic Information Content (PIC) criterion, is used to evaluate them. Definitely, less uncertainty should be preferred for decision-making. However, is the probability measure generated from a belief function with less uncertainty always rational or beneficial for decision-making? To answer this, i.e., to illustrate the irrationality of the over-emphasis of PIC or entropy, another probability transformation based on a constrained entropy minimization [14] is used and analyzed through examples. When using entropy or PIC for evaluation, the probability measure with the least uncertainty seems the best one. Unfortunately, some risky and unexpected results may be also obtained. [14] shows that either entropy or PIC is not a comprehensive measure.

Comprehensive evaluation of probability transformation is desired and motivates this paper. PIC only emphasizes the clarity of the transformed probability, which is only from the

aspect of the clarity of decision-making. On the other hand, the transformed probability should be consistent with the original belief function in some sense meaning that comprehensive evaluation should also consider the fidelity of the transformed probability to the original BBA. Higher degree of fidelity means the less loss of information caused by probability transformation. Our comprehensive evaluation aims to make a balance between clarity and fidelity, i.e., a probability with higher clarity and bigger fidelity should be preferred. Then, how to quantify the degree of fidelity? The distance of evidence [15] is used to measure the dissimilarity between two BBAs. Since the probability can be considered as a particular BBA, we can simply use the distance of evidence [15] between the original BBA and the transformed probability to quantify the degree of fidelity (smaller distance means higher degree of fidelity). So, in this paper we evaluate the probability transformations jointly by PIC and the distance of evidence. This joint use of the two criteria is implemented by using value based fusion (via the values of PIC and distance) and rank based fusion (via the ranks of the values of PIC and distance). We also propose an application-oriented evaluation approach for probability transformations, such as the application of pattern classification. Besides the evaluation criteria, some desired properties (qualitative evaluations) of probability transformations are also helpful. In [16], some desired properties of probability transformations were proposed and analyzed including upper and lower bound consistency and combination consistency. In this paper, some new desired properties of a probability transformation are also proposed. This paper extends our previous ideas briefly introduced in [14], where we preliminarily pointed out that entropy or PIC is not enough to evaluate a probability transformation. However, in [14], comprehensive evaluation was not proposed, which are the main contribution of this paper.

The rest of this paper is organized as follows. In Section II, evidence theory is briefly introduced. The decision-making methods in evidence theory including belief based approaches and probability transformations are briefly summarized in Section III. The definitions and pertinent analysis of the commonly used probability transformations are given in Section IV. In Section V, the evaluation of the probability transformation is discussed. The irrationality of using entropy alone as an evaluation criterion is clearly shown by simple examples. In Section VI, we propose to evaluate a probability transformation based on two criteria (PIC and distance). The joint use of them is implemented either directly at their values, or at their ranks. Some supporting examples are provided in Section VII. In Section VIII, an application-oriented evaluation approach is proposed. In Section IX, some desired properties of probability transformations are proposed and analyzed. Conclusions are drawn in Section X.

## II. BASICS OF EVIDENCE THEORY

In Dempster-Shafer theory [2], the elements in the frame of discernment (FOD)  $\Theta$ , which is a discrete finite set, are mutually exclusive and exhaustive. Let  $2^\Theta$  be the power set

of the FOD. The function  $m : 2^\Theta \rightarrow [0, 1]$  defines a basic belief assignment (BBA), also called a mass function, which satisfies:

$$\sum_{A \subseteq \Theta} m(A) = 1, \text{ and } m(\emptyset) = 0. \quad (1)$$

Then, the belief function and the plausibility function are defined as in (2) and (3), respectively,  $\forall A \in 2^\Theta$ :

$$Bel(A) = \sum_{B \subseteq A} m(B), \quad (2)$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B), \quad (3)$$

where  $Bel(A)$  and  $Pl(A)$  can be interpreted as the lower and the upper bounds of the probability  $P(A)$ .

Dempster's rule of combination, which is used to fuse  $n$  distinct<sup>1</sup> bodies of evidence (BOEs), is:

$$m(A) = \begin{cases} 0, & \forall A = \emptyset, \\ \frac{\sum_{\cap_{A_i=A} \prod_{1 \leq i \leq n} m_i(A_i)}}{\sum_{\cap_{A_i \neq \emptyset} \prod_{1 \leq i \leq n} m_i(A_i)}}, & \forall A \neq \emptyset, \end{cases} \quad (4)$$

where  $m_1, m_2, \dots, m_n$  are  $n$  BBAs.

Distances of evidence [15], [17] measures the dissimilarity between BOEs. One of the most commonly used distance of evidence is the Jousselme's distance  $d_J(\cdot, \cdot)$  [15]:

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2}(m_1 - m_2)^T \mathbf{Jac} (m_1 - m_2)}, \quad (5)$$

where the element  $J_{ij} = \mathbf{Jac}(A_i, B_j)$  of Jaccard's weighting matrix  $\mathbf{Jac}$  is defined as:

$$\mathbf{Jac}(A_i, B_j) = \frac{|A_i \cap B_j|}{|A_i \cup B_j|}. \quad (6)$$

For example, two BBAs are defined over the FOD  $\Theta = \{\theta_1, \theta_2\}$ :

$$\begin{aligned} m_1(A_1) &= 0.2, & m_1(A_2) &= 0.8, \\ m_2(B_1) &= 0.5, & m_2(B_2) &= 0.5. \end{aligned}$$

where

$$\begin{aligned} A_1 &= \{\theta_1\}, & A_2 &= \{\theta_1, \theta_2\}, \\ B_1 &= \{\theta_2\}, & B_2 &= \{\theta_1, \theta_2\}. \end{aligned}$$

We have

$$\begin{aligned} \mathbf{Jac}(A_1, B_1) &= |\emptyset|/|\{\theta_1, \theta_2\}| = 0; \\ \mathbf{Jac}(A_2, B_1) &= |\{\theta_2\}|/|\{\theta_1, \theta_2\}| = 0.5; \\ \mathbf{Jac}(A_1, B_2) &= |\{\theta_1\}|/|\{\theta_1, \theta_2\}| = 0.5; \\ \mathbf{Jac}(A_2, B_2) &= |\{\theta_1, \theta_2\}|/|\{\theta_1, \theta_2\}| = 1. \end{aligned}$$

Although there are other distance definitions for belief functions, they either have some limitations or are not strict distance metrics [18]. Jousselme's distance has been proved to be a strict distance metric [19].

The aim of the evidential reasoning is for decision-making. Several decision-making approaches in evidence theory are briefly reviewed next.

## III. DECISION-MAKING IN EVIDENCE THEORY

There are two major types of decision-making approaches under the evidence theory framework: directly using belief functions [20], [21] and using probability transformations of belief functions [22].

<sup>1</sup>i.e., cognitively independent.

### A. Decision-making using belief functions

There exist three main decision-making rules using  $Bel$  and  $Pl$ .

1) *Max Bel*: One chooses the proposition  $A$  with the maximum  $Bel(A)$ .  $Bel(\cdot)$  describes the lowest trust degree of a given proposition. So it is also called pessimistic decision-making in DST [20].

2) *Max Pl*: One chooses the proposition  $A$  with the maximum  $Pl(A)$ .  $Pl(\cdot)$  describes the highest trust degree of a given proposition. So it is also called optimistic decision-making in DST [20].

3) *Joint use of Bel and Pl*:  $Bel$  and  $Pl$  measure the degree of trust of a given proposition from two points of view. So it is not comprehensive to make a decision based on only one of them. An extension is the “final belief” defined below [21].

$$FB(A) = Bel(A) + \alpha(Pl(A) - Bel(A)), \quad (7)$$

where  $\alpha = Bel(A)/(Bel(A) + Bel(\bar{A}))$ . The proposition  $A$  with the maximum  $FB(A)$  is preferred.

Note that the proposition  $A$  can be either a singleton or a compound proposition (containing more than one singleton).

### B. Decision-making using probability transformations

Probability-based decision rules are the main stream of decision-making based on evidence theory [21], because the two-level reasoning and decision structure (i.e., the credal and pignistic levels) proposed by Smets in his TBM is quite appealing. In this type of decision-making approach, the belief function (or BBA, plausibility function) is transformed into a probability measure  $P$  first and then the decision can be made as  $\theta_i^* = \arg \max_{\theta_i} P(\theta_i)$ , where  $\theta_i$  is a singleton of the FOD. As we will see next, the probability transformation is crucial for this type of decision-making.

## IV. PROBABILITY TRANSFORMATIONS

A probability transformation is a mapping  $PT : Bel_{\Theta} \rightarrow Pr_{\Theta}$ .  $Bel_{\Theta}$  is a belief function defined on  $\Theta$  and  $Pr_{\Theta}$  is a probability measure (in fact a probability mass function, pmf) defined on  $\Theta$ . Major probability transformations (PTs) are summarized below.

1) *Pignistic transformation*: The classical pignistic probability was proposed in TBM framework [3], which is a subjective and non-probabilistic interpretation of evidence theory. At the credal level of TBM, beliefs are entertained [3], combined and updated. While at the pignistic level, decisions are made by applying the pignistic probability transformation (PPT).

Suppose that FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  in the sequel. The PPT [3] for singletons is defined as:

$$BetP(\theta_i) = \sum_{\theta_j \in B, B \in 2^{\Theta}} \frac{m(B)}{|B|}. \quad (8)$$

PPT is designed according to an idea similar to uncertainty maximization [14]. In PPT, masses are assigned uniformly to different singletons involved.

2) *Sudano's probabilities*: Sudano [5]–[7] proposed probability transformation proportional to plausibilities (PrPl) [5], probability transformation proportional to beliefs (PrBel) [5], probability transformation proportional to all plausibilities (PraPl) [5], hybrid probability transformation (PrHyb) [6], and an iterative version of probability transformation (PrScP) [5].

They are defined by different types of mappings as follows:

$$PrPl(\theta_i) = Pl(\{\theta_i\}) \cdot \sum_{Y \in 2^{\Theta}, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} Pl(\{\theta_j\})}, \quad (9)$$

$$PrBel(\theta_i) = Bel(\{\theta_i\}) \cdot \sum_{Y \in 2^{\Theta}, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} Bel(\{\theta_j\})}, \quad (10)$$

$$PraPl(\theta_i) = Bel(\{\theta_i\}) + \frac{1 - \sum_j Bel(\{\theta_j\})}{\sum_j Pl(\{\theta_j\})} \cdot Pl(\{\theta_i\}), \quad (11)$$

$$PrHyb(\theta_i) = PraPl(\theta_i) \cdot \sum_{Y \in 2^{\Theta}, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} PraPl(\theta_j)}, \quad (12)$$

$$PrScP(\theta_i) = \sum_{\theta_i \in Y} \left( \frac{PrScP(\theta_i)}{\sum_j PrScP(\theta_j)} \right) \cdot m(Y). \quad (13)$$

Note that the iterative PrScP should be initiated by some other transformation.

3) *Cobb-Shenoy's normalization of plausibility*: This probability transformation is defined as the normalized plausibility function of singletons [8].

$$PnPl(\theta_i) = \frac{Pl(\{\theta_i\})}{\sum_j Pl(\{\theta_j\})}. \quad (14)$$

4) *Cuzzolin's intersection probability*: From a geometric interpretation of evidence theory, an intersection probability measure [12] was proposed using the proportional repartition of the total nonspecific mass (TNSM =  $\sum_{A \in 2^{\Theta}, |A| > 1} m(A)$ ) for each contribution of the nonspecific masses involved.

$$CuzzP(\theta_i) = m(\{\theta_i\}) + \frac{Pl(\{\theta_i\}) - m(\{\theta_i\})}{\sum_j (Pl(\{\theta_j\}) - m(\{\theta_j\}))} \cdot TNSM. \quad (15)$$

5) *DSmP*: The  $DSmP_{\varepsilon}(\theta_i)$  [9] can be directly obtained by:

$$DSmP_{\varepsilon}(\theta_i) = m(\{\theta_i\}) + (m(\{\theta_i\}) + \varepsilon) \cdot \left( \sum_{\substack{X \in 2^{\Theta} \\ \theta_i \in X, |X| \geq 2}} \frac{m(X)}{\sum_{\substack{Y \in 2^{\Theta} \\ Y \subset X, |Y|=1}} m(Y) + \varepsilon \cdot |X|} \right). \quad (16)$$

In DSmP, both the mass assignments and the cardinality of focal elements are used in the proportional redistribution. DSmP makes an improvement compared with Sudano's, Cuzzolin's and BetP, in that DSmP makes a more judicious redistribution of the ignorance masses to the singletons involved. DSmP works for both DST and DSMT.

6) *HDSmP*: HDSmP is a hierarchical version of DSMP (See [13] for details). When the mass for the focal elements with the same cardinality are zero, HDSmP<sub>0</sub> can not be computed due to its hierarchical nature. Therefore, the parameter  $\epsilon$  is necessary to improve the applicability of HDSmP [13].

7) *PrBP1*: The proportional transformation hypothesis used in PrBP1 [11] assumes that the masses are distributed proportionally to the product of  $Bel(\theta_i)$  and  $Pl(\theta_i)$  among each singleton element of  $\theta_i \in Y$  with  $Y \subseteq \Theta$ .

$$\text{PrBP1}(\theta_i) = \sum_{Y, \theta_i \in Y} \left( \frac{Bel(\theta_i)Pl(\theta_i)}{\sum_{j, \theta_j \in Y} Bel(\theta_j)Pl(\theta_j)} \right) \cdot m(Y). \quad (17)$$

8) *PrBP2*: The PrBP2 [11] assumes that the masses are distributed proportionally to some given parameters  $s_i = \frac{Bel(\theta_i)}{1-Pl(\theta_i)}$  or  $s_i = \frac{Pl(\theta_i)}{1-Bel(\theta_i)}$ .

$$\text{PrBP2}(\theta_i) = \sum_{Y, \theta_i \in Y} \left( \frac{s_i}{\sum_{j, \theta_j \in Y} s_j} \right) \cdot m(Y). \quad (18)$$

A probability transformation outputs a Bayesian BBA (having only singleton focal elements) corresponding to a given (non-Bayesian) BBA. That is why the probability transformation is also called the Bayesian transformation. A Bayesian BBA is not a probability measure, but if  $m(\cdot)$  is a Bayesian BBA, then its corresponding  $Bel(\cdot)$  and  $Pl(\cdot)$  coincide with a probability measure, i.e.,  $Bel(\cdot) = Pl(\cdot) = P(\cdot)$ . Due to the tradition, it is still called the “probability transformation” in this paper.

## V. QUESTIONING OF TRADITIONAL EVALUATION OF PROBABILITY TRANSFORMATION

The evaluation of different probability transformations is important for analysis and their applications. It is also important for the design of new transformations. In this section, we will provide some comments on traditional evaluation approaches for probability transformations.

### A. Traditional Evaluation approaches for probability transformation

Qualitative evaluation approaches were proposed. In [13], three desired properties of a probability transformation are introduced, including:

1) *p-consistency*: A probability transformation  $PT$  is  $p$ -consistent (probability consistent) if  $PT(m) = m$  for any Bayesian BBA  $m$ .

2) *ULB-consistency*: A probability transformation is ULB-consistent (upper and lower bound consistent) if its resulting transformed probability  $P = PT(m)$  satisfies  $Bel(X) \leq P(X) \leq Pl(X)$ .

3) *Combination-consistency*: The combination-consistency means that we will obtain the same result either, if we combine two BBAs  $m_1$  and  $m_2$  using the combination rule first and perform probability transformation thereafter, or perform probability transformation to both input BBAs  $m_1$  and  $m_2$  first and combine them thereafter. It is defined through commutation property of combination rule and probability transformation. It is difficult to be satisfied, and PnPl [8] is the only one known

to the authors that can satisfy it when using Dempster’s rule of combination.

There also exist some quantitative metrics measuring the strength of a critical decision based on a probability measure:

1) *Normalized Shannon Entropy*: Suppose that  $P(\theta)$  is a pmf, where  $\theta \in \Theta$ ,  $|\Theta| = N$ . An evaluation criterion for the pmf transformed is as follows [11]:

$$E_H = \frac{-\sum_{\theta \in \Theta} P(\theta) \log_2(P(\theta))}{\log_2 N}, \quad (19)$$

i.e., the ratio of Shannon entropy [23] and the maximum Shannon entropy. Clearly  $E_H$  is normalized. The larger (smaller)  $E_H$  gets, the larger (smaller) the degree of uncertainty gets. When  $E_H = 0$ , one singleton proposition will have probability 1 and the others will have zero probabilities. Therefore, the agent or system can make decision without error if the probability  $P(\cdot)$  corresponds to the real probability of the events. When  $E_H = 1$ , it is unlikely to make a correct decision, because  $P(\theta)$  are equal, for all  $\theta \in \Theta$ , i.e., one has a uniform pmf.

2) *Probabilistic Information Content*: The Probabilistic Information Content (PIC) criterion [5] is the dual of the normalized Shannon entropy. The PIC value of a pmf obtained from a probability transformation indicates the total knowledge to make a correct decision:

$$\text{PIC}(P) = 1 + \frac{1}{\log_2 N} \cdot \sum_{\theta \in \Theta} P(\theta) \log_2(P(\theta)). \quad (20)$$

Obviously,  $\text{PIC} = 1 - E_H$ . A PIC value of zero indicates that the knowledge to make a correct decision is not informative enough (all propositions have equal probabilities, i.e., one has the maximal entropy).

Less uncertainty means that the corresponding probability transformation result is more helpful in making a decision. According to such an idea, the probability transformation should attempt to enlarge the belief differences among all the propositions and thus to achieve a clearer decision result. Is this rational? Is uncertainty degree always judicious at all to evaluate a probability transformation for decision-making purpose? If this is true, a probability transformation approach based on direct uncertainty minimization should be the best choice. Is that true? In the next section, we examine the legitimacy of using uncertainty degree as a criterion to evaluate a probability transformation.

### B. Probability transformation based on uncertainty minimization

As mentioned above, the “best” probability transformation can be obtained by directly minimizing  $E_H$  (or equivalently maximizing PIC) as follows.

$$\begin{aligned} & \min_{\{P(\theta) | \theta \in \Theta\}} \left\{ -\sum_{\theta \in \Theta} P(\theta) \log_2(P(\theta)) \right\}, \\ & \text{s.t.} \left\{ \begin{array}{l} Bel(B) \leq \sum_{\theta \in B} P(\theta) \leq Pl(B), \\ 0 \leq P(\theta) \leq 1, \forall \theta \in \Theta \\ \sum_{\theta \in \Theta} P(\theta) = 1. \end{array} \right. \end{aligned} \quad (21)$$

where the objective function is the Shannon entropy and the constraints are the ulb-consistency and the property of probability. Given a belief function, the solution of (21) is guaranteed to have the least uncertainty and is thus seemingly more preferable in decision-making. This so called the “best” transformation is denoted by  $Un_{min}$ .

Clearly, the problem of finding a minimum-entropy pmf does not have a unique solution in general. We use the Quasi-Newton method followed by a global optimization algorithm [24] to solve (21) to alleviate the effect of the local extremum problem. Other optimization algorithms [25], [26] can also be used, e.g., Genetic Algorithm (GA) and Particle Swarm Optimization (PSO).

C. Analysis of probability transformation based on uncertainty minimization

To compare different probability transformations, the following two examples drawn from [6] and [11] are considered, where PIC is used for evaluation.

Example 1: For FOD  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , the corresponding BBA is as follows:

$$\begin{aligned} m(\{\theta_1\}) &= 0.16, m(\{\theta_2\}) = 0.14, m(\{\theta_3\}) = 0.01, \\ m(\{\theta_4\}) &= 0.02, m(\{\theta_1, \theta_2\}) = 0.20, \\ m(\{\theta_1, \theta_3\}) &= 0.09, m(\{\theta_1, \theta_4\}) = 0.04, \\ m(\{\theta_2, \theta_3\}) &= 0.04, m(\{\theta_2, \theta_4\}) = 0.02, \\ m(\{\theta_3, \theta_4\}) &= 0.01, m(\{\theta_1, \theta_2, \theta_3\}) = 0.10, \\ m(\{\theta_1, \theta_2, \theta_4\}) &= 0.03, m(\{\theta_1, \theta_3, \theta_4\}) = 0.03, \\ m(\{\theta_2, \theta_3, \theta_4\}) &= 0.03, m(\Theta) = 0.08. \end{aligned}$$

TABLE I  
PROBABILITY TRANSFORMATION RESULTS FOR EXAMPLE 1.

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	PIC
PnPl	0.3614	0.3168	0.1931	0.1287	0.0526
CuzzP	0.3860	0.3382	0.1607	0.1151	0.0790
BetP	0.3983	0.3433	0.1533	0.1050	0.0926
PraPl	0.4021	0.3523	0.1394	0.1062	0.1007
PrPl	0.4544	0.3609	0.1176	0.0671	0.1638
PrHyb	0.4749	0.3749	0.0904	0.0598	0.2014
PrBel	0.5176	0.4051	0.0303	0.0470	0.3100
DSmP <sub>0</sub>	0.5176	0.4051	0.0303	0.0470	0.3100
DSmP <sub>0.001</sub>	0.5162	0.4043	0.0319	0.0477	0.3058
HDSmP <sub>0</sub>	0.5293	0.3960	0.0310	0.0437	0.3161
HDSmP <sub>0.001</sub>	0.5258	0.3943	0.0344	0.0455	0.3064
PrScP	0.5420	0.3870	0.0324	0.0386	0.3247
PrBP1	0.5419	0.3998	0.0243	0.0340	0.3480
PrBP2	0.5578	0.3842	0.0226	0.0353	0.3529
$Un_{min}$	0.7300	0.2300	0.0100	0.0300	0.4813

Based on the probability transformations defined in (8)–(18) and (21), respectively, the BBA can be transformed into different probabilities as illustrated in Table I. Their corresponding PIC’s can be calculated using (20), which are also listed in Table I. The  $Un_{min}$  provides the maximum PIC as expected.

Example 2: For FOD  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ , the corresponding BBA is as follows:

$$\begin{aligned} m(\{\theta_1\}) &= 0.05, m(\{\theta_2\}) = 0.00, m(\{\theta_3\}) = 0.00, \\ m(\{\theta_4\}) &= 0.00, m(\{\theta_1, \theta_2\}) = 0.39, \\ m(\{\theta_1, \theta_3\}) &= 0.19, m(\{\theta_1, \theta_4\}) = 0.18, \\ m(\{\theta_2, \theta_3\}) &= 0.04, m(\{\theta_2, \theta_4\}) = 0.02, \\ m(\{\theta_3, \theta_4\}) &= 0.01, m(\{\theta_1, \theta_2, \theta_3\}) = 0.04, \\ m(\{\theta_1, \theta_2, \theta_4\}) &= 0.02, m(\{\theta_1, \theta_3, \theta_4\}) = 0.03, \\ m(\{\theta_2, \theta_3, \theta_4\}) &= 0.03, m(\Theta) = 0.00. \end{aligned}$$

By using the probability transformations defined in (8) – (18) and (21), respectively, we can transform the BBA into different probabilities as illustrated in Table II. Their corresponding PIC’s can be calculated using (20), which are also listed in Table II. In this example, the masses for some singletons are zero, so some probability transformations can not be applied as shown in Table II, where N/A means “Not available”. The notation DSmP<sub>0</sub>, DSmP<sub>0.001</sub>, HDSmP<sub>0</sub> and HDSmP<sub>0.001</sub> mean that the values of the parameter  $\varepsilon$  in DSmP and HDSmP transformations are chosen to 0 and 0.001.

From the experimental results in Tables I and II, the pmf obtained from the proposed  $Un_{min}$  approach has the least uncertainty (and thus the greatest PIC) when compared with the others. That is, the difference among all propositions of the existing probability transformation approaches can be further enlarged, which is seemingly helpful for more consolidated and clearer decision-making.

TABLE II  
PROBABILITY TRANSFORMATION RESULTS FOR EXAMPLE 2.

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	PIC
PrBel	N/A due to 0 value of singletons				
PrScP	N/A due to 0 value of singletons				
PrBP1	N/A due to 0 value of singletons				
DSmP <sub>0</sub>	N/A due to 0 value of singletons				
HDSmP <sub>0</sub>	N/A due to 0 value of singletons				
PnPl	0.4348	0.2609	0.1643	0.1401	0.0733
CuzzP	0.4498	0.2540	0.1599	0.1364	0.0822
PraPl	0.4630	0.2478	0.1561	0.1331	0.0907
BetP	0.4600	0.2550	0.1533	0.1317	0.0910
PrPl	0.6161	0.2160	0.0960	0.0719	0.2471
PrBP2	0.6255	0.2109	0.0936	0.0700	0.2572
PrHyb	0.6368	0.2047	0.0909	0.0677	0.2698
DSmP <sub>0.001</sub>	0.8820	0.0486	0.0400	0.0294	0.6464
HDSmP <sub>0.001</sub>	0.8646	0.0604	0.0454	0.0296	0.6106
$Un_{min}$	0.9000	0.0900	0.0000	0.0100	0.7420

**Remark:** However, there exist serious deficiencies associated with  $Un_{min}$ , as shown in the following example.

Example 3: The BBA defined on FOD  $\Theta = \{\theta_1, \theta_2\}$  is:  
 $m(\{\theta_1\}) = 0.3, m(\{\theta_2\}) = 0.1, m(\{\theta_1, \theta_2\}) = 0.6.$

The experimental results of different approaches are listed in Table III. Is the probability transformation based on PIC maximization (i.e., entropy minimization) rational?

TABLE III  
PROBABILITY TRANSFORMATION RESULTS FOR EXAMPLE 3.

	$\theta_1$	$\theta_2$	PIC
PnP1	0.5625	0.4375	0.0113
CuzzP	0.6000	0.4000	0.0291
BetP	0.6000	0.4000	0.0291
PrP1	0.6375	0.3625	0.0553
PraP1	0.6375	0.3625	0.0553
PrHyb	0.6825	0.3175	0.0984
DSmP <sub>0</sub>	0.7500	0.2500	0.1887
DSmP <sub>0.001</sub>	0.7493	0.2507	0.1875
HDSmP <sub>0</sub>	0.7500	0.2500	0.1887
HDSmP <sub>0.001</sub>	0.7485	0.2515	0.1864
PrBel	0.7500	0.2500	0.1887
PrScP	0.7500	0.2500	0.1887
PrBP1	0.7765	0.2235	0.2334
PrBP2	0.8400	0.1600	0.3657
Un <sub>min</sub>	0.9000	0.1000	0.5310

In this simple example, all of the mass 0.6 committed to  $\{\theta_1, \theta_2\}$  is only redistributed to the singleton  $\{\theta_1\}$  when the Un<sub>min</sub> transformation is used in order to achieve the maximum PIC.

It can be also shown that for the Un<sub>min</sub>, the mass assignments  $m(\{\theta_1, \theta_2\}) > 0$  is always completely redistributed to  $\{\theta_1\}$  as long as  $m(\{\theta_1\}) > m(\{\theta_2\})$  in order to achieve the maximum PIC.

This is also true in the situations where the difference between masses of singletons is very small as demonstrated by the following BBA defined on FOD  $\Theta = \{\theta_1, \theta_2\}$ :

$$m(\{\theta_1\}) = 0.1000001, m(\{\theta_2\}) = 0.1, m(\{\theta_1, \theta_2\}) = 0.79999999.$$

Such a case shows that  $m(\{\theta_1\})$  is almost the same as  $m(\{\theta_2\})$  and there is no specific reason to obtain a very high probability for  $\theta_1$  and a small one for  $\theta_2$ . Therefore, the decision based on the result from Un<sub>min</sub> transformation appears to be very risky or dogmatic. In some applications, a decision has to be made and we cannot avoid to make one (good or bad). However, when the time to make a decision is not too limited or rejection decision is permitted, it is better to collect more observations (information) or to make a rejection rather than to take high risk to make an erroneous decision. So the criterion of uncertainty minimization, which can bring such risky results, is not always judicious to evaluate a probability transformation for decision-making purpose. Furthermore, when we use Un<sub>min</sub>, there are also some other problems. See the next example for details.

*Example 4:* The BBA defined on the FOD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  is:

$$m(\{\theta_1, \theta_2\}) = m(\{\theta_2, \theta_3\}) = m(\{\theta_1, \theta_3\}) = 1/3.$$

Using Un<sub>min</sub>, we can obtain six different pmf's yielding

the same minimal entropy, which are listed as follows:

$$\begin{aligned} P(\{\theta_1\}) &= 1/3, P(\{\theta_2\}) = 2/3, P(\{\theta_3\}) = 0; \\ P(\{\theta_1\}) &= 1/3, P(\{\theta_2\}) = 0, P(\{\theta_3\}) = 2/3; \\ P(\{\theta_1\}) &= 0, P(\{\theta_2\}) = 1/3, P(\{\theta_3\}) = 2/3; \\ P(\{\theta_1\}) &= 0, P(\{\theta_2\}) = 2/3, P(\{\theta_3\}) = 1/3; \\ P(\{\theta_1\}) &= 2/3, P(\{\theta_2\}) = 1/3, P(\{\theta_3\}) = 0; \\ P(\{\theta_1\}) &= 2/3, P(\{\theta_2\}) = 0, P(\{\theta_3\}) = 1/3. \end{aligned}$$

It is clear that the problem of finding a pmf with the minimal entropy does not have a unique solution in general. Then how to choose a unique one? In this example, different admissible pmf yields different decision result. This is a serious problem for decision-making.

From all the above examples, we conclude that the maximization of PIC (or minimization of Shannon entropy) is not satisfactory for evaluation. Therefore, more comprehensive evaluation methods are needed, which is an open and challenging problem.

## VI. A NEW BI-CRITERIA SOLUTION FOR PROBABILITY TRANSFORMATION EVALUATION

To design a single criterion for the comprehensive evaluation of probability transformations is always difficult. Jointly using multiple criteria is one option meaning that besides entropy or PIC, another measure, which describes some other aspects of a probability transformation, may be incorporated for evaluation.

The level of PIC characterizes the clarity of a given pmf. Indeed, higher PIC (lower entropy) means that the pmf tends to concentrate on a specific hypothesis of the FOD, which makes the decision easier for the decider. Also, the pmf is transformed from a given BBA representing the original information source. If the obtained pmf is in some sense closer to the original BBA, it will be preferred. This is because such a pmf has high degree of fidelity to the original BBA (i.e., with less loss of information in the transformation). The clarity and fidelity can always be balanced. So by adding one more criterion representing the degree of fidelity to the evaluation measure, it will be more comprehensive. How to characterize the closeness between the obtained pmf and the original BBA? The answer is to use the distance (or dissimilarity) of evidence [15]. In summary, we attempt to propose bi-criteria evaluation approaches by jointly using the distance of evidence and PIC.

By treating the pmf obtained from some probability transformation as a special BBA (not strict), a distance of evidence can be used to describe the dissimilarity between the pmf and the original BBA. We use the distance of evidence together with PIC (or entropy) as the elements of a two-tuple:

$$\langle \text{PIC}(P), d_J(P, m) \rangle, \quad \text{or} \quad \langle \text{Entropy}(P), d_J(P, m) \rangle, \quad (22)$$

where  $P$  is the transformed probability (i.e., pmf) and  $m$  is the original BBA.

A two-tuple can provide more comprehensive information; however, how to jointly use them to evaluate a probability transformation? Larger PIC represents greater clarity and smaller  $d_J$  represents greater fidelity. Over-emphasizing on

any single criterion is not preferred. As aforementioned, if we want to choose a better probability transformation approach, there should be a tradeoff between the two criteria. Here, we propose two approaches to jointly use the two criteria. The first one is directly based on the arithmetic operations over PIC and  $d_J$ . The second one first sorts probability transformations to be evaluated according to PIC and  $d_J$ , respectively, to obtain two corresponding ranks. And then, the two ranks are fused through some rank fusion rule to obtain a comprehensive rank for evaluation.

#### A. Comprehensive evaluation with value based fusion

Suppose that there are  $N$  pmf's denoted by  $P_1, P_2, \dots, P_N$ , which are obtained from  $N$  different probability transformations. Their entropies are  $Ent_1, Ent_2, \dots, Ent_N$  and their  $d_J$ 's (between each pmf and the original BBA) are  $d_1, d_2, \dots, d_N$ . To jointly use these two criteria, we calculate the comprehensive scores for different probability transformations as follows.

$$C_{joint}(P_i) = \alpha \cdot Ent'(i) + (1 - \alpha) \cdot d'(i), \quad (23)$$

where  $i = 1, \dots, N$  and  $\alpha$  denotes a weight representing the degree of preference on entropy. Distances (and entropies) usually take different values depending on the probability transformation. Therefore, to be consistent, we need to first normalize them by their ranges as:

$$\begin{cases} Ent'(i) = \frac{Ent(i) - \min(Ent)}{\max(Ent) - \min(Ent)}; \\ d'(i) = \frac{d(i) - \min(d)}{\max(d) - \min(d)}. \end{cases} \quad (24)$$

where the vector  $d = [d(1), d(2), \dots, d(N)]$  and the vector  $Ent = [Ent(1), Ent(2), \dots, Ent(N)]$  are the vectors of distances and entropies corresponding to the probability transformations  $P_1, P_2, \dots, P_N$ .

Smaller entropy (larger PIC, i.e., bigger clarity) and smaller distance (bigger fidelity) are desired. Then, by sorting the values of  $C_{joint}(P_i)$  in ascending order, we can obtain the rank as:

$$\Lambda_C = (r_C(P_1), r_C(P_2), \dots, r_C(P_N)). \quad (25)$$

The probability transformations with the best rank, i.e., those having the smallest rank value, are preferred.

#### B. Comprehensive evaluation with rank based fusion

Let us consider  $N$  pmf's denoted by  $P_1, P_2, \dots, P_N$ , which are obtained from  $N$  different probability transformations. A comprehensive rank can then be obtained from the rank based fusion (rank fusion for short) [27], [28] implemented in following steps.

- **Step 1:** Obtain the PIC-based rank

Sort the pmf's in descending order according to their PICs (this is because higher PIC value is desired). Then the rank of all the pmf's is:

$$\Lambda_{PIC} = (r_{PIC}(P_1), r_{PIC}(P_2), \dots, r_{PIC}(P_N)). \quad (26)$$

- **Step 2:** Obtain the distance-based rank

Sort the pmf's in ascending order according to the Jousselme distances ( $d_J$ ) (this is because smaller  $d_J$  is desired), then the rank for all the pmf's can be obtained as:

$$\Lambda_d = (r_d(P_1), r_d(P_2), \dots, r_d(P_N)). \quad (27)$$

- **Step 3:** Obtain the global rank by a rank fusion

To find the joint (or comprehensive) rank of  $\Lambda_{PIC}$  and  $\Lambda_d$ , a rank fusion is applied as:

$$\Lambda_f = f(\Lambda_{PIC}, \Lambda_d), \quad (28)$$

where  $f$  is a rank fusion rule and  $\Lambda_f$  is:

$$\Lambda_f = (r_j(P_1), r_j(P_2), \dots, r_j(P_N)). \quad (29)$$

The probability transformations with the best rank, i.e., those having the smallest rank, are preferred.

The selection of rank fusion rule is crucial, which is discussed below.

- 1) *Min rule:*

$$r_j(P_k) = \min(r_{PIC}(P_k), r_d(P_k)), \forall k = 1, \dots, N; \quad (30)$$

- 2) *Max rule:*

$$r_j(P_k) = \max(r_{PIC}(P_k), r_d(P_k)), \forall k = 1, \dots, N; \quad (31)$$

- 3) *Arithmetic averaging rule:*

$$r_j(P_k) = w_1 \cdot r_{PIC}(P_k) + w_2 \cdot r_d(P_k), \forall k = 1, \dots, N, \quad (32)$$

where  $w_1, w_2$  are weights for the two different ranks to be fused.

- 4) *Optimization rule:* The optimization based rank fusion rule is:

$$\Lambda^* = \arg \min_{\Lambda} \frac{1}{L} \sum_{j=1}^L d_r(\Lambda, \Lambda_j), \quad (33)$$

where  $\Lambda_1, \dots, \Lambda_L$  are the  $L$  different ranks to fuse and  $d_r(\cdot, \cdot)$  is a distance between two ranks that will be presented in the sequel.

Here we have two ranks  $\Lambda_{PIC}$  and  $\Lambda_d$ . The above equation could be rewritten as:

$$\Lambda_f = \arg \min_{\Lambda} \frac{1}{2} [d_r(\Lambda, \Lambda_{PIC}) + d_r(\Lambda, \Lambda_d)]. \quad (34)$$

$d_r(\cdot, \cdot)$  could be any rank distance including the footrule distance [29], the Kendall distance [30], [31] and the Spearman distance [32] as introduced below.

Suppose that  $\Lambda_1, \Lambda_2$  are two ranks. Let  $X = \{x_1, x_2, \dots, x_N\}$  be a set of items to be ranked.  $\Lambda_j(i)$  is the rank associated with the item  $x_i$ , where  $j = 1, 2$  and  $i = 1, 2, \dots, N$ .

- 1) *Footrule distance:*

$$F(\Lambda_1, \Lambda_2) = \sum_{i=1}^N |\Lambda_1(i) - \Lambda_2(i)|. \quad (35)$$

2) *Kendall distance*:

$$K(\Lambda_1, \Lambda_2) = \sum_{\{i,j\} \in P} K_{i,j}^*(\Lambda_1, \Lambda_2), \quad (36)$$

where  $P$  is the set of the unordered pairs of distinct items in  $X$  and

$$K_{i,j}^*(\Lambda_1, \Lambda_2) = \begin{cases} 0, & \text{if } x_i \text{ and } x_j \text{ are in the} \\ & \text{same order in } \Lambda_1 \text{ and } \Lambda_2, \\ 1, & \text{if } x_i \text{ and } x_j \text{ are in the} \\ & \text{reverse order in } \Lambda_1 \text{ and } \Lambda_2. \end{cases}$$

3) *Spearman distance*:

$$\rho(\Lambda_1, \Lambda_2) = 1 - \frac{6 \cdot \sum_{i=1}^N (\Lambda_1(i) - \Lambda_2(i))^2}{N(N^2 - 1)}. \quad (37)$$

Clearly,  $\rho \in [-1, 1]$ .  $\rho = 1$  means a total positive correlation between the ranks and  $\rho = -1$  means a total negative one.

Here  $\Lambda_1$  and  $\Lambda_2$  could be  $\Lambda_{PIC}$  and  $\Lambda_d$ , respectively.

### C. An illustrative example of rank fusion

Suppose that  $X = \{x_1, x_2, x_3, x_4\}$  corresponds to a set of possible choices in a decision-making problem.  $\Lambda_1 = [1, 2, 3, 4]$  and  $\Lambda_2 = [1, 3, 4, 2]$  are two ranks provided by two experts for  $X$ .

When using different rank fusion rules, the results (denoted by  $\Lambda_f$ ) are as follows.

1) *Min rule*:

$$\begin{aligned} \Lambda_f &= [\min(1, 1), \min(2, 3), \min(3, 4), \min(4, 2)] \\ &= [1, 2, 3, 2]. \end{aligned}$$

This rule has a tie.

2) *Max rule*:

$$\begin{aligned} \Lambda_f &= [\max(1, 1), \max(2, 3), \max(3, 4), \max(4, 2)] \\ &= [1, 3, 4, 4]. \end{aligned}$$

This rule also has a tie.

3) *Arithmetic averaging rule*:

$$\begin{aligned} \Lambda_f &= [0.5 \cdot 1 + 0.5 \cdot 1, 0.5 \cdot 3 + 0.5 \cdot 2, 0.5 \cdot 3 \\ &\quad + 0.5 \cdot 4, 0.5 \cdot 4 + 0.5 \cdot 2] \\ &= [1, 2.5, 3.5, 3]. \end{aligned}$$

As we can see that both weights are equal to 0.5.

We encounter the non-integer rank value. This doesn't matter. What we care is only the relative value of a rank. Therefore, we obtain the final result as  $[1, 2, 4, 3]$ , which is the rank obtained by ordering  $[1, 2.5, 3.5, 3]$  in ascending order.

4) *Optimization rule*: Here we use footrule distance. Suppose that  $\Lambda_f = [r_f(1), r_f(2), r_f(3), r_f(4)]$ . We try to find a  $\Lambda_f$  which minimizes

$$\begin{aligned} F(\Lambda_f, \Lambda_1) + F(\Lambda_f, \Lambda_2) &= |r_f(1) - 1| + |r_f(2) - 2| \\ &\quad + |r_f(3) - 3| + |r_f(4) - 4| \\ &\quad + |r_f(1) - 1| + |r_f(2) - 3| \\ &\quad + |r_f(3) - 4| + |r_f(4) - 2|. \end{aligned}$$

By using the optimization rank fusion rule in (33), one gets  $\Lambda_f^* = [1, 2, 4, 3]$ .

## VII. EXPERIMENTS FOR THE BI-CRITERIA EVALUATION APPROACH

In this section, we examine the previous Examples 1–3 using the new bi-criteria evaluation approach.

### A. Example 1 revisited

Table IV shows the evaluation results of different probability transformations (the initial pmf for PrScP used here is BetP), their distances and PIC's, their ranks obtained using two criteria ( $\Lambda_{PIC}$  and  $\Lambda_d$ ), and the joint rank using value based fusion ( $\Lambda_{Value}$ ). Table IV also provides the evaluation results using rank fusion, where  $\Lambda_{min}$  denotes the fused rank using min rule;  $\Lambda_{max}$  denotes the fused rank using max rule;  $\Lambda_{ave}$  denotes the fused rank using arithmetic averaging rule;  $\Lambda_{opt}$  denotes the fused rank using optimization. The weight for value based fusion is  $\alpha = 0.5$  while the weights for arithmetic averaging rank fusion are  $w_1 = w_2 = 0.5$ . In optimization-based rank fusion, the distance used is the Spearman distance in (37) due to its quadric form, which is mathematically more tractable for optimization. The comparisons among evaluation results of different criteria are also shown in Fig. 1. Note that in all experiments here, smaller value of rank represents higher rank.

In Table IV there exist cases of tie. Our strategy for the tie is as follows. When a tie happens, the alternatives in the tie will be assigned the same rank. The rank of the closest following alternative of the tie will be increased by the number of alternatives in the tie. For example, in Table IV, the PIC's of PrBel and DSMP<sub>0</sub> are the same, so their ranks are both 6. The PIC value of HDSMP<sub>0.001</sub> is the closest following alternative, so its rank becomes 8. That is, there is no rank 7 here because rank 6 appeared twice.

From Table IV, although the pmf obtained from Un<sub>min</sub> has the maximum PIC (thus it seems to be the best choice), it also has the maximum  $d_J$ . Therefore, it is not the best choice according to  $d_J$ . The joint evaluation results show that Un<sub>min</sub> is not preferred. The bi-criteria evaluation appears more natural and helpful than PIC alone.

Also, PnPI has the minimum  $d_J$ . Thus according to  $\Lambda_d$ , it is the best. But PnPI has the lowest PIC which is not good for making a clear or solid decision. From this angle, it is the worst. As we can see, the evaluation based on PIC or  $d_J$  alone is not satisfactory. In Table IV, PnPI has obtained the worst score based our proposed bi-criteria evaluation approach. This new evaluation approach can assure a "good" score for both elements of the two-tuple and meanwhile it can also counteract the exaggeration of a single factor.

It can be seen from the experimental results listed in Table IV and Fig. 1, although there are some differences among different evaluation approaches, BetP, PraPI, DSMP and HDSMP all perform pretty well in this example, which make a good balance between PIC and  $d_J$ , i.e., the clarity and the fidelity.

TABLE IV  
EVALUATIONS OF PROBABILITY TRANSFORMATION RESULTS OF BBA IN EXAMPLE 1 USING DIFFERENT CRITERIA.

	PIC	$d_J$	$\Lambda_{PIC}$	$\Lambda_d$	$\Lambda_{Value}$	$\Lambda_{min}$	$\Lambda_{max}$	$\Lambda_{ave}$	$\Lambda_{opt}$
PnPI	0.0526	0.2504	15	4	15	7	4	15	15
CuzzP	0.0790	0.24651	14	3	14	5	12	14	14
BetP	0.0926	0.2462	13	1	13	1	10	1	1
PraPI	0.1007	0.24647	12	2	12	3	8	1	2
PrPI	0.1638	0.2524	11	5	11	9	6	15	13
PrHyb	0.2014	0.2589	10	6	10	11	5	5	7
PrBel	0.3100	0.28084	6	9	6	11	2	3	3
DSmP <sub>0</sub>	0.3100	0.28084	6	9	6	11	2	3	3
DSmP <sub>0.001</sub>	0.3058	0.2801	9	7	8	14	2	5	5
HDSmP <sub>0</sub>	0.3161	0.2827	5	11	5	9	6	5	8
HDSmP <sub>0.001</sub>	0.3064	0.28082	8	8	9	15	1	5	6
PrScP	0.3247	0.2853	4	12	4	7	8	5	9
PrBP1	0.3480	0.2887	3	13	1	5	10	5	12
PrBP2	0.3529	0.2917	2	14	2	3	12	5	10
Un <sub>min</sub>	0.4813	0.3676	1	15	3	1	4	5	10

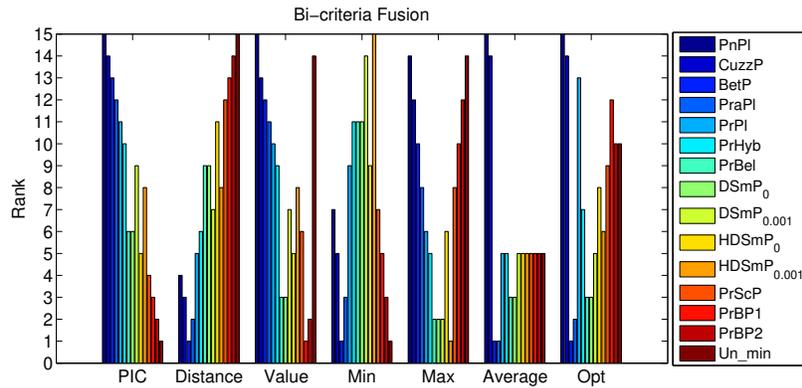


Fig. 1. Evaluation results for Example 1 using different criteria.

### B. Example 2 revisited

The parameters are the same as in revisited Example 1. The results are shown in Table V and Fig. 2.

In this experiment, we can see that four probability transformations (PrBel, PrBP1, HDSmP<sub>0</sub> and DSmP<sub>0</sub>) cannot provide results due to the zero value for singletons. This shows that they are not robust and have more requirements for the original BBA. DSmP and HDSmP have the parameter which can counteract this negative effect.

From the comparisons among the remaining eleven probability transformations in Table V, we see that although there are some differences among different evaluation approaches, those exaggerated transformations e.g., PnPI and Un<sub>min</sub>, over-emphasizing only one criterion, do not perform that well. The bi-criteria evaluation appears more natural and helpful than using PIC alone.

### C. Example 3 revisited

The results are shown in Table VI and Fig. 3.

In this experiment, those transformations (PrPI, PrHyb, PraPI, DSmP<sub>0.001</sub>, HDSmP<sub>0.001</sub>) making a good balance between clarity and fidelity always perform well using different evaluation approaches.

The above results show that DSmP and HDSmP can always generate a probability measure with less uncertainty. At the same time, this is not too risky, i.e., they can achieve a better tradeoff between PIC and risk in decision-making.

We prefer the evaluation approaches using rank fusion when compared with that using value based fusion. This is because in the evaluation approaches using rank fusion, the values are not that important. The rank fusion is not sensitive to the ranges of different criteria. In this sense, it is relatively more robust. Although in the evaluation approaches using value based fusion, we added the step of normalization to counteract the sensitivity to the value ranges, it cannot be avoided but suppressed to some extent.

Among all the evaluation approaches using rank fusion, we prefer the arithmetic averaging and the optimization based ones. This is because they are moderate, i.e., neither too

pessimistic (like min rule) nor too optimistic (like max rule).

$$PT(i_1) - PT(i_2) \geq \tau, \quad (42)$$

where  $\tau$  is the threshold for decision-making. If (42) is not satisfied,  $i_1$  will be rejected. The threshold  $\tau$  is selected from

$$\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 0.9, 1.0\}.$$

### VIII. APPLICATION-ORIENTED EVALUATION OF PROBABILITY TRANSFORMATIONS

The performance evaluation approaches for decision-making applications such as classification, are available. Therefore, we can use the evaluation of decision-making under the evidence theory framework to indirectly evaluate probability transformations, which is illustrated in Example 5.

*Example 5:* In this example, a pattern classification application is considered. We consider only three classes of samples in this example, which are illustrated in Fig. 4. The 2D-dataset is artificially generated. Samples of each class are uniformly distributed around three different centers. Abscissa and ordinate in Fig. 5 represent two feature dimensions of each sample.

The classifier used in this example is the  $K$ -nearest neighbor ( $K$ -NN) [33]. For each test sample, the output of the classifier is represented by a BBA. The corresponding BBA for each test sample is generated as follows.

a) The class space is  $C = \{1, 2, 3\}$ . For a test sample, find its  $K$  nearest neighbors. In the  $K$  nearest neighbors, calculate the ratio of the samples belonging to each class as follows:

$$P(i) = \frac{k(i)}{\sum_{j=1}^3 k(j)}, \quad (38)$$

where  $P(i)$  represents the ratio of class  $i$  and  $k(i)$  represents the number of samples belonging to class  $i$  in the  $K$  nearest neighbors,  $i = 1, 2, 3$ . Obviously,  $K = \sum_{j=1}^3 k(j)$ .

b) For the two classes  $s$  and  $t$  ( $s, t \in \{1, 2, 3\}, s \neq t$ ) with the top two values of  $k(i), i = 1, 2, 3$ , the corresponding mass assignments are generated according to [34]:

$$m(\{i\}) = P(i), \forall i = s, t. \quad (39)$$

The remaining mass is assigned to the total set  $\Theta$ :

$$m(\Theta) = 1 - m(\{s\}) - m(\{t\}). \quad (40)$$

For example, for a test sample  $x_q$ , among its 7 nearest neighbors, 4 belong to class 1, 2 belong to class 2, and one belongs to class 3. The class distribution is then  $P(1) = 4/7, P(2) = 2/7, P(3) = 1/7$ . The dominant class is class 1 and class 2 is at the second place. The corresponding BBA is  $m(\{1\}) = 4/7, m(\{2\}) = 2/7$  and  $m(\{1, 2, 3\}) = 1/7$ .

There are 200 samples for each one class with a total of 600 samples. In each experiment cycle, the samples are randomly selected from each class with 100 samples for training (300 training samples in total) and the remaining samples are used for testing (300 test samples in total).

For a probability transformation  $PT$ , the decision result will be class  $i_1$  if

$$i_1 = \arg \max_j PT(j), i_2 = \arg \max_{j, j \neq i_1} PT(j), \quad (41)$$

For each of these tested thresholds, the above experiment procedure is repeated 100 times to calculate the average performance of different probability transformations<sup>2</sup>. Based on the simulation results, the average rejection rates (average over all threshold values) of different probability transformations are shown in Fig. 5.

As we can see from Fig. 5, the rejection rate of  $Un_{min}$  is the minimum one. This is because  $Un_{min}$  emphasizes on the clarity. Therefore, the difference between different alternatives is relatively large. We propose to use “rejection-error” curve as shown in Fig. 6 to evaluate probability transformations.

The abscissas and ordinates of the points on rejection-error curves are respectively the average rejection rate values and the average error rate values for different probability transformations at each threshold  $\tau$ . For any probability transformation at each threshold value, the average rejection rate and the average error rate are the mean of the repeated 100 times simulations. The deviations of the rejection rate and error rate of each probability transformation at different thresholds are listed in Tables VII and VIII, respectively.

From the results in Fig. 6, it can be seen that given the same rejection rate, the classification error rate of  $Un_{min}$  is always the highest. The decision results based on  $Un_{min}$  are the worst. Although  $Un_{min}$  has the least uncertainty degree and the minimum rejection rate, it is not the winner. The smaller rejection rate is at the price of higher classification error rate. The rejection-error curves can be used as a comprehensive and indirect application-oriented evaluation approach for probability transformations.

The performance of other probability transformations (except for  $Un_{min}$ ) is similar when using application-oriented performance evaluation (rejection-error curves).

### IX. DESIRED PROPERTIES OF PROBABILITY TRANSFORMATIONS

In this section, we discuss some desired properties of probability transformations.

#### A. Order preservation

It is preferred that a probability transformation can maintain some order before and after the transformation, e.g., the order of the uncertainty degree. Given  $N$  BBAs:  $m_1, \dots, m_N$ , we can obtain the rank or order of their uncertainty degree according

<sup>2</sup>The probability transformations HDSMP<sub>0</sub>, PrBP1, PrBP2 are not included in our simulation because they cannot be computed when zero masses occur.

TABLE V  
EVALUATIONS OF PROBABILITY TRANSFORMATION RESULTS OF BBA IN EXAMPLE 2 USING DIFFERENT CRITERIA.

	PIC	$d_J$	$\Lambda_{PIC}$	$\Lambda_d$	$\Lambda_{Value}$	$\Lambda_{min}$	$\Lambda_{max}$	$\Lambda_{ave}$	$\Lambda_{opt}$
PrBel	N/A due to 0 value of singletons								
DSmP <sub>0</sub>	N/A due to 0 value of singletons								
HDSmP <sub>0</sub>	N/A due to 0 value of singletons								
PrBP1	N/A due to 0 value of singletons								
PnPI	0.0733	0.3128	11	4	9	7	10	11	11
CuzzP	0.0822	0.3123	10	3	7	5	8	10	10
BetP	0.0910	0.3121	8	1	5	1	4	1	1
PraPI	0.907	0.3122	9	2	6	3	6	2	2
PrPI	0.2471	0.3373	7	5	1	9	2	3	3
PrHyb	0.2698	0.3440	5	7	3	9	2	3	3
DSmP <sub>0.001</sub>	0.6464	0.4684	2	10	11	3	8	3	3
HDSmP <sub>0.001</sub>	0.6094	0.4572	3	9	10	5	6	3	3
PrScP	0.5987	0.4309	4	8	4	7	4	3	3
PrBP2	0.2572	0.3402	6	6	2	11	1	3	3
Un <sub>min</sub>	0.7421	0.4764	1	11	8	1	10	3	3

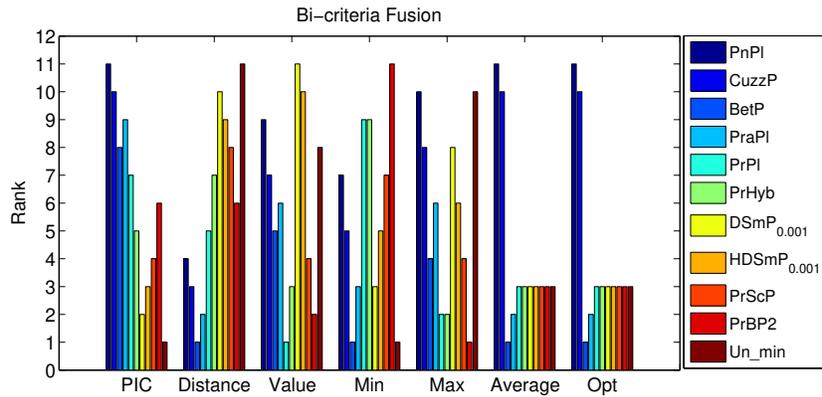


Fig. 2. Evaluation results for Example 2 using different criteria.

TABLE VI  
EVALUATION OF PROBABILITY TRANSFORMATION RESULTS OF BBA IN EXAMPLE 3 USING DIFFERENT CRITERIA.

	PIC	$d_J$	$\Lambda_{PIC}$	$\Lambda_d$	$\Lambda_{Value}$	$\Lambda_{min}$	$\Lambda_{max}$	$\Lambda_{ave}$	$\Lambda_{opt}$
PnPI	0.0113	0.3023	15	3	15	5	14	15	15
CuzzP	0.0290	0.3000	13	1	11	1	10	1	1
BetP	0.0290	0.3000	13	1	11	1	10	1	1
PraPI	0.0553	0.3023	11	3	2	5	8	1	3
PrPI	0.0553	0.3023	11	3	2	5	8	1	3
PrHyb	0.0984	0.3111	10	6	1	13	3	9	14
PraBel	0.1887	0.3354	5	10	7	9	3	5	5
DSmP <sub>0</sub>	0.1887	0.3354	5	10	7	9	3	5	5
DSmP <sub>0.001</sub>	0.1875	0.3351	8	8	5	15	1	9	9
HDSmP <sub>0</sub>	0.1887	0.3354	5	10	7	9	3	5	5
HDSmP <sub>0.001</sub>	0.0553	0.3023	9	7	4	14	2	9	10
PrScP	0.1887	0.3354	5	10	7	9	3	5	5
PrBP1	0.2334	0.3481	3	13	10	5	10	9	11
PrBP2	0.3657	0.3842	2	14	13	4	13	9	12
Un <sub>min</sub>	0.5310	0.4243	1	15	14	1	14	9	12

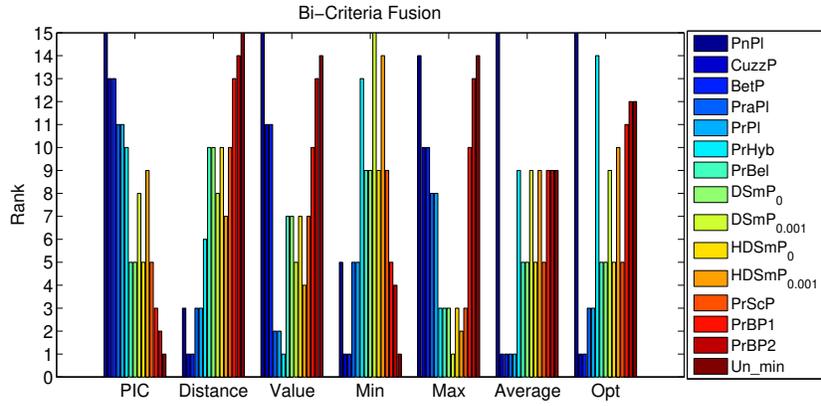


Fig. 3. Evaluation results for Example 3 using different criteria.

TABLE VII  
DEVIATIONS OF THE REJECTION RATE OF EACH PROBABILITY TRANSFORMATION AT DIFFERENT  $\tau$ .

$\tau$	BetP	Un <sub>min</sub>	DSMP <sub>0.001</sub>	PnP1	DSMP <sub>0</sub>	PrP1	PrBel	PraP1	HybP1	CuzzP	HDSMP <sub>0.001</sub>
0	0	0	0	0	0	0	0	0	0	0	0
0.1	0.0406	0.0130	0.0382	0.0434	0.0360	0.0406	0.0360	0.0406	0.0382	0.0406	0.0406
0.2	0.0401	0.0332	0.0421	0.0356	0.0421	0.0412	0.0421	0.0412	0.0421	0.0401	0.0412
0.3	0.0374	0.0496	0.0418	0.0312	0.0418	0.0411	0.0418	0.0411	0.0418	0.0374	0.0387
0.4	0.0284	0.0453	0.0309	0.0238	0.0330	0.0309	0.0330	0.0309	0.0309	0.0276	0.0298
0.5	0.0265	0.0320	0.0282	0.0267	0.0284	0.0274	0.0284	0.0274	0.0282	0.0265	0.0274
0.6	0.0209	0.0247	0.0214	0.0199	0.0214	0.0209	0.0214	0.0209	0.0214	0.0209	0.0209
0.7	0.0208	0.0213	0.0213	0.0206	0.0213	0.0213	0.0213	0.0213	0.0213	0.0208	0.0208
0.8	0.0185	0.0192	0.0192	0.0185	0.0192	0.0192	0.0192	0.0192	0.0192	0.0195	0.0195
0.9	0.0165	0.0167	0.0167	0.0165	0.0167	0.0167	0.0167	0.0167	0.0167	0.0165	0.0165
1.0	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131	0.0131

TABLE VIII  
DEVIATIONS OF THE ERROR RATE OF EACH PROBABILITY TRANSFORMATION AT DIFFERENT  $\tau$ .

$\tau$	BetP	Un <sub>min</sub>	DSMP <sub>0.001</sub>	PnP1	DSMP <sub>0</sub>	PrP1	PrBel	PraP1	HybP1	CuzzP	HDSMP <sub>0.001</sub>
0	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0204	0.0199
0.1	0.0318	0.0182	0.0308	0.0319	0.0300	0.0318	0.0300	0.0318	0.0308	0.0318	0.0318
0.2	0.0283	0.0238	0.0322	0.0233	0.0322	0.0311	0.0322	0.0311	0.0322	0.0283	0.0311
0.3	0.0191	0.0268	0.0228	0.0127	0.0228	0.0214	0.0228	0.0214	0.0228	0.0191	0.0197
0.4	0.0099	0.0236	0.0118	0.0071	0.0144	0.0118	0.0144	0.0118	0.0118	0.0094	0.0108
0.5	0.0059	0.0140	0.0080	0.0057	0.0085	0.0070	0.0085	0.0070	0.0080	0.0059	0.0070
0.6	0.0036	0.0075	0.0039	0.0027	0.0039	0.0039	0.0039	0.0039	0.0039	0.0036	0.0036
0.7	0.0005	0.0008	0.0008	0.0005	0.0008	0.0008	0.0008	0.0008	0.0008	0.0005	0.0005
0.8	3e-17	0.0005	0.0005	3e-17	0.0005	0.0005	0.0005	0.0005	0.0005	3e-17	3e-17
0.9	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17
1.0	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17	3e-17

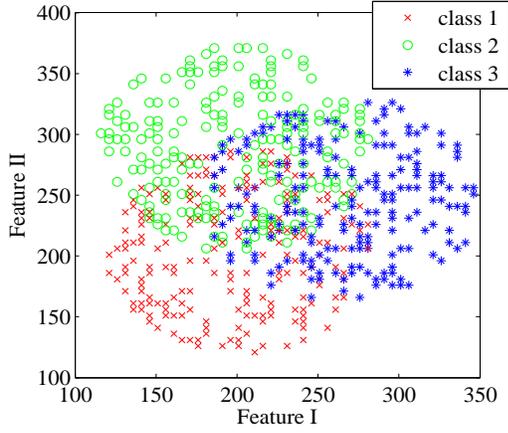


Fig. 4. The samples for classification.

to some uncertainty measure in evidence theory as introduced in the sequel:

$$Rank_m = [r_m(1), \dots, r_m(N)].$$

After applying the probability transformations, the rank of  $P_1, \dots, P_N$  in terms of uncertainty degree in probability theory<sup>3</sup> is:

$$Rank_P = [r_P(1), \dots, r_P(N)].$$

It is very hard to completely maintain the order after the transformation. However, the degree of accordance or consistency between  $Rank_m$  and  $Rank_P$ , i.e., the degree of order-preserving, can be used to evaluate different probability transformations. Such a degree of order-preservation can be defined using the distance between ranks as introduced before. Less degree of order-preserving represents more twist or loss of information in the procedure of probability transformations.

When calculating the degree of order-preservation, we need to specify an uncertainty measure of belief functions. Some uncertainty measures in evidence theory are as follows.

1) *Non-specificity*: Non-specificity [35] is defined as:

$$N(m) = \sum_{A \subseteq \Theta} m(A) \log_2 |A|. \quad (43)$$

2) *Confusion*: Confusion [36] is defined using the BBA  $m$  and the  $Bel$  in the spirit of entropy as:

$$Conf(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 (Bel(A)). \quad (44)$$

3) *Dissonance*: Dissonance [37] is defined using the BBA  $m$  and the  $Pl$  in the spirit of entropy as:

$$Diss(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 (Pl(A)). \quad (45)$$

<sup>3</sup>Here we only consider Shannon's entropy as the ranking criterion for convenience.

4) *Aggregate Uncertainty measure (AU)*: Let  $Bel$  be a belief measure on the FOD  $\Theta$ . The AU [38] associated with  $Bel$  is measured by:

$$AU(Bel) = \max_{\mathcal{P}_{Bel}} \left[ - \sum_{\theta \in \Theta} p_\theta \log_2 p_\theta \right], \quad (46)$$

where the maximum is taken over all probability distributions that are consistent with the given belief function.  $\mathcal{P}_{Bel}$  consists of all probability distributions  $\langle p_\theta | \theta \in \Theta \rangle$  satisfying:

$$\begin{cases} p_\theta \in [0, 1], \forall \theta \in \Theta, \\ \sum_{\theta \in \Theta} p_\theta = 1 \\ Bel(A) \leq \sum_{\theta \in A} p_\theta \leq 1 - Bel(\bar{A}), \forall A \subseteq \Theta. \end{cases} \quad (47)$$

AU is an aggregated total uncertainty (ATU) measure.

AU satisfies all the requirements of uncertainty measure [39] including probability consistency, set consistency, value range, sub-additivity and additivity for the joint BPA in Cartesian space. AU also has the drawbacks [3] of high computing complexity, high insensitivity to the changes of evidence, etc.

5) *Ambiguity Measure (AM)*: Let  $m$  be a BBA defined over the FOD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . AM (ambiguity measure) [40] is defined as:

$$AM(m) = - \sum_{\theta \in \Theta} BetP_m(\theta) \log_2 (BetP_m(\theta)), \quad (48)$$

where  $BetP_m(\theta) = \sum_{B \in \mathcal{B}, B \subseteq \Theta} m(B) / |B|$  is the pignistic probability. Jousselme *et al* [40] declared that the AM satisfies the requirements of uncertainty measure and at the same time it overcomes the defects of AU, but in fact AM does not satisfy the sub-additivity [41]. Moreover in [39], AM has been proved to be logically non-monotonic under some conditions.

6) *Contradiction Measure (CM)*: The contradiction measure [42] is defined as:

$$CM(m) = \sqrt{\frac{2n}{n-1}} \cdot \sum_{X \in \mathcal{X}} m(X) \cdot d_J(m, m_X), \quad (49)$$

where  $\mathcal{X}$  denotes the set of all focal elements of  $m$  and  $d_J$  is Jousselme's distance. There exists  $CM \in [0, 1]$ .

#### B. Simulation of degree of order-preservation

Our simulation consists of the following steps:

- **Step 1**: Randomly generate 10 BBAs and calculate the degree of uncertainty in each BBA to generate  $Rank_m$ ;
- **Step 2**: Apply the transformation using  $N$  types of PTs. We can obtain various  $Rank_p^i$ ,  $i = 1, \dots, N$ ;
- **Step 3**: Calculate the distance ( $d_p(i)$ ) between  $Rank_m$  and each  $Rank_p^i$ ,  $i = 1, \dots, N$ ;
- **Step 4**: Repeat Step 1 - 3 a hundred times. Calculate the average distance and the corresponding standard deviation as follows:

$$d_{pm}(i) = \frac{1}{100} \sum_{j=1}^{100} d_p^j(i), \quad (50)$$

$$d_{p-std}(i) = \sqrt{\frac{1}{100-1} \sum_{j=1}^{100} (d_p^j(i) - d_{pm}(i))^2}. \quad (51)$$

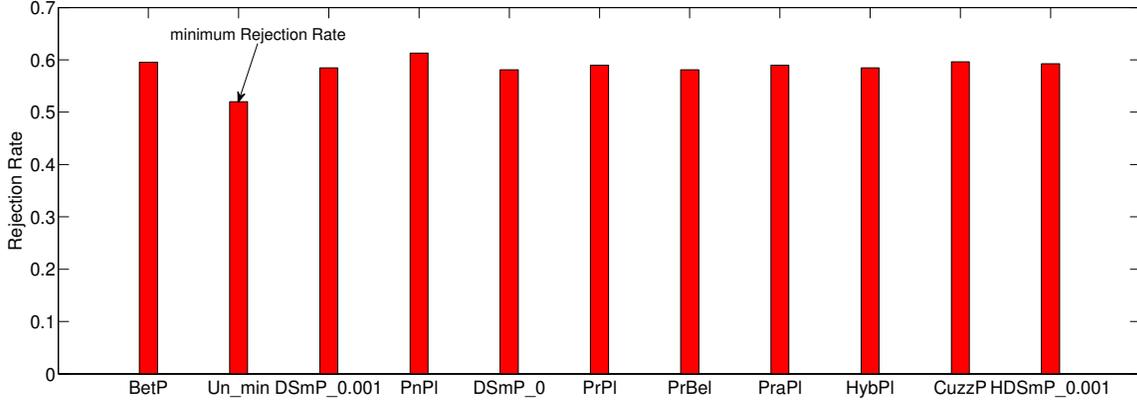


Fig. 5. Comparison of average rejection rate over all thresholds among different probability transformations.

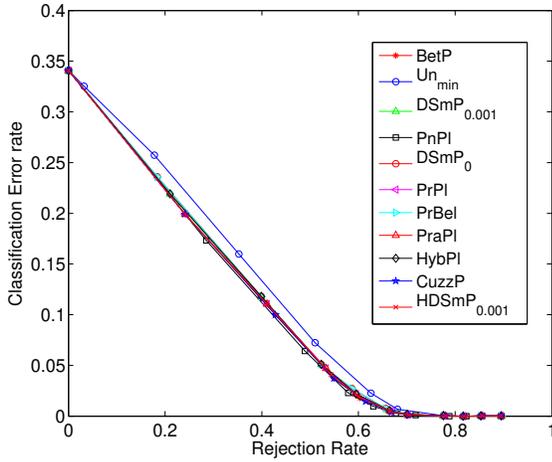


Fig. 6. Rejection-error curves comparison of different probability transformations.

The PTs with smaller  $d_{pm}(i)$  and  $d_{p-std}(i)$  are preferred.

TABLE IX

Algorithm 1. RANDOM GENERATION OF BBA.

---

**Input:**  $\Theta$ : Frame of discernment;  
 $N_{max}$ : Maximum number of focal elements  
**Output:** **Output:**  $m$ : BBA  
 Generate  $\mathcal{P}(\Theta)$ , which is the power set of  $\Theta$ ;  
 Generate a random permutation of  $\mathcal{P}(\Theta) \rightarrow \mathcal{R}(\Theta)$ ;  
 Generate an integer between 1 and  $N_{max} \rightarrow l$ ;  
**FOReach** First  $k$  elements of  $\mathcal{R}(\Theta)$  do  
 Generate a value within  $[0, 1] \rightarrow m_i, i = 1, \dots, l$ ;  
**END**  
 Normalize the vector  $m = [m_1, \dots, m_l] \rightarrow m'$ ;  
 $m(A_i) = m'_i$ ;

---

In Step 1, BBAs are generated using Algorithm 1 [17] in Table IX. All six types of uncertainty measures introduced above are used in the simulations. The distance between  $Rank_m$  and  $Rank_p$  used is the Kendall distance in (36). The evaluation results are shown in Tables X, XI and Fig. 7.

From Tables X, XI, and Fig. 7,  $Un_{min}$  provides the worst (with the lowest degree of order preservation). Although all the degree of order change based on different uncertainty measures are all listed in Tables X, XI, and Fig. 7, we prefer to use the ones based on AU. The reasons are as follows. First, in all the uncertainty measures used here, only AU is a strict total uncertainty measure, which can describe both the discord and non-specificity in a body of evidence. Second, as we can see from the subfigure using AM, BetP's degree of order change is zero. However, this does not make sense, because AM is defined based on BetP. Therefore, it is partial (or non-neutral) when BetP is also included for evaluation. AU is relative more appropriate to be used here. According to the subfigure based on AU,  $Un_{min}$  is the worst. HDSmP and DSMP are also not that good in terms of order preservation. Other probability transformations perform similar to each other in terms of order preservation.

Even AU is still not absolutely impartial (or neutral), because AU is designed also based on some probability transformation (maximization of entropy). If such an entropy maximization-based probability transformation is also included for evaluation, it will be partial (or non-neutral). So some new uncertainty measure for BBA not related to probability transformation is needed for an impartial evaluation. We think that the contradiction measure in (49) is a good attempt. It is an uncertainty measure which is not based on a probability transformation, although its strictness (satisfying the requirements of the uncertainty measure) still deserves further research.

## X. CONCLUSIONS

In this paper, we focus on the evaluations of probability transformations of a belief function. The existing transformations are briefly reviewed and compared. Our experimental results and analysis show that PIC criterion alone is insufficient to truly measure the quality of a probability transformation. A compromise between fidelity and clarity is achieved by the joint use of PIC and the distance of evidence. We have also proposed an application-oriented evaluation approach for

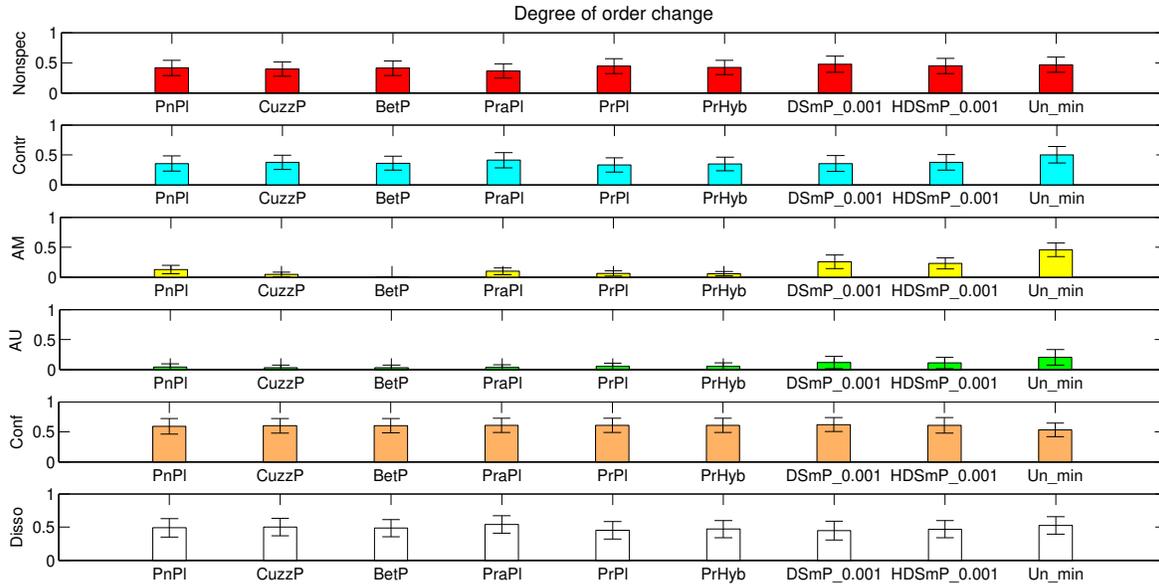


Fig. 7. Evaluation in terms of the degree of order change.

TABLE X  
EVALUATION OF DEGREE OF ORDER CHANGE (AVERAGE VALUE).

Prob Trans	PnPI	CuzzP	BetP	PraPI	PrPI	PrHyb	DSMP <sub>0.001</sub>	HDSMP <sub>0.001</sub>	Un <sub>min</sub>
Nonspecificity	0.4140	0.3958	0.4100	0.3640	0.4411	0.4240	0.4760	0.4476	0.4671
Contradiction	0.3567	0.3756	0.3618	0.4111	0.3320	0.3502	0.3551	0.3767	0.5022
AM	0.1282	0.0476	0	0.0989	0.0622	0.0589	0.2571	0.2284	0.4560
AU	0.0416	0.0304	0.0311	0.0324	0.0489	0.0507	0.1151	0.1082	0.2029
Confusion	0.5896	0.6009	0.6004	0.6084	0.6064	0.6071	0.6213	0.6076	0.5302
Dissonance	0.4882	0.4998	0.4836	0.5389	0.4511	0.4696	0.4453	0.4673	0.5247

TABLE XI  
EVALUATION OF DEGREE OF ORDER CHANGE (STANDARD DEVIATION).

Prob Trans	PnPI	CuzzP	BetP	PraPI	PrPI	PrHyb	DSMP <sub>0.001</sub>	HDSMP <sub>0.001</sub>	Un <sub>min</sub>
Nonspecificity	0.1278	0.1183	0.1207	0.1177	0.1237	0.1177	0.1335	0.1263	0.1271
Contradiction	0.1289	0.1193	0.1170	0.1249	0.1176	0.1127	0.1317	0.1302	0.1387
AM	0.0711	0.0378	0	0.0567	0.0417	0.0372	0.1127	0.0915	0.1119
AU	0.0545	0.0404	0.0400	0.0432	0.0542	0.0579	0.1033	0.0945	0.1282
Confusion	0.1297	0.1206	0.1192	0.1200	0.1187	0.1188	0.1175	0.1282	0.1131
Dissonance	0.1380	0.1309	0.1308	0.1294	0.1317	0.1296	0.1409	0.1296	0.1309

probability transformations. Furthermore, we have evaluated probability transformations by their robustness to preserve the uncertainty order of the original BBAs. The simulation results show that our proposed evaluation approaches are able to make rational comparison of different probability transformations. Future work includes the development of general and direct measures of uncertainty of a BBA, which do not depend on the choice of probability transformations. This is important for the property of uncertainty order preservation.

Note that the evaluations for the issues in evidence theory (like the probability transformations, the evidence combination, the determination of BBA, etc) lack solid theoretical foundation so far. In the future, we will attempt to propose more rational and useful criteria for probability transformations and try to establish more theoretically sound evaluation approaches for probability transformations, which are important and challenging problems.

## REFERENCES

- [1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of mathematical statistics*, vol. 38, no. 4, pp. 325–339, 1967.
- [2] G. Shafer, *A mathematical theory of evidence*. Princeton, NJ, USA: Princeton university press Princeton, 1976, vol. 1.
- [3] P. Smets and R. Kennes, "The transferable belief model," *Artificial intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [4] F. Smarandache and J. Dezert, Eds., *Applications and Advances of DSMT for Information Fusion (Vol III)*. Rehoboth, DE, USA: Amer. Res. Press, 2009. [Online]. Available: <http://www.gallup.unm.edu/~smarandache/DSMT-book3.pdf>
- [5] J. J. Sudano, "Pignistic probability transforms for mixes of low- and high-probability events," in *Proceedings of 4th International Conference on Information Fusion*, Montréal, Canada, August 2001, pp. 23–27.
- [6] —, "The system probability information content (PIC) relationship to contributing components, combining independent multi-source beliefs, hybrid and pedigree pignistic probabilities," in *Proceedings of the 5th International Conference on Information Fusion*, Annapolis, MD, USA, July 2002, pp. 1277–1283.
- [7] —, "Yet another paradigm illustrating evidence fusion (YAPIEF)," in *Proceedings of the 9th International Conference on Information Fusion*, Florence, Italy, July 2006, pp. 1–7.
- [8] B. R. Cobb and P. P. Shenoy, "On the plausibility transformation method for translating belief function models to probability models," *International Journal of Approximate Reasoning*, vol. 41, no. 3, pp. 314–330, 2006.
- [9] J. Dezert and F. Smarandache, "A new probabilistic transformation of belief mass assignment," in *Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, June – July 2008, pp. 1–8.
- [10] J. J. Sudano, "Belief fusion, pignistic probabilities, and information content in fusing tracking attributes," in *Proceedings of the IEEE 2004 Radar Conference*. Philadelphia, PA, USA: IEEE, April 2004, pp. 218–224.
- [11] W. Pan and H. Yang, "New methods of transforming belief functions to pignistic probability functions in evidence theory," in *Proceedings of the 2009 International Workshop on Intelligent Systems and Applications*, 2009, pp. 1–5.
- [12] F. Cuzzolin, "On the properties of the intersection probability," in *Proceedings of 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, Verona, Italy, July 2007, pp. 287–298.
- [13] J. Dezert, D. Han, Z. Liu, and J.-M. Tacnet, "Hierarchical proportional redistribution for bba approximation," in *Belief functions: theory and applications*, ser. Advances in Intelligent and Soft Computing, T. Denoeux and M.-H. Masson, Eds. Compiegne, France: Springer, 2012, vol. 164, pp. 275–283.
- [14] D. Han, J. Dezert, C. Han, and Y. Yang, "Is entropy enough to evaluate the probability transformation approach of belief function?" in *Proceedings of the 13th Conference on Information Fusion*, Edinburgh, UK, July 2010, pp. 1–7.
- [15] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.
- [16] M. Daniel, "Probabilistic transformations of belief functions," in *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*. Barcelona, Spain: Springer, July 2005, pp. 539–551.
- [17] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, no. 2, pp. 118–145, 2012.
- [18] D. Han, Y. Deng, C. Han, and Y. Yang, "Some notes on betting commitment distance in evidence theory," *Science China - Information Sciences*, vol. 55, no. 3, pp. 558–565, 2012.
- [19] M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the jaccard index matrix," *International Journal of Approximate Reasoning*, vol. 54, no. 5, pp. 615–626, 2013.
- [20] R. R. Yager, "Decision making under Dempster-Shafer uncertainties," *International Journal of General System*, vol. 20, no. 3, pp. 233–245, 1992.
- [21] W. Lv, "Decision-making rules based on belief interval with DS evidence theory," in *Fuzzy Information and Engineering*, ser. Advances in Soft Computing, B. Cao, Ed. Springer, 2007, vol. 40, pp. 619–627.
- [22] P. Smets, "Decision making in the TBM: the necessity of the pignistic transformation," *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 133–147, 2005.
- [23] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [24] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*, 2nd ed. New York, NY, USA: Kluwer Academic Publishers, 2000.
- [25] D. T. Pham and D. Karaboga, *Intelligent Optimisation Techniques*. New York, NY, USA: Springer, 2000.
- [26] J. Kennedy, J. F. Kennedy, and R. C. Eberhart, *Swarm Intelligence*. San Francisco, CA, USA: Morgan Kaufmann, 2001.
- [27] M. E. Renda and U. Straccia, "Web metasearch: rank vs. score based rank aggregation methods," in *Proceedings of the 18th ACM Symposium on Applied Computing*. Melbourne, FL, USA: ACM, March 2003, pp. 841–846.
- [28] A. Argentini, "Ranking aggregation based on belief function theory," Ph.D. dissertation, University of Trento, 2012.
- [29] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation methods for the web," in *Proceedings of the 10th international conference on World Wide Web*. Hong Kong, China: ACM, May 2001, pp. 613–622.
- [30] M. G. Kendall, *Rank Correlation Methods*. London, UK: Charles Griffen, 1948.
- [31] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, vol. 17, no. 1. SIAM, 2003, pp. 134–160.
- [32] J. L. Myers, A. Well, and R. F. Lorch, *Research Design and Statistical Analysis*, 3rd ed. Hove, UK: Routledge, 2010.
- [33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: John Wiley & Sons, 2012.
- [34] Y. Bi, D. Bell, and J. Guan, "Combining evidence from classifiers in text categorization," in *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*. Wellington, New Zealand: Springer, Sept 2004, pp. 521–528.
- [35] D. Dubois and H. Prade, "A note on measures of specificity for fuzzy sets," *International Journal of General Systems*, vol. 10, no. 4, pp. 279–283, 1985.
- [36] U. Höhle, "Entropy with respect to plausibility measures," in *Proceedings of the 12th IEEE International Symposium on Multiple-Valued Logic*, Paris, France, May 1982, pp. 167–169.
- [37] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *International Journal of General Systems*, vol. 9, no. 4, pp. 249–260, 1983.
- [38] D. Harmanec and G. J. Klir, "Measuring total uncertainty in Dempster-Shafer theory: a novel approach," *International Journal of General Systems*, vol. 22, no. 4, pp. 405–419, 1994.

- [39] J. Abellán and A. Masegosa, "Requirements for total uncertainty measures in Dempster–Shafer theory of evidence," *International Journal of General Systems*, vol. 37, no. 6, pp. 733–747, 2008.
- [40] A.-L. Jousselme, C. Liu, D. Grenier, and É. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 36, no. 5, pp. 890–903, 2006.
- [41] G. J. Klir and H. W. Lewis, "Remarks on "measuring ambiguity in the evidence theory"," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 38, no. 4, pp. 995–999, 2008.
- [42] F. Smarandache, D. Han, and A. Martin, "Comparative study of contradiction measures in the theory of belief functions," in *Proceedings of the 15th International Conference on Information Fusion*, Singapore, July 2012, pp. 271–277.



# Decision-Making with Belief Interval Distance

Jean Dezert<sup>a</sup>, Deqiang Han<sup>b</sup>, Jean-Marc Tacnet<sup>c</sup>, Simon Carladous<sup>c,d,e</sup>, Yi Yang<sup>f</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

<sup>c</sup>UGA, Irstea, UR ETGR, 2 rue de la Papeterie-BP 76, F-38402 St-Martin-d'Hères, France.

<sup>d</sup>AgroParisTech, 19 avenue du Maine, 75732 Paris, France.

<sup>e</sup>ENSMSE - DEMO, 29, rue Ponchardier 42100 Saint-Etienne, France.

<sup>f</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong Univ, Xi'an, China 710049.

Emails: jean.dezert@onera.fr, deqhan@xjtu.edu.cn, jean-marc.tacnet@irstea.fr,  
simon.carladous@irstea.fr, jiafeiyi@mail.xjtu.edu.cn

Originally published as: J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-Making with Belief Interval Distance*, in Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21–23, 2016, and reprinted with permission.

**Abstract**—In this paper we propose a new general method for decision-making under uncertainty based on the belief interval distance. We show through several simple illustrative examples how this method works and its ability to provide reasonable results.

**Keywords:** belief functions, decision-making, distance between BBAs.

## I. INTRODUCTION

Dempster-Shafer Theory (DST), also known as the Mathematical Theory of Evidence or the Theory of Belief Functions (BF), was introduced by Shafer in 1976 [1] based on Dempster's previous works [2]. This theory offers an elegant theoretical framework for modeling uncertainty, and provides a method for combining distinct bodies of evidence collected from different sources. In the past more than three decades, DST has been used in many applications, in fields including information fusion, pattern recognition, and decision making [3]. Although belief functions are very appealing for modeling epistemic uncertainty, the two main important questions related to them remain still open:

- 1) How to combine efficiently several independent belief functions?

This open question is out of the scope of this paper and it has been widely disputed by many experts [4]–[14]. In this short paper, we focus on the second question below.

- 2) How to take a final decision from a belief function?

This second question is also very crucial in many problems involving epistemic uncertainty where the final step (after beliefs elicitation, and beliefs combination) is to make a decision.

In the sequel, we assume that the reader is familiar with Dempster-Shafer Theory of belief functions [1] and its notations. Due to space restriction, we will not recall the definitions of basic belief assignment  $m(\cdot)$ , belief  $Bel(\cdot)$  (also called credibility by some authors), and plausibility functions  $Pl(\cdot)$  functions defined over a given finite discrete frame of discernment (FoD)  $\Theta$ . For any focal element  $X$  of the powerset of  $\Theta$ , denoted by  $2^\Theta$ , the interval  $BI(X) \triangleq [Bel(X), Pl(X)]$

is called the belief interval of  $X$ . Its length  $Pl(X) - Bel(X)$  characterizes the uncertainty on  $X$  (also called ambiguity in [15]). This paper is organized as follows. In section 2, we recall the common decision-making techniques used so far to make a decision from belief functions. In section 3 we recall the new distance measure based on Belief interval, and we present a new general method for decision-making with belief functions. Finally, examples of this new approach are given in section 4, with concluding remarks in section 5.

## II. CLASSICAL DECISION-MAKING USING BF

We assume a given FoD  $\Theta = \{\theta_1, \dots, \theta_n\}$  and a given BBA  $m(\cdot)$  defined on  $2^\Theta$ . We want to make a decision from  $m(\cdot)$ . It consists in choosing a particular element of the FoD that solves the problem under consideration, which is represented by the set of potential solutions (choices)  $\theta_i$ ,  $i = 1, \dots, n$ . How to do this in an effective manner is the fundamental question of decision-making under epistemic uncertainty. Many decision-making criteria have been proposed in the literature. Some advanced techniques developed in the 1990s [15]–[19] have not been widely used so far in the BF community, probably because of their complexity of implementation. In this section, we only present briefly the simplest ones frequently used.

- 1) Decision based on maximum of credibility:

This decision-making scheme is the so-called prudent (or pessimistic) scheme. It consists in choosing the element of the FoD  $\Theta$  that has the maximum of credibility. In other terms, one will decide  $\hat{\theta} = \theta_{i^*}$  with<sup>1</sup>

$$\theta_{i^*} = \arg \max_i Bel(\theta_i). \quad (1)$$

- 2) Decision based on maximum of plausibility:

On the contrary, if we prefer to adopt a more optimistic decision-making (less prudent) attitude, one will choose

<sup>1</sup>The notation with hat indicates the decision taken. Here  $\hat{\theta}$  specifies that the decision taken is only a singleton of  $\Theta$ .

the element of the FoD  $\Theta$  that has the maximum of plausibility. In other terms, one will decide  $\hat{\theta} = \theta_{i^*}$  with

$$\theta_{i^*} = \arg \max_i Pl(\theta_i). \quad (2)$$

### 3) Decision based on maximum of probability:

Usually decision-makers prefer to adopt a more balanced decisional attitude making a compromise between the aforementioned pessimistic and optimistic attitudes. For this, the BBA  $m(\cdot)$  is transformed into a subjective probability measure  $P(\cdot)$  compatible with the belief interval  $[Bel(\cdot), Pl(\cdot)]$ , and one will choose the element of the FoD  $\Theta$  that has the maximum of probability. In other terms, one will decide  $\hat{\theta} = \theta_{i^*}$  with

$$\theta_{i^*} = \arg \max_i P(\theta_i). \quad (3)$$

In practice, many probabilistic transformations are available to approximate (or transform) a BBA  $m(\cdot)$  in a probability measure  $P(\cdot)$ . By example, the pignistic transformation [20], the plausibility transformation [21], the DSMP transformation and other ones presented in [22], etc.

Of course, in case of multiple maximum values, no decision can be clearly drawn. Usually if only one decision must be made, a random sample between elements  $\theta_i$  generating the maximal decision-making criterion value is used to make a unique final decision  $\hat{\theta}$ . Another more prudent decision scheme is to use the disjunction of all elements generating the maximal decision-making criterion value, to provide a less specific final decision (if it is allowed for the problem under concern).

Our main criticism about using these decision-making schemes is that they do not use the whole information contained in the original BBA, which is in fact expressed by the whole belief interval. The pessimistic attitude uses only the credibility values, whereas the optimistic attitude uses only the plausibility values. The prudent attitude based on the criteria (3) requires a particular choice of probabilistic transformation which is often disputed by users. Making a decision from the  $P(\cdot)$  measure is theoretically not satisfactory at all because the transformation is lossy since we cannot retrieve  $m(\cdot)$  from  $P(\cdot)$  when some focal elements of  $m(\cdot)$  are not singletons. In the next section, we propose a better justified decision scheme based on the belief interval distance [23], [24].

### III. DECISION-MAKING USING BELIEF INTERVAL DISTANCE

In our previous works [23], [24], we have defined a Euclidean belief interval distance between two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined on the powerset of a given FoD  $\Theta = \{\theta_1, \dots, \theta_n\}$  as follows

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))}, \quad (4)$$

where  $N_c = 1/2^{n-1}$  is a normalization factor to have  $d_{BI}(m_1, m_2) \in [0, 1]$ , and  $d_W(BI_1(X), BI_2(X))$

is the Wasserstein's distance [25] between belief intervals  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)] = [a_1, b_1]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)] = [a_2, b_2]$ . More specifically,

$$d_W([a_1, b_1], [a_2, b_2]) \triangleq \left[ \left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2 \right]^{\frac{1}{2}}. \quad (5)$$

In [23], we have proved that  $d_{BI}(x, y)$  is a true distance metric because it satisfies the properties of non-negativity ( $d(x, y) \geq 0$ ), non-degeneracy ( $d(x, y) = 0 \Leftrightarrow x = y$ ), symmetry ( $d(x, y) = d(y, x)$ ), and the triangle inequality ( $d(x, y) + d(y, z) \geq d(x, z)$ ), for any BBAs  $x, y$  and  $z$  defined on  $2^\Theta$ . The choice of Wasserstein's distance in  $d_{BI}$  definition is justified by the fact that Wasserstein's distance is a true distance metric and it fits well with our needs because we have to compute a distance between  $[Bel_1(X), Pl_1(X)]$  and  $[Bel_2(X), Pl_2(X)]$ .

For notation convenience, we denote  $m_X$  the categorical BBA having only  $X$  as focal element, where  $X \neq \emptyset$  is an element of the powerset of  $\Theta$ . More precisely,  $m_X$  is the particular (categorical) BBA defined by  $m_X(X) = 1$  and  $m_X(Y) = 0$  for any  $Y \neq X$ . Such basic BBA plays an important role in our new decision scheme because its corresponding belief interval reduces to the degenerate interval  $[1, 1]$  which represents the certainty on  $X$ . The basic principle of the new decision scheme we propose is very simple and intuitively makes sense. It consists in selecting as the final decision (denoted by  $\hat{X}$ ) the element of the powerset for which the belief interval distance between the BBA  $m(\cdot)$  and  $m_X$ ,  $X \in 2^\Theta \setminus \{\emptyset\}$  is the smallest one<sup>1</sup>. Therefore, take as the final decision  $\hat{X}$  given by

$$\hat{X} = \arg \min_{X \in 2^\Theta \setminus \{\emptyset\}} d_{BI}(m, m_X), \quad (6)$$

where  $d_{BI}(m, m_X)$  is computed according to (4).  $m(\cdot)$  is the BBA under test and  $m_X(\cdot)$  the categorical BBA focused on  $X$  defined above.

This decision scheme is very general in the sense that the decision making can be done on any type of element<sup>2</sup> of the power-set  $2^\Theta$ , and not necessarily only on the elements (singletons) of the FoD (see examples in the next section). This method not only provides the final decision  $\hat{X}$  to make, but also it evaluates how good this decision is with respect to its alternatives if we define the quality indicator  $q(\hat{X})$  as follows

$$q(\hat{X}) \triangleq 1 - \frac{d_{BI}(m, m_{\hat{X}})}{\sum_{X \in 2^\Theta \setminus \{\emptyset\}} d_{BI}(m, m_X)}. \quad (7)$$

One sees that the quality indicator  $q(\hat{X})$  of the decision  $\hat{X}$  made will become maximum (equal to one) when the distance

<sup>1</sup>This simple principle has also been proposed by Essaid et al. [26] using Jousselme's distance.

<sup>2</sup>empty set excluded.

between the BBA  $m(\cdot)$  and  $m_{\hat{X}}$  is zero, which means that the BBA  $m(\cdot)$  is focused in fact only on the element  $\hat{X}$ . The higher  $q(\hat{X})$  is, the more confident in the decision  $\hat{X}$  we should be.

Of course, if a decision must be made with some extra constraint<sup>3</sup> defined by a (or several) condition(s), denoted  $c(X)$ , then we must take into account  $c(X)$  in Eq. (6), that is  $\hat{X} = \arg \min_{X \in 2^\Theta \setminus \{\emptyset\} \text{ s.t. } c(X)} d_{BI}(m, m_X)$ , and also in the derivation of quality indicator by taking  $\sum_{X \in 2^\Theta \setminus \{\emptyset\} \text{ s.t. } c(X)} d_{BI}(m, m_X)$  as denominator in (7).

Theoretically any other strict distance metric, for instance Jousselme's distance [27]–[29], could be used instead of  $d_{BI}(\cdot, \cdot)$ . We have chosen  $d_{BI}$  distance because of its ability to provide good and reasonable behavior [23] as will be shown. When there exists a tie between multiple decisions  $\{\hat{X}_j, j > 1\}$ , then the prudent decision corresponding to their disjunction  $\hat{X} = \cup_j \hat{X}_j$  should be preferred (if allowed), otherwise the final decision  $\hat{X}$  is made by a random selection of elements  $\hat{X}_j$ .

#### IV. EXAMPLES AND COMPARISONS

In this section we present several examples when the cardinality of the FoD  $|\Theta|$  is only 2 and 3 because it is easier to see whether the decision-making results make sense or not. We compare and discuss decisions only made with the belief interval distance  $d_{BI}$  and Jousselme's distance  $d_J$  because the other lossy decision schemes do not exploit both credibility and plausibility values. The examples corresponding to cases where the BBA  $m(\cdot)$  is focused on a single element  $X$  of  $2^\Theta$  are not presented because one trivially gets  $\hat{X} = X$  using either  $d_{BI}$  or  $d_J$  distances.

The next tables present several BBAs from which a decision has to be made. By convention, and since we work with normal BBAs satisfying  $m(\emptyset) = 0$ , the empty set is not included in the tables. The rows for  $d_{BI}^{\min}(m_i, m_X)$  and for  $d_J^{\min}(m_i, m_X)$  list the minimal values obtained for  $d_{BI}(m_i, m_X)$  and  $d_J(m_i, m_X)$ . The rows for  $\hat{X}^{d_{BI}}$  and for  $\hat{X}^{d_J}$  list the decision(s)  $\hat{X}$  made when using  $d_{BI}(m_i, m_X)$  and  $d_J(m_i, m_X)$  respectively. The rows for  $q(\hat{X}^{d_{BI}})$  and  $q(\hat{X}^{d_J})$  list the quality indicators of decision(s) made using  $d_{BI}(m_i, m_X)$  and  $d_J(m_i, m_X)$  respectively. Depending on the BBA, it is possible to have multiple decisions  $\{\hat{X}_j\}$  in case of a tie. If a tie occurs either a random sampling of  $\{\hat{X}_j\}$  must be drawn, or (if allowed) the disjunction of decisions  $\hat{X}_j$  is preferred. In the next subsections, we present results in free-constraint case (i.e.  $c(X) = \emptyset$ ), as well as when the decisions are restricted to be singletons (i.e.  $c(X) \equiv " |X| = 1 "$ ).

##### A. Examples with $\Theta = \{A, B\}$

Table I shows the decisions made when there is no constraint on the cardinality of the decision  $\hat{X}$ .

One sees that methods based on min of  $d_{BI}(m, m_X)$  and on min of  $d_J(m, m_X)$  yield the same reasonable decisions

in almost all cases. With  $m_2$ , one has multiple decisions  $\hat{X}^{d_J} = \{A, B, A \cup B\}$  with quality 0.6667 when using  $d_J$ , which is a bit surprising in our opinion because there is a real tie between  $A$  and  $B$ . Consequently, the decision  $A \cup B$  should be preferred when there is no constraint on the cardinality of decisions. For this  $m_2$  case, one gets a unique decision  $\hat{X}^{d_{BI}} = A \cup B$  with a better quality 0.776 which seems more reasonable. We see also that all minimal distance values obtained with  $d_{BI}$  are less (or equal in case  $m_1$ ) to the minimal values obtained with  $d_J$ . In fact, when the mass function is distributed symmetrically, it is naturally expected that no real decision can be easily taken (as illustrated for BBA's  $m_2(\cdot)$  and  $m_5(\cdot)$  in Table I). Here, the decision  $A \cup B$  for BBA's  $m_2(\cdot)$  and  $m_5(\cdot)$  can be interpreted as a no proper decision, in the sense that  $A \cup B$  is the whole universe of discourse, hence we are merely selecting anything (and discarding nothing). Such kind of no proper decision may however be very helpful in some fusion systems because it warns that input information is not rich enough, and that one needs more information to take a proper decision (by including more sensors or more experts reports in the system for instance). For symmetrical mass function, the decision drawn from the new proposed decision rule is consistent with what we can reasonably get because. To make a proper decision we will always need to introduce some possibly arbitrary additional constraints.

Table II shows the decisions made for same examples when we force the decision to be a singleton, that is when the constraint is  $c(X) \equiv " |X| = 1 "$ . One sees that the decisions restricted to the set of singletons using  $d_{BI}(m, m_X)$  or  $d_J(m, m_X)$  are the same but the quality indicators are a bit better when using  $d_{BI}(m, m_X)$  with respect to  $d_J(m, m_X)$ . The values of the quality indicators in Table II are different to those of Table I which is normal because we use the constraint  $c(X)$  in the denominator of the formula (7).

##### B. Examples with $\Theta = \{A, B, C\}$

Table III shows the decisions made when there is no constraint on the cardinality of the decision  $\hat{X}$ , whereas Table IV shows the results for the same examples when the decisions made are restricted to singletons. As shown in the tables all minimal distance values obtained with  $d_{BI}$  are less (or equal) to the minimal values obtained with  $d_J$  and the quality indicator decisions is better when computed with  $d_{BI}$  (except in case  $m_1$  of Table III). The decisions results obtained with  $d_J$  are mostly consistent with those obtained with  $d_{BI}$  (except in case  $m_2$  and  $m_3$  of Table III) where a larger set of decisions (tie) is obtained using  $d_J$ .

If the decisions are restricted to singletons (see Table IV), then the decision-making based on  $d_{BI}$  and on  $d_J$  provides the same results with a better quality of decisions using  $d_{BI}$ .

<sup>3</sup>for instance, making a choice only among the singletons of  $2^\Theta$ .

Table I  
EXAMPLES OF SEVERAL BBA'S AND DECISIONS MADE (NO CONSTRAINT CASE).

$X \in 2^\Theta$	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$	$m_4(\cdot)$	$m_5(\cdot)$	$m_6(\cdot)$	$m_7(\cdot)$
A	0.9	0.5	0.8	0.1	0.4	0.9	0.1
B	0.1	0.5	0.1	0.1	0.4	0	0
$A \cup B$	0	0	0.1	0.8	0.2	0.1	0.9
$d_{BI}^{\min}(m_i, m_X)$	0.1000	0.2887	0.1528	0.0577	0.2309	0.0577	0.0577
$q(\hat{X}^{d_{BI}})$	0.9330	0.7760	0.8939	0.9502	0.8134	0.9622	0.9513
$\hat{X}^{d_{BI}}$	A	$A \cup B$	A	$A \cup B$	$A \cup B$	A	$A \cup B$
$d_J^{\min}(m_i, m_X)$	0.1000	0.5000	0.1581	0.1000	0.4000	0.0707	0.0707
$q(\hat{X}^{d_J})$	0.9390	0.6667	0.8999	0.9276	0.6409	0.9574	0.9501
$\hat{X}^{d_J}$	A	A, B, $A \cup B$	A	$A \cup B$	$A \cup B$	A	$A \cup B$

Table II  
EXAMPLES OF SEVERAL BBA'S AND DECISIONS MADE (RESTRICTED TO SINGLETONS).

$X \in 2^\Theta$	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$	$m_4(\cdot)$	$m_5(\cdot)$	$m_6(\cdot)$	$m_7(\cdot)$
A	0.9	0.5	0.8	0.1	0.4	0.9	0.1
B	0.1	0.5	0.1	0.1	0.4	0	0
$A \cup B$	0	0	0.1	0.8	0.2	0.1	0.9
$d_{BI}^{\min}(m_i, m_X)$	0.1000	0.5000	0.1528	0.5508	0.5033	0.0577	0.5196
$q(\hat{X}^{d_{BI}})$	0.9000	0.5000	0.8477	0.5000	0.5000	0.9427	0.5393
$\hat{X}^{d_{BI}}$	A	A, B	A	A, B	A, B	A	A
$d_J^{\min}(m_i, m_X)$	0.1000	0.5000	0.1581	0.6403	0.5099	0.0707	0.6364
$q(\hat{X}^{d_J})$	0.9000	0.5000	0.8434	0.5000	0.5000	0.9308	0.5276
$\hat{X}^{d_J}$	A	A, B	A	A, B	A, B	A	A

Table III  
EXAMPLES OF SEVERAL BBA'S AND DECISIONS MADE (NO CONSTRAINT CASE).

$X \in 2^\Theta$	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$	$m_4(\cdot)$	$m_5(\cdot)$	$m_6(\cdot)$	$m_7(\cdot)$	$m_8(\cdot)$
A	0.9	0.5	1/3	0.5	0	0	0	0.2
B	0.1	0.5	1/3	0	0	0	0	0.1
$A \cup B$	0	0	0	0.5	0.5	2/3	1/3	0.05
C	0	0	1/3	0	0	0	0	0.05
$A \cup C$	0	0	0	0	0	0	1/3	0.1
$B \cup C$	0	0	0	0	0.5	1/3	1/3	0.2
$A \cup B \cup C$	0	0	0	0	0	0	0	0.3
$d_{BI}^{\min}(m_i, m_X)$	0.1000	0.2887	0.4082	0.2887	0.2887	0.1925	0.2357	0.2227
$q(\hat{X}^{d_{BI}})$	0.9776	0.9242	0.8787	0.9271	0.9120	0.9421	0.9241	0.9280
$\hat{X}^{d_{BI}}$	A	$A \cup B$	$2^\Theta \setminus \{\emptyset, A, B, C\}$	A, $A \cup B$	$A \cup B,$ $B \cup C, \emptyset$	$A \cup B$	$A \cup B \cup C$	$A \cup B \cup C$
$d_J^{\min}(m_i, m_X)$	0.1000	0.5000	0.5774	0.3536	0.4082	0.2722	0.3333	0.3149
$q(\hat{X}^{d_J})$	0.9798	0.8870	0.8571	0.9225	0.8989	0.9337	0.9111	0.9152
$\hat{X}^{d_J}$	A	A, B, $A \cup B$	$2^\Theta \setminus \{\emptyset\}$	A, $A \cup B$	$A \cup B,$ $B \cup C, \emptyset$	$A \cup B$	$A \cup B \cup C$	$A \cup B \cup C$

Table IV  
EXAMPLES OF SEVERAL BBA'S AND DECISIONS MADE (RESTRICTED TO SINGLETONS).

$X \in 2^\Theta$	$m_1(\cdot)$	$m_2(\cdot)$	$m_3(\cdot)$	$m_4(\cdot)$	$m_5(\cdot)$	$m_6(\cdot)$	$m_7(\cdot)$	$m_8(\cdot)$
A	0.9	0.5	1/3	0.5	0	0	0	0.2
B	0.1	0.5	1/3	0	0	0	0	0.1
$A \cup B$	0	0	0	0.5	0.5	2/3	1/3	0.05
C	0	0	1/3	0	0	0	0	0.05
$A \cup C$	0	0	0	0	0	0	1/3	0.1
$B \cup C$	0	0	0	0	0.5	1/3	1/3	0.2
$A \cup B \cup C$	0	0	0	0	0	0	0	0.3
$d_{BI}^{\min}(m_i, m_X)$	0.1000	0.5000	0.5774	0.2887	0.5000	0.5092	0.6236	0.5770
$q(\hat{X}^{d_{BI}})$	0.9488	0.7321	0.6667	0.8531	0.7388	0.7364	0.6667	0.6855
$\hat{X}^{d_{BI}}$	A	A, B	A, B, C	A	B	B	A, B, C	A
$d_J^{\min}(m_i, m_X)$	0.1000	0.5000	0.5774	0.3536	0.5774	0.5932	0.6667	0.6117
$q(\hat{X}^{d_J})$	0.9488	0.7321	0.6667	0.8300	0.7257	0.7229	0.6667	0.6836
$\hat{X}^{d_J}$	A	A, B	A, B, C	A	B	B	A, B, C	A

## V. CONCLUSIONS

We have presented a new method for decision-making with belief functions which truly exploits the belief interval value of each focal element of a BBA. It is easy to implement and can be applied with any strict distance metric between two BBAs. We have considered and compared the well-known Jousselme's distance and the recent belief interval distance. This method is general because the decision can be made not only on singletons, but also on any other compound focal elements (if needed and allowed). It also provides a quality indicator of the decision made.

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [2] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. Math. Statist., Vol. 38, pp. 325–339, 1967.
- [3] P. Smets, *Practical uses of belief functions*, In: K. B. Laskey and H. Prade, Editors, *Uncertainty in Artificial Intelligence*, Vol. 15, pp. 612–621, Stockholm, Sweden, 1999.
- [4] L.A. Zadeh, *On the validity of Dempster's rule of combination*, Memo M79/24, Univ. of California, Berkeley, U.S.A., 1979.
- [5] J. Lemmer, *Confidence factors, empiricism and the Dempster-Shafer theory of evidence*, In Proc. of 1st Conf. on Uncertainty in Artificial Intelligence (UAI-85), pp. 160–176, 1985.
- [6] J. Pearl, *Reasoning with belief functions: An analysis of compatibility*, IJAR, Vol. 4, pp. 363–389, 1990, (with Rejoinder in IJAR, Vol. 6, pp. 425–443, 1992).
- [7] F. Voorbraak, *On the justification of Dempster's rule of combination*, Dept. of Philosophy, Univ. of Utrecht, The Netherlands, Logic Group Preprint Series, No. 42, 1988.
- [8] P. Wang, *A defect in Dempster-Shafer theory*, in Proc. of 10th Conf. on Uncertainty in AI, pp. 560–566, 1994.
- [9] A. Gelman, *The boxer, the wrestler, and the coin flip: a paradox of robust Bayesian inference and belief functions*, American Statistician, Vol. 60, No. 2, pp. 146–150, 2006.
- [10] J. Dezert, P. Wang, A. Tchamova, *On The Validity of Dempster-Shafer Theory*, in Proc. of Int. Conf. on Information Fusion (Fusion 2012), Singapore, 2012.
- [11] A. Tchamova, J. Dezert, *On the Behavior of Dempster's Rule of Combination and the Foundations of Dempster-Shafer Theory*, in Proc. of IEEE IS2012 Conf., Sofia, Bulgaria, 2012.
- [12] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's rule doesn't behave as Bayes rule with informative priors*, in Proc. of 2013 IEEE INISTA 2013, Albena, Bulgaria, 2013.
- [13] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, IJIS, Vol. 29, Issue 3, pp. 223–252, 2014.
- [14] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information fusion (Collected Works)*, Volumes 1–4, American Research Press, USA 2004–2015. <https://www.onera.fr/fr/staff/jean-dezert/references>
- [15] R.P. Srivastava, *Decision Making Under Ambiguity: A Belief-Function Perspective*, Archives of Control Sciences, Vol. 6 (XLII), No. 1–2, pp. 5–27, 1997.
- [16] H.T. Nguyen, E.A. Walker, *On Decision Making Using Belief Functions*, in *Advances in the Dempster-Shafer Theory of Evidence*. Edited by R.R. Yager, M. Fedrizzi, and J. Kacprzyk, John Wiley & Sons, New York, NY, USA, 1994.
- [17] J.-Y. Jaffray, *Utility Theory for Belief Functions*, Oper. Res. Letters, Vol. 8, pp. 107–112, 1989.
- [18] T.M. Strat, *Decision Analysis Using Belief Functions*, IJAR, Vol. 4, No. 5, p. 6, 1990.
- [19] R.R. Yager, *Decision making under Dempster-Shafer uncertainties*, Int. J. General Systems, Vol. 20, pp. 233–245, 1992.
- [20] P. Smets, R. Kennes, *The transferable belief model*, Artif. Int., Vol. 66, pp. 191–234, 1994.
- [21] B.R. Cobb, P.P. Shenoy, *On the plausibility transformation method for translating belief function models to probability models*, IJAR, Vol. 41, No. 3, pp. 314–330, 2006.
- [22] J. Dezert, F. Smarandache, *A new probabilistic transformation of belief mass assignment*, Proc. of Int. Conf. on Information Fusion (Fusion 2008), Cologne, Germany, 2008.
- [23] D. Han, J. Dezert, Y. Yang, *New Distance Measures of Evidence based on Belief Intervals*, in Proc. of Belief 2014, Oxford, UK, 2014.
- [24] D. Han, J. Dezert, Y. Yang, *Belief Interval Based Distance Measures in the Theory of Belief Functions*, IEEE Trans. on Systems, Man, and Cybernetics: Systems, Vol. 48(6), pp. 833–850, June 2018.
- [25] A. Irpino, R. Verde, *Dynamic Clustering of Interval Data Using a Wasserstein-based Distance*, Pattern Recognition Letters, Vol. 29, pp. 1648–1658, 2008.
- [26] A. Essaid, A. Martin, G. Smits, B. Ben Yaghlane, *A distance-based decision in the credal level*, in Proc. of Int. Conf. AISC 2014, Sevilla, Spain, 2014.
- [27] A.-L. Jousselme, D. Grenier, É. Bossé, *A new distance between two bodies of evidence*, Information Fusion, Vol. 2, No. 2, pp. 91–101, 2001.
- [28] A.-L. Jousselme, P. Maupin, *Distances in evidence theory: Comprehensive survey and generalizations*, IJAR, Vol. 53, No. 2, pp. 118–145, 2012.
- [29] M. Bouchard, A.-L. Jousselme, P.-E. Doré, *A proof for the positive definiteness of the Jaccard index matrix*, IJAR, Vol. 54, pp. 615–626, 2013.



# The BF-TOPSIS Approach for Solving Non-Classical MCDM Problems

Jean Dezert<sup>a</sup>, Deqiang Han<sup>b</sup>, Jean-Marc Tacnet<sup>c</sup>, Simon Carladous<sup>c</sup>, Hanlin Yin<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>b</sup>CIESR, Xi'an Jiaotong Univ, Xi'an, China.

<sup>c</sup>UGA, Irstea, UR ETGR, 2 rue de la Papeterie, St-Martin-d'Hères, France.

Emails: jean.dezert@onera.fr, deqhan@xjtu.edu.cn, jean-marc.tacnet@irstea.fr,  
simon.carladous@irstea.fr, iverlon1987@stu.xjtu.edu.cn

Originally published as: J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, H. Yin, *The BF-TOPSIS Approach for Solving Non-Classical MCDM Problems*, in Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21–23, 2016, and reprinted with permission.

**Abstract**—In this paper we show how the Belief-Function based Technique for Order Preference by Similarity to Ideal Solution (BF-TOPSIS) approach can be used for solving non-classical multi-criteria decision-making (MCDM) problems. We give simple examples to illustrate our presentation.

**Keywords:** multi-criteria decision-making, belief functions, TOPSIS, BF-TOPSIS.

## I. INTRODUCTION

Classical Multi-Criteria Decision-Making (MCDM) consists in choosing an alternative among a known set of alternatives based on their quantitative evaluations (numerical scores) obtained with respect to different criteria. A typical example could be the selection of a car to buy among a given set of cars based on different criteria (cost, engine robustness, fuel economy, CO<sub>2</sub> emission, etc). The classical MCDM problem, although easily formulated, have no solution at all in general due to the fact that no alternative exists that optimizes all criteria jointly. Thus MCDM problems are generally not *solved*, but a decision is found by means of ranking, compromises etc. The difficulty of MCDM problems is also because the scores are usually expressed in different (physical) units with different scales which generally necessitates an ad-hoc choice of a normalization step that may lead to rank reversal.

Many MCDM methods have been developed, like AHP<sup>1</sup> [1] and its extensions [2]–[6], ELECTRE<sup>2</sup> [7], TOPSIS<sup>3</sup> [8], [9] which are widely used in applications. They have already been extended in the belief function framework in our previous works [2], [10], [11] to take into account epistemic uncertainty, missing scores<sup>7</sup> values as well as conflicting information between sources<sup>4</sup>. In this work, we show how the BF-TOPSIS methods proposed recently in [11] (with application in [12]), can be directly used for solving also non-classical multicriteria decision-making problems where not only alternatives are scored (with possibly missing values), but also any element of the power set of alternatives.

<sup>1</sup>Analytic hierarchy process.

<sup>2</sup>Elimination and choice translating reality.

<sup>3</sup>Technique for order preference by similarity to ideal solution.

<sup>4</sup>In the MCDM context, a source of information consists in the list of scores values of alternatives related to a given criterion.

In the sequel, we assume the reader to be familiar with the theory of belief functions [13] and its definitions and notations, mainly the basic belief assignment (BBA)  $m(\cdot)$ , the belief function  $Bel(\cdot)$  and the plausibility function  $Pl(\cdot)$  defined with respect to a discrete finite frame of discernment (FoD).

## II. NON-CLASSICAL MCDM PROBLEM FORMULATION

We consider a given set of alternatives denoted by  $\mathcal{A} \triangleq \{A_1, A_2, \dots, A_M\}$  ( $M > 2$ ) representing the FoD of our problem under consideration, and we denote by  $2^{\mathcal{A}}$  the power set<sup>5</sup> of  $\mathcal{A}$ . In our approach, we work with Shafer's classical model of FoD and we do not allow the empty set to be a focal element<sup>6</sup> because in our opinion it does not make sense to compare an alternative with respect to the empty set from the decision-making standpoint. The cardinality of the (non empty) elements of the power set varies from 1 to  $2^M - 1$ . We also consider a given set of criteria  $\mathcal{C} \triangleq \{C_1, C_2, \dots, C_N\}$  ( $N \geq 1$ ), where each criterion  $C_j$  is characterized by a relative importance weighting factor  $w_j \in [0, 1]$ ,  $j = 1, \dots, N$  such that  $\sum_{j=1}^N w_j = 1$ . The set of normalized weighting factors is denoted by  $\mathbf{w} = \{w_1, w_2, \dots, w_N\}$ . The score<sup>7</sup> value is a number  $S_{ij} = S_j(X_i)$  related to the evaluation of an element  $X_i \in 2^{\mathcal{A}} \setminus \{\emptyset\}$  from a given criterion  $C_j$ . If the score value  $S_j(X_i)$  is not available (or missing), we denote it by the “varnothing” symbol  $\emptyset$ . The non-classical MCDM problem can be formulated as follows in the worst case (i.e. when scores apply to all elements of  $2^{\mathcal{A}}$ ): given the  $(2^M - 1) \times N$  score matrix  $\mathbf{S} = [S_j(X_i)]$  whose elements take either a numerical value or a  $\emptyset$  value (if the value is not available) and knowing the set  $\mathbf{w}$  of the relative importance weights of criteria, how to rank the elements of  $2^{\mathcal{A}} \setminus \{\emptyset\}$  to make the final decision?

<sup>5</sup>The power set  $2^{\mathcal{A}}$  is the set of all subsets of  $\mathcal{A}$ , empty set  $\emptyset$  and  $\mathcal{A}$  included.

<sup>6</sup>as proposed in Smets Transferable Belief Model for instance.

<sup>7</sup>Depending on the context, the score can be interpreted either as a cost/expense or as a reward/benefit. In the sequel, by convention and without loss of generality, we will interpret the score as a reward having monotonically increasing preference. Thus, the best alternative with respect to a given criterion will be the one providing the highest reward/benefit.

**Example:** Let us consider the ranking of five students  $A_1$ ,  $A_2$ ,  $A_3$ ,  $A_4$ , and  $A_5$  based on two criteria  $C_1$  and  $C_2$ . The criterion  $C_1$  is their long jump performance (in meters), and the criteria  $C_2$  is a realization of a small project to collect funds (in euros) to help a bigger nature protection project. Highest scores values mean better results in this particular context. Let us assume that students were allowed to realize their project in joint collaboration (no more than three students are allowed in a group), or alone. At the end term of the project, suppose that one has the two following evaluations (scoring)

$$\mathbf{S}_{C_1} = \begin{matrix} & C_1 \\ A_1 & \begin{bmatrix} 3.7 m \\ 3.6 m \\ 3.8 m \\ 3.7 m \end{bmatrix} \end{matrix} \quad \text{and} \quad \mathbf{S}_{C_2} = \begin{matrix} & C_2 \\ A_5 & \begin{bmatrix} 640 \text{€} \\ 600 \text{€} \\ 650 \text{€} \end{bmatrix} \\ A_1 \cup A_2 & \\ A_3 \cup A_4 & \end{matrix} \quad (1)$$

The scores' values listed in  $\mathbf{S}_{C_1}$  indicate in fact that the student  $A_2$  has not been able to pass the long jump test for some reason (medical, familial or whatever), so his score is missing. The scores' values listed in  $\mathbf{S}_{C_2}$  indicate that  $A_5$  did choose to realize his project alone with a pretty good performance, and the project realized by the collaboration of students  $A_3$  with  $A_4$  has obtained the best performance (the highest amount of collected funds). In this very simple example, one sees that the score evaluation can be done not only on single alternatives (as for criterion  $C_1$ ) but also on a subset of elements of  $2^A$  (as for criterion  $C_2$ ). All the elements having a score are called scoring focal elements. In general, these focal elements can differ from one criterion  $C_j$  to another criterion  $C_k$  for  $k \neq j$  and the score matrix cannot be built by a simple (horizontal) stacking of scoring lists. In general, one must identify all focal elements of each scoring list to determine the minimum number of rows necessary to define the scoring matrix. As mentioned, we use the symbol  $\emptyset$  to identify all values that are missing in the scoring matrix. Note that we do not set missing values to zero number (or any other chosen number) to make explicit distinction between the known precise numerical value zero and a missing value. In this example, the scoring matrix will be defined as

$$\mathbf{S} = \begin{matrix} & C_1 & C_2 \\ A_1 & \begin{bmatrix} 3.7 m \\ 3.6 m \\ 3.8 m \\ 3.7 m \\ \emptyset \\ \emptyset \end{bmatrix} & \begin{bmatrix} \emptyset \\ \emptyset \\ \emptyset \\ 640 \text{€} \\ 600 \text{€} \\ 650 \text{€} \end{bmatrix} \\ A_3 & \\ A_4 & \\ A_5 & \\ A_1 \cup A_2 & \\ A_3 \cup A_4 & \end{matrix} \quad (2)$$

The question we want to address is how to rank the students based on such a kind of scoring information including disjunctions of alternatives and missing values, taking into account the relative importance weight of each criterion. Is it possible to solve such type of non-classical MCDM problems, and how?

### III. THE BF-TOPSIS APPROACH

The BF-TOPSIS approach has been proposed recently in [11] in a classical MCDM context where the focal elements of the scoring function  $S_j(\cdot)$  ( $j = 1, \dots, N$ ) are only the singletons  $A_i$  ( $i = 1, \dots, M$ ) of the frame of discernment  $\mathcal{A}$ . BF-TOPSIS is initially based on belief functions for MCDM support which exploits only the  $M \times N$  score matrix  $\mathbf{S} = [S_j(A_i)]$  and the relative importance weighting factors of criteria. The first main step of BF-TOPSIS is the construction of an  $M \times N$  BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$  from the score matrix  $\mathbf{S}$ , and then the combination of components of  $\mathbf{M}$  to make a final decision thanks to the Euclidean belief interval distance, denoted by  $d_{BI}$ , defined in [14], [15].

In fact, the BF-TOPSIS approach can also be directly applied to solve the non-classical MCDM problems because the belief interval  $[Bel_{ij}(X_i), Pl_{ij}(X_i)]$  of each proposition (i.e. each focal element which is not necessarily a singleton)  $X_i$  based on a criteria  $C_j$  can be established in a consistent manner<sup>8</sup> from the score matrix  $\mathbf{S} = [S_j(X_i)]$  as follows

$$[Bel_{ij}(X_i); Pl_{ij}(X_i)] \triangleq \left[ \frac{Sup_j(X_i)}{X_{\max}^j}; 1 - \frac{Inf_j(X_i)}{X_{\min}^j} \right] \quad (3)$$

where the  $Sup_j(X_i)$  and  $Inf_j(X_i)$  are computed from the score matrix  $\mathbf{S}$  by

$$Sup_j(X_i) \triangleq \sum_{Y \in 2^A | S_j(Y) \leq S_j(X_i)} |S_j(X_i) - S_j(Y)| \quad (4)$$

$$Inf_j(X_i) \triangleq - \sum_{Y \in 2^A | S_j(Y) \geq S_j(X_i)} |S_j(X_i) - S_j(Y)| \quad (5)$$

$Sup_j(X_i)$  is called the “positive support” of  $X_i$  because it measures how much  $X_i$  is better than other propositions according to criterion  $C_j$ , and  $Inf_j(X_i)$  is called the “negative support” of  $X_i$  because it measures how much  $X_i$  is worse than other propositions according to criterion  $C_j$ . The length of interval  $[0, Sup_j(X_i)]$  measures the support in favor of  $X_i$  as being the best proposition with respect to all other ones, and the length of  $[Inf_j(X_i), 0]$  measures the support against  $X_i$  based on the criterion  $C_j$ .

The denominators involved in (3), are defined by  $X_{\max}^j \triangleq \max_i Sup_j(X_i)$  and  $X_{\min}^j \triangleq \min_i Inf_j(X_i)$ , and they are supposed different from zero<sup>9</sup>. From the belief interval  $[Bel_{ij}(X_i); Pl_{ij}(X_i)]$ , we obtain the BBA  $m_{ij}(\cdot)$  defined by

$$m_{ij}(X_i) \triangleq Bel_{ij}(X_i) \quad (6)$$

$$m_{ij}(\bar{X}_i) \triangleq Bel_{ij}(\bar{X}_i) = 1 - Pl_{ij}(X_i) \quad (7)$$

$$m_{ij}(X_i \cup \bar{X}_i) \triangleq Pl_{ij}(X_i) - Bel_{ij}(X_i) \quad (8)$$

If a numerical value  $S_j(X_i)$  is missing in the score matrix  $\mathbf{S}$  (it is equal to  $\emptyset$ ), one chooses  $m_{ij}(\cdot)$  equals  $(0, 0, 1)$ , i.e., one takes a vacuous belief assignment. In [11], we have proposed four methods (called BF-TOPSIS1, ..., BFTOPSIS4) to make a decision from the BBA matrix  $\mathbf{M} = [m_{ij}(\cdot)]$ . Due to space

<sup>8</sup>Indeed,  $Bel_{ij}(X_i)$  and  $Bel_{ij}(\bar{X}_i)$  (where  $\bar{X}_i$  is the complement of  $X_i$  in the FoD  $\mathcal{A}$ ) belong to  $[0, 1]$ . They are consistent because the equality  $Pl_{ij}(X_i) = 1 - Bel_{ij}(\bar{X}_i)$  holds. The proof is similar to the one in [11].

restriction constraint, we just recall the principle of the BF-TOPSIS1 method because it is the simplest one. Applications of BFTOPSIS2–BFTOPSIS4 methods to non-classical MCDM problems is also possible without difficulty. The proposed transformation of score values to BBAs and basis of BF-TOPSIS method are theoretically justified in [11].

Before presenting succinctly the BF-TOPSIS1 method, we need to recall the definition of Belief Interval-based Euclidean distances  $d_{BI}(m_1, m_2)$  introduced in [14] between two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined on a same FoD  $\Theta$ . Mathematically,  $d_{BI}(m_1, m_2)$  is a true distance defined by [14]

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))} \quad (9)$$

where  $N_c = 1/2^{|\Theta|-1}$  is a normalization factor to have  $d_{BI}(m_1, m_2) \in [0, 1]$ , and  $d_W(BI_1(X), BI_2(X))$  is the Wasserstein distance [16] between belief intervals  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)] = [a_1, b_1]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)] = [a_2, b_2]$ . More specifically,

$$d_W([a_1, b_1], [a_2, b_2]) \triangleq \left( \left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2 \right)^{1/2} \quad (10)$$

**Principle of BF-TOPSIS1:** From the BBA matrix  $\mathbf{M}$  and for each proposition (focal element)  $X_i$ , one computes the Belief Interval-based Euclidean distances  $d_{BI}(m_{ij}, m_{ij}^{\text{best}})$  defined in (9) between the BBA  $m_{ij}(\cdot)$  and the ideal best BBA defined by  $m_{ij}^{\text{best}}(X_i) = 1$ , and the distance  $d_{BI}(m_{ij}, m_{ij}^{\text{worst}})$  between  $m_{ij}(\cdot)$  and the ideal worst BBA defined by  $m_{ij}^{\text{worst}}(\bar{X}_i) = 1$ .

Then, one computes the weighted average of  $d_{BI}(m_{ij}, m_{ij}^{\text{best}})$  values with relative importance weighting factor  $w_j$  of criteria  $C_j$ . Similarly, one computes the weighted average of  $d_{BI}(m_{ij}, m_{ij}^{\text{worst}})$  values. More specifically, one computes

$$d^{\text{best}}(X_i) \triangleq \sum_{j=1}^N w_j \cdot d_{BI}(m_{ij}, m_{ij}^{\text{best}}) \quad (11)$$

$$d^{\text{worst}}(X_i) \triangleq \sum_{j=1}^N w_j \cdot d_{BI}(m_{ij}, m_{ij}^{\text{worst}}) \quad (12)$$

The relative closeness of the proposition  $X_i$  with respect to ideal best solution  $X^{\text{best}}$  defined by

$$C(X_i, X^{\text{best}}) \triangleq \frac{d^{\text{worst}}(X_i)}{d^{\text{worst}}(X_i) + d^{\text{best}}(X_i)} \quad (13)$$

is used to make the preference ordering according to the descending order of  $C(X_i, X^{\text{best}}) \in [0, 1]$ , where a larger  $C(X_i, X^{\text{best}})$  value means a better proposition  $X_i$ .

Note that once the BBA matrix is computed from Eqs. (6)–(8), we can also apply (if we prefer) BF-TOPSIS2, BF-TOPSIS3 or BFTOPSIS4 methods to make the final decision. Their presentation is out of the scope of this paper.

#### IV. BF-TOPSIS APPLIED TO A NON-CLASSICAL MCDM

We present the results of the BF-TOPSIS1 method for two simple non-classical MCDM problems.

**Example 1:** This example is given by the score matrix of Eq. (2). We consider the relative importance weights  $w_1 = 1/3$  and  $w_2 = 2/3$  of criteria  $C_1$  and  $C_2$  respectively. Applying BBA construction formulas (6)–(8) for this example<sup>9</sup>, we get the BBA matrix  $\mathbf{M} = [(m_{ij}(X_i), m_{ij}(\bar{X}_i), m_{ij}(X_i \cup \bar{X}_i))]$  with

$$\mathbf{M} = \begin{matrix} & C_1 & C_2 \\ \begin{matrix} A_1 \\ A_3 \\ A_4 \\ A_5 \\ A_1 \cup A_2 \\ A_3 \cup A_4 \end{matrix} & \begin{bmatrix} (0.25, 0.25, 0.50) \\ (0, 1, 0) \\ (1, 0, 0) \\ (0.25, 0.25, 0.50) \\ (0, 0, 1) \\ (0, 0, 1) \end{bmatrix} & \begin{bmatrix} (0, 0, 1) \\ (0, 0, 1) \\ (0, 0, 1) \\ (0.6667, 0.1111, 0.2222) \\ (0, 1, 0) \\ (1, 0, 0) \end{bmatrix} \end{matrix} \quad (14)$$

From this matrix  $\mathbf{M}$ , we compute the distances  $d_{BI}(\cdot, \cdot)$  with respect to ideal best and worst solutions shown in Table I. Table II provides  $d^{\text{best}}(X_i)$ ,  $d^{\text{worst}}(X_i)$  and  $C(X_i, X^{\text{best}})$  values computed from the formulas (11)–(13). Based on  $C(X_i, X^{\text{best}})$  values sorted in descending order, we finally get the preference order  $(A_3 \cup A_4) \succ A_5 \succ A_4 \succ A_1 \succ (A_1 \cup A_2) \succ A_3$ . If we restrict the preference order to only singletons, we will get  $A_5 \succ A_4 \succ A_1 \succ A_3$  (i.e. student  $A_5$  is the best one). Note that student  $A_2$  alone cannot be ranked with respect to the other students, which is normal based on the non-specific input (scoring) information one has for him. Of course ad-hoc ranking solutions to rank all five students can always be developed<sup>10</sup>, but without necessarily preserving the compatibility with the rank obtained previously.

**Example 2:** In mountains, protecting housing areas against torrential floods is based on a lot of alternatives at the watershed scale such as check dams' series, sediment traps, dikes, and individual protections [12]. Moreover, alternatives can be the maintenance of existing structures or the construction of new ones to increase the protection level. Final propositions generally involve several of previous individual alternatives. We propose here a simplified case of application. Within a given watershed, a check-dams' series already exists. Older than one century years old, its maintenance (alternative  $A_1$ ) is questioned. Some experts propose to abandon it and to build a sediment trap upstream the alluvial fan (alternative  $A_2$ ) or to limit damage on buildings through individual protections (alternative  $A_3$ ). The Decision-Maker (DM), here the local municipality, must decide the best proposition taking into account several criteria: the investment cost ( $C_1$  in €, in negative values), the risk reduction in 50 years between the current situation and the expected situation after each proposition implementation ( $C_2$  in €), the impact on environment ( $C_3$  is a grade from 1 to 10), and the land-use areas needed in privates ( $C_4$  in  $m^2$ , in negative values). For each criterion, the higher is the score, the better is the proposition. The DM gives

<sup>9</sup>When a score value is missing for some proposition  $X_i$  (i.e. if  $S_j(X_i) = \emptyset$ ), then we take the vacuous BBA  $m_{ij}(X_i \cup \bar{X}_i) = 1$ .

<sup>10</sup>for instance by normalizing the  $C(X_i, X^{\text{best}})$  values (the most right column of Table II) and interpret it as a BBA, and then apply a decision method described in [15].

Focal elem. $X_i$	$d_{BI}(m_{i1}, m^{best})$	$d_{BI}(m_{i1}, m^{worst})$	$d_{BI}(m_{i2}, m^{best})$	$d_{BI}(m_{i2}, m^{worst})$
$A_1$	0.6016	0.2652	0.7906	0.2041
$A_3$	0.8416	0	0.7906	0.2041
$A_4$	0	0.8416	0.7906	0.2041
$A_5$	0.6016	0.2652	0.2674	0.5791
$A_1 \cup A_2$	0.5401	0.3536	0.6770	0
$A_3 \cup A_4$	0.5401	0.3536	0	0.6770

Table I  
DISTANCES TO IDEAL BEST AND WORST SOLUTIONS.

Focal elem. $X_i$	$d^{best}(X_i)$	$d^{worst}(X_i)$	$C(X_i, X^{best})$	Ranking
$A_1$	0.7276	0.2245	0.2358	4
$A_3$	0.8076	0.1361	0.1442	6
$A_4$	0.5270	0.4166	0.4415	3
$A_5$	0.3788	0.4745	0.5561	2
$A_1 \cup A_2$	0.6314	0.1179	0.1573	5
$A_3 \cup A_4$	0.1800	0.5692	0.7597	1

Table II  
AVERAGE DISTANCES AND RELATIVE CLOSENESS INDICATORS.

the same importance weight to  $C_1$  and  $C_2$  ( $w_1 = w_2 = 0.33$ ), but they are more important than  $C_3$  ( $w_3 = 0.20$ ) which is more important than  $C_4$  ( $w_4 = 0.14$ ). The score matrix is given in Eq. (15). In this case, the problem is not to have no knowledge on some scores but is that they are not cumulative in the same way for each criterion. For  $C_1$  and  $C_4$ , the score of the disjunction of two alternatives is the sum of individual scores whereas it is not the case for  $C_2$  and  $C_3$ .

$$\mathbf{S} = \begin{matrix} & C_1 & C_2 & C_3 & C_4 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_1 \cup A_2 \\ A_1 \cup A_3 \\ A_2 \cup A_3 \\ A_1 \cup A_2 \cup A_3 \end{matrix} & \begin{bmatrix} -150000 & 100000 & 10 & 0 \\ -500000 & 200000 & 2 & -20000 \\ -550000 & 250000 & 10 & -5000 \\ -650000 & 230000 & 2 & -20000 \\ -700000 & 250000 & 10 & -5000 \\ -1050000 & 250000 & 2 & -25000 \\ -1200000 & 250000 & 2 & -25000 \end{bmatrix} \end{matrix} \quad (15)$$

The BBA matrix based on  $\mathbf{S}$  using (3)-(8) (rounded to 2 decimal points) is

$$\mathbf{M} = \begin{matrix} & C_1 & C_2 & C_3 & C_4 \\ \begin{bmatrix} (1, 0, 0) \\ (0.44, 0.10, 0.46) \\ (0.37, 0.13, 0.50) \\ (0.27, 0.21, 0.52) \\ (0.23, 0.26, 0.51) \\ (0.04, 0.75, 0.21) \\ (0, 1, 0) \end{bmatrix} & \begin{bmatrix} (0, 1, 0) \\ (0.45, 0.28, 0.27) \\ (1, 0, 0) \\ (0.73, 0.10, 0.17) \\ (1, 0, 0) \\ (1, 0, 0) \\ (1, 0, 0) \end{bmatrix} & \begin{bmatrix} (1, 0, 0) \\ (0, 1, 0) \\ (1, 0, 0) \\ (0, 1, 0) \\ (1, 0, 0) \\ (0, 1, 0) \\ (0, 1, 0) \end{bmatrix} & \begin{bmatrix} (1, 0, 0) \\ (0.10, 0.67, 0.23) \\ (0.70, 0.07, 0.23) \\ (0.10, 0.67, 0.23) \\ (0.70, 0.07, 0.23) \\ (0, 1, 0) \\ (0, 1, 0) \end{bmatrix} \end{matrix}$$

The weighted distances to the ideal best and worst solutions and the relative closeness indicator are listed in Table III. Based on relative closeness indicator sorted in descending order, the final preference order is  $(A_1 \cup A_3) \succ A_3 \succ A_1 \succ (A_1 \cup A_2) \succ (A_2 \cup A_3) \succ A_2 \succ (A_1 \cup A_2 \cup A_3)$ : maintaining the existing check dams' series and implementing individual protections is the best option. If the preferences are restricted to single alternatives, one will get as final preference order  $A_3 \succ A_1 \succ A_2$ , i.e. option  $A_3$  (only individual protections) should be preferred by the DM.

### V. CONCLUSIONS

In this paper, we have shown how the BF-TOPSIS approach can be exploited to solve non-classical MCDM problems. This method is relatively easy to use. It does not require the normalization of data and offers a consistent construction of basic belief assignments from the available scoring values. It can also deal with missing scoring values and different criteria weights as well. In this paper only the BF-TOPSIS1 method has been presented, but other more sophisticate BF-TOPSIS methods could be also used to solve non-classical problems, but at the price of a higher complexity. The application of this new BF-TOPSIS approach to solve non-classical MCDM problems for natural risk prevention is currently under evaluation, and it will be reported in a forthcoming publication.

### REFERENCES

- [1] T. Saaty, *The Analytic Hierarchy Process*, McGraw-Hill, 1980.
- [2] J. Dezert, J.-M. Tacnet, M. Batton-Hubert, F. Smarandache, *Multi-criteria decision making based on DSMT/AHP*, in Proc. of Int. Workshop on Belief Functions, Brest, France, 2010.
- [3] J. Dezert, J.-M. Tacnet, *Evidential Reasoning for Multi-Criteria Analysis based on DSMT-AHP*, in Proc. of ISAFP 2011, Italy, 2011.
- [4] J.-M. Tacnet, J. Dezert, M. Batton-Hubert, *AHP and Uncertainty Theories for Decision Making using the ER-MCDA Methodology*, in Proc. of ISAFP 2011, Italy, 2011.
- [5] A. Ennaceur, Z. Elouedi, E. Lefevre, *Multi-criteria decision making method with belief preference relations*, Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems (IJUFKS), Vol. 22(4), pp. 573–590, 2014.
- [6] A. Ennaceur, Z. Elouedi, E. Lefevre, *Belief AHP method: AHP method with the belief function framework*, Int. J. of Information Technology & Decision Making (IJITDM), Vol. 15(3), pp. 553–573, 2016.
- [7] X. Wang, E. Triantaphyllou, *Ranking irregularities when evaluating alternatives by using some ELECTRE methods*, Omega, Vol. 36(1), pp. 45–63, 2008.
- [8] C.L. Hwang, K. Yoon, *Multiple Attribute Decision Making*, Lecture Notes in Economics and Math. Syst., Vol. 186, Springer-Verlag, Berlin, 1981.
- [9] Y.J. Lai, T.Y. Liu, C.L. Hwang, *TOPSIS for MODM*, European Journal of Operational Research, Vol. 76(3), pp. 486–500, 1994.
- [10] J. Dezert, J.-M. Tacnet, *Soft ELECTRE TRI outranking method based on belief functions*, in Proc. of the 15th Int. Conf on Information Fusion (Fusion 2012), Singapore, 2012.

Focal elem. $X_i$	$d^{\text{best}}(X_i)$	$d^{\text{worst}}(X_i)$	$C(X_i, X^{\text{best}})$	Ranking
$A_1$	0.3012	0.6116	0.6700	3
$A_2$	0.5668	0.3677	0.3935	6
$A_3$	0.1830	0.7483	0.8035	2
$A_1 \cup A_2$	0.4476	0.4901	0.5226	4
$A_1 \cup A_3$	0.1555	0.7775	0.8333	1
$A_2 \cup A_3$	0.5562	0.3614	0.3938	5
$A_1 \cup A_2 \cup A_3$	0.8328	0.2694	0.2444	7

Table III

AVERAGE DISTANCES AND RELATIVE CLOSENESS INDICATORS.

- [11] J. Dezert, D. Han, H. Yin, *A New Belief Function Based Approach for Multi-Criteria Decision-Making Support*, in Proc. of the 19th Int. Conf on Information Fusion (Fusion 2016), Heidelberg, Germany, 2016.
- [12] S. Carladous, J.-M. Tacnet, J. Dezert, D. Han, M. Batton-Hubert, *Evaluation of Efficiency of Torrential Protective Structures With New BF-TOPSIS Methods*, in Proc. of the 19th Int. Conf on Information Fusion (Fusion 2016), Heidelberg, Germany, 2016.
- [13] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [14] D. Han, J. Dezert, Y. Yang, *New Distance Measures of Evidence based on Belief Intervals*, in Proc. of the 3rd Belief Int. Conf., Oxford, UK, 2014.
- [15] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-Making with Belief Interval Distance*, in Proc. of the 4th Belief Int. Conf., Prague, Czech Republic, 2016.
- [16] A. Irpino, R. Verde, *Dynamic Clustering of Interval Data Using a Wasserstein-based Distance*, Pattern Recognition Letters, Vol.29, pp. 1648–1658, 2008.



# Examples Where Dempster’s Rule is Insensitive to the Conflict Level Between the Sources of Evidence

Florentin Smarandache<sup>a</sup>, Jean Dezert<sup>b</sup>, Albena Tchamova<sup>c</sup>

<sup>a</sup>Department of Mathematics, University of New Mexico, Gallup, NM, USA.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>c</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

Emails: smarand@unm.edu, jean.dezert@onera.fr, tchamova@bas.bg

**Abstract**—In this short note, we present two classes of examples showing that Dempster’s rule of combination is insensitive to the conflict level between the sources of evidence. This behavior is intuitively not satisfying because the amount of dissonance between sources should have an impact in the fusion result when the basic belief assignments (BBA) to combine are truly informative (not vacuous).

**Keywords:** Dempster’s rule, Information fusion, belief functions.

## I. INTRODUCTION

In this short note, we discuss the behavior of Dempster’s rule of combination used in Dempster-Shafer Theory (DST) [1], [2] to combine basic belief assignments provided by distinct sources of evidences. After a brief introduction of belief functions in Section II and a recall of Dempster’s rule of combination in section III, we provide in section IV two classes of examples showing the counter-intuitive behavior of Dempster’s rule. These new classes of examples generalize examples presented in [3]. The conclusion is made in section V.

## II. BELIEF FUNCTIONS IN SHORT

Belief functions have been introduced by Shafer in [1] to model epistemic uncertainty. We assume that the answer<sup>1</sup> of the problem under concern belongs to a known (or given) finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are exclusive<sup>2</sup>. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$  and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . A basic belief assignment (BBA) associated with a given source of evidence is defined [1] as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are respectively defined by

$$Bel(A) = \sum_{\substack{B \subseteq A \\ B \in 2^\Theta}} m(B), \quad \text{and} \quad Pl(A) = 1 - Bel(\bar{A}). \quad (1)$$

<sup>1</sup>i.e. the solution, or the decision to take.

<sup>2</sup>This is so-called Shafer’s model of FoD [2].

If  $m(A) > 0$ ,  $A$  is called a focal element of  $m(\cdot)$ . The set of focal elements of a BBA  $m$  is denoted  $\mathcal{F}(m)$ . When all focal elements are singletons then  $m(\cdot)$  is called a *Bayesian BBA* [1] and its corresponding  $Bel(\cdot)$  function is homogeneous to a (subjective) probability measure. The vacuous BBA, or VBBA for short, representing a totally ignorant source is defined as<sup>3</sup>  $m(\Theta) = 1$ .

Shafer [1] proposed to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD with Dempster’s rule. The justification and behavior of Dempster’s rule have been disputed over the years from many counter-examples involving high or low conflicting sources (from both theoretical and practical standpoints) as reported in [4]–[7]. After a brief recall of Dempster’s rule of combination in section II, we present new interesting examples showing the counter-intuitive behavior of this rule in section IV.

## III. DEMPSTER’S RULE OF COMBINATION

Dempster’s rule of combination can be seen as a normalized version of the conjunctive rule. So, let’s recall at first what is the conjunctive rule (CR) of combination. Mathematically, CR of  $s \geq 2$  BBAs  $m_i(\cdot)$ ,  $i = 1, \dots, s$  defined with respect to same FoD  $\Theta$  is defined for any  $X \in 2^\Theta$  by

$$m_{12\dots s}^{CR}(X) \triangleq \sum_{\substack{X_1, \dots, X_s \in 2^\Theta \\ X_1 \cap X_2 \cap \dots \cap X_s = X}} \prod_{i=1}^s m_i(X_i). \quad (2)$$

The conjunction (intersection) of two (or more) sources of evidence only keeps the items of information asserted by both (all) sources. This rule has been justified by Dempster [8] in statistical terms on the basis of the independence of the sources which provide  $m_i$  with  $\mathcal{F}(m_i)$ ,  $i = 1, \dots, s$ . The set of focal elements of  $m_{12\dots s}^{CR}(\cdot)$  is given by

$$\mathcal{F}(m_{12\dots s}^{CR}) = \{X_1 \cap \dots \cap X_s \mid X_i \in \mathcal{F}(m_i), i = 1, \dots, s\}.$$

The term  $m_{12\dots s}^{CR}(\emptyset)$  reflects the amount of dissonance between the sources [9] (also called the level or degree of conflict between the sources of evidence). Its management

<sup>3</sup>The complete ignorance is denoted  $\Theta$  in Shafer’s book [1].

gives rise to many debates on the choice of possible rules to combine distinct and reliable sources of evidence. In DST, Shafer proposed Dempster's rule<sup>4</sup> in which the positive value  $m_{12\dots s}^{CR}(\emptyset)$  (if any) committed to the empty set (impossible event) is removed through a simple normalization technique. Mathematically Dempster's rule of combination of  $s \geq 2$  basic belief assignments is defined by  $m_{12\dots s}^{DS}(\emptyset) = 0$ , and for any  $X \neq \emptyset \in 2^\Theta$

$$m_{12\dots s}^{DS}(X) = [m_1 \oplus \dots \oplus m_s](X) \triangleq \frac{m_{12\dots s}^{CR}(X)}{1 - m_{12\dots s}^{CR}(\emptyset)}. \quad (3)$$

Dempster's rule is commutative and associative and preserves the neutrality of vacuous BBA in the fusion process, which makes Dempster's rule an appealing method to fuse BBAs from implementation standpoint, even if the validity of its result has been highly disputed since its first criticism made by Zadeh in [10] over last decades in case of high conflicting situations, and more recently in [4]–[7] for the case of low conflicting situations.

In the next section we present two classes of examples where Dempster's rule is insensitive to the conflict level.

#### IV. NEW CLASSES OF EXAMPLES

##### A. First class of examples

Let's consider a finite frame of discernment  $\Theta$  and two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  with focal elements in  $2^\Theta$  given by

$$\begin{aligned} \mathcal{F}(m_1) &= \{A_1, A_2, \dots, A_n\}, \\ \mathcal{F}(m_2) &= \{A, B_1, \dots, B_m\}, \end{aligned}$$

where  $A_i \subseteq A$  for  $1 \leq i \leq n$ , and  $A_i \cap B_j = \emptyset$  for  $1 \leq i \leq n$  and  $1 \leq j \leq m$ .

The mass of each focal element is denoted by its corresponding lowercase letter, that is  $m_1(A_i) = a_i$  for  $1 \leq i \leq n$ , and  $m_2(A) = a$  and  $m_2(B_j) = b_j$  for  $1 \leq j \leq m$ . Because  $m_1(\cdot)$  and  $m_2(\cdot)$  are normalized BBAs, one has  $\sum_{i=1}^n a_i = 1$  and  $a + \sum_{j=1}^m b_j = 1$ .

In applying the conjunctive rule of combination of  $m_1(\cdot)$  with  $m_2(\cdot)$ , we get

$$m_{12}^{CR}(A_i) = a \cdot a_i \quad \text{for } 1 \leq i \leq n,$$

and

$$m_{12}^{CR}(\emptyset) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j.$$

Obviously  $m_{12}^{CR}(\emptyset) + \sum_{i=1}^n m_{12}^{CR}(A_i) = 1$ , which means that the following equality holds

$$m_{12}^{CR}(\emptyset) = 1 - \sum_{i=1}^n a \cdot a_i = 1 - a \sum_{i=1}^n a_i = 1 - a, \quad (4)$$

because  $\sum_{i=1}^n a_i = 1$ .

<sup>4</sup>This rule has been introduced by Dempster in [8]. It has been denoted and popularized with the operator symbol  $\oplus$  by Shafer in [1].

To get Dempster's rule result, we need to normalize the BBA  $m_{12}^{CR}(\cdot)$  by dividing the masses  $m_{12}^{CR}(A_i)$  by  $1 - m_{12}^{CR}(\emptyset)$ , or equivalently just by dividing  $m_{12}^{CR}(A_i)$  by the value  $a$  because from (4) one always has  $1 - m_{12}^{CR}(\emptyset) = 1 - (1 - a) = a$ .

After the normalization by division of masses  $m_{12}^{CR}(A_i)$  by  $a$ , one gets as Dempster-Shafer fusion result

$$m_{12}^{DS}(A_i) = [m_1 \oplus m_2](A_i) = m_1(A_i) = a_i. \quad (5)$$

Therefore, it is clear in such class of examples that the BBA  $m_2(\cdot)$  has absolutely no impact in Dempster-Shafer fusion result even if  $m_2(\cdot)$  is truly informative (not vacuous) and conflicting with the BBA  $m_1(\cdot)$ .

The conflict level  $m_{12}^{CR}(\emptyset) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j$  can be as high (close to one) or as low (close to zero) as we want, Dempster's rule provides in this class of examples always the same result  $m_{12}^{DS}(\cdot) = m_1(\cdot)$ , which is a counter-intuitive behavior not very recommended for fusion applications.

##### B. Second class of examples

This second class of example is a bit more general than the previous one. We consider a finite frame of discernment  $\Theta$  and two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  with focal elements in  $2^\Theta$  given by

$$\begin{aligned} \mathcal{F}(m_1) &= \{A_1, A_2, \dots, A_n, B\}, \\ \mathcal{F}(m_2) &= \{B, C_1, \dots, C_m\}, \end{aligned}$$

such that

- $A_i \subseteq B$  for  $1 \leq i \leq n$ ,
- $B \cap C_j = \emptyset$  for  $1 \leq j \leq m$ , and
- $A_i \cap C_j = \emptyset$  for  $1 \leq i \leq n$  and  $1 \leq j \leq m$ .

The mass of each focal element is denoted by its corresponding lowercase letter for all elements  $A_i$ , that is  $m_1(A_i) = a_i$  for  $1 \leq i \leq n$  and by  $m_1(B) = b_1$ . Similarly  $m_2(C_j) = c_j$  for  $1 \leq j \leq m$  and  $m_2(B) = b_2$ . Because  $m_1(\cdot)$  and  $m_2(\cdot)$  are normalized BBAs, one has  $b_1 + \sum_{i=1}^n a_i = 1$  and  $b_2 + \sum_{j=1}^m c_j = 1$ .

In applying the conjunctive rule of combination of  $m_1(\cdot)$  with  $m_2(\cdot)$ , we get

$$\begin{aligned} m_{12}^{CR}(A_i) &= b_2 a_i \quad \text{for } 1 \leq i \leq n \\ m_{12}^{CR}(B) &= b_1 b_2 \end{aligned}$$

and

$$m_{12}^{CR}(\emptyset) = \sum_{i=1}^n \sum_{j=1}^m a_i c_j + \sum_{j=1}^m b_1 c_j.$$

Obviously  $m_{12}^{CR}(\emptyset) + \sum_{i=1}^n m_{12}^{CR}(A_i) + m_{12}^{CR}(B) = 1$ , which means that the following equality holds

$$\begin{aligned} m_{12}^{CR}(\emptyset) &= 1 - \sum_{i=1}^n m_{12}^{CR}(A_i) - m_{12}^{CR}(B) \\ &= 1 - \sum_{i=1}^n b_2 \cdot a_i - b_1 b_2 \\ &= 1 - b_2(b_1 + \sum_{i=1}^n a_i) \\ &= 1 - b_2, \end{aligned}$$

because  $b_1 + \sum_{i=1}^n a_i = 1$ .

To get Dempster's rule result, we need to normalize the BBA  $m_{12}^{CR}(\cdot)$  by dividing the masses  $m_{12}^{CR}(A_i)$  and  $m_{12}^{CR}(B)$  by  $1 - m_{12}^{CR}(\emptyset) = 1 - (1 - b_2) = b_2$ .

After the normalization by division of masses  $m_{12}^{CR}(A_i)$  and  $m_{12}^{CR}(B)$  by  $b_2 \neq 0$ , one gets the Dempster-Shafer fusion result for  $i = 1, \dots, n$

$$m_{12}^{DS}(A_i) = [m_1 \oplus m_2](A_i) = m_1(A_i) = a_i, \quad (6)$$

and

$$m_{12}^{DS}(B) = [m_1 \oplus m_2](B) = m_1(B) = b_1. \quad (7)$$

Therefore, it is clear in such second class of examples that the BBA  $m_2(\cdot)$  has also absolutely no impact in Dempster-Shafer fusion result even if  $m_2(\cdot)$  is truly informative (not vacuous) and conflicting with the BBA  $m_1(\cdot)$ .

The conflict level  $m_{12}^{CR}(\emptyset) = \sum_{i=1}^n \sum_{j=1}^m a_i c_j + \sum_{j=1}^m b_1 c_j = 1 - b_2$  can be as high (close to one) or as low (close to zero) as we want. Dempster's rule provides in this second class of examples always the same result  $m_{12}^{DS}(\cdot) = m_1(\cdot)$ , which is a counter-intuitive behavior not recommended for fusion applications.

## V. CONCLUSIONS

We have given two classes of counter-examples to Dempster's Rule, where this rule is insensitive to the fusion, in the sense that combining two different conflicting sources of information characterized by the basic belief assignments  $m_1(\cdot)$  and  $m_2(\cdot)$ , the fusion result is equal to  $m_1(\cdot)$ . Therefore  $m_2(\cdot)$  has no impact in the fusion, although  $m_2(\cdot)$  is different from the uninformative source characterized by the vacuous basic belief assignment  $m(\Theta) = 1$ . Numerical counter-examples to Dempster's Rule can also be found in [11].

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [2] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSmt for information fusion*, American Research Press, Rehoboth, NM, U.S.A., Vol. 1-4, 2004-2015. <http://fs.gallup.unm.edu/DSmt.htm>
- [3] J. Dezert, A. Tchamova, *On the behavior of Dempster's rule of combination*, presented in Poster Session of Spring School on Belief Functions Theory and Applications, April 4-8, 2011, Autrans, France. Its extended version has been published in [4]. This paper is available at <https://hal.archives-ouvertes.fr/hal-00577983/fr/>
- [4] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Proc. of Fusion 2012, Singapore, July 9-12, 2012.
- [5] A. Tchamova, J. Dezert, *On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory*, 6th IEEE Int. Conf. on Int. Syst., Sofia, Bulgaria, Sept. 6-8, 2012.
- [6] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., Vol. 29, No. 3, pp. 223-252, March 2014 (with erratum in DSmt Book Vol. 5).
- [7] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, Proc. of Fusion 2013, Istanbul, Turkey, July 2013.
- [8] A.P. Dempster, *Upper and lower probabilities induced by a multi valued mapping*, Annals of Mathematical Statistics, Vol. 38, pp. 325-339, 1967.
- [9] D. Dubois, H. Prade, *A set-theoretic view of belief functions*, Int. J. General Systems, Vol. 12, pp. 193-226, 1976.
- [10] L.A. Zadeh, *On the validity of Dempster's rule of combination*, Memo M79/24, Univ. of California, Berkeley, U.S.A., 1979.
- [11] F. Smarandache, J. Dezert, V. Kroumov, *Examples where the conjunctive and Dempster's rules are insensitive*, Proc. of 2013 International Conference on Advanced Mechatronic Systems, Luoyang, China, Sept. 25-27, 2013 (also available in [2], Vol. 4).



# Determination of Basic Belief Assignment Using Fuzzy Numbers

Zhe Zhang<sup>a</sup>, Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>

<sup>a</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>c</sup>School of Aerospace, Xi'an Jiaotong University, Xi'an, China.

Emails: zhezhang@stu.xjtu.edu.cn, deqhan@xjtu.edu.cn, jean.dezert@onera.fr, jiafeiyi@mail.xjtu.edu.cn

Originally published as: Z. Zhang, D. Han, J. Dezert, Y. Yang, *Determination of Basic Belief Assignment Using Fuzzy Numbers*, in Proc. of 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10–13, 2017, and reprinted with permission.

**Abstract**—Dempster-Shafer evidence theory (DST) is a theoretical framework for uncertainty modeling and reasoning. The determination of basic belief assignment (BBA) is crucial in DST, however, there is no general theoretical method for BBA determination. In this paper, a method of generating BBA using fuzzy numbers is proposed. First, the training data are modeled as fuzzy numbers. Then, the dissimilarities between each test sample and the training data are measured by the distance between fuzzy numbers. In the final, the BBAs are generated from the normalized dissimilarities. The effectiveness of this method is demonstrated by an application of classification problem. Experimental results show that the proposed method is robust to outliers.

**Keywords**—Evidence theory, basic belief assignment (BBA), fuzzy numbers, outliers.

## I. INTRODUCTION

The theory of belief functions also called Dempster-Shafer evidence theory (DST) [1], [2], is a theoretical framework for uncertainty modeling and reasoning. The expression of uncertainty, i.e., the determination of basic belief assignment (BBA) is one of the most crucial problems to deal with. BBA is a kind of random set in nature and its determination is actually the problem of modeling the distribution of random set, which is still unsolved in mathematics [3]. Therefore, the determination of BBA is a challenging problem in DST and has aroused widespread concerns.

One category for generating BBA is the application-based empirical approach. Shafer [1] generates BBA based on statistic evidence. Selzer [12] generates BBA according to the class number and the neighborhood of the target for automatic target classification. Bi [13] proposed focal element triplet for text categorization. Valente [4] proposed several BBA determination methods for speech recognition based on the membership. Zhang [5] generates BBA based on evidential Markov random field for image segmentation. Salzenstein [14] proposed an iterative estimation method to generate BBA based on the Gaussian model for multisensor image segmentation. Dezert [6] generates BBA to describe the uncertainty of threshold choosing in edge detection. Han [7] generates BBA based on the intervals of the expected payoffs for different alternatives to deal with multi-criteria decision making problems.

The another category for generating BBA is the application-free approach. Boudraa [8] proposed a method based on fuzzy membership functions. Deng [9] generates BBA based on the similarity measure described by the radius of gyration. Han [10] proposed a method based on uncertain optimization. Kang [11] proposed a method based on interval numbers.

In Kang's method [11], the training data are modeled as interval numbers determined by their lower and upper bound values. Since the interval number is a special case of the fuzzy number and only keeps minimum and maximum values, other important information, such as mean value and median, are lost when modeling the data. To deal with this, other types of fuzzy numbers are used to model the training data in this paper, i.e., the mean value and median are also kept to describe the training data. Then the BBAs are generated from the dissimilarities between the test sample and the training data using the distance between fuzzy numbers. Compared with the distance between interval numbers in Kang's method, the distance between fuzzy numbers is more robust when there exist outliers in training data. To verify the effectiveness of the proposed BBA determination method, we consider its application on the classification problem. The experiment results show that the proposed method can achieve high classification accuracies.

## II. BASIS OF EVIDENCE THEORY

Dempster-Shafer evidence theory (DST) [1], [2] is a theoretical framework for uncertainty modeling and reasoning. In DST, the frame of discernment (FOD)  $\Theta$  contains  $l$  mutually exclusive and exhaustive elements:  $\Theta = \{\theta_1, \theta_2, \dots, \theta_l\}$ . The power set of  $\Theta$  (the set of all subsets of  $\Theta$ ) is denoted by  $2^\Theta$ . The basic belief assignment (BBA, also called a mass function)  $m$  is defined from  $2^\Theta$  to  $[0, 1]$  satisfying

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad m(\emptyset) = 0 \quad (1)$$

$m(A)$  represents the evidence support to the proposition  $A$ . If  $m(A) > 0$ ,  $A$  is called a focal element.

The plausibility function ( $Pl$ ) and belief function ( $Bel$ ) are defined respectively as:

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (2)$$

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (3)$$

Dempster's rule of combination [1], used for combining two distinct sources of evidence in the DST framework, is defined as

$$m_1 \oplus m_2(A) = \begin{cases} 0, & A = \emptyset \\ \frac{1}{1-K} \sum_{B \cap C = A} m_1(B)m_2(C), & A \neq \emptyset \end{cases} \quad (4)$$

where  $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$  represents the total conflict or contradictory mass assignments.

For a probabilistic decision-making based on the BBA, Smets defined the pignistic probability transformation [15] to transform a BBA into a probability measure  $BetP$ :

$$BetP(\theta_i) \triangleq \sum_{\theta_i \in A} \frac{m(A)}{|A|} \quad \forall \theta_i \in \Theta \quad (5)$$

where  $|A|$  denotes the cardinality of  $A$ . The final decision is often made by choosing the element in FOD which has the highest  $BetP$  value.

### III. THE DETERMINATION OF BBA BASED ON INTERVAL NUMBERS

In DST, the expression of uncertainty is the process of generating BBA. Therefore, the determination of BBA is the first step and crucial in the applications of DST. However, BBA is a kind of random set and its determination is actually the problem of modeling the distribution of random set, which is still unsolved in mathematics [3]. Kang [11] proposed a BBA determination method based on interval numbers (IN). The basis of interval numbers is briefly introduced first.

#### A. Basis of interval numbers

An interval number  $\tilde{a}$  in  $\mathbb{R}$  is a set of real numbers that lie between two real numbers, i.e.,  $\tilde{a} = [a_1, a_2] = \{x | a_1 \leq x \leq a_2\}$ ,  $a_1, a_2 \in \mathbb{R}$  and  $a_1 \leq a_2$ .

The dissimilarity between two interval numbers  $\tilde{a} = [a_1, a_2]$  and  $\tilde{b} = [b_1, b_2]$  can be measured by the distance between them [16]:

$$\begin{aligned} D^2(\tilde{a}, \tilde{b}) &= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \left\{ \left[ \left( \frac{a_1 + a_2}{2} \right) + x(a_2 - a_1) \right] \right. \\ &\quad \left. - \left[ \left( \frac{b_1 + b_2}{2} \right) + y(b_2 - b_1) \right] \right\}^2 dx dy \\ &= \left[ \left( \frac{a_1 + a_2}{2} \right) - \left( \frac{b_1 + b_2}{2} \right) \right]^2 \\ &\quad + \frac{1}{3} \left[ \left( \frac{a_2 - a_1}{2} \right)^2 + \left( \frac{b_2 - b_1}{2} \right)^2 \right] \end{aligned} \quad (6)$$

The larger  $D(\tilde{a}, \tilde{b})$  is, the larger dissimilarity between  $\tilde{a}$  and  $\tilde{b}$  is.

#### B. IN-based method

In IN-based method, the training data belonging to the same focal element  $A \subseteq \Theta$  are modeled as an interval number  $\tilde{a} = [a_1, a_2]$ , where  $a_1$  and  $a_2$  are the minimum and maximum values of the training data respectively. For a single test sample, it is also modeled as a degenerate interval number  $\tilde{t} = [t, t]$ , where  $t$  is its value. If the test sample  $\tilde{t}$  is similar to the training data  $\tilde{a}$ , the corresponding proposition (the test sample belongs to  $A$ ) should be assigned a large belief.

The similarity between  $\tilde{a}$  and  $\tilde{t}$  is defined as:

$$S(\tilde{a}, \tilde{t}) = \frac{1}{1 + \alpha D(\tilde{a}, \tilde{t})} \quad (7)$$

where  $\alpha > 0$  is a parameter to control the degree of dispersion of the normalized similarities and  $D(\tilde{a}, \tilde{t})$  is the distance between the interval numbers  $\tilde{a}$  and  $\tilde{t}$ . Finally, the BBA can be generated from the normalized similarities.

In IN-based method, when modeling the training data, only the minimum and maximum values are kept and used to calculate similarities. However, when the distribution of the data is not uniform, the extreme values are insufficient to well describe the data. Actually, any interval number is a special case of a fuzzy number. Other types of fuzzy numbers, such as triangular fuzzy number (TFN) and trapezoidal fuzzy number (TrFN), can keep more useful information of the data, such as the mean value and median. Thus, TFN and TrFN are used to model the data in this paper.

### IV. BBA CONSTRUCTION FROM FUZZY NUMBERS

#### A. Basis of fuzzy numbers

The generalized left right fuzzy number (GLRFN)  $\tilde{b} = [b_1, b_2, b_3, b_4]$  is a special case of a convex, normalized fuzzy set of the real line when its membership function is defined by [17]:

$$\mu(x) = \begin{cases} L\left(\frac{b_2 - x}{b_2 - b_1}\right) & \text{for } b_1 \leq x \leq b_2 \\ 1 & \text{for } b_2 \leq x \leq b_3 \\ R\left(\frac{x - b_3}{b_4 - b_3}\right) & \text{for } b_3 \leq x \leq b_4 \\ 0 & \text{else} \end{cases} \quad (8)$$

where  $L$  and  $R$  are strictly decreasing functions defined on  $[0, 1]$  and satisfy the conditions:

$$\begin{aligned} L(x) = R(x) = 1 & \quad \text{if } x \leq 0, \\ L(x) = R(x) = 0 & \quad \text{if } x \geq 1. \end{aligned} \quad (9)$$

The interval number is a special case of GLRFN with  $b_1 = b_2$  and  $b_3 = b_4$ . The triangular fuzzy number (TFN) and trapezoidal fuzzy number (TrFN) [16] are two of the most common fuzzy numbers encountered in applications involving fuzzy numbers.

For TrFN,  $L(x) = R(x) = 1 - x$ . The distance between two TrFNs  $\tilde{a} = [a_1, a_2, a_3, a_4]$  and  $\tilde{b} = [b_1, b_2, b_3, b_4]$  is defined as:

$$\begin{aligned}
 D^2(\tilde{a}, \tilde{b}) &= \frac{1}{4}[(a_2 + a_3) - (b_2 + b_3)]^2 \\
 &+ \frac{1}{4}[(a_2 + a_3) - (b_2 + b_3)] \\
 &\times (a_4 - a_3 - a_2 + a_1 - b_4 + b_3 + b_2 - b_1) \\
 &+ \frac{1}{12}(a_3 - a_2)^2 + \frac{1}{12}(b_3 - b_2)^2 \\
 &+ \frac{1}{12}(a_3 - a_2)[a_4 - a_3 + a_2 - a_1] \\
 &+ \frac{1}{12}(b_3 - b_2)[b_4 - b_3 + b_2 - b_1] \\
 &+ \frac{1}{9}[(a_4 - a_3)^2 + (a_2 - a_1)^2] \\
 &+ \frac{1}{9}[(b_4 - b_3)^2 + (b_2 - b_1)^2] \\
 &- \frac{1}{9}[(a_2 - a_1)(a_4 - a_3) + (b_2 - b_1)(b_4 - b_3)] \\
 &+ \frac{1}{6}[(a_4 - a_3)(b_2 - b_1) + (a_2 - a_1)(b_4 - b_3)] \\
 &- \frac{1}{6}[(a_4 - a_3)(b_4 - b_3) + (a_2 - a_1)(b_2 - b_1)]
 \end{aligned} \tag{10}$$

The larger  $D(\tilde{a}, \tilde{b})$  is, the larger dissimilarity between  $\tilde{a}$  and  $\tilde{b}$  is.

For TFN,  $L(x) = R(x) = 1 - x$  and  $b_2 = b_3$ . The distance between two TFNs  $\tilde{a} = [a_1, a_2, a_3]$  and  $\tilde{b} = [b_1, b_2, b_3]$  is defined as:

$$\begin{aligned}
 D^2(\tilde{a}, \tilde{b}) &= (a_2 - b_2)^2 + \frac{1}{2}(a_2 - b_2)[(a_3 + a_1) - (b_3 + b_1)] \\
 &+ \frac{1}{9}[(a_3 - a_2)^2 + (a_2 - a_1)^2] \\
 &+ \frac{1}{9}[(b_3 - b_2)^2 + (b_2 - b_1)^2] \\
 &- \frac{1}{9}[(a_2 - a_1)(a_3 - a_2) + (b_2 - b_1)(b_3 - b_2)] \\
 &+ \frac{1}{6}(2a_2 - a_1 - a_3)(2b_2 - b_1 - b_3)
 \end{aligned} \tag{11}$$

The larger  $D(\tilde{a}, \tilde{b})$  is, the larger dissimilarity between  $\tilde{a}$  and  $\tilde{b}$  is.

### B. Fuzzy-number-based methods

1) *Data modeling*: To generate BBAs, the fuzzy numbers are used to model the training data and test samples in this paper. For the training data belonging to  $A \subseteq \Theta$  and the test sample  $t$ , we can use three different kinds of fuzzy numbers to model them:

- (1) TFNmean: the training data are modeled as a triangular fuzzy number  $\tilde{a} = [a_1, a_2, a_3]$ , where  $a_1$  and  $a_3$  are the minimum and maximum values of the training data respectively and  $a_2$  is the mean value. The test sample is modeled as  $\tilde{t} = [t, t, t]$ .
- (2) TFNmed: the training data are modeled as a triangular fuzzy number  $\tilde{b} = [b_1, b_2, b_3]$ , where  $b_1$  and  $b_3$  are the minimum and maximum values of the training data respectively and  $b_2$  is the median. The test sample is modeled as  $\tilde{t} = [t, t, t]$ .
- (3) TrFN: the training data are modeled as a trapezoidal fuzzy number  $\tilde{c} = [c_1, c_2, c_3, c_4]$ , where  $c_1$  and  $c_4$  are the minimum and maximum values of the training data respectively,  $c_2$  is either the mean value or median,

whichever is smaller and  $c_3$  is either the mean value or median, whichever is larger. The test sample is modeled as  $\tilde{t} = [t, t, t, t]$ .

In these ways, besides the maximum and minimum values, the mean value and (or) median can be also kept to describe the training data.

2) *Calculate the similarities*: Similar to the IN-based method, the similarity between the training data and test sample are measured from the distance between them (Eq. (11) for TFN or Eq. (10) for TrFN) using Eq. (7). Actually, other normalization functions can be used here.

3) *Generate the BBAs*: The BBAs are generated from the normalized similarities. If the test sample  $\tilde{t}$  is similar to the training data  $\tilde{a}$ , the corresponding proposition ( $\tilde{t}$  belongs to the same focal element with  $\tilde{a}$ ) should be assigned with a large belief.

In the next section, we consider the classification problem to verify the effectiveness of our proposed BBA determination method.

## V. CLASSIFICATION EXAMPLE BASED ILLUSTRATION OF THE PROPOSED BBA DETERMINATION METHOD

We give a classification example on a set of artificial data to illustrate the process of our BBA determination method and verify its effectiveness.

### A. Artificial training data

Suppose there are three classes in a set of artificial data:  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . Each sample has three features,  $f_1$ ,  $f_2$  and  $f_3$ , and each feature is correspondent to a normal distribution. The deviation parameters for each class are 0.25, 1 and 0.25 respectively and the mean parameters for each feature of each class are given in Table I.

TABLE I  
THE MEAN PARAMETERS FOR EACH FEATURE OF EACH CLASS

Class	$f_1$	$f_2$	$f_3$
$\theta_1$	9	5	10
$\theta_2$	10	9	5
$\theta_3$	5	10	9

We generate 60 training data for each class. Among the 60 samples belonging to class  $\theta_1$ , there is an outlier whose value of feature  $f_1$  is much larger than others belonging to class  $\theta_1$ . The generated training data are shown in Fig. 1.

In this case, each class can be distinguished easily from other classes using one feature (when its mean parameter is 5), but are difficult distinguished from other classes using other features.

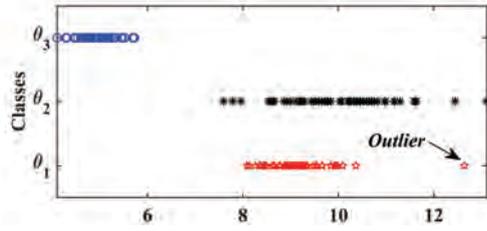
### B. The process of classification

For a given test sample, the process of labeling its class can be outlined below:

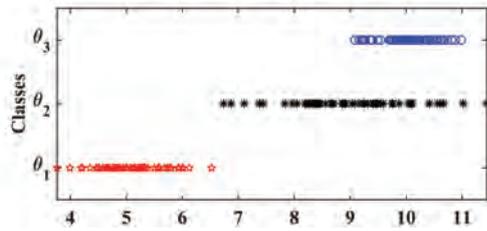
- Step 1 Generate three mass functions  $m_1$ ,  $m_2$  and  $m_3$  according to the corresponding features of the training data respectively.

TABLE II  
MODELING THE TRAINING DATA ON FEATURE  $f_1$

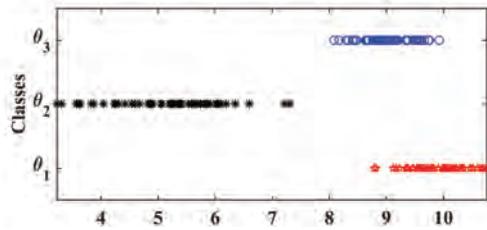
Focal element	IN	TFNmean	TFNmed	TrFN
$\{\theta_1\}$	[8.1, 12.6]	[8.1, 9.1, 12.6]	[8.1, 9.0, 12.6]	[8.1, 9.0, 9.1, 12.6]
$\{\theta_2\}$	[7.6, 13.1]	[7.6, 10.0, 13.1]	[7.6, 10.0, 13.1]	[7.6, 10.0, 10.0, 13.1]
$\{\theta_3\}$	[4.1, 5.7]	[4.1, 5.0, 5.7]	[4.1, 4.8, 5.7]	[4.1, 4.8, 5.0, 5.7]
$\{\theta_2, \theta_3\}$	[8.1, 12.6]	[8.1, 9.6, 12.6]	[8.1, 9.3, 12.6]	[8.1, 9.3, 9.6, 12.6]



(a) Values of the training data for feature  $f_1$ .



(b) Values of the training data for feature  $f_2$ .



(c) Values of the training data for feature  $f_3$ .

Fig.1. Values of the training data.

- Step 2 Combine  $m_1, m_2$  and  $m_3$  using Eq. (4) to obtain the combined mass function  $m$ .
- Step 3 Transform  $m$  into the probability measure  $BetP$  using Eq. (5).
- Step 4 The class of the test sample is labeled as class  $\theta_i \in \Theta$  which has the highest  $BetP$  value.

We take a test sample  $t = (t_1, t_2, t_3) = (11.8, 9.8, 3.9)$  (whose class is  $\theta_2$ ) as an example to explain how to generate  $m_1$  based on fuzzy numbers in Step 1 in detail. The result of interval-number-based method is also given for comparison.

1) *Data modeling*: For feature  $f_1$ , the training data belonging to each focal element  $A \in \Theta$  can be modeled as an interval number (IN) or a fuzzy number (TFNmean, TFNmed or TrFN), as shown in Table II. The test sample can be modeled as  $\tilde{t}_1 = [11.8, 11.8]$  (IN),  $\tilde{t}_1 = [11.8, 11.8, 11.8]$  (TFNmean or TFNmed) or  $\tilde{t}_1 = [11.8, 11.8, 11.8, 11.8]$  (TrFN).

In this case, the training data from class  $\theta_1$  has an overlapped region with the data from  $\theta_2$ . For a test sample

belonging to this region, it is difficult to distinguish whether class  $\theta_1$  or  $\theta_2$  it should be labeled as and its belief assigned to focal element  $\{\theta_1, \theta_2\}$  ( $m_1\{\theta_1, \theta_2\}$ ) should also be considered. Thus, the training data belonging to the overlapped region are also modeled.

2) *Calculate the distance between fuzzy numbers*: The distances between the test sample  $\tilde{t}_1$  and the training data from different focal elements are calculated using Eq. (6) (for IN), Eq. (11) (for TFNmean and TFNmed) or Eq. (10) (for TrFN), as given in Table III.

TABLE III  
THE DISTANCE BETWEEN THE TEST SAMPLE AND TRAINING DATA

Focal element	IN	TFNmean	TFNmed	TrFN
$\{\theta_1\}$	1.928	3.439	3.541	2.217
$\{\theta_2\}$	2.134	2.552	2.529	1.838
$\{\theta_3\}$	6.875	9.660	9.738	6.862
$\{\theta_1, \theta_2\}$	1.928	2.977	3.265	2.040

In Table III, according to the IN-based method, the test sample is closer to  $\{\theta_1\}$  than  $\{\theta_2\}$ . However, without the outlier, the actual range of the training data from  $\{\theta_1\}$  is [8.1, 10.4] and the test sample 11.8 should be assigned a smaller distance to  $\{\theta_2\}$ , whose range is [7.6, 13.1]. By only considering the minimum and maximum values, the IN-based method can easily get counterintuitive distances, especially when there are outliers. However, the mean value and median are relatively insensitive to outliers, so that the fuzzy-number-based methods can obtain more reasonable distances. In this case, the fuzzy-number-based methods assign the test sample a smaller distance to  $\{\theta_2\}$  than  $\{\theta_1\}$ .

3) *Calculate the similarities*: The similarities between the test sample  $\tilde{t}_1$  and the training data from different focal elements are calculated from the above distances using Eq. (7), where  $\alpha$  is taken as 5, as shown in Table IV.

TABLE IV  
THE SIMILARITIES BETWEEN THE TEST SAMPLE AND TRAINING DATA

Focal element	IN	TFNmean	TFNmed	TrFN
$\{\theta_1\}$	0.094	0.055	0.054	0.083
$\{\theta_2\}$	0.086	0.073	0.073	0.098
$\{\theta_3\}$	0.028	0.020	0.020	0.028
$\{\theta_1, \theta_2\}$	0.094	0.063	0.058	0.089

4) *Generate  $m_1$* :  $m_1$  is generated from the normalized similarities, as shown in Table V. Our fuzzy-number-based

methods assign the largest mass of belief to  $\{\theta_2\}$  rather than  $\{\theta_1\}$  or  $\{\theta_1, \theta_2\}$ , which is more reasonable compared with the IN-based method.

TABLE V  
THE GENERATED  $m_1$

Focal element	IN	TFNmean	TFNmed	TrFN
$\{\theta_1\}$	0.311	0.261	0.261	0.277
$\{\theta_2\}$	0.284	0.345	0.358	0.329
$\{\theta_3\}$	0.094	0.096	0.099	0.095
$\{\theta_1, \theta_2\}$	0.311	0.298	0.282	0.299

In the same way,  $m_2$  and  $m_3$  can be generated from feature  $f_2$  and  $f_3$  respectively, as shown in Table VI and Table VII.

TABLE VI  
THE GENERATED  $m_2$

Focal element	IN	TFNmean	TFNmed	TrFN
$\{\theta_1\}$	0.062	0.038	0.039	0.047
$\{\theta_2\}$	0.177	0.173	0.176	0.184
$\{\theta_3\}$	0.380	0.378	0.375	0.378
$\{\theta_2, \theta_3\}$	0.381	0.411	0.410	0.391

TABLE VII  
THE GENERATED  $m_3$

Focal element	IN	TFNmean	TFNmed	TrFN
$\{\theta_1\}$	0.160	0.138	0.141	0.147
$\{\theta_2\}$	0.484	0.547	0.537	0.522
$\{\theta_3\}$	0.183	0.162	0.167	0.171
$\{\theta_1, \theta_3\}$	0.173	0.153	0.155	0.160

After generating  $m_1$ ,  $m_2$  and  $m_3$ , the combined mass function  $m$  can be obtained by using the Dempster's rule of combination (Eq. (4)) and then the probability measure  $BetP$  can be obtained using Eq. (5), as given in Table VIII.

TABLE VIII  
THE GENERATED  $BetP$

Class	IN	TFNmean	TFNmed	TrFN
$\theta_1$	0.065	0.026	0.027	0.038
$\theta_2$	0.808	0.873	0.866	0.853
$\theta_3$	0.127	0.101	0.107	0.109

Finally, the test sample  $t = (11.8, 9.8, 3.9)$  is labeled as class  $\theta_2$  since it has the highest  $BetP$  value.

## VI. EXPERIMENTS

To further compare the effectiveness of the proposed BBA determination methods with the IN-based method, we did the classification experiments on three UCI data sets (Iris, Wine and Wdbc).

In each experiment, the amounts of the samples from different classes are equal. Among the samples from the same class, 60% samples are used as the training data and the rest

40% samples are used as the test samples. We generate BBAs from all the features (one BBA generated from one feature) and the final classification result is obtained from the combined mass function. The value of  $\alpha$  in Eq. (7) is set as 5. The accuracy of each classification is calculated from 100 runs of the Monte-Carlo experiments. The classification accuracies<sup>1</sup> are given in Table IX.

TABLE IX  
THE ACCURACIES OF THE CLASSIFICATIONS (%)

Data set	IN	TFNmean	TFNmed	TrFN
Iris	92.67	93.83	93.85	93.92
Wine	91.48	93.29	94.23	92.79
Wdbc	67.71	86.91	88.32	81.27

From Table IX we can see, the proposed fuzzy-number-based methods can achieve higher accuracies than IN-based method.

Furthermore, we compared the robustness of our proposed method with IN-based method. We add one outlier to the training data for each class, whose values on each feature are set as:

$$O(f_i) = \max(f_i) + 0.2 \times (\max(f_i) - \min(f_i)) \quad (12)$$

where  $\max(f_i)$  and  $\min(f_i)$  are the maximum and minimum values of the training data respectively on feature  $f_i$ . The accuracies are given in Table X.

TABLE X  
THE ACCURACIES OF THE CLASSIFICATIONS WITH OUTLIERS (%)

Data set	IN	TFNmean	TFNmed	TrFN
Iris	88.72	93.08	93.07	92.13
Wine	80.89	91.75	92.53	90.06
Wdbc	61.80	82.87	84.28	73.69

From Table IX and Table X we can see, the accuracies of IN-based method drop significantly when the outliers are added while the accuracies of our fuzzy-number-based methods drop slightly. Therefore, the proposed fuzzy-number-based methods are more robust for outliers than IN-based method.

## VII. CONCLUSION

In this paper we have proposed new methods for generating BBA based on fuzzy numbers. The experiments on its application of classification show that our proposed method is effective and robust for outliers and can achieve higher accuracies than the IN-based method.

In future work, we will focus on the distance between fuzzy numbers. More types of distance will be used and compared to describe the dissimilarity between the test sample and training data. Other normalization functions to establish similarities will be evaluated, as well as other possible decision-making strategies. Also, other evidence combination rules will be tested to make comparisons.

<sup>1</sup>The accuracy is defined as the percentage of correct classifications.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation (Nos. 61573275, 61671370), Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), Science and Technology Project of Shaanxi Province (No. 2013KJXX-46), Postdoctoral Science Foundation of China (No. 2016M592790), and Fundamental Research Funds for the Central Universities (No. xjj2014122, xjj2016066).

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
- [2] A. P. Dempster, "Upper and lower probabilities induced by a multiple valued mapping," *The Annals of Mathematical Statistics*, vol. 38, no. 2, pp. 325-339, 1967.
- [3] D. Han, Y. Yang and C. Han, "Advances in DS evidence theory and related discussions," *Control and Decision*, vol. 29, no. 1, pp. 1-11, 2014.
- [4] F. Valente and H. Hermansky, "Combination of acoustic classifiers based on Dempster-Shafer theory of evidence," *Proc of IEEE Int Conf on Acoustics, Speech and Signal Processing*, IV-1129-IV-1132, 2007.
- [5] Z. Zhang, D. Han and Y. Yang, "Image segmentation based on evidential Markov random field model," *IEEE International Conference on Control, Automation and Information Sciences*, Changshu, pp. 239-244, October 2015.
- [6] J. Dezert, Z. Liu and G. Mercier, "Edge detection in color images based on DSMT," *The 14th Int Conf on Information Fusion*, Chicago, pp. 969-976, 2011.
- [7] D. Han, J. Dezert and J. M. Tacnet, "A fuzzy-cautious OWA approach with evidential reasoning," *The 15th Int Conf on Information Fusion*, Singapore, pp. 278-285, 2012.
- [8] A. O. Boudraa, A. Bentabet and F. Salzenstein, "Dempster-Shafer's basic probability assignment based on fuzzy membership functions," *Electronic Letters on Computer Vision and Image Analysis*, vol. 4, no. 1, pp. 1-9, 2004.
- [9] Y. Deng, W. Jiang and X. Xu, "Determining BPA under uncertainty environments and its application in data fusion," *Journal of Electronics (China)*, vol. 26, no. 1, pp. 13-17, 2009.
- [10] D. Han, Y. Deng and C. Han, "Novel approaches for the transformation of fuzzy membership function into basic probability assignment based on uncertain optimization," *Int J of Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 21, no. 2, pp. 289-322, 2013.
- [11] B. Kang, Y. Li and Y. Deng, "Determination of basic probability assignment based on interval numbers and its application," (in Chinese) *Acta Electronica Sinica*, vol. 40, no. 6, pp. 1092-1096, 2012.
- [12] F. Selzer and D. Gutfinger, "LADAR and FLIR based sensor fusion for automatic target classification," *Proc of SPIE*, Montreal, pp. 236-246, 1988.
- [13] Y. X. Bi, D. Bell and J. W. Guan, "Combining evidence from classifiers in text categorization," *Proc of the 8th Int Conf on KES*, Wellington, pp. 521-528, 2004.
- [14] F. Salzenstein and A. O. Boudraa, "Iterative estimation of Dempster Shafer's basic probability assignment: Application to multisensor image segmentation," *Optical Engineering*, vol. 43, no. 6, pp. 1293-1299, 2004.
- [15] P. Smets, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191-234, 1994.
- [16] L. Tran and L. Duckstein, "Comparison of fuzzy numbers using a fuzzy distance measure," *Fuzzy Sets and Systems*, vol. 130, no. 3, pp. 331-341, 2002.
- [17] D. Dubois, H. Prade, "Fuzzy Sets and Systems: Theory and Applications," *Academic Press*, New York, 1980.

# Comparative Study on BBA Determination Using Different Distances of Interval Numbers

Jiankun Ding<sup>a</sup>, Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>

<sup>a</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>c</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an, China.

Emails: d4574b@163.com, deqhan@xjtu.edu.cn, jean.dezert@onera.fr, jiafeiy@mail.xjtu.edu.cn

Originally published as: J. Ding, D. Han, J. Dezert, Y. Yang, *Comparative Study on BBA Determination Using Different Distances of Interval Numbers*, in Proc. of 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10–13, 2017, and reprinted with permission.

**Abstract**—Dempster-Shafer theory (DST) is an important theory for information fusion. However, in DST how to determine the basic belief assignment (BBA) is still an open issue. The interval number based BBA determination method is simple and effective, where the features of different classes' samples are modeled using the interval numbers, i.e., an interval number model is constructed for each focal element. Then, the distances of interval numbers are used for measuring the similarity degrees between the testing sample and each focal element, and the similarity degrees are used for determining the BBA. The definition of interval numbers' distance is crucial for the effectiveness of the interval number based BBA determination methods. In this paper, we use different interval numbers' distances for determining BBAs. By using the artificial data set and the Iris date set of open UCI data base, respectively, we compare and analyze the determination of BBAs with different distances.

**Index Terms**—Dempster-Shafer theory, basic belief assignment, distance of interval numbers, information fusion, classification.

## I. INTRODUCTION

Dempster-Shafer theory (DST) [1] was proposed by Dempster in 1960s, and was developed by Shafer [2]. In DST, the basic beliefs are assigned to the power set of the frame of discernment (FOD), which is used to describe the uncertainty of sources of evidence. The evidences (i.e., basic belief assignments, BBAs) originated from different sources can be fused using the Dempster's combination rule [1]. DST has been widely used in the information fusion fields [3]–[5].

Using DST, the first step is to determine the BBAs, which is still an open issue. The determination of BBAs can mainly categorized into two branches [6]: (1) The experts give the BBAs directly according to their personal experiences; (2) The BBAs are determined based on the samples using some special determination rules. In the first branch, the determination of BBAs relies on the experts' subjective points of view. In this paper, we focus on the second branch approaches, i.e., the BBAs are determined based on available samples. Researchers have proposed many approaches in this branch. Selzer et al. [3] determined the BBAs based on the number of classes and the environmental weighting coefficient. Shafer [2] proposed a BBA determination method based on statistical evidences. Bi et al. [7] designed a kind of triple focal elements BBA in dealing with the text classification problem. Szluzen-stein et al. [8] used the Gaussian model getting the BBAs

through iterative estimation. Deng et al. [9] defined a similarity measure based on radius of gravity, and then the similarity measure is used for determining the BBAs. Boudraa et al. [10] and Florea et al. [11] determines the BBAs based on the membership functions. Han et al. [12] proposed a method for the transformation of fuzzy membership function into BBAs by solving a constrained maximization or minimization optimization problem. Recently, Kang et al. [6] designed a BBA determination method using the interval numbers.

Kang's interval number based BBA determination method is simple and effective. Kang's method first constructs the interval number [14] models for each focal element (including the singleton focal elements with single class and the compound focal elements with multiple classes) based on the set of training samples. In Kang's method, the Tran and Duckstein's [14], [16] interval number distance (TD-IND) is used for measuring the similarity degree of the testing samples compared with different focal elements' interval number models. In the final, the similarities are normalized to get the values of BBA. The definitions of the interval numbers' distances (INDs) are crucial for the performance of the interval number based BBA determination method. There exist many possible choices for INDs, e.g., the Gowda and Ravi's distance [15] (GR-IND), the Tran and Duckstein's distance [16] (TD-IND), the Hausdorff distance [17] (H-IND) and the De Carvalho's norm-q distance [18] (Nq-IND). In this paper, we implement the Kang's interval number based method using different INDs. We analyze the differences of the BBAs determined using different INDs based on numerical examples. Furthermore, we use Monte-Carlo experiments for comparing the performances of interval number based methods with different INDs by classifying an artificial set and the iris set<sup>1</sup>.

## II. BASIC OF DEMPSTER-SHAFER THEORY

Dempster-Shafer theory (DST) (also known as the Evidence Theory) is an appealing mathematical framework which can effectively describe the uncertainty information for the state of nature. In DST, the frame of discernment (FOD) is denoted by  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The elements in  $\Theta$  are mutually

<sup>1</sup><http://archive.ics.uci.edu/ml/datasets/Iris>

and exhaustive. The basic belief assignment (BBA) function assigns basic beliefs on the power set of  $\Theta$ , i.e.,  $2^\Theta$ . The BBA is also called the mass function which satisfies:

$$\sum_{A \subseteq \Theta} m(A) = 1, m(\emptyset) = 0 \quad (1)$$

If  $A \subseteq \Theta, m(A) > 0$ ,  $A$  is called a focal element.

The Belief ( $Bel$ ) and Plausibility ( $Pl$ ) of  $A$  are defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

$$Pl(A) = \sum_{B \cap A = \emptyset} m(B) = 1 - Bel(\bar{A}) \quad (3)$$

The interval  $[Bel(A), Pl(A)]$  is called the belief interval, which represents the uncertainty of the support degree of  $A$ .

Different information sources can provide different evidences, i.e., the BBAs. In DST, two BBAs associated with two distinct sources of evidence can be combined according to the Dempster's rule, as in Eq. (4).

$$m(A) = \begin{cases} \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} & A \neq \emptyset \\ 0 & A = \emptyset \end{cases} \quad (4)$$

where  $K = \sum_{B \cap C = \emptyset} m(B) m(C)$  denotes the conflicting coefficient. Dempster's combination rule is both commutative and associative.

To make a probabilistic decision, the fused BBA can be transformed into the probability using the Pignistic probability transformation:

$$Betp(\theta_i) = \sum_{\theta_i \in A, A \subseteq \Theta} \frac{m(A)}{|A|}, \forall \theta_i \in \Theta \quad (5)$$

where  $|A|$  denotes the cardinality of  $A$ .

### III. KANG'S BBA DETERMINATION METHOD BASED ON THE INTERVAL NUMBERS' DISTANCES

Using the DST, the determination of the BBAs is the first step, which is still a challenging task. Interval number, which can describe the uncertainty or insufficient information, is useful for determining the BBAs. The definition of interval numbers is as follows: An *interval number*  $\tilde{a}$  in  $\mathbb{R}$  is a set of real numbers that lie between two real numbers, i.e.,  $\tilde{a} = [a^-, a^+] = \{x | a^- \leq x \leq a^+\}$ ,  $a^-, a^+ \in \mathbb{R}$  and  $a^- \leq a^+$ . Kang et al. [6] proposed a BBA determination method based on the interval number models, where the basic beliefs assigned to different focal elements are determined based on the interval numbers' distances between the testing sample and the interval number models of focal elements. Here, we recall the Kang's interval number based BBA determination method first.

Kang's method determines BBAs on different single features respectively. In a single feature, Kang's method models different focal elements (including the focal elements with single class and the focal elements with multiple classes) using interval numbers, and the testing sample is treated as

a degenerate interval (a precise number) with a zero length. Kang's method measures the distances between the testing sample and different interval number models of the focal elements. The testing sample should have a higher similarity degree with the focal element when the distance is small, and the corresponding focal element is assigned a higher basic belief. The steps of Kang's method are described as follows:

- 1) The interval number models of the focal elements with single class are constructed by finding the minimum and the maximum of the corresponding classes' training samples. Then, the interval number models of the focal elements with mixture classes are obtained by finding the overlapping region of the corresponding single classes' interval number models. The interval number models of different focal elements are denoted by  $\tilde{b}_f, f \in 2^\Theta$ .
- 2) Calculate the distances between the testing sample (denoted by  $\tilde{a}$ ) and different focal elements' interval number models, i.e.,  $D(\tilde{a}, \tilde{b}_f), \forall f \in 2^\Theta$ . Note that the length of  $\tilde{a}$  is 0, i.e.,  $a^+ = a^-$ .
- 3) Calculate the similarity degree based on the distances according to Eq. (6).

$$S(\tilde{a}, \tilde{b}_f) = \frac{1}{1 + \alpha D(\tilde{a}, \tilde{b}_f)} \quad (6)$$

where  $\alpha > 0$  is the support coefficient. Empirically, it is proper to set  $\alpha = 5$  [6].

- 4) The BBA is determined by normalizing the similarity degrees of all the focal elements.

Kang's method defines the similarity degrees using interval numbers' distance, and the BBAs are obtained by normalizing the similarity degrees. Thus, the definition of the IND (i.e., the  $D(\tilde{a}, \tilde{b}_f)$ ) is crucial for this method. The differences of the BBAs determined by Kang's method using different INDs are compared in the next section.

### IV. COMPARISONS OF INTERVAL NUMBER BASED BBA DETERMINATION METHOD USING DIFFERENT INDs

As aforementioned, the definition of the IND is crucial for the interval number based BBA determination methods. Many INDs have been proposed. Here, we introduce four widely used INDs.

#### A. Introduction of the interval number's distances

Suppose  $\tilde{a} = [a^-, a^+]$  and  $\tilde{b} = [b^-, b^+]$  are two interval numbers. Then [13], [14],  $\tilde{c} = \tilde{a} \oplus \tilde{b} = [c^-, c^+]$ , where  $c^- = \min(a^-, b^-)$  and  $c^+ = \max(a^+, b^+)$ . The length (or width) of the interval number  $\tilde{a}$  is  $\mu(\tilde{a}) = a^+ - a^-$ .  $D_d$  is the length of the domain [14] of the interval numbers. To measure the difference between two interval numbers, many interval numbers' distances (INDs) have been proposed. Here, we introduce four widely used INDs, which are introduced as follows:

**Gowda and Ravi (1995) [15]:** In 1995 Gowda and Ravi proposed a metric (denoted by GR-IND) combining a position and a size component, as follows

$$D_{GR}(\tilde{a}, \tilde{b}) = D_p(\tilde{a}, \tilde{b}) + D_s(\tilde{a}, \tilde{b}) \quad (7)$$

where the position component is defined as,

$$D_p(\tilde{a}, \tilde{b}) = \cos \left[ \left( 1 - \frac{|a^- - b^-|}{\mu(D_d)} \right) \times \frac{\pi}{2} \right] \quad (8)$$

and the size component is defined as

$$D_s(\tilde{a}, \tilde{b}) = \cos \left[ \frac{\mu(\tilde{a}) + \mu(\tilde{b})}{2 \times \mu(\tilde{a} \oplus \tilde{b})} \times \frac{\pi}{2} \right] \quad (9)$$

**Tran and Duckstein (2002) [16]:** In the framework of fuzzy data analysis, Tran and Duckstein proposed the interval numbers' distance (TD-IND):

$$\begin{aligned} D_{TD}^2(\tilde{a}, \tilde{b}) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ \left[ \frac{1}{2} (a^+ + a^-) + x(a^+ - a^-) \right] \right. \\ &\quad \left. - \left[ \frac{1}{2} (b^+ + b^-) + y(b^+ - b^-) \right] \right\}^2 dx dy \quad (10) \\ &= \frac{1}{4} [(a^- + a^+) - (b^- + b^+)]^2 \\ &\quad + \frac{1}{12} [(a^+ - a^-)^2 + (b^+ - b^-)^2] \end{aligned}$$

**Hausdorff distance [17]:** Considering two sets  $A$  and  $B$  of points of  $\mathbb{R}^n$ , and a distance  $d(x, y)$ , where  $x \in A$  and  $y \in B$ . The Hausdorff distance (H-IND) is defined as follows:

$$D_H(A, B) = \max \left( \sup_{x \in A} \inf_{y \in B} d(x, y), \sup_{y \in B} \inf_{x \in A} d(x, y) \right) \quad (11)$$

If  $d(x, y)$  is the Manhattan distance (also called the City block distance), i.e.,  $d(x, y) = |x - y|$ , then Chavent et al. (2002) proved that

$$D_H(\tilde{a}, \tilde{b}) = \max(|a^- - b^-|, |a^+ - b^+|) \quad (12)$$

**De Carvalho et al. (2006) [18]:** A family of distances between interval numbers has been proposed by De Carvalho et al. based on the bounds of interval numbers. The metric of norm- $q$  (Nq-IND) is defined as:

$$D_{N_q}(\tilde{a}, \tilde{b}) = (|a^- - b^-|^q + |a^+ - b^+|^q)^{\frac{1}{q}} \quad (13)$$

### B. Numerical example

Different INDs can be used for implementing the BBA determinations. Here, we use a numerical example for comparing the interval number based BBA determination methods using different INDs. The BBA determination methods using different INDs are applied on a three-classes classification problem. In this numerical example, we give the features' ranges of different classes directly, as shown in Figure 1, where the feature's range of class 1 ( $\theta_1$ ) is [1, 4], class 2 ( $\theta_2$ ) is [3, 7] and class 3 ( $\theta_3$ ) is [5, 8].

From the Figure 1, the interval numbers models of focal elements can be constructed, which is listed in Table I. Note

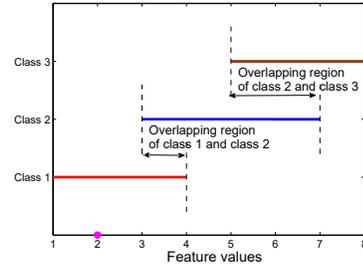


Fig. 1. Feature values' ranges of different classes

TABLE I  
THE INTERVAL NUMBERS MODELS OF FOCAL ELEMENTS.

Focal elements	Interval number model
$\{\theta_1\}$	[1, 4]
$\{\theta_2\}$	[3, 7]
$\{\theta_3\}$	[5, 8]
$\{\theta_1, \theta_2\}$	[3, 4]
$\{\theta_2, \theta_3\}$	[5, 7]
$\{\theta_1, \theta_3\}$	N/A
$\{\theta_1, \theta_2, \theta_3\}$	N/A

that in this example  $\{\theta_1, \theta_3\}$  and  $\{\theta_1, \theta_2, \theta_3\}$  do not have interval number models, because the  $\{\theta_1\}$ 's and  $\{\theta_3\}$ 's interval number models do not have overlapping region.

Suppose we have a testing sample whose feature value is 2, i.e.,  $\tilde{a} = [2, 2]$ , as the purple dot on X-axis of Figure 1. Then we use different INDs, i.e., the GR-IND as in Eq. (7), the TD-IND as in Eq. (10), the H-IND as in Eq. (12), and the Nq-IND as in Eq. (13) (with  $q = 2$  in Nq-IND), for measuring the distance between the  $\tilde{a}$  and different focal elements' interval number models, respectively. The distances are listed in Table II.

TABLE II  
THE INDs BETWEEN THE  $\tilde{a}$  AND FOCAL ELEMENTS' INTERVAL NUMBER MODELS.

Focal elements	GR-IND	TD-IND	H-IND	Nq-IND
$\{\theta_1\}$	0.9296	1.0000	2.0000	2.2361
$\{\theta_2\}$	1.0315	3.2146	5.0000	5.0990
$\{\theta_3\}$	1.5474	4.5826	6.0000	6.7082
$\{\theta_1, \theta_2\}$	1.1464	1.5275	2.0000	2.2361
$\{\theta_2, \theta_3\}$	1.5745	4.0415	5.0000	5.8310

Then, using the distances the similarity degrees are calculated according to Eq. (6), where the support coefficient is set to  $\alpha = 5$ . By normalizing the similarity degrees the BBAs are obtained as listed in Table III.

As the BBAs in Table III, the basic beliefs assigned to different focal elements have small differences using GR-IND compared with that using TD-IND, H-IND and Nq-IND. For example, using GR-IND the basic beliefs assigned to  $\{\theta_1\}$  and  $\{\theta_2\}$  are 0.2552 and 0.2305, which have small differences. Using TD-IND, the basic beliefs of  $\{\theta_1\}$  and  $\{\theta_2\}$  are 0.4086 and 0.1289, whose difference is larger. The BBAs

TABLE III  
THE BBAs DETERMINATED BASED ON DIFFERENT INDS.

Focal elements	BBAs			
	GR-IND	TD-IND	H-IND	Nq-IND
$\{\theta_1\}$	0.2552	0.4086	0.3184	0.3163
$\{\theta_2\}$	0.2305	0.1289	0.1281	0.1394
$\{\theta_3\}$	0.1546	0.0906	0.1069	0.1061
$\{\theta_1, \theta_2\}$	0.2078	0.2693	0.3184	0.3162
$\{\theta_2, \theta_3\}$	0.1519	0.1026	0.1282	0.1220

determined based on H-IND and Nq-IND are similar to each other.

Here, we use the Pignistic probability transformation (as in Eq. (5)) for transforming the BBAs to probabilities for decision making. The probabilities of the testing sample belonging to different classes are listed in Table IV.

TABLE IV  
THE PIGNISTIC PROBABILITIES OBTAINED BASED ON DIFFERENT INDS.

Classes	Pignistic probabilities			
	GR-IND	TD-IND	H-IND	Nq-IND
Class 1 ( $\theta_1$ )	0.3591	0.5433	0.4777	0.4744
Class 2 ( $\theta_2$ )	0.4103	0.3148	0.3514	0.3585
Class 3 ( $\theta_3$ )	0.2306	0.1419	0.1709	0.1671

Intuitively, the testing sample belongs more likely to class 1, as shown in Fig. 1. According to Table IV, the methods using the TD-IND, H-IND and Nq-IND all can make right classifications. According to the probabilities originated from the GR-IND, the testing sample should be classified to class 2. Revisiting the BBA determined based on GR-IND, the basic beliefs assigned to the focal elements with single class has the right tend, i.e.,  $m(\{\theta_1\}) > m(\{\theta_2\}) > m(\{\theta_3\})$ . However, the Pignistic probabilities originated from the GR-IND is counter-intuitive, where the beliefs assigned to the focal elements with multiple classes are counted together. From this perspective, the BBA determined based on GR-IND is not so good. In this numerical example, the interval number based methods using the TD-IND, H-IND and Nq-IND perform more proper for the BBA determination than that using the GR-IND if the decision-making is based on max of BetP.

V. EXPERIMENT

To compare the interval number based BBA determination method using different INDS, we use Monte-Carlo experiments on the classification of the artificial set and the iris set. The information fusion based classification is implemented as follows. In each classification, the interval number based method is used for determining the BBA in each single feature. Then these multiple BBAs are combined using Dempster’s combination rule as in Eq. (4). Then the combined BBA is transformed into probabilities using Pignistic probability transformation as in Eq. (5). The testing sample is classified as the class which has the largest Pignistic probability.

In the experiment, the interval number based methods using different INDS are used for determining the BBAs respec-

tively. In the Nq-IND, we have taken  $q = 2$ . The parameter  $\alpha$  in the generation of the similarity degrees in the interval number based BBA determination method (as in Eq. (6)) is set to 5. The Monte-Carlo classification experiments are repeated 100 times with random testing samples. The effectiveness of the interval number based BBA determination methods using different INDS are compared using the average accuracy of the 100 runs.

A. Experiment on artificial set

The artificial set generated contains 3 classes. Each class has 50 samples, and each sample has 3 features. The features of different classes are generated according to Gaussian distribution, i.e.,  $G(\mu, \sigma^2)$ . The standard deviations ( $\sigma$ ) of different classes’ different features are all set as  $\sigma = 1$ . The mean ( $\mu$ ) settings of different classes’ different features are listed in Table V.

TABLE V  
THE MEAN ( $\mu$ ) SETTINGS OF DIFFERENT CLASSES’ DIFFERENT FEATURES.

Classes	Mean ( $\mu$ )		
	Feature 1	Feature 2	Feature 3
Class 1 ( $\theta_1$ )	8	5	10
Class 2 ( $\theta_2$ )	10	9	6
Class 3 ( $\theta_3$ )	5	11	9

The features of different classes in the artificial set we generated are shown in Figures 2–4.

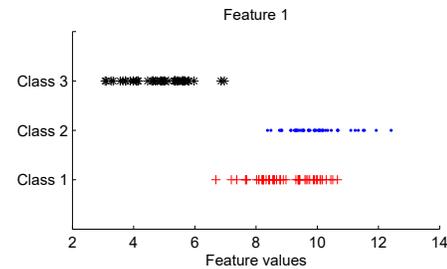


Fig. 2. Artificial samples’ feature 1 of different classes.

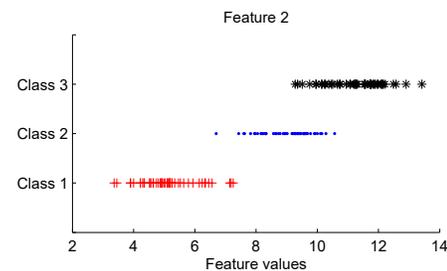


Fig. 3. Artificial samples’ feature 2 of different classes.

As shown in Figures 2–4, the class 3 is linearly separable from class 1 and class 2, and class 1 and class 2 are not linearly separable from each other in feature 1. Similarly, class 2 and

class 3 are not linearly separable from each other in feature 2, and class 1 and class 3 are not linearly separable from each other in feature 3.

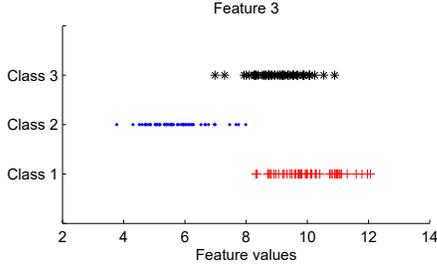


Fig. 4. Artificial samples' feature 3 of different classes.

In each Monte-Carlo run, we randomly select 25 samples from each class (75 samples in total) as the set of training samples, and the remaining samples are used as the testing samples. We first classify the testing sample according to the BBA determined based on each single feature, respectively. Then, we combine the BBAs determined based on the 3 features, and use the combined BBA for classifying the testing sample. The results of the methods based on different INDs are listed in Table VI.

TABLE VI  
THE RESULTS OF THE METHODS BASED ON DIFFERENT INDs.

INDs	Classification correct rate (%)			
	Feature 1	Feature 2	Feature 3	Combined
GD-IND	44.70	64.86	42.62	80.95
TD-IND	67.71	84.13	61.66	94.84
H-IND	64.66	80.24	56.01	89.66
Nq-IND	65.86	81.68	55.84	91.97

In Table VI, the columns “Feature 1”, “Feature 2” and “Feature 3” are the results of the methods using different INDs based on each single features. The column “Combined” are the results obtained by combining the BBAs determined on different features with Dempster’s rule of combination. According to Table VI, the classifications of the methods using different INDs based on each single feature does not perform well. However, the BBAs determined based on different features reflect different aspects’ information of the samples. By fusing the BBAs based on different features, better classification performances are obtained. Comparing the results of the methods based on different INDs, the method based on GD-IND performs the worst. The performances of the methods based on TD-IND, H-IND and Nq-IND are similar, where the one based on TD-IND is the best. The BBA built using the GD-IND is not recommended for the BBA determination.

### B. Experiment on iris set

The iris set contains 3 classes. Each class has 50 samples, and each sample has 4 features. In this experiment, we randomly select different numbers of samples as the training samples (the number of the samples selected from different

classes are the same), and all the samples are used as the testing samples. The results of the interval number based BBA determination methods based on different INDs are shown in Figure 5.

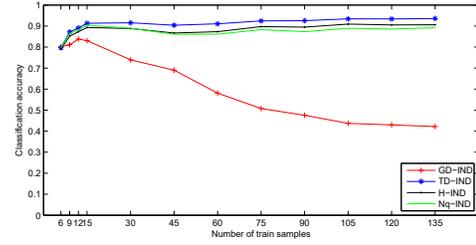


Fig. 5. Performances of the interval number based methods using different INDs with different scales of training samples on iris data set.

According to Figure 5, the methods using TD-IND, H-IND and Nq-IND perform well in both the cases with small number of training samples and large number of training samples. The method using TD-IND performs the best compared with the methods using other three INDs. The results of the method using GD-IND have a counter-intuitive behavior, since its accuracy decreases with the increasing of the number of the training samples. When the number of training samples is large, the interval numbers generated can better model the features of corresponding classes, especially, for the mixture classes’ focal elements (i.e., the overlapping range of corresponding classes’ interval number models). However, as discussed in the numerical example in section IV-B, the interval number based method using GD-IND is not recommended for determining the BBA, especially, counting the mixture class focal elements together. That is why the method using GD-IND performs bad when the number of training samples is large.

## VI. CONCLUSION

In this paper, we have tested different INDs for implementing the interval number based BBA determination method. The effectiveness of the BBAs are compared based on the information fusion based classification problems. The experiments validate that combining the BBAs determined using interval number based methods with different INDs performs well for the classification problems. The methods using the TD-IND, H-IND and Nq-IND provide quasi similar performances, where the one using TD-IND is the best one. Using the GD-IND, the basic beliefs construction is not very effective. With GD-IND, the differences of the basic beliefs assigned to different focal elements are small, which is not discriminant enough for making decisions, especially, counting the mixture classes’ focal elements. Therefore, the method using the GD-IND is not recommended.

Up to now, the interval number based BBA determination methods are implemented on the single feature. In future work, we will try to use the interval numbers for determining the BBAs on the multiple features spaces, and compare the effectiveness of the ones using different INDs. We will explore

also different decision-making strategies (i.e. DSMP, min of  $d_{BI}$ , etc.), and test other rules of combination as well to see if we can improve classification performances.

#### ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation (Nos. 61573275, 61671370), Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), Science and Technology Project of Shaanxi Province (No. 2013KJXX-46), Postdoctoral Science Foundation of China (No. 2016M592790), and Fundamental Research Funds for the Central Universities (No. xjj2014122, xjj2016066).

#### REFERENCES

- [1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann Mathematical Statistics*, vol. 38, pp. 325–339, 1967.
- [2] G. Shafer, *A Mathematical Theory of Evidence*, Princeton: Princeton University, 1976.
- [3] F. Selzer and G. Dan, "LADAR and FLIR based sensor fusion for automatic target classification," *Robotics Conferences International Society for Optics and Photonics*, 1989, pp. 236–246.
- [4] F. Valente, H. Hermansky, "Combination of acoustic classifiers based on Dempster-Shafer theory of evidence," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Martigny, Switzerland, 2007, pp. IV–1129–IV–1132.
- [5] J. Dezert, Z. G. Liu, G. Mercier, "Edge detection in color images based on DSMT," *The 14th Int Conf on Information Fusion*, Chicago, USA, 2011, pp. 969–976.
- [6] B. Y. Kang, L. I. Ya, Y. Deng, et al., "Determination of basic probability assignment based on interval numbers and its application," (in Chinese) *Acta Electronica Sinica*, vol. 40, no. 6, pp. 1092–1096, 2012.
- [7] Y. X. Bi, D. Bell, J. W. Guan, "Combining evidence from classifiers in text categorization," *Proc of the 8th Int Conf. on KES*, Wellington, New Zealand, 2004, pp. 521–528.
- [8] F. Salzenstein, A. O. Boudraa, "Iterative estimation of Dempster Shafers basic probability assignment: Application to multisensor image segmentation", *Optical Engineering*, vol. 43, no. 6, pp. 1293–1299, 2004.
- [9] Y. Deng, W. Jiang, X. B. Xu, et al., "Determining BPA under uncertainty environments and its application in data fusion", *Journal of Electronics*, vol. 26, no. 1, pp. 13–17, 2009.
- [10] A. O. Boudraa, A. Bentabet, F. Salzenstein, et al. "Dempster-Shafers basic probability assignment based on fuzzy membership functions", *Electronic Letters on Computer Vision and Image Analysis*, vol. 4, no. 1, pp. 1–9, 2004.
- [11] M. C. Florea, A. L. Jusselme, D. Grenier, et al., "Approximation techniques for the transformation of fuzzy sets into random sets," *Fuzzy Sets Systems* vol. 159, no. 3, pp. 270–288, 2008.
- [12] D. Q. Han, Y. Deng, C. Z. Han, "Novel approaches for the transformation of fuzzy membership function into basic probability assignment based on uncertain optimization." *Int J of Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 21, no. 2, pp. 289–322, 2013.
- [13] R. Moore, R. Kearfott, and M. Cloud, *Introduction to interval analysis*, Philadelphia: Society for Industrial and Applied Mathematics, 2009.
- [14] A. Irpino and R. Verde, "Dynamic clustering of interval data using a Wasserstein-based distance," *Pattern Recognition Letters*, vol. 29, pp. 1648–1658, April 2008.
- [15] K. C. Gowda and T. V. Ravi, "Agglomerative clustering of symbolic objects using the concepts of both similarity and dissimilarity," *Pattern Recognition Letters*, vol. 16, no. 6, pp. 647–652, 1995.
- [16] L. Tran and L. Duckstein, "Comparison of fuzzy numbers using a fuzzy distance measure," *Fuzzy Sets & Systems*, vol. 130, no. 3, pp. 331–341, 2002.
- [17] M. Chavent and Y. Lechevallier, "Dynamical clustering of interval data: optimization of an adequacy criterion based on Hausdorff distance," *Journal of Classification*, vol. 90, no. 8, pp. 53–60, 2002.
- [18] F. D. Carvalho, P. Brito, H. H. Bock, "Dynamic clustering for interval data based on  $L_2$  distance," *Computational Statistics*, vol. 21, no. 2, pp. 231–250, 2006.

# Time-series 3D Building Change Detection Based on Belief Functions

Jiaojiao Tian<sup>a</sup>, Jean Dezert<sup>b</sup>, Rongjun Qin<sup>c</sup>

<sup>a</sup>Remote Sensing Technology Institute, German Aerospace Center, 82234 Oberpfaffenhofen, Germany.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>c</sup>The Ohio State University, Columbus, OH 43210, USA.

Emails: jiaojiao.tian@dlr.de, jean.dezert@onera.fr, qin.324@osu.edu

Originally published as: J. Tian, J. Dezert, R. Qin, *Time-series 3D Building Change Detection Based on Belief Functions*, in Proc. of Int. Conf. on Information Fusion (Fusion 2018), Cambridge, UK, July 10–13, 2018, and reprinted with permission.

**Abstract**—One of the challenges of remote sensing image based building change detection is distinguishing building changes from other types of land cover alterations. Height information can be a great assistance for this task but its performance is limited to the quality of the height. Yet, the standard automatic methods for this task are still lacking. We propose a very high resolution stereo series data based building change detection approach that focuses on the use of time series information. In the first step, belief functions are explored to fuse the change features from the 2D and height maps to obtain an initial change detection result. In the second step, the building probability maps (BPMs) from the series data are adopted to refine the change detection results based on Dempster-Shafer theory. The final step is to fuse the series building change detection results in order to obtain a final change map. The advantages of the proposed approach are demonstrated by testing it on a set of time series data captured in North Korea.

**Keywords:** Change detection, belief functions, DST, DSM..

## I. INTRODUCTION

Building change detection is one of the fundamental remote sensing research topics. Although many approaches are available, it is still very difficult to select one standard approach that works for all situations. Especially along with the improvement of the image resolution, besides building changes many irrelevant changes may also be visible in the remote sensing images, which makes the 2D building change detection more challenge. 3D building change detection has gained a great attention and is able to provide more accurate results. Due to the unprecedented technology development of sensor, platforms and algorithms for 3D data acquisition and generation, the 3D data become more accessible than before. Stereo time series data will allow a better understanding of the building change types and further increase the change accuracy.

Many research works have proved the advantages of introducing Digital Surface Models (DSM) to building change detection [1]–[2]. However, the performance of the 3D change detection approaches rely heavily on the quality of the DSMs. And the DSMs from satellite images do not always provide reliable height information, due to the occlusion and matching errors. In the case of large regions have incorrect height values, it is very difficult to avoid false detections. In our previous research, time-series information worked well to improve the building detection results. In this research, we will further adopt this information to improve the change detection results.

In paper [3], the belief functions introduced in the Dempster-Shafer Theory (DS) [4]–[5], and extended in Dezert-Smarandache Theory (DSMT) [6] were used to deal with the uncertainty information delivered from the DSMs. In [3] the possibility of using Dempster’s fusion rule and the Proportional Conflict Redistribution Rule #6 (PCR6) of DSMT in our application were tested. Though improvements have been proven by comparing with the method stated in [7], the results delivered under DS and DSMT frameworks were rather similar. Therefore, in this paper, only the DS fusion rule is used to get an initial change detection result.

This paper is organized as follow: firstly, the belief functions and building change detection fusion models are briefly reviewed. Then, the series image based fusion model together with the building extraction method are introduced. In the end, these refined fusion models are tested on the satellite real data.

## II. DS BELIEF FUNCTION BASED BUILDING CHANGE DETECTION

### A. Basics of DST

Dempster-Shafer fusion (DST) is one of the fundamental decision fusion theory. It allows the combination of evidence from individual experts or any data sources. The general introduction of DST can be found in [4], [6], and [8].

Let  $\Theta$  be a frame of discernment of a problem under consideration.  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  consists of a list of  $N$  exhaustive and mutually exclusive elements  $\theta_i$ ,  $i = 1, 2, \dots, N$ . Each  $\theta_i$  represents a possible state related to the problem we want to solve. The assumption of exhaustivity and mutual exclusivity of elements of  $\Theta$  is classically referred as *Shafer’s model* of the frame  $\Theta$ . A basic belief assignment (BBA) also called a belief mass function (or just a mass for short), is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  from the power set<sup>1</sup> of  $\Theta$  denoted  $2^\Theta$  to  $[0, 1]$ , that verifies [4]:

$$m(\emptyset) = 0, \quad \text{and} \quad \sum_{X \in 2^\Theta} m(X) = 1. \quad (1)$$

$m(X)$  represents the mass of belief exactly committed to  $X$ . An element  $X \in 2^\Theta$  is called a focal element (FE) if and only if  $m(X) > 0$ . In DST, the combination (fusion) of several independent sources of evidences is done with Dempster-Shafer<sup>2</sup> (DS) rule of combination, assuming that

<sup>1</sup>The power set is the set of all subsets of  $\Theta$ , including empty set.

<sup>2</sup>Although the rule has been proposed originally by Dempster, we call it Dempster-Shafer rule because it has been widely promoted by Shafer in DST.

the sources are not in total conflict<sup>3</sup>. DS combination of two independent BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$ , denoted symbolically by  $DS(m_1, m_2)$ , are defined by  $m^{DS}(\emptyset) = 0$ , and for all  $X \in 2^\Theta \setminus \{\emptyset\}$  by:

$$m^{DS}(X) = \frac{1}{1 - K^{DS}} \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2), \quad (2)$$

where the total degree of conflict  $K^{DS}$  is given by

$$K^{DS} \triangleq \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2). \quad (3)$$

### B. Building change detection

1) *Choice of the frame of discernment*: As noted above, the accuracy of 2D change detection is limited due to the misdetections caused by irrelevant changes. These irrelevant changes have a greater effect on very high resolution (VHR) images since more detail on land-cover objects is visible. To solve this problem, in the decision fusion based 3D change detection framework, three classes have been considered. They are,

$$\begin{aligned} \Theta = \{ & \theta_1 \triangleq \text{Pixel} \in \text{BuildingChange}, \\ & \theta_2 \triangleq \text{Pixel} \in \text{OtherChange}, \\ & \theta_3 \triangleq \text{Pixel} \in \text{NoChange} \}, \end{aligned} \quad (4)$$

and

$$\theta_1 \cap \theta_2 \cap \theta_3 = \emptyset. \quad (5)$$

Based on the three classes, the set of focal elements  $FE$  that are of interest in our application is:

$$FE = \{\theta_1, \theta_2, \theta_3, \theta_1 \cup \theta_2, \theta_2 \cup \theta_3, \theta_1 \cup \theta_2 \cup \theta_3\}. \quad (6)$$

Two change indicators, one from images and one from DSMs were involved in the fusion model. Changes from spectral images are highlighted by using the Iteratively Reweighted Multivariate Alteration Detection (IRMAD) [9]. Consequently height changes from DSMs are shown after robust height differencing [7].

2) *BBAs construction*: In [3], the sigmoidal model for both concordance and discordance indexes are constructed by projecting the change values to a sigmoid curve  $f_{\tau, T}$ .  $T$  represents the symmetry point of the sigmoid curve, while the  $\tau$  control the slope of it. The concordance index measures the concordance of change indicator and BBA in the assertion, while the discordance measures the opposition of change indicator to the BBAs in the assertion. The symmetry point of the concordance and discordance sigmoid curves can be automatically calculated with multi-level thresholding method proposed by Otsu [10].

Thus, using height change index as example, the BBAs for discordance and concordance height change index are functions of values  $a_{\Delta H}$  and  $b_{\Delta H}$  defined by

$$a_{\Delta H} = f_{\tau, T_1}(\Delta H), \quad \text{and} \quad b_{\Delta H} = f_{-\tau, T_2}(\Delta H). \quad (7)$$

<sup>3</sup>Otherwise DS rule is mathematically undefined because of 0/0 indeterminacy.

The factor  $\tau$  could be calculated with a sample value ( $\Delta H = 1$ ,  $a_{\Delta H} = 0.1$ ), which means 1 meter height change indicates 10% probability to be building changes. The BBAs for discordance and concordance image change index are built similarly. Differences appearing in 2D images give a concordance indication for all changes, which include the building changes and other changes ( $\theta_1 \cup \theta_2$ ). In this paper the changes from images are named  $\Delta Img$ .

In [3], the fusion models have been described in detail. Here we only explain the fusion model of the height changes as an example. In Table I the construction of the BBAs from the sources of evidence based DS rule of combination for the height change indicator (i.e. the first source of evidence). In Table I,  $m_1(\cdot)$  and  $m'_1(\cdot)$  represent the concordance and discordance BBAs from  $\Delta H$ .

TABLE I. BBA CONSTRUCTION FOR HEIGHT CHANGE INDICATOR  $\Delta H$ .  
[ $K_{\Delta H} = a_{\Delta H}b_{\Delta H}$ ]

Focal Elem.	$m_1(\cdot)$	$m'_1(\cdot)$	$m^{DS}(\cdot)$
$\theta_1$	$a_{\Delta H}$	0	$\frac{a_{\Delta H}(1-b_{\Delta H})}{1-K_{\Delta H}}$
$\theta_2$	0	0	0
$\theta_3$	0	0	0
$\theta_1 \cup \theta_2$	0	0	0
$\theta_2 \cup \theta_3$	0	$b_{\Delta H}$	$\frac{(1-a_{\Delta H})b_{\Delta H}}{1-K_{\Delta H}}$
$\theta_1 \cup \theta_2 \cup \theta_3$	$1 - a_{\Delta H}$	$1 - b_{\Delta H}$	$\frac{(1-a_{\Delta H})(1-b_{\Delta H})}{1-K_{\Delta H}}$

### III. TIME SERIES FUSION MODEL

To further improve the accuracy of the change detection map, the pre- and post-event building probability maps are introduced to the decision fusion model. The building probability maps are prepared using our previous research results.

#### A. Time-series based building probability map extraction

The building extraction method based on spatiotemporal inferences are adopted to prepare the building probability maps (BPM) [11]. The approach is mainly composed of three steps: (1) training sample selection; (2) feature extraction and classification; (3) spatiotemporal based BPM refinement.

1) *training sample selection*: Training sample selection is a time consuming and tedious process, which should be avoided for the automatic image processing chain. In this step, we are trying to produce the training data automatically from history database. More precisely, only one set of training data containing of building, ground&road, shadow and trees was manually annotated. Training data for the images captures of other dates can be automatically generated by using a decision based change detection approach. The normalized DSM is used to separate the above ground object from ground and road. And the height changes, shadow index changes and the normalized difference vegetation index (NDVI) changes are used to generate a coarse change maps, thus update the training data. More details are described in paper [11].

2) *feature extraction and classification*: Based on the training data, the Random Forests (RF) [12] supervised classifier was adopted. The features extracted for the classification task include: 1) Principal component analysis [13] transformation components of the multispectral channels; 2) Differential morphological profile [14] of the panchromatic image; 3)

Normalized DSM. The RF classifier took the normalized feature vectors as input for a pixel-based classification. Besides the final classification results, the confidence values of each class label were given to each pixel. Thus a BPM map were generated for each dataset among the time-series images.

3) *Spatiotemporal based BPM refinement*: As it was mentioned in [11], the BPMs across all dates may be not consistent due to some potentially imprecise training samples, or unwanted objects on optical images, such as cloud/snow covered regions. In addition, as a general drawback of the pixel-based classification approaches the salt-and-pepper effect exists in the results. Thus a consistency check through spatial and temporal domain would be helpful to improve the final result. The basic idea of this approach could be explained through Eq. (8).

$$P_f(x, y, t) = \frac{1}{\sum_{m=x+l}^{m=x-l} \sum_{n=y+l}^{n=y-l} \sum_{k=1}^h w(m, n, k) P(m, n, t)}, \quad (8)$$

where  $P(m, n, t)$  is the BPM at time  $t$ . The refined BPM is recorded as  $P_f(x, y, t)$ . A window size  $(2 \times l + 1)^2$  is used for the spatial consistence check. We have used  $l = 7$  in [11].  $h$  is the number of temporal data set.  $w(m, n, k)$  is the 3D adaptive kernel, which aims to balance the similarity and distance of the neighboring pixels in three dimensions.

### B. DS based change map refinement

One of the main advantages of DST lies in the handling indicators from various sources flexibly. Benefit from the previous steps, the refined BPMs from each time would be delivered. And a building change probability map can be calculated by using the approach from Section II. Based on the building change detection approach, we obtain a building change probability map in which all pixels represent a probability that pixel were classified as building change. Thus, when comparing two datasets the available indicators would be,

- pre-event BPM ( $P_{pre}$ )
- post-event BPM ( $P_{after}$ )
- initial building change probability map ( $P_{BC}$ )

four To model this situation more precisely, we categorize the change situations into four groups, which are *buildings* to *buildings* (BB), *non-building* to *buildings* (NB), *buildings* to *non-building* (BN) and *non-building* to *non-building* (NN). Based on these four classes and the indicators, the *FE* set that are of interest in this fusion model is,

$$FE = \{BB, NB, BN, NN, BB \cup BN, BB \cup NB, \Theta\}. \quad (9)$$

The probability masses  $\{P_1, P_2, P_3\}$  obtained respectively from these three indicators are assigned to *FE* as shown in Table. II. One of the basis principles is that all newly built buildings should have a lower value in the pre-event BPM, and a higher value in the post-event BPM. Based on the DS fusion rule, the fused masses are listed in the last column.

TABLE II. DS FUSION MODEL FOR RESULT REFINEMENT  
[ $K = P_1 * P_3$ ].

FE.	$P_{pre}$	$P_{after}$	$P_{BC}$	Fused mass
BB	0	0	0	$\frac{P_1 * P_2 * (1 - P_3)}{1 - K}$
NB	0	0	$P_3$	$\frac{(1 - P_1) * P_3}{1 - K}$
BN	0	0	0	0
NN	0	0	0	0
BB $\cup$ BN	$P_1$	0	0	$\frac{P_1 * (1 - P_2) * (1 - P_3)}{1 - K}$
BB $\cup$ NB	0	$P_2$	0	$\frac{(1 - P_1) * P_2 * (1 - P_3)}{1 - K}$
$\Theta$	$1 - P_1$	$1 - P_2$	$1 - P_3$	$\frac{(1 - P_1) * (1 - P_2) * (1 - P_3)}{1 - K}$

### C. Time-series fusion model

The previous fusion steps are performed to each multi-temporal data pair separately. They can be then combined in the time-series fusion model. In this section, we use three datasets captured from three dates as an example to describe our fusion model. Three datasets are notated as  $d_1$ ,  $d_2$  and  $d_3$ , respectively. They are arranged according to the acquisition time.  $d_1$  is the oldest dataset. Then the building change detection (BC) outcomes among these datasets can be recoded as  $BC_{12}$ ,  $BC_{23}$  and  $BC_{13}$ , respectively.

By referring to the fusion model in Table II, a global mass  $\{m_{BB}, m_{NB}, m_{BN}, m_{NN}, m_{BB \cup BN}, m_{BB \cup NB}, m_{\Theta}\}$  can be obtained. We will transform the global mass to a three-classes FE mass on  $NB$ ,  $\overline{NB}$ , and  $\Theta$  by coarsening the original set of focal elements. For this, we apply the following transformation which can be seen as a partial pignistic transformation. In the pignistic probabilities [4], global masses of joint elements are averagely redistributed to each class. Since the full ignorance  $\Theta$  is one focal elements in the three-classes FE, we will keep this value to  $m(\Theta)$  and only take the partial ignorance when calculating the  $m'(\cdot)$  as shown in Eq. (10).

$$\begin{aligned} m'(NB) &= m_{NB} + \frac{1}{2}m_{BB \cup NB}, \\ m'(BB) &= m_{BB} + \frac{1}{2}m_{BB \cup NB} + \frac{1}{2}m_{BB \cup BN}, \\ m'(BN) &= m_{BN} + \frac{1}{2}m_{BB \cup BN}, \\ m'(NN) &= 0, \\ m'(\Theta) &= m(\Theta). \end{aligned} \quad (10)$$

To clarify the notation, we use  $a$ ,  $b$  and  $c$  to represent the mass values for  $NB$ ,  $\overline{NB}$  and  $\Theta$ , respectively. With  $m'(NN) = 0$ , they can be calculated with Eq. (11). Thus the building change detection results from  $NB_{12}$ ,  $NB_{23}$  and  $NB_{13}$ , denoted as  $P_{NB12}$ ,  $P_{NB23}$  and  $P_{NB13}$ , can be fused according to the fusion model shown in Table III.

TABLE III. THE INDICATORS FOR TIME SERIES FUSION MODEL.

FE.	$P_{NB12}$	$P_{NB23}$	$P_{NB13}$
$C_1$	$P_{1a}$	0	0
$C_2$	0	$P_{2a}$	0
$C_3$	0	0	$P_{3b}$
$C_1 \cup C_2$	0	0	$P_{3a}$
$C_2 \cup C_3$	$P_{1b}$	0	0
$C_1 \cup C_3$	0	$P_{2b}$	0
$\Theta$	$P_{1c}$	$P_{2c}$	$P_{3c}$

In this fusion model, the focal three change classes would be, changes happened between  $d_1$  and  $d_2$ , notated as  $C_1$ ; changes happened between  $d_2$  and  $d_3$ , notated as  $C_2$ ;

and no-building change, notated as  $C_3$ .

$$\begin{aligned} a &= m'(NB), \\ c &= m'(\Theta), \\ b &= m'(\overline{NB}) = 1 - m'(NB) - m'(\Theta). \end{aligned} \tag{11}$$

With the conjunction rules, the following belief masses will be obtained,

$$\begin{aligned} m(C_1) &= P_{1a}P_{2c}P_{3a} + P_{1a}P_{2c}P_{3c} + P_{1a}P_{2b}P_{3a} \\ &\quad + P_{1a}P_{2b}P_{3c} + P_{1c}P_{2b}P_{3a}, \\ m(C_2) &= P_{1b}P_{2a}P_{3a} + P_{1b}P_{2a}P_{3c} + P_{1b}P_{2c}P_{3a} \\ &\quad + P_{1c}P_{2a}P_{3a} + P_{1c}P_{2a}P_{3c}, \\ m(C_3) &= P_{1b}P_{2b}P_{3b} + P_{1b}P_{2c}P_{3b} + P_{1c}P_{2b}P_{3b} \\ &\quad + P_{1c}P_{2c}P_{3b} + P_{1b}P_{2b}P_{3c}, \\ m(C_1 \cup C_2) &= P_{1c}P_{2c}P_{3a}, \\ m(C_2 \cup C_3) &= P_{1b}P_{2c}P_{3c}, \\ m(C_1 \cup C_3) &= P_{1c}P_{2b}P_{3c}, \\ m(\Theta) &= P_{1c}P_{2c}P_{3c}, \\ K &= 1 - m(C_1) - m(C_2) - m(C_3) - m(C_1 \cup C_2) \\ &\quad - m(C_2 \cup C_3) - m(C_1 \cup C_3) - m(\Theta). \end{aligned} \tag{12}$$

in which  $K$  represents the mass of conflict.

Based on the DS fusion rule, the final mass will be calculated by

$$m_{DS}(X) = \frac{m(X)}{1 - K}, \tag{13}$$

for  $X \in \{m_{BB}, m_{NB}, m_{BN}, m_{NN}, m_{BB \cup BN}, m_{BB \cup NB}, m_{\Theta}\}$ , and  $X \neq \emptyset$ ; and  $m_{DS}(\emptyset) = 0$ .

#### IV. EXPERIMENTS

The improved building change detection fusion models have been tested on satellite images. The datasets and the experiments are described in this section.

##### A. Datasets

The experimental datasets consist of five pairs of IKONOS and one pair of GeoEye-1 stereo imagery captured from 2006 to 2011. The detailed capture dates of these data are shown in Table IV. The true color images of the earliest and latest datasets are shown in Fig. 1 (a) and 1 (b), respectively. Within these five years, many new buildings are constructed in this test region. As a data preparation procedure, DSMs have been generated based on the method explained in [15].

TABLE IV. TIME SERIES DATASETS DESCRIPTION.

No.	Satellite	Capture date	Resolution (m)	
			PAN	MS
1	IKONOS	23-02-2006	1	4
2	GeoEye-1	20-12-2009	0.5	2
3	IKONOS	12-01-2010	1	4
4	IKONOS	13-05-2010	1	4
5	IKONOS	07-01-2011	1	4
6	IKONOS	02-05-2011	1	4

The sub-pixel co-registration among these data is performed based on the camera model parameters correction [11]. The radiometric co-registration method is described in [7].

##### B. Results and evaluation

In the first change detection step, the data from 2006 are used as the pre-event test data. The rest five datasets are the post-event dataset. Thus five change detection case studies are prepared and named as  $C_{06-09}$ ,  $C_{06-1001}$ ,  $C_{06-1005}$ ,  $C_{06-1101}$ ,  $C_{06-1105}$ , correspondingly. The change indicators from DSMs and images are detected respectively by using robust height differences and IRMAD. The change maps are recorded as  $H_{diff}$  and  $Img_{diff}$ . At the same time, BPMs from all six datasets are calculated and refined with the approach described in section III-A. Then the change detection result between each pairs of datasets is refined by using the pre-event and post-event BPMs.

To evaluate quantitatively the performances of the different fusion approaches, the extracted BBAs from both approaches (original and refined) are compared to the manually extracted change reference masks. The results are analyzed in terms of Receiver Operating Characteristic (ROC) curve [16]. A larger area under the ROC curve (AUC) indicates a better accuracy of the building change map. The numerical evaluation results are described in Table V. The obtained AUC values prove an obvious accuracy improvement after the proposed fusion model is applied. The  $m'(NB)$  is used as the first-step change detection results, and listed as  $Refined_1$ .

TABLE V. BUILDING CHANGE DETECTION ACCURACY COMPARISON.

Change maps	$C_{06-09}$	$C_{06-1001}$	$C_{06-1005}$	$C_{06-1101}$	$C_{06-1105}$
$H_{diff}$	0.9267	0.9233	0.9016	0.8289	0.8211
$Img_{diff}$	0.9049	0.5937	0.9004	0.8283	0.8610
<i>Fusion</i>	0.9540	0.9271	0.9474	0.8885	0.8862
<i>Refined<sub>1</sub></i>	0.9771	0.9744	0.9668	0.9241	0.9442

In the time-series fusion model, we have tested the data from 2006, 2009 and 2010 May as a test combination. The further improved building change probability map ( $C_{06-09}$ ) with (AUC=0.9795) is delivered. The differences between the original change detection result and the refined one can be observed in Fig. 2.

#### V. CONCLUSIONS

Detecting building changes is an important but difficult topic. Many approaches have been proposed for specific building types or for certain types of data sets. In addition, the image quality and the existing of some unwanted objects may also influence the effectiveness of some approaches. Our previous research has evidenced the performance of the belief functions in DSM assisted change detection [3]. In this paper, we have further explored in more detail the belief functions for building change detection. Time-series data were used for this purpose. They were firstly adopted to provide the BPMs after checking the temporal consistency for each date. Then, pre- and post-event BPMs are used to improve the accuracy of the DS-based building detection result. In the last step, the time-series change detection results can be fused again according to the DS fusion rule, which results a further improved building change map. However, in the time-series fusion model, only three sets of data are involved. As part of our further work, this fusion model will be further refined that may accommodate more datasets as inputs.

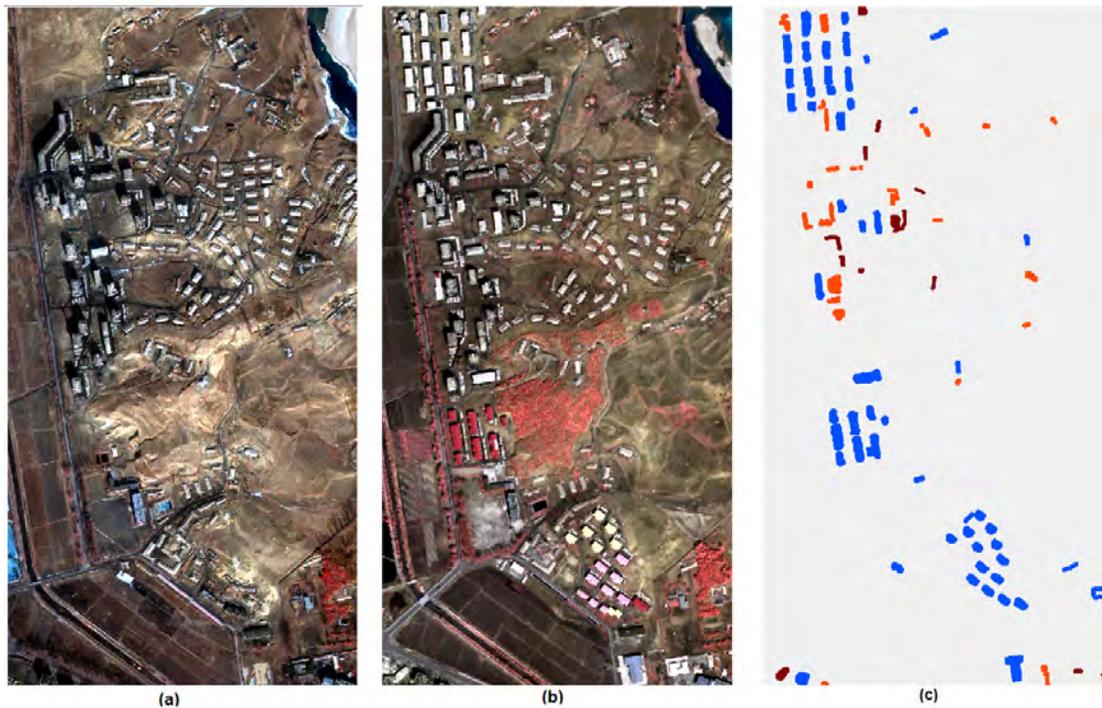


Fig. 1. The true color images of the first (a) and sixth (b) experimental dataset and (c) the change reference map (Blue: built before 2009; Orange: built before January 2011; Red: built before May 2011)

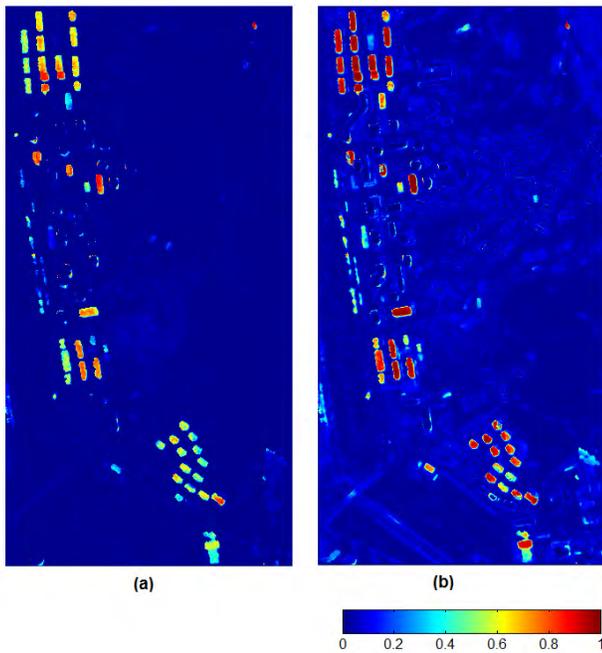


Fig. 2. The building change detection results between 2006 and 2009 based on (a) initial fusion model (b) time-series fusion model.

REFERENCES

[1] R. Qin, J. Tian, P. Reinartz, *3D change detection—approaches and applications*, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 122, pp. 41–56, 2016.  
 [2] J. Tian, P. Reinartz, P. d’Angelo, M. Ehlers, *Region-based automatic building and forest change detection on cartosat-1 stereo imagery*, ISPRS J. Photogramm. and Remote Sens., Vol. 79, pp. 226–239, 2013.  
 [3] J. Tian, P. Reinartz, J. Dezert, *Building change detection in satellite*

*stereo imagery based on belief functions*, in Urban Remote Sensing Event (JURSE), 2015 Joint. IEEE, pp. 1–4, 2015.  
 [4] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, NJ, USA, 1976.  
 [5] A. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Annals of Mathematical Stat., Vol. 38(2), pp. 325–339, 1967.  
 [6] F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion*, Vol. 1–4, American Research Press, Rehoboth, NM, USA, 2004–2015. <http://www.onera.fr/staff/jean-dezert?page=2>  
 [7] J. Tian, S. Cui, P. Reinartz, *Building change detection based on satellite stereo imagery and digital surface models*, IEEE Trans. Geosci. Remote Sens., Vol. 52(1), pp. 406–417, 2014.  
 [8] J. Dezert, A. Tchamova, *On the validity of Dempster’s fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. Intell. Syst., Vol. 29(3), pp. 223–252, 2014.  
 [9] A.A. Nielsen, *The regularized iteratively reweighted MAD method for change detection in multi- and hyperspectral data*, IEEE Trans. Image Process., Vol. 16(2), pp. 463–478, 2007.  
 [10] N. Otsu, *A threshold selection method from gray-level histograms*, IEEE Trans. Syst., Man, Cybern., Vol. 9(1), pp. 62–66, 1975.  
 [11] R. Qin, J. Tian, P. Reinartz, *Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images*, International Journal of Remote Sensing, Vol. 37(15), pp. 3455–3476, 2016.  
 [12] L. Breiman, *Random forests*, Machine learning, Vol. 45(1), pp. 5–32, 2001.  
 [13] I.T. Jolliffe, *Principal component analysis and factor analysis*, in Principal component analysis, Springer, pp. 115–128, 1986.  
 [14] J.A. Benediktsson, M. Pesaresi, K. Amason, *Classification and feature extraction for remote sensing images from urban areas based on morphological transformations*, IEEE Trans. on Geoscience and Remote Sensing, Vol. 41(9), pp. 1940–1949, 2003.  
 [15] P. d’Angelo, P. Reinartz, *DSM based orientation of large stereo satellite image blocks*, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. 39(B1), pp. 209–214, 2012.  
 [16] M.H. Zweig, G. Campbell, *Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine*, Clinical chemistry, Vol. 39(4), pp. 561–577, 1993.



# Belief Interval Based Distance Measures in the Theory of Belief Functions

Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>

<sup>a</sup>School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049 China.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>c</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an, Shaanxi 710049 China.

Emails: deqhan@gmail.com, jean.dezert@onera.fr, jiafeiyy@mail.xjtu.edu.cn

Originally published as: D. Han, J. Dezert, Y. Yang, *Belief interval Based Distances Measures in the Theory of Belief Functions*, IEEE Trans. on SMC, Vol. 48(6), pp. 833–850, June 2018, and reprinted with permission.

**Abstract**—In belief functions related fields, the distance measure is an important concept, which represents the degree of dissimilarity between bodies of evidence. Various distance measures of evidence have been proposed and widely used in diverse belief function related applications, especially in performance evaluation. Existing definitions of strict and non-strict distance measures of evidence have their own pros and cons. In this paper, we propose two new strict distance measures of evidence (Euclidean and Chebyshev forms) between two basic belief assignments (BBAs) based on the Wasserstein distance between belief intervals of focal elements. Illustrative examples, simulations, applications and related analyses are provided to show the rationality and efficiency of our proposed measures for distance of evidence.

**Keywords:** distance of evidence, belief functions, evidence theory, dissimilarity, belief interval.

## I. INTRODUCTION

The theory of belief functions, also called Dempster-Shafer evidence theory (DST) [1], is an important mathematical framework for uncertainty modelling and reasoning. It has been applied to information fusion [2], pattern recognition [3] [4], multiple-attribute decision making [5], fault diagnosis [6], etc. DST has some limitations, see [7]–[9] for discussions. Generalized or refined theories were proposed including transferable belief model (TBM) [10] and Dezert-Smarandache Theory (DSMT) [7], [11], etc.

In DST, the basic belief assignment (BBA) is a common way for modeling (epistemic) uncertainty. The distance of evidence is a crucial metric for measuring the distance between two BBAs. It indicates a BBA is “far” from or “close” to another one. In many belief functions related applications, the distance of evidence is required. Such belief function-related applications can be categorized into two types. The first type is the performance evaluation or optimization [12]–[16]. For example, in the performance evaluation of BBA approximation [16], which aims to simplify the BBA to reduce the computational complexity, the distance of evidence is needed to measure accuracy of an approximated BBA (the one closer to the original BBA is better). Furthermore, some BBA approximation approach is directly based on the distance minimization [17], therefore, the distance of evidence is indispensable. The second type of applications is to determine the

agreement between sources of information. For example, in the clustering analysis [4], [18], [19] and the determination of discounting factors [20], [21], the distance of evidence is required.

Since the distance of evidence is a very crucial concept in many applications, it has attracted increasing research interest recently in the belief functions community. Many definitions of distance (or dissimilarity) measures have been proposed in the past two decades [22]. Some of them are non-strict distance metrics, although they are often called “distance”. In practice, Josselme’s (strict) distance of evidence [13] and Tessem’s (non strict) betting commitment distance [23] (also called the pignistic probability distance) are most frequently used ones. A fuzzy set based distance of evidence was also proposed in our previous work [24]. Josselme et al. provided an excellent survey [22] on available works on the distance of evidence, where many definitions are introduced and compared.

Various types of distance of evidence have been proposed under the geometric interpretation [25] of the DST, where a basic belief assignment (BBA) is considered as a vector of a Cartesian-alike space and each focal element is deemed as a base of the space [22]. However, all existing distances of evidence have their own limitations. First, a strict distance metric should satisfy the requirements including the non-negativity, non-degeneracy, symmetry, and triangular inequality. None of the existing distances of evidence except for Josselme’s distance can satisfy all the requirements, i.e., they are not strict distance metrics. This is due to the switch between theoretical frameworks. For example, Tessem’s betting commitment distance [23] first transforms BBAs into pignistic probabilities, and fuzzy set based distance of evidence [24] first transforms BBAs into fuzzy membership functions. Such switches between different frameworks lead to the loss of information, thus the distance between BBAs cannot be described precisely using these measures. Therefore, their strictness cannot be assured and they may encounter counter-intuitive results when measuring the distance between different BBAs. Although Josselme’s distance is a strict metric and performs well in many cases, it still has some unsatisfactory behaviors based on our experiments, e.g., the lack of discriminability in some cases and the maximum value

problem as pointed out and analyzed in this paper. Due to the limitations of existing distance measures, we are motivated to propose better strict distance measures of evidence. We propose to use belief intervals [1]  $[Bel(A), Pl(A)]$  of each focal element  $A$  to describe the closeness between BBAs, where  $Bel(A)$  and  $Pl(A)$  are respectively the belief and plausibility of a focal element  $A$  computed from the given BBA defined on a known frame of discernment. If we consider that a BBA is used to model the uncertainty as a whole for all focal elements, then the belief interval of each focal element in a BBA represents the uncertainty of the corresponding proposition. If we use all belief intervals of a BBA as a whole as its “feature” vector, then the distance between the “feature” vectors of different BBAs describes the difference between them. Since a belief interval is an interval number, the distance between the same focal element’s two belief intervals in two BBAs can be calculated by Wasserstein’s distance of interval numbers [26]. Based on all the distance values between belief intervals, we design a Euclidean-family distance using the sum of squares of all belief intervals’ distance values, and a Chebyshev-family distance using the maximum of all belief intervals’ distance values, respectively, to measure the distance between two different “feature” vectors of belief intervals, and thus to measure the distance between two BBAs. Our new definitions directly use the belief intervals defined in the DST, i.e., there is no switch between different theoretical frameworks. It can be proved that our new proposed measures of distance of evidence are strict distance metrics satisfying the requirements of non-negativity, non-degeneracy, symmetry and triangle inequality. This paper extends our preliminary results in [27], where the basic idea of the belief interval based distance is briefly introduced and a few illustrative examples are provided. In this paper, the limitations of existing distances are summarized more specifically, and the causes of these limitations are analyzed. More detailed formulations, proofs, and theoretical analyses of the new proposed distance measures are provided. More examples, simulations, and related analyses are provided for comparison between our proposed distances and the existing ones. An application of the proposed distances of evidence in the BBA approximation and an application of multiple criteria decision making (MCDM) using the proposed distance of evidence is also provided. These are all added values (contributions) of this paper.

The rest of this paper is organized as follows. Basics of the theory of belief functions are briefly introduced in Section II. The geometric interpretation and some commonly used distance measures of evidence are reviewed in Section III. Limitations of existing measures are explained based on illustrative examples in Section III. In Section IV, two new distance metrics in DST are proposed based on the belief intervals and the distance between interval numbers. The proof of our proposed distance metrics’ strictness, and the comparisons between our measures and distance bounds are also provided in Section IV. In Section V, examples, simulations, applications and related analyses are provided based on the comparison between new metrics and some existing ones from different

aspects to show the rationality and efficiency of our new metrics. Section VI concludes this paper.

## II. BASICS OF THEORY OF BELIEF FUNCTIONS

The theory of belief functions was first proposed by Dempster and then further developed by Shafer, therefore, it is usually called Dempster-Shafer evidence theory (DST) [1]. It has become an important theory and tool for uncertainty modeling and reasoning.

The basic concept of the theory of belief functions is the frame of discernment (FOD), which represents the discourse domain of the problem we are interested in. Under the closed-world assumption, the FOD:  $\Theta = \{\theta_1, \dots, \theta_n\}$  is defined as a set of  $n$  mutually exclusive and exhaustive elements. If a set function  $m : 2^\Theta \rightarrow [0, 1]$ , where  $2^\Theta$  is the powerset of  $\Theta^1$ , satisfies

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad m(\emptyset) = 0, \quad (1)$$

and if  $m(A) \geq 0$  holds, then  $m$  is called a basic belief assignment (BBA, or mass function) over the FOD  $\Theta$ . All the sets  $A \in 2^\Theta$  satisfying  $m(A) > 0$  are called the focal elements. Each focal element represents a proposition in the FOD. Given a BBA, a body of evidence (BOE) [1] can be determined, which is defined as the set of focal elements and their corresponding mass assignments.

A belief function over the FOD  $\Theta$ , denoted by  $Bel : 2^\Theta \rightarrow [0, 1]$ , is defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B), \forall A \subseteq \Theta. \quad (2)$$

A plausibility function over the FOD  $\Theta$ , denoted by  $Pl : 2^\Theta \rightarrow [0, 1]$ , is defined as:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \forall A \subseteq \Theta. \quad (3)$$

The plausibility function and the belief function satisfy [1]:

$$Pl(A) = 1 - Bel(\bar{A}), \quad (4)$$

where  $\bar{A}$  is the complementary proposition of  $A \in 2^\Theta$ . The plausibility  $Pl(A)$  and the belief  $Bel(A)$  constitute a belief interval  $[Bel(A), Pl(A)]$ . The length of the belief interval  $[Bel(A), Pl(A)]$  represents the degree of imprecision for the proposition or focal element  $A$ . The non-null mass value assigned to  $\Theta$  represents the degree of ignorance, i.e., the “unknown” state. Furthermore, in DST, different uncertainty measures have been proposed such as Non-specificity [28], Ambiguity Measure (AM) [29], Aggregated Uncertainty (AU) [30] and distance-based uncertainty measures [31].

The evidence combination rules are for uncertainty reasoning, e.g., Dempster’s rule of combination is used to combine different distinct bodies of evidence (BOEs). Suppose that there are two independent BBAs:  $m_1$  and  $m_2$ . The conflict coefficient [1] is defined as

$$K \triangleq \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j). \quad (5)$$

<sup>1</sup>The powerset is the set of all subsets of  $\Theta$  including the empty set  $\emptyset$ .

If  $K < 1$ , then the combined BBA  $\mathbf{m}$  can be obtained using Dempster's rule of combination:

$$m(A) = \begin{cases} 0, & A = \emptyset, \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}, & A \neq \emptyset, \end{cases} \quad (6)$$

where  $A_1, \dots, A_k$  and  $B_1, \dots, B_l$  are focal elements of  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , respectively. Note that Dempster's rule of combination is both commutative and associative, i.e., symmetric.

The obtained BBA is in fact the orthogonal sum of the original BBAs. Dempster's rule of combination has been criticized for its counter-intuitive behaviors [9], [32], especially in high conflict cases. Accordingly, many alternative combination rules have emerged. See [7], [33] for details.

### III. TRADITIONAL MEASURES OF DISTANCE OF EVIDENCE

How to measure the closeness between two BBAs? This is crucial for performance evaluation, algorithm optimization and other belief functions based applications. The answer is the distance of evidence. The conflict coefficient  $K$  (defined in Eq. (5)) in Dempster's rule of combination was the only means to quantify the interaction between BBAs for about two decades (from 1967 to 1990). However, this coefficient  $K$  (denoted by  $d_C$  in the sequel) may be inappropriate to quantify the closeness between two BBAs as the conflict between two identical BBAs might not equal to 0.

**Example 1.** Suppose that the FOD is  $\Theta = \{\theta_1, \dots, \theta_n\}$ . Two BBAs defined on  $\Theta$  are  $m_1(\{\theta_1\}) = \dots = m_1(\{\theta_n\}) = 1/n$  and  $m_2(\{\theta_1\}) = \dots = m_2(\{\theta_n\}) = 1/n$ .

Obviously, they are two identical BBAs and  $d_C = 1 - 1/n$ . When  $n$  becomes large,  $d_C$  approximates to its upper bound (i.e., 1). If one considered  $d_C$  as a distance, such a result would be somewhat counter-intuitive.

A strict distance metric defined on the set  $\mathcal{E} d(\cdot, \cdot) : \mathcal{E} \times \mathcal{E} \rightarrow \mathbb{R}$ ,  $(x, y) \mapsto d(x, y)$  should satisfy

- 1) Non-negativity:  $d(x, y) \geq 0$ ;
- 2) Non-degeneracy:  $d(x, y) = 0 \Leftrightarrow x = y$ ;
- 3) Symmetry:  $d(x, y) = d(y, x)$ ;
- 4) Triangle inequality:  $d(x, y) + d(y, z) \geq d(x, z), \forall z \in \mathcal{E}$ .

Obviously,  $d_C$  violates the Non-degeneracy condition. It is not difficult to verify that  $d_C$  only satisfies the Non-negativity and Symmetry conditions. Therefore, it is not a strict distance metric.

Many other definitions<sup>2</sup> of distance of evidence were proposed in the past two decades as reported in Jousselme's survey [22]. Most of them can be considered as being established under the framework of the geometrical interpretation of the DST.

<sup>2</sup>To be rigorous, only those definitions satisfying the four requirements can be called "distance". The ones that do not satisfy these four requirements can only be called "dissimilarity" or "closeness" measures. In the sequel, for the convenience, all dissimilarity definitions are called "distance" when no ambiguity should occur.

#### A. Geometric interpretation of the theory of belief functions

The geometrical interpretation of the DST [25] is as follows.

Suppose that the FOD is  $\Theta$  with  $|\Theta| = n$ . Let  $\mathcal{E}_\Theta$  be the  $2^n$ -dimensional Cartesian space<sup>3</sup> spanned by the set of column vectors  $\{\mathbf{e}_A, A \subseteq \Theta\}$ . Each vector  $\mathbf{v}$  of  $\mathcal{E}_\Theta$  could be rewritten as  $\mathbf{v} = \sum_{A \subseteq \Theta} \alpha_A \cdot \mathbf{e}_A$ . Here  $\alpha_A \in \mathbb{R}$  can be considered as the coordinate of  $\mathbf{v}$  along the direction of  $\mathbf{e}_A$ .

A BBA  $\mathbf{m}$  is a vector of  $\mathcal{E}_\Theta$ , which should satisfy  $\sum_{A \subseteq \Theta} \alpha_A = 1, \alpha_\emptyset = 0$ , with  $\alpha_A \geq 0$  and  $\alpha_A \triangleq m(A)$  due to the properties of unity and non-negativity for mass values, as illustrated in Eq. (1).

For example, suppose that the FOD  $\Theta = \{\theta_1, \theta_2\}$ . A BBA  $\mathbf{m}$  on  $\Theta$  is  $m(\{\theta_1\}) = 0.3, m(\{\theta_2\}) = 0.2, m(\{\theta_1, \theta_2\}) = 0.5$ . Under the closed-world assumption,  $\mathbf{m}$  is illustrated in Fig. 1.

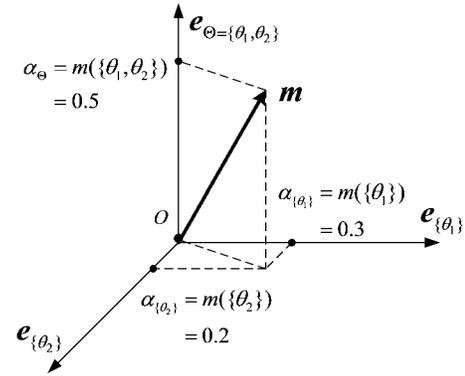


Figure 1. Geometrical interpretation of a BBA.

According to the geometrical interpretation of DST, two BBAs  $\mathbf{m}_1$  and  $\mathbf{m}_2$  are two vectors. That is,  $\mathbf{m}_1$  and  $\mathbf{m}_2$  are two "points" in the evidential Cartesian space. In the past thirty years, people use all kinds of distance for Cartesian space like Euclidean distances, Chebyshev distances, Minkowski distances, Manhattan distances, etc, to define the distance between BOEs according to the geometrical interpretation [22]. Note that many available definitions are non-strict distance metrics [22]. A few typical measures are reviewed in details in the following. Many other definitions can be found in Jousselme's survey [22].

#### B. Selected existing distance measures of evidence

The earliest distance of evidence is the Tessem's distance of betting commitment [23], which is proposed for the evaluation of BBA approximations.

1) *Tessem's betting commitment distance:* The pignistic probability corresponding to a BBA  $\mathbf{m}$  is defined by [34]:

$$\text{BetP}(A) \triangleq \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} m(B), \quad (7)$$

<sup>3</sup>Note that whether the geometric interpretation of the DST satisfies the strict requirements or properties of the geometric space needs further justifications. Here we call it as the evidential Cartesian space.

which is a probabilistic transformation [35] from a BBA for the probabilistic decision making in DST. The betting commitment distance (or Tessem's distance)  $d_T$  is computed by [23]:

$$d_T(\mathbf{m}_1, \mathbf{m}_2) \triangleq \max_{A \subseteq \Theta} \{ |\text{BetP}_1(A) - \text{BetP}_2(A)| \}. \quad (8)$$

It can be reformulated according to the evidential Cartesian space as

$$cd_T(\mathbf{m}_1, \mathbf{m}_2) = \max \left\{ \begin{array}{c} |\text{BetP}_1(A_1) - \text{BetP}_2(A_1)| \\ |\text{BetP}_1(A_2) - \text{BetP}_2(A_2)| \\ \vdots \\ |\text{BetP}_1(A_{2^n}) - \text{BetP}_2(A_{2^n})| \end{array} \right\} \\ = \max_{A \subseteq \Theta} \{ |\mathbf{BetP}_1' \cdot \mathbf{e}_A - \mathbf{BetP}_2' \cdot \mathbf{e}_A| \}, \quad (9)$$

where  $\mathbf{BetP}_i = [\text{BetP}_i(A_1), \text{BetP}_i(A_2), \dots, \text{BetP}_i(A_{2^n})]'$ ,  $i = 1, 2$ .

$d_T$  is a Chebyshev  $L_\infty$  alike distance. From the definition of  $d_T$ , we can see that there is a switch from the DST framework to the probability framework when calculating this distance. The inconsistency between different theoretical frameworks leads to the loss of information and some unexpected results, therefore, it is not recommended.

Actually, due to the switch between different frameworks, Tessem's distance is not a strict distance metric [36]. It violates the non-degeneracy condition as shown in Example 2.

**Example 2.** Suppose that FOD is  $\Theta = \{\theta_1, \dots, \theta_n\}$ . There are two BBAs  $\mathbf{m}_1$  and  $\mathbf{m}_2$  defined on  $\Theta$  including  $m_1(\{\theta_1\}) = \dots = m_1(\{\theta_n\}) = 1/n$  and  $m_2(\Theta) = 1$ . Their corresponding pignistic probabilities are both  $P(\theta_1) = \dots = P(\theta_n) = 1/n$ . Therefore,  $d_T(\mathbf{m}_1, \mathbf{m}_2) = 0$ , although they are different BBAs. Thus,  $d_T$  does not satisfy the non-degeneracy condition.  $d_T$  also does not satisfy the triangle inequality and has other drawbacks. See details in [36].

2) *Fuzzy membership function (FMF) based dissimilarity:* First transform BBAs  $\mathbf{m}_1(\cdot)$  and  $\mathbf{m}_2(\cdot)$  into FMFs<sup>4</sup>:  $\mu^{(1)}$  and  $\mu^{(2)}$  as for  $i = 1, 2$

$$\mu^{(i)} = [ \mu^{(i)}(\theta_1), \mu^{(i)}(\theta_2), \dots, \mu^{(i)}(\theta_n) ] \\ = [ Pl^{(i)}(\theta_1), Pl^{(i)}(\theta_2), \dots, Pl^{(i)}(\theta_n) ]. \quad (10)$$

According to the dissimilarity definition between FMFs,  $d_F$  is defined as [24]:

$$d_F(\mathbf{m}_1, \mathbf{m}_2) = 1 - \frac{\sum_{i=1}^n (\mu^{(1)}(\theta_i) \wedge \mu^{(2)}(\theta_i))}{\sum_{i=1}^n (\mu^{(1)}(\theta_i) \vee \mu^{(2)}(\theta_i))}. \quad (11)$$

In (11), the operator  $\wedge$  represents the conjunction (min) and  $\vee$  represents the disjunction (max).

<sup>4</sup>The FMF quantifies the membership grade of the element to the fuzzy set. It is a generalization of the characteristic function in classical set and can take its values in the interval  $[0, 1]$ .

It can be reformulated according to the evidential Cartesian space as

$$d_F(\mathbf{m}_1, \mathbf{m}_2) \triangleq \\ 1 - \frac{\sum_{i=1}^n \min((\mathbf{Int} \cdot \mathbf{m}_1)' \cdot \mathbf{e}_{\theta_i}, (\mathbf{Int} \cdot \mathbf{m}_2)' \cdot \mathbf{e}_{\theta_i})}{\sum_{i=1}^n \max((\mathbf{Int} \cdot \mathbf{m}_1)' \cdot \mathbf{e}_{\theta_i}, (\mathbf{Int} \cdot \mathbf{m}_2)' \cdot \mathbf{e}_{\theta_i})}, \quad (12)$$

where  $\mathbf{Int}$  is the intersection matrix, whose element is  $\text{Int}(A, B) = 1$ , if  $A \cap B \neq \emptyset$ ;  $\text{Int}(A, B) = 0$ , if  $A \cap B = \emptyset$ . One has  $\mathbf{Pl} = \mathbf{Int} \cdot \mathbf{m}$ , where  $\mathbf{Pl}$  is the corresponding plausibility vector of  $\mathbf{m}$ .

$d_F$  in fact indirectly represents the distance between two BBAs using the distance between their corresponding FMFs. Note that  $d_F$  is not a strict distance metric. First,  $d_F$  does not satisfy the non-degeneracy condition due to the switch from the DST framework to the fuzzy set framework. Given two different BBAs, their corresponding fuzzy membership functions (FMFs) (singleton plausibility) might be the same as shown in Example 3.

**Example 3.** Suppose that FOD is  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . Two BBAs  $\mathbf{m}_1$  and  $\mathbf{m}_2$  defined on  $\Theta$  are

$$m_1(\{\theta_1, \theta_3\}) = 0.3, m_1(\{\theta_1, \theta_2\}) = 0.7. \\ m_2(\{\theta_1\}) = 0.3, m_2(\{\theta_1, \theta_2\}) = 0.4, m_2(\Theta) = 0.3.$$

Their corresponding singleton plausibilities are the same:

$$\mu^{(1)}(\theta_1) = Pl_1(\{\theta_1\}) = 1.0, \mu^{(1)}(\theta_2) = Pl_1(\{\theta_2\}) = 0.7, \\ \mu^{(1)}(\theta_3) = Pl_1(\{\theta_3\}) = 0.3. \\ \mu^{(2)}(\theta_1) = Pl_2(\{\theta_1\}) = 1.0, \mu^{(2)}(\theta_2) = Pl_2(\{\theta_2\}) = 0.7, \\ \mu^{(2)}(\theta_3) = Pl_2(\{\theta_3\}) = 0.3.$$

Therefore,  $d_F(\mathbf{m}_1, \mathbf{m}_2) = 0$ , although  $\mathbf{m}_1$  and  $\mathbf{m}_2$  are different BBAs.

3) *Jousselme's distance:* By borrowing the  $L_2$  Euclidean distance with weighting matrix in Cartesian space, Jousselme's distance [13] is defined as:

$$d_J(\mathbf{m}_1, \mathbf{m}_2) \triangleq \sqrt{0.5 \cdot (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{Jac} (\mathbf{m}_1 - \mathbf{m}_2)}, \quad (13)$$

where the elements  $Jac(A, B)$  of Jaccard's weighting matrix  $\mathbf{Jac}$  are defined as

$$\mathbf{Jac}(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (14)$$

It has been proved to be a strict distance metric in [37] and has become the most commonly used one so far; however, it might cause some unsatisfactory results as shown in Example 4.

**Example 4.** Suppose that the FOD is  $\Theta = \{\theta_1, \dots, \theta_6\}$ . Three groups of BBAs are as follows.

$$\begin{cases} m_1(\{\theta_1\}) = 1; \\ m_2(\{\theta_2\}) = 1. \\ m_3(\{\theta_1, \theta_2\}) = 1; \\ m_4(\{\theta_3, \theta_4\}) = 1. \\ m_5(\{\theta_1, \theta_2, \theta_3\}) = 1; \\ m_6(\{\theta_4, \theta_5, \theta_6\}) = 1. \end{cases}$$

Using Jousselme's distance, one gets  $d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_3, \mathbf{m}_4) = d_J(\mathbf{m}_5, \mathbf{m}_6) = 1$ , that is, they all reach the

maximum value 1. The six BBAs here are all categorical BBAs<sup>5</sup>.  $m_1$  and  $m_2$  each has a unique singleton focal element. The opinions of  $m_1$  and  $m_2$  are totally different, and both of them are specific, i.e., with no ambiguity. The opinions of  $m_3$  and  $m_4$  are totally different, however, both of them are not specific and with ambiguity. The BBAs  $m_5$  and  $m_6$  carry larger ambiguity. Intuitively, it makes sense that the distance between  $m_1$  and  $m_2$  should be larger than the distance between  $m_3$  and  $m_4$ ; also, the distance between  $m_3$  and  $m_4$  should be larger than the distance between  $m_5$  and  $m_6$ . Jousselme's distance does not provide this expected behavior.

Furthermore, Jousselme's distance is relatively insensitive to the change of BBA in some cases as shown in Example 5.

**Example 5.** Suppose that FOD is  $\Theta = \{\theta_1, \dots, \theta_3\}$ . Consider the following three BBAs  $m_1$ ,  $m_2$  and  $m_3$ :

$$\begin{aligned} m_1(\{\theta_1\}) &= m_1(\{\theta_2\}) = m_1(\{\theta_3\}) = 1/3; \\ m_2(\{\theta_1\}) &= m_2(\{\theta_2\}) = m_2(\{\theta_3\}) = 0.1, m_2(\Theta) = 0.7; \\ m_3(\{\theta_1\}) &= m_3(\{\theta_2\}) = 0.1, m_3(\{\theta_3\}) = 0.8. \end{aligned}$$

Since both  $m_1$  and  $m_2$  have no preference on any singleton  $\{\theta_i\}$  and  $m_3$  commits more belief to  $\{\theta_3\}$ , it is intuitively expected that the distance between  $m_1$  and  $m_2$  should be smaller than that between  $m_1$  and  $m_3$ . However, Jousselme's distance leads to  $d_J(m_1, m_2) = d_J(m_1, m_3) = 0.4041$ , which shows that  $d_J$  does not discriminate them well.

In summary, many existing distance measures of evidence have evident limitations, even for the strict Jousselme's distance metric. For Tessem's distance and FMF-based distance, there exist the switches between different theoretical frameworks. With Tessem's distance, there is a switch from the framework of DST to the framework of the probability theory; with FMF-based distance, there is a switch from the framework of DST to the framework of the fuzzy sets theory. These switches bring the undesired loss of information, which should be avoided. Jousselme's distance borrows the distance metric from the traditional Cartesian space to the evidential Cartesian space. The strictness of the evidential Cartesian space, i.e., the geometrical interpretation of DST needs further verification. Therefore, it is not uncommon to obtain some unsatisfactory results when using Jousselme's distance.

Since traditional distances in DST have limitations (or unsatisfactory behaviors), we propose new strict distance measures of evidence with better behaviors.

#### IV. DISTANCE OF EVIDENCE USING BELIEF INTERVALS

As aforementioned, the limitations and non-strictness of some existing distances of evidence are caused by the switches between theoretical frameworks, therefore in our design of the new distances, no such switch is allowed. In Jousselme's distance, there is no switch between different theoretical frameworks, where only the focal elements and the corresponding mass values are used. Given a BBA, the mass value for a proposition (or focal element)  $A$  represents the basic belief assigned to  $A$ . Besides the mass value  $m(A)$ , other

values, like  $Bel(A)$  and  $Pl(A)$ , are optional. Furthermore, the belief interval  $[Bel(A), Pl(A)]$  can be used to represent the degree of imprecision of  $A$ . Therefore, the belief interval  $[Bel(A), Pl(A)]$  carries more information of a given proposition  $A$  than the mass value  $m(A)$ , which is a scalar. Therefore, we propose to use the belief interval (with more information) to replace the mass value for achieving better performance.

In DST, besides the BBA ( $m$ ), the belief function ( $Bel$ ) and plausibility function ( $Pl$ ), there also exist the doubt function ( $Dou$ ) and the commonality function ( $Q$ ) [1]. Given one function, it can be transformed to any other one of these five functions according to their definitions and the Möbius transformations [1]. That is, any one of the five functions has one-to-one correspondence to the other, therefore, one can also try to jointly use other functions like the commonality and doubt for designing new distance measures. In this paper, we choose the belief interval  $[Bel(A), Pl(A)]$ ,  $\forall A \subseteq \Theta$ , since the belief and plausibility are more familiar to people and more widely used in practice than the doubt and the commonality. Furthermore,  $[Bel(A), Pl(A)]$  has intuitive physical meaning, i.e., the degree of imprecision for the proposition  $A$ .

Suppose that two BBAs  $m_1$  and  $m_2$  are defined on  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . For each focal element  $A_i \subseteq \Theta$  ( $i = 1, \dots, 2^n - 1$ ), we can calculate the belief interval of  $A_i$  for  $m_1$  and  $m_2$ , respectively, which are denoted by  $[Bel_1(A_i), Pl_1(A_i)]$  and  $[Bel_2(A_i), Pl_2(A_i)]$ . That is, each BBA  $m_j$  ( $j = 1, 2$ ) can also equivalently be modeled by a matrix with the size of  $(2^n - 1) \times 2$ :

$$\begin{bmatrix} [Bel_j(A_1), Pl_j(A_1)] \\ \vdots \\ [Bel_j(A_{2^n-1}), Pl_j(A_{2^n-1})] \end{bmatrix}.$$

A belief interval can be regarded as a classical interval number<sup>6</sup> included in  $[0, 1]$ . Then the above matrix can be regarded as a vector of interval numbers (belief intervals):

$$\begin{aligned} Fe_j &= \begin{bmatrix} [Bel_j(A_1), Pl_j(A_1)] \\ \vdots \\ [Bel_j(A_{2^n-1}), Pl_j(A_{2^n-1})] \end{bmatrix} \\ &= \begin{bmatrix} BI_j(A_1) \\ \vdots \\ BI_j(A_{2^n-1}) \end{bmatrix}. \end{aligned} \quad (15)$$

Here  $Fe_j$  can be considered as a generalized feature vector describing the BBA  $m_j$ . If we can define the distance between  $Fe_1$  and  $Fe_2$ , then the distance between  $m_1$  and  $m_2$  is readily obtained. Here  $Fe_1$  and  $Fe_2$  are two generalized vectors whose elements are intervals<sup>7</sup>.

We can borrow the definition of the distance metric for the vectors in Cartesian space to define the distance of evidence here: 1) define the distance between two feature vectors in each dimension; 2) combine the distance value for each dimension into a scalar.

<sup>6</sup>An interval number  $[a, b]$  with  $a \leq b$  is actually an interval with the lower bound  $a$  and the upper bound  $b$ , where  $a, b \in \mathbb{R}$ . When  $a = b$ , an interval number degenerates to a real number.

<sup>7</sup>The  $Fe_j$  can be also considered in the evidential Cartesian-alike space, however, the coordinate of each direction  $e_A$  is a generalized real number, i.e., an interval number.

<sup>5</sup>A categorical BBA is a BBA only has one focal element.

Therefore, in the step 1, the distance in each dimension should be defined, i.e., we must define the distance between two interval numbers. Irpino et al [26] proposed a Wasserstein distance for interval numbers as briefly introduced below.

Suppose that  $F$  and  $G$  are the corresponding distribution functions of the random variables  $f$  and  $g$ , respectively, Wasserstein  $L_2$  metric is defined as [26]

$$d_{\text{Wass}}(F, G) \triangleq \sqrt{\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt}. \quad (16)$$

For a uniform distribution of points, an interval of reals  $x_i = [a_i, b_i]$  can be expressed as a function of [26]:

$$x_i(t) = [a_i, b_i] = a_i + t(b_i - a_i), \forall 0 \leq t \leq 1. \quad (17)$$

If one further considers a description of the interval using its midpoint  $(a_i + b_i)/2$  and radius  $(b_i - a_i)/2$ ,  $x_i$  can be rewritten as

$$x_i(t) = \frac{a_i + b_i}{2} + \frac{b_i - a_i}{2} (2t - 1), \forall 0 \leq t \leq 1 \quad (18)$$

Then, Euclidean distance between homologous points of two intervals  $x_1 = [a_1, b_1]$  and  $x_2 = [a_2, b_2]$  is defined as [26]

$$\begin{aligned} d_{BI}([a_1, b_1], [a_2, b_2]) &= d_{\text{Wass}}(x_1, x_2) \\ &= \sqrt{\int_0^1 [x_1(t) - x_2(t)]^2 dt} \\ &= \sqrt{\int_0^1 \left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} + \left( \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right) (2t - 1) \right]^2 dt} \\ &= \sqrt{\left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2}. \end{aligned} \quad (19)$$

Note that there are also other types of distance between interval numbers [26]. We choose the Wasserstein distance in Eq. (19) to calculate the distance between belief intervals, because it is a strict distance metric, which is very crucial for defining distance measures of evidence. Furthermore, it has a simple form, and is easy to compute.

According to Eq. (19), the distance between two feature vectors  $\mathbf{Fe}_1$  and  $\mathbf{Fe}_2$  in terms of each dimension  $i$  ( $i = 1, \dots, 2^n - 1$ ), i.e., the distance between two belief intervals  $BI_1(A_i) : [Bel_1(A_i), Pl_1(A_i)]$  and  $BI_2(A_i) : [Bel_2(A_i), Pl_2(A_i)]$  can be obtained.  $d_{BI}(BI_1(A_i), BI_2(A_i))$  can be regarded as the distance between  $\mathbf{m}_1$  and  $\mathbf{m}_2$  when considering the focal element  $A_i$  only.

Therefore, we can obtain in total  $2^n - 1$  belief interval distance values for all  $A_i \subseteq \Theta$ .

In step 2, we combine all the  $2^n - 1$  distance values into one scalar, i.e., to get the total distance between  $\mathbf{Fe}_1$  and  $\mathbf{Fe}_2$ .

In Cartesian space, if we try to measure the distance between two points, we also calculate the dissimilarity between each dimension of the two points, and then use some way to combine the dissimilarity values of different dimensions to a scalar, i.e., the distance value. Euclidean family and Chebyshev family are two commonly used ways to generate such a scalar in the Cartesian space. We can borrow this idea to generate a scalar from the above mentioned  $2^n - 1$  focal elements' corresponding dissimilarity values. Therefore, two

commonly used distance definitions — the Euclidean family and the Chebyshev family — are used to combine the distance values of all dimensions into a scalar, i.e., the distance value. Two new distances of evidence are presented next.

#### A. Euclidean-family Belief Interval-based Distance $d_{BI}^E$

Given two BBAs  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , our proposed Euclidean-family belief interval-based distance is a combination of each focal element's belief interval distance value. To be specific, it is a normalized root squared summation of the distance value between belief intervals in each dimension (focal element) as shown in Eq. (20)

$$d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) \triangleq \sqrt{N_c \cdot \sum_{i=1}^{2^n-1} [d_{BI}(BI_1(A_i), BI_2(A_i))]^2}. \quad (20)$$

Here  $N_c$  denotes the normalization factor to make  $d_{BI}^E \in [0, 1]$ . Eq. (20) can be re-written as

$$\begin{aligned} d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) &\triangleq \sqrt{N_c \cdot \mathbf{d}_{BI} \cdot \mathbf{I}^{(2^n-1)} \cdot \mathbf{d}_{BI}^T} \\ &= \sqrt{N_c \cdot \mathbf{d}_{BI} \cdot \mathbf{d}_{BI}^T}, \end{aligned} \quad (21)$$

where  $\mathbf{I}^{(2^n-1)}$  is an identity matrix with rank  $2^n - 1$ , and

$$\mathbf{d}_{BI} = \begin{bmatrix} d_{BI}(BI_1(A_1), BI_2(A_1)) \\ \vdots \\ d_{BI}(BI_1(A_{2^n-1}), BI_2(A_{2^n-1})) \end{bmatrix}.$$

The normalization factor for Euclidean-family Belief Interval-based Distance  $d_{BI}^E$  is  $N_c = 1/2^{n-1}$ .

Suppose that the FOD is  $\{\theta_1, \theta_2, \dots, \theta_n\}$ .  $\mathbf{m}_1$  and  $\mathbf{m}_2$  are two BOEs.  $m_1(\{\theta_l\}) = 1, l \in \{1, \dots, n\}$  is a categorical BBA, which represents the most certain case, i.e., there is no uncertainty when assign the belief to the singleton proposition  $\{\theta_l\}$ . The two BBAs:

$$m_1(\{\theta_i\}) = 1, m_2(\{\theta_j\}) = 1, \forall i \neq j, \quad i, j \in \{1, \dots, n\}. \quad (22)$$

are two different and the most certain cases. They have no common part, i.e., they fully support different singletons, therefore, the dissimilarity (distance) between them reaches the maximum.

Assume that  $A$  is a focal element.

When  $|A| = 1$ , only two belief intervals have distance value  $d_{BI}$  of 1 (i.e.,  $d_{BI}(BI_1(\theta_i), BI_2(\theta_i)) = 1$  and  $d_{BI}(BI_1(\theta_j), BI_2(\theta_j)) = 1$ ). The other values are 0.

When  $|A| > 1$ ,  $d_{BI} = 1$  for those focal elements including  $\theta_i$  or  $\theta_j$  (but not including both  $\theta_i$  and  $\theta_j$ ) are 1.  $d_{BI} = 0$  for the rest.

To be specific,

when  $|A| = 2$ ,  $d_{BI} = 1$  only for  $2 \times C_{n-2}^1$  focal elements<sup>8</sup>.  $d_{BI} = 0$  for the rest;

when  $|A| = 3$ ,  $d_{BI} = 1$  only for  $2 \times C_{n-2}^2$  focal elements.  $d_{BI} = 0$  for the rest;

<sup>8</sup>Choose one element  $\theta_k$  out of the  $\Theta' = \Theta - \{\theta_i, \theta_j\}$  ( $|\Theta'| = n-2$ ). Then, together with  $\theta_i$  and  $\theta_j$ , respectively, to constitute focal element  $\{\theta_k, \theta_i\}$  and  $\{\theta_k, \theta_j\}$ , respectively. So, the number of focal elements with  $d_{BI}$  values of 1 is  $2 \times C_{n-2}^1$ . Similarly, we can obtain the values in other cases for  $A > 1$ .

⋮

when  $|A| = n - 1$ ,  $d_{BI} = 1$  only for  $2 \times C_{n-2}^{n-2}$  focal elements.  $d_{BI} = 0$  for the rest;

when  $|A| = n$ , the  $d_{BI}$  value of unique focal element, i.e., total set ( $\Theta$ ) is 0.

So, the summation  $S_c$  of all  $(d_{BI})^2$  is

$$\begin{aligned} S_c &= 2 \times 1 + 2 \times C_{n-2}^1 + 2 \times C_{n-2}^2 + \dots \\ &\quad + 2 \times C_{n-2}^{n-2} + 0 \\ &= 2 \times (C_{n-2}^0 + C_{n-2}^1 + C_{n-2}^2 + \dots + C_{n-2}^{n-2}) \\ &= 2 \times 2^{n-2} = 2^{n-1}. \end{aligned} \quad (23)$$

So, the normalization factor  $N_c = 1/S_c = 1/2^{n-1}$ .

### B. Chebyshev-family Belief Interval-based Distance $d_{BI}^C$

Given two BBAs  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , our proposed Chebyshev-family belief interval-based distance is the maximum of all belief interval distance values.

$$d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2) \triangleq \max_{A_i \subseteq \Theta, i=1, \dots, 2^n-1} \{d_{BI}(BI_1(A_i), BI_2(A_i))\}. \quad (24)$$

Actually, we use the distance of belief intervals for focal elements instead of their mass assignments to define the distances of evidence when compared with traditional definitions. Euclidean-family belief interval-based distance  $d_{BI}^E$  and Chebyshev-family belief interval-based distance  $d_{BI}^C$  are strict distance metrics.  $d_{BI}^E$  and  $d_{BI}^C$  are defined over belief intervals. Given a BBA  $(m(A_i), i = 1, \dots, 2^n - 1)$ , we can generate a set of belief intervals  $([Bel(A_i), Pl(A_i)])$ . On the other hand, given a set of belief intervals  $([Bel(A_i), Pl(A_i)])$ , according to the Möbius transformation, we can generate a unique BBA  $(m(A_i), i = 1, \dots, 2^n - 1)$  from  $Pl(A_i), i = 1, \dots, 2^n - 1$  or  $Bel(A_i), i = 1, \dots, 2^n - 1$ . As we know [1], there is a one-to-one mapping between a set of belief intervals  $([Bel(A_i), Pl(A_i)])$  and a BBA  $(m(A_i), i = 1, \dots, 2^n - 1)$ .

According to Eqs. (20)-(21) and (24), it is easy to verify that  $d_{BI}^E$  and  $d_{BI}^C$  satisfy non-negativity, non-degeneracy and symmetry conditions. We need to prove the property of triangle inequality of  $d_{BI}^E$ .

Suppose that there are three BBAs  $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$  defined on the same FOD with size of  $n$ . Because  $d_{BI}$  defined in Eq. (19) is a strict distance metric, so, for each  $A_i$  ( $i = 1, \dots, s, s = 2^n - 1$ ) there exists

$$\begin{aligned} d_{BI}^E(m_1(A_i), m_2(A_i)) + d_{BI}^E(m_2(A_i), m_3(A_i)) \\ \geq d_{BI}^E(m_1(A_i), m_3(A_i)). \end{aligned}$$

Suppose that

$$\begin{aligned} d_{BI}^E(m_1(A_i), m_2(A_i)) &= a_i; \quad d_{BI}^E(m_2(A_i), m_3(A_i)) = b_i; \\ d_{BI}^E(m_1(A_i), m_3(A_i)) &= c_i. \end{aligned}$$

One has

$$\begin{aligned} a_i + b_i &\geq c_i \\ \Rightarrow (a_i + b_i)^2 &\geq c_i^2 \\ \Rightarrow a_i^2 + b_i^2 + 2a_i b_i &\geq c_i^2 \\ \Rightarrow \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sum_{i=1}^s a_i b_i &\geq \sum_{i=1}^s c_i^2. \end{aligned} \quad (25)$$

According to Cauchy-Schwarz inequality,

$$\sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \geq \sum_{i=1}^s a_i b_i. \quad (26)$$

So,

$$\begin{aligned} \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \\ \geq \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sum_{i=1}^s a_i b_i \geq \sum_{i=1}^s c_i^2 \\ \Rightarrow \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \geq \sum_{i=1}^s c_i^2. \end{aligned} \quad (27)$$

Therefore,

$$\begin{aligned} \sum_{i=1}^s a_i^2 + \sum_{i=1}^s b_i^2 + 2 \sqrt{\sum_{i=1}^s a_i^2 \sum_{i=1}^s b_i^2} \\ = \left( \sqrt{\sum_{i=1}^s a_i^2} + \sqrt{\sum_{i=1}^s b_i^2} \right)^2 \\ = (d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) + d_{BI}^E(\mathbf{m}_2, \mathbf{m}_3))^2 \\ \Rightarrow (d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) + d_{BI}^E(\mathbf{m}_2, \mathbf{m}_3))^2 \geq (d_{BI}^E(\mathbf{m}_1, \mathbf{m}_3))^2 \\ \Rightarrow d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) + d_{BI}^E(\mathbf{m}_2, \mathbf{m}_3) \geq d_{BI}^E(\mathbf{m}_1, \mathbf{m}_3). \end{aligned} \quad (28)$$

So, the triangle inequality for  $d_{BI}^E$  is satisfied.

For  $d_{BI}^C$ , we have

$$\begin{aligned} d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2) + d_{BI}^C(\mathbf{m}_2, \mathbf{m}_3) &= \max_{i=1, \dots, s} a_i + \max_{i=1, \dots, s} b_i, \\ d_{BI}^C(\mathbf{m}_1, \mathbf{m}_3) &= \max_{i=1, \dots, s} c_i = a_k + b_k, \quad k = \arg \max_{i=1, \dots, s} c_i. \end{aligned} \quad (29)$$

There exists

$$a_k + b_k \leq \max_{i=1, \dots, s} a_i + \max_{i=1, \dots, s} b_i = d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2) + d_{BI}^C(\mathbf{m}_2, \mathbf{m}_3), \quad (30)$$

i.e.,  $d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2) + d_{BI}^C(\mathbf{m}_2, \mathbf{m}_3) \geq d_{BI}^C(\mathbf{m}_1, \mathbf{m}_3)$ . Consequently,  $d_{BI}^C$  satisfies triangle inequality.

In summary,  $d_{BI}^E$  and  $d_{BI}^C$  are strict distance metrics.

In the traditional geometric interpretation of DST introduced in section III, the coordinates of different bases are represented by mass values (real numbers), while for our new proposed distances, the coordinates are represented by belief intervals (interval numbers). Therefore, our new distances are under a generalized geometric interpretation of evidence theory.

### C. An illustrative example

**Example 6.** Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . Two BBAs  $\mathbf{m}_1, \mathbf{m}_2$  over the FOD are:

$$\begin{aligned} m_1(\{\theta_1\}) &= 0.1, m_1(\{\theta_2\}) = 0.1, m_1(\{\theta_3\}) = 0.05, \\ m_1(\{\theta_1, \theta_2\}) &= 0.1, m_1(\{\theta_1, \theta_3\}) = 0.05, \\ m_1(\{\theta_2, \theta_3\}) &= 0.1, m_1(\Theta) = 0.5. \end{aligned}$$

$$\begin{aligned} m_2(\{\theta_1\}) &= 0.2, m_2(\{\theta_2\}) = 0.3, m_2(\{\theta_3\}) = 0.1, \\ m_2(\{\theta_1, \theta_2\}) &= 0.05, m_2(\{\theta_1, \theta_3\}) = 0.1, \\ m_2(\{\theta_2, \theta_3\}) &= 0.05, m_2(\Theta) = 0.2. \end{aligned}$$

First, the belief intervals are calculated for each focal element of  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , respectively:

$$\begin{aligned} BI_1(\{\theta_1\}) &: [0.10, 0.75], \\ BI_1(\{\theta_2\}) &: [0.10, 0.80], \\ BI_1(\{\theta_3\}) &: [0.05, 0.70], \\ BI_1(\{\theta_1, \theta_2\}) &: [0.30, 0.95], \\ BI_1(\{\theta_1, \theta_3\}) &: [0.20, 0.90], \\ BI_1(\{\theta_2, \theta_3\}) &: [0.25, 0.90], \\ BI_1(\Theta) &: [1.00, 1.00]. \end{aligned}$$

$$\begin{aligned} BI_2(\{\theta_1\}) &: [0.20, 0.55], \\ BI_2(\{\theta_2\}) &: [0.30, 0.60], \\ BI_2(\{\theta_3\}) &: [0.10, 0.45], \\ BI_2(\{\theta_1, \theta_2\}) &: [0.55, 0.90], \\ BI_2(\{\theta_1, \theta_3\}) &: [0.40, 0.70], \\ BI_2(\{\theta_2, \theta_3\}) &: [0.45, 0.80], \\ BI_2(\Theta) &: [1.00, 1.00]. \end{aligned}$$

Second, use Eq. (19) to compute the distance between belief intervals of each corresponding focal element in  $\mathbf{m}_1$  and  $\mathbf{m}_2$ :

$$d_{BI} = \begin{bmatrix} d_{BI}(BI_1(\{\theta_1\}), BI_2(\{\theta_1\})) \\ d_{BI}(BI_1(\{\theta_2\}), BI_2(\{\theta_2\})) \\ d_{BI}(BI_1(\{\theta_3\}), BI_2(\{\theta_3\})) \\ d_{BI}(BI_1(\{\theta_1, \theta_2\}), BI_2(\{\theta_1, \theta_2\})) \\ d_{BI}(BI_1(\{\theta_1, \theta_3\}), BI_2(\{\theta_1, \theta_3\})) \\ d_{BI}(BI_1(\{\theta_2, \theta_3\}), BI_2(\{\theta_2, \theta_3\})) \\ d_{BI}(BI_1(\Theta), BI_2(\Theta)) \end{bmatrix} = \begin{bmatrix} 0.1000 \\ 0.1155 \\ 0.1323 \\ 0.1323 \\ 0.1155 \\ 0.1000 \\ 0.0000 \end{bmatrix}.$$

Then, according to Eq. (20),  $d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2)$  is computed by

$$\begin{aligned} d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) &= \sqrt{2^{3-1} \times \left( \begin{array}{l} 0.1000^2 + 0.1155^2 + 0.1323^2 + \\ 0.1323^2 + 0.1155^2 + 0.1000^2 + 0^2 \end{array} \right)} \\ &= 0.1429. \end{aligned}$$

According to Eq. (24),  $d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2)$  is computed by

$$\begin{aligned} d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2) &= \max \left\{ \begin{array}{l} 0.1000, \quad 0.1155, \quad 0.1323, \\ 0.1323, \quad 0.1155, \quad 0.1000, \quad 0 \end{array} \right\} \\ &= 0.1323. \end{aligned}$$

#### D. On distance bounds

Here, the distance bounds are analyzed. In Antonucci's work [38], a lower bound and an upper bound of a distance of evidence were proposed based on the distance of consistent probabilities. For a BBA  $\mathbf{m}$ , its consistent set of probability mass functions (PMF) is

$$K_{\mathbf{m}} = \left\{ P \mid \begin{array}{l} \sum_{\theta \in \Theta} P(\theta) = 1 \\ \sum_{\theta \in A} P(\theta) \geq \mathbf{Bel}(A), \forall A \in 2^\Theta \end{array} \right\}, \quad (31)$$

where  $P$  is a consistent PMF. Given two PMFs  $P_1$  and  $P_2$ , their Manhattan distance is

$$\delta(P_1, P_2) \triangleq \frac{1}{2} \cdot \sum_{\theta \in \Theta} |P_1(\theta) - P_2(\theta)|. \quad (32)$$

Given two BBAs  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , the lower bound  $\underline{\delta}$  and upper bound  $\bar{\delta}$  are defined as

$$\begin{cases} \underline{\delta}(\mathbf{m}_1, \mathbf{m}_2) = \min_{P_1 \in K_{\mathbf{m}_1}, P_2 \in K_{\mathbf{m}_2}} \delta(P_1, P_2), \\ \bar{\delta}(\mathbf{m}_1, \mathbf{m}_2) = \max_{P_1 \in K_{\mathbf{m}_1}, P_2 \in K_{\mathbf{m}_2}} \delta(P_1, P_2). \end{cases} \quad (33)$$

We calculate  $d_{BI}^E$ ,  $d_{BI}^C$  and the strict distance measure  $d_J$  together with the upper and lower bounds to check whether these measures are beyond the bounds or not. We set  $|\Theta| = 3$  and randomly generate 1000 BBA pairs according to the BBA generation algorithm [39] in Table I.

Table I

ALGORITHM 1: RANDOM BBA GENERATION - UNIFORM SAMPLING FROM ALL FOCAL ELEMENTS.

---

**Input:**  $\Theta$ : Frame of discernment;  
 $N_{max}$ : Maximum number of focal elements  
**Output:**  $\mathbf{m}$ : BBA  
 Generate  $\mathcal{P}(\Theta)$ , which is the power set of  $\Theta$ ;  
**FOReach**  $1 \leq i \leq |\mathcal{P}(\Theta)|$  **do**  
 Generate a value according to the Gamma distribution  $\mathcal{G}(1, 1) \rightarrow m_i$ ,  
**END**  
 Normalize the vector  $m = [m_1, \dots, m_{|\mathcal{P}(\Theta)|}] \rightarrow m'$ ;  
 $m(A_i) = m'_i$ ;

---

The results are shown in Figs. 2 and 3 (zoom in around lower bound). Results are sorted by increasing values of  $d_J$ . It is experimentally shown that  $d_J$  and our proposed  $d_{BI}^E$  and  $d_{BI}^C$  are not beyond the lower and upper bounds as shown in Figs. 2 and 3 in this simulation.

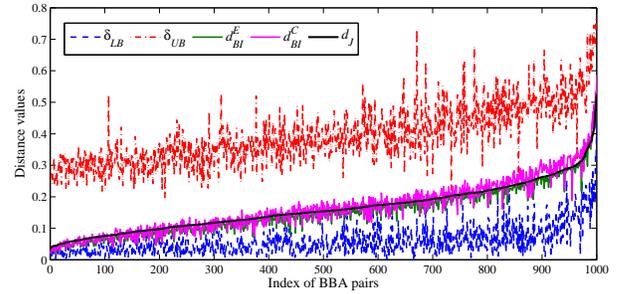


Figure 2. Comparisons between bounds,  $d_J$ ,  $d_{BI}^E$  and  $d_{BI}^C$ .

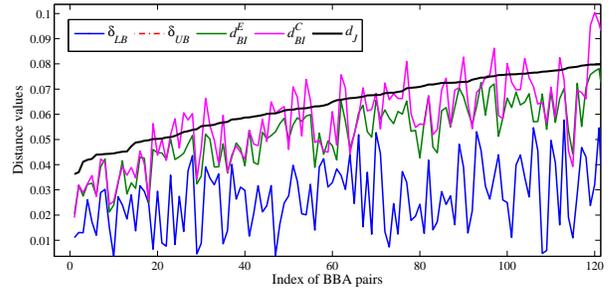


Figure 3. Comparisons between bounds,  $d_J$ ,  $d_{BI}^E$  and  $d_{BI}^C$  (Zoom in around the lower bound).

In the next section, experiments and simulations are provided to show the rationalities of our proposed distance measures of evidence based on the comparisons with available measures.

## V. EXPERIMENTS, SIMULATIONS AND APPLICATIONS

To verify the rationality of the proposed distances, numerical examples, simulations and the applications related to BBA approximations and MCDM are provided.

In each example below,  $d_J$ ,  $d_T$ ,  $d_F$ ,  $d_C$ ,  $d_{BI}^E$  and  $d_{BI}^C$  are compared<sup>9</sup>.

## A. Example 7

Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ .  $m_1$  has relatively large mass value for  $\{\theta_2\}$  as shown in Table II. Therefore, intuitively, for  $m_i$ ,  $i = 2, \dots, 7$  listed in Table III, if the mass assignment for  $\{\theta_2\}$  is relative large, the distance between  $m_1$  and  $m_i$  intuitively should be relatively small. For  $m_5$  and  $m_6$ , the mass of focal elements containing  $\{\theta_2\}$  (i.e.,  $\{\theta_1, \theta_2\}$  and  $\{\theta_2, \theta_3\}$ ) is 0.8. It makes more sense if the distance value with respect to  $m_5$  and  $m_6$  decreases.

 Table II  
BBA  $m_1$ 

Focal element	Mass assignment
$\{\theta_1\}$	0.1
$\{\theta_2\}$	0.8
$\{\theta_3\}$	0.1
$\{\theta_1, \theta_2\}$	0
$\{\theta_2, \theta_3\}$	0
$\{\theta_1, \theta_3\}$	0
$\{\theta_1, \theta_2, \theta_3\}$	0

 Table III  
BBAs  $m_i$ ,  $i = 2, \dots, 7$ 

Focal el. \ BBAs	$m_2$	$m_3$	$m_4$	$m_5$	$m_6$	$m_7$
$\theta_1$	0.8	0	0	0	0	0
$\theta_2$	0	0.8	0	0	0	0
$\theta_3$	0	0	0.8	0	0	0
$\theta_1 \cup \theta_2$	0	0	0	0.8	0	0
$\theta_2 \cup \theta_3$	0	0	0	0	0.8	0
$\theta_1 \cup \theta_3$	0	0	0	0	0	0.8
$\theta_1 \cup \theta_2 \cup \theta_3$	0.2	0.2	0.2	0.2	0.2	0.2

Calculate the distance between  $m_1$  and  $m_i$ ,  $i = 2, \dots, 7$  using different distance definitions as illustrated in Fig. 4. All the distance measures perform similarly in all seven cases and agree with the expected behavior as we can see in Fig. 4.

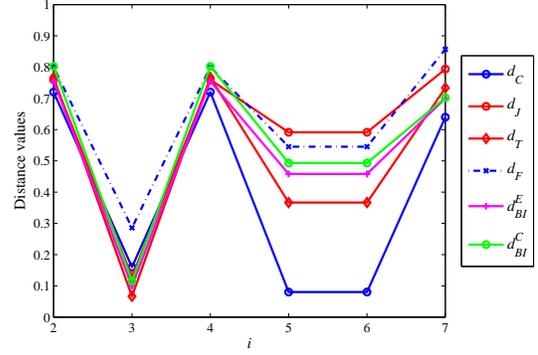
The following Examples 8 - 12 drawn from [13] are used for comparing our proposed measures and available ones.

## B. Example 8

Suppose that three BBAs  $m_1$ ,  $m_2$ , and  $m_3$  are defined on the FOD  $\Theta = \{\theta_1, \dots, \theta_n\}$  as follows:

$$\begin{aligned} m_1(\{\theta_1\}) &= m_1(\{\theta_2\}) = \dots = m_1(\{\theta_n\}) = 1/n; \\ m_2(\Theta) &= 1; \\ m_3(\{\theta_k\}) &= 1, \text{ for some } k \in \{1, \dots, n\}. \end{aligned}$$

<sup>9</sup> $d_C$  corresponds to the conflict coefficient  $K$  defined in Eq. (4).

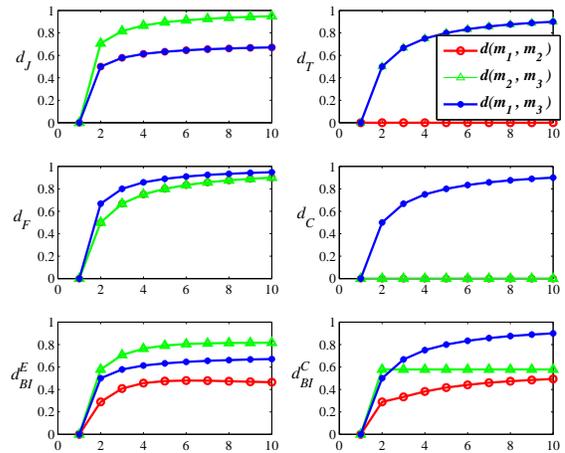

 Figure 4. Distance between  $m_1$  and  $m_i$ ,  $i = 2, \dots, 7$ .

The change of the distance values with the increase of the size  $n$  of FOD are illustrated in Fig. 5.

$d_T$  provides undesired result, i.e., with the increase of  $n$ , there always exists  $d_T(m_1, m_2) = 0$ .  $d_C$  cannot discriminate  $m_1$  and  $m_2$ , and also  $m_2$  and  $m_3$ .

In this example,  $m_1$  is a Bayesian BBA, which only has singleton focal elements;  $m_2$  is a vacuous BBA, which only has the total set  $\Theta$  as the unique focal element;  $m_3$  is a categorical BBA with one singleton focal element, which is absolutely confident in  $\{\theta_k\}$ .

$m_1$  represents the case with only discord and with zero non-specificity;  $m_3$  represents the crispest case;  $m_2$  represents the most ambiguous case. So, the distance between  $m_2$  and  $m_3$  represents the dissimilarity between the most ambiguous case and the crispest case; the distance between  $m_1$  and  $m_3$  represents the dissimilarity between the case with zero non-specificity and the crispest case; the distance between  $m_1$  and  $m_2$  represents the dissimilarity between the case with zero non-specificity and the most ambiguous case.


 Figure 5. Dissimilarities between  $m_1$ ,  $m_2$  and  $m_3$  for Example 8.

Therefore, intuitively, the distance between  $\mathbf{m}_2$  and  $\mathbf{m}_3$  should be the largest one. As we can see in Fig. 5,  $d_{BI}^E(\mathbf{m}_2, \mathbf{m}_3)$  and  $d_J(\mathbf{m}_2, \mathbf{m}_3)$  provide satisfactory results, i.e.,

$$d_{BI}^E(\mathbf{m}_2, \mathbf{m}_3) = \max_{\substack{i,j \in \{1,2,3\} \\ i \neq j}} d_{BI}^E(\mathbf{m}_i, \mathbf{m}_j).$$

From the decision standpoint,  $\mathbf{m}_1$  has no inclination to any choice  $\theta_i$ ;  $\mathbf{m}_2$  also has no inclination to any choice  $\theta_i$ ;  $\mathbf{m}_3$  has a clear inclination to the choice  $\theta_k$ . Therefore, intuitively, the dissimilarity between  $\mathbf{m}_1$  and  $\mathbf{m}_2$  should be smaller than that between  $\mathbf{m}_1$  and  $\mathbf{m}_3$ . As shown in Fig. 5, our proposed  $d_{BI}^E$ ,  $d_{BI}^C$ , and  $d_F$  provide satisfactory results according to this standpoint, i.e.,  $d_{BI}^E(\mathbf{m}_1, \mathbf{m}_2) < d_{BI}^E(\mathbf{m}_1, \mathbf{m}_3)$ ,  $d_{BI}^C(\mathbf{m}_1, \mathbf{m}_2) < d_{BI}^C(\mathbf{m}_1, \mathbf{m}_3)$  and  $d_F(\mathbf{m}_1, \mathbf{m}_2) < d_F(\mathbf{m}_1, \mathbf{m}_3)$ .

As we observed,  $d_J$  cannot discriminate this since  $d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_1, \mathbf{m}_3)$  with the increase of  $n$ . This is because one has

$$d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_1, \mathbf{m}_3) = \sqrt{\frac{1}{2} \left(1 - \frac{1}{n}\right)}.$$

according to Jousselme's distance defined in Eq. (13).

Based on the analyses above,  $d_{BI}^E$  provides rational behaviors in this example.

### C. Example 9 (Example 5 Revisited)

The values of the different distances between  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , and between  $\mathbf{m}_1$  and  $\mathbf{m}_3$  are given in Table IV.

Table IV  
EXAMPLE 9: RESULTS BASED ON DIFFERENT DISTANCES OF EVIDENCE.

Distance	$d_J$	$d_T$	$d_F$	$d_C$	$d_{BI}^E$	$d_{BI}^C$
$d(\mathbf{m}_1, \mathbf{m}_2)$	0.4041	0	0.5833	0.2000	0.2858	0.2333
$d(\mathbf{m}_1, \mathbf{m}_3)$	0.4041	0.4667	0.6364	0.6667	0.4041	0.4667

As aforementioned, both  $\mathbf{m}_1$  and  $\mathbf{m}_2$  have no preference on any singleton  $\{\theta_i\}$  and  $\mathbf{m}_3$  commits more belief to  $\{\theta_3\}$ , therefore, it is intuitively expected that the distance between  $\mathbf{m}_1$  and  $\mathbf{m}_2$  should be smaller than that between  $\mathbf{m}_1$  and  $\mathbf{m}_3$ .

Using Jousselme's distance, one obtains  $d_J(\mathbf{m}_1, \mathbf{m}_2) = d_J(\mathbf{m}_1, \mathbf{m}_3) = 0.4041$  which is unsatisfactory for such a case. Table IV shows that when using  $d_T$ ,  $d_C$ ,  $d_F$ ,  $d_{BI}^E$  and  $d_{BI}^C$ , one obtains  $d(\mathbf{m}_1, \mathbf{m}_2) < d(\mathbf{m}_1, \mathbf{m}_3)$ , which is more reasonable. However, Tessem's distance leads to  $d_T(\mathbf{m}_1, \mathbf{m}_2) = 0$ , and it is counter-intuitive.

### D. Example 10

Suppose that the FOD is  $\Theta = \{\theta_1, \dots, \theta_{10}\}$ . A BBA  $\mathbf{m}_t$  defined on  $\Theta$  is

$$m_t(\Theta) = 0.1, m_t(\{\theta_2, \theta_3, \theta_3\}) = 0.05, m_t(\{\theta_7\}) = 0.05, \\ m_t(A_t) = 0.8.$$

where  $A_t$  is a varying focal element from  $\{\theta_1\}$  to  $\Theta$ . One singleton  $\{\theta_i\}$  is added at each step. All the  $A_t$ ,  $\forall t = 1, \dots, 10$  are as shown in Table V. The second BBA  $\mathbf{m}^*$  has only one focal element, and it is defined as

$$m^*(\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}) = 1.$$

Table V  
EXAMPLE 10: DISTANCE VALUE CHANGES WITH  $A_t$ .

Step	$A_t$
1	$\{\theta_1\}$
2	$\{\theta_1, \theta_2\}$
3	$\{\theta_1, \theta_2, \theta_3\}$
4	$\{\theta_1, \theta_2, \theta_3, \theta_4\}$
5	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$
6	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$
$\vdots$	$\vdots$
10	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \dots, \theta_{10}\}$

We use different distance measures including  $d_J$ ,  $d_T$ ,  $d_C$ ,  $d_F$ ,  $d_{BI}^E$  and  $d_{BI}^C$  to calculate the distance between  $\mathbf{m}^*$  and  $\mathbf{m}_t$ . Their behaviors are shown in Fig. 6.

Intuitively, when  $A_t$  starts from the focal element  $\{\theta_1\}$  to the focal element  $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , the distance between  $\mathbf{m}_t$  and  $\mathbf{m}^*$  should become smaller. When  $A_t = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , the distance should reach the minimum value. Then, when the size of  $A_t$  becomes larger and departs from  $\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , the distance value should become larger. As shown in Fig. 6,  $d_J$ ,  $d_T$ ,  $d_F$  and our proposed  $d_{BI}^E$  provide expected behaviors.

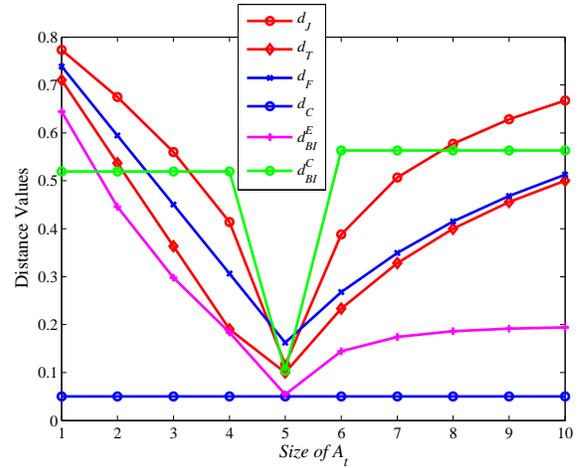


Figure 6. Distance between  $\mathbf{m}_t$  and  $\mathbf{m}^*$  for Example 10.

Since the conflict between  $\mathbf{m}_t$  and  $\mathbf{m}^*$  are fixed, i.e.,  $d_C(\mathbf{m}_t, \mathbf{m}^*) = m_t(\{\theta_7\}) \cdot m^*(\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}) = 0.05$ , the value of  $d_C$  is fixed to 0.05. Therefore,  $d_C$  is not a proper distance. As shown in Fig. 6, our proposed  $d_{BI}^E$  performs well, however  $d_{BI}^C$  does not provide a satisfactory behavior. Although  $d_{BI}^C$  reaches its minimum value when  $A_t = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$ , it cannot detect the change of  $A_t$  when the size of  $A_t$  is smaller than 5 or the size of  $A_t$  is larger than 5.

### E. Example 11

Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$ . In each case of this example, we set a fixed BBA  $\mathbf{m}_2$ , re-

spectively, where  $m_2(B) = 1$ .  $B$  can be considered as a “desired” focal element. Another BBA  $m_1$  is also set, where  $m_1(A) = 1/63, \forall A \subseteq \Theta$ . Let  $m_1$  approximate to  $m_2$  in some way. To implement this, at each step, we increase  $m_1(B)$ 's value of  $\Delta = 0.02$  and the mass value of other focal elements ( $A \neq B, \forall A \subseteq \Theta$ ) is decreased of  $\Delta/62$ .

We also let  $m_1$  go away from  $m_2$ . To implement this, at each step,  $m_1(C), C \neq B, \forall C \subseteq \Theta$  has an increase of  $\Delta = 0.02$  and the mass value of other focal elements ( $A \neq C, \forall A \subseteq \Theta$ ) has a decrease of  $\Delta/62$ . Therefore,  $C$  can be considered as an “undesired” focal element.

We use different distances between  $m_1$  and  $m_2$  at each step. Their behaviors with varying  $m_1$  are analyzed.

1) *Case A*: Here  $B = \{\theta_2\}$ , i.e., the desired focal element  $B$  is a singleton. With the change of  $m_1(B)$ ,  $m_1$  is gradually close to  $m_2$ . Therefore, if a distance measure becomes smaller with the change of  $m_1(B)$ , then it behaves as intuitively expected.

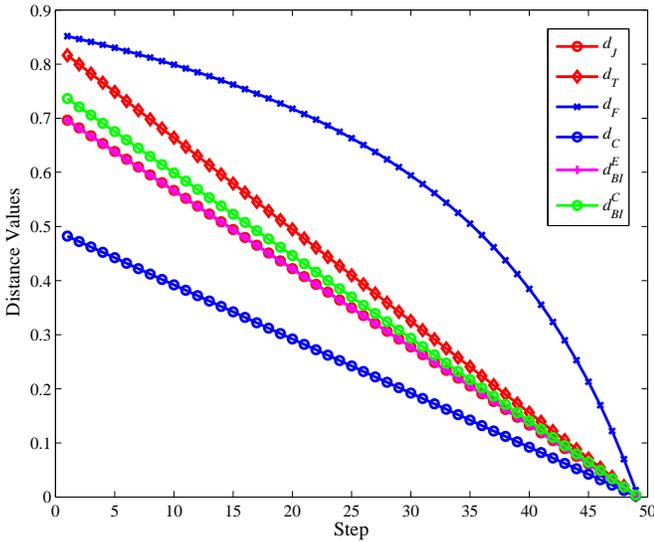


Figure 7. Distance between  $m_1$  and  $m_2$  for Example 11-Case A.

The changes of the different distance measures in the above procedure are shown in Fig. 7. All the distance measures used here provide expected behaviors.

2) *Case B*: Here  $|B| > 1$ , e.g.,  $B = \{\theta_1, \theta_2\}$  or  $B = \{\theta_1, \theta_2, \theta_3\}$ . That is, the desired focal element  $B$  is a compound focal element. With the change of  $m_1(B)$ ,  $m_1$  is gradually close to  $m_2$ .

Given different  $|B|$ , the changes of different distances in the above procedure are shown in Fig. 8, where all distances used here provide expected behaviors when  $B$  is a compound focal element.

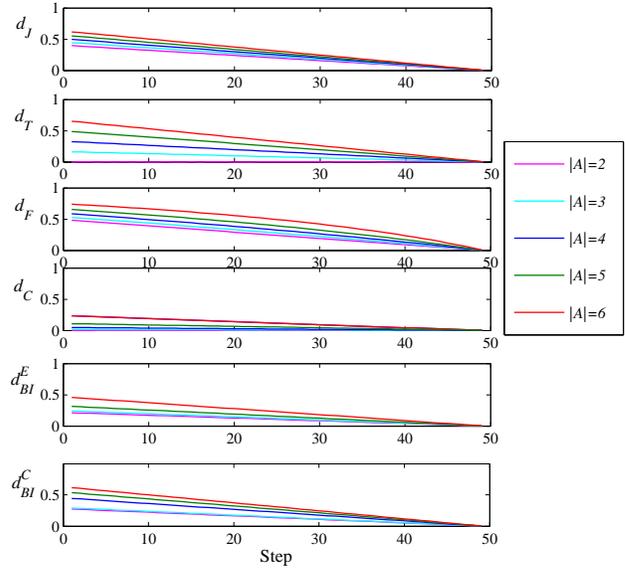


Figure 8. Distance between  $m_1$  and  $m_2$  for Example 11-Case B.

3) *Case C*: Here  $B = \{\theta_4\}$  and  $C = \{\theta_5\}$ , where the undesired focal element  $C$  is a singleton.

With the change of  $m_1(C)$ ,  $m_1$  is gradually away from  $m_2$ . If a distance measure becomes larger with the change of  $m_1(C)$ , then it behaves reasonably (i.e., as intuitively expected).

The changes of the different distance measures in the above procedure are shown in Fig. 9, where all the distance measures tested here provide expected behaviors.

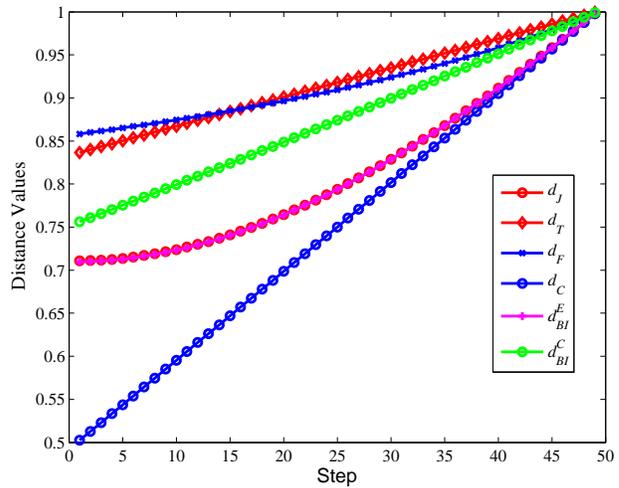


Figure 9. Distance between  $m_1$  and  $m_2$  for Example 11-Case C.

4) *Case D*: Here  $B = \{\theta_6\}$  and the undesired focal element  $C = \Theta$ . With the change of  $m_1(C)$ ,  $m_1$  is gradually away from

$m_2$ . If a distance measure becomes larger with the change of  $m_1(C)$ , then it behaves as intuitively expected. Fig. 10 shows the changes of the different distance measures in the above procedure.

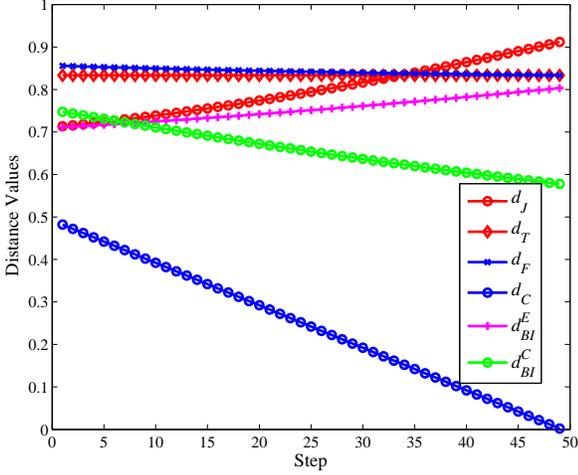


Figure 10. Distance between  $m_1$  and  $m_2$  for Example 11-CaseD.

As seen in Fig. 10, only  $d_{BI}^E$  and  $d_J$  behave as expected.  $d_T$  never changes in the whole procedure, because the corresponding pignistic probability never change with the increase of  $m_1(\Theta)$ .  $d_C$  diminishes significantly with the increase of  $m_1(\Theta)$ , because as the mass assignment is increasing for the total set, the conflict between  $m_1$  and  $m_2$  becomes smaller. Therefore,  $d_C$  is only a conflict degree and must not be used as a proper distance measure.

#### F. Example 12

Suppose that the FOD is  $\Theta = \{\theta_1, \dots, \theta_{10}\}$ . A BBA  $m_t$  defined on  $\Theta$  is

$$m_t(\Theta) = 0.1, m_t(\{\theta_2, \theta_3, \theta_4\}) = 0.05, m_t(\{\theta_7\}) = 0.05, \\ m_t(B_t) = 0.8.$$

where  $B_t$  is a varying focal element from  $\{\theta_1\}$  to  $\Theta$ . One singleton  $\theta_i$  is added at each step (step 1-10).  $B_t, \forall t = 1, \dots, 10$  equals to  $A_t$  as shown in Table V in Example 10.

From the step 11 - 19,  $B_t$  is pruned from its first element until attaining the singleton  $\{\theta_{10}\}$  at step 19. All the  $B_t$  at different steps are shown in Table VI. The second BBA  $m^*$  is  $m^*(\{\theta_{10}\}) = 1$ . We test different distance measures including  $d_J, d_T, d_C, d_F, d_{BI}^E$  and  $d_{BI}^C$  to calculate the distance between  $m^*$  and  $m_t$ . Their behaviors are shown in Fig. 11.

From the Step 1 to 9,  $B_t$  does not include  $\{\theta_{10}\}$ . At the step 10,  $B_t = \{\theta_1, \dots, \theta_{10}\}$ , which first includes  $\{\theta_{10}\}$ . After the Step 10, all distance values diminish to reach their minimum values when  $B_t = \{\theta_{10}\}$ . This is what we expect intuitively.

Table VI  
EXAMPLE 10: DISTANCE VALUE CHANGES WITH  $A_t$ .

Step	$B_t$
1-10	$A_t$
11	$\{\theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
12	$\{\theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
13	$\{\theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
14	$\{\theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
15	$\{\theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}\}$
...	...
19	$\{\theta_{10}\}$

At the first stage (Step 1 - Step 9),  $d_C$  does not change when  $B_t$  changes, because the conflict between  $m_t$  and  $m^*$  never changes before the step 10, where

$$d_C(m_t, m^*) = (m_t(\{\theta_1, \theta_2, \theta_3\}) \\ + m_t(\{\theta_7\}) + m_t(B_t)) \cdot m^*(\{\theta_{10}\}) = 0.9.$$

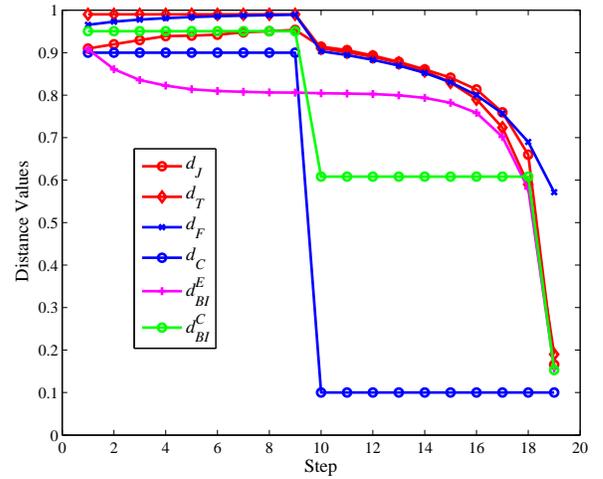


Figure 11. Distance between  $m_t$  and  $m^*$  for Example 12.

At the second stage (Step 10 - Step 19),  $d_C$  does not change with the change of  $B_t$ . Although with the emergence of  $\{\theta_{10}\}$ ,  $d_C$  diminishes, however, its value is fixed up to the final step, because the degree of conflict never changes after the decreasing at the Step 10, where

$$d_C(m_t, m^*) = (m_t(\{\theta_1, \theta_2, \theta_3\} + m_t(\{\theta_7\})) \cdot m^*(\{\theta_{10}\}) \\ = 0.1.$$

Therefore,  $d_C$  must not be used as a proper distance measure. It is just a degree of conflict between two BBAs.

At the first stage,  $d_F$  provides unsatisfactory behavior. It slightly increases with the change of  $B_t$ , that is, it is insensitive to the change of  $B_t$  in the first stage. At the second stage,  $d_F$  provides an expected behavior, i.e., it decreases and reaches its minimum value at the final Step 19.

$d_{BI}^C$  is insensitive to the change of  $B_t$  in both the first and the second stages. Its value never changes in the first stage

and after a decreasing at the step 10, it remains unchanged in the second stage.

The major difference between the behaviors of  $d_J$  and  $d_{BI}^E$  is in the first stage, where  $d_J$  increases while  $d_{BI}^E$  decreases. We think that the decrease makes more sense in fact, and the reason is as follows. In the first stage, the size of  $B_t$  becomes larger, and thus, the degree of uncertainty, i.e., the ambiguity of  $\mathbf{m}_t$  increases. For two focal elements  $\{\theta_1\}$  and  $\{\theta_1, \theta_2\}$ , although they both do not include  $\{\theta_{10}\}$ , the distance from  $\{\theta_{10}\}$  to a more ambiguous case, i.e.,  $\{\theta_1, \theta_2\}$  intuitively should be smaller than the distance from  $\{\theta_{10}\}$  to a more specific case. We can make an analogy here.  $\{\theta_{10}\}$  is our desired result, while  $\{\theta_1\}$  and  $\{\theta_1, \theta_2\}$  are two undesired results. A more ambiguous undesired result should be more preferred than a clear undesired result, i.e., the distance from the desired result to the more ambiguous undesired result should be intuitively smaller.

With the increase of  $|B_t|$ , such a distance should intuitively further decrease. Therefore,  $d_{BI}^E$  provides the correct expected behavior in this example.

### G. Example 13

Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_{2n}\}$ . Two BBAs defined on  $\Theta$  are

$$\begin{aligned} \mathbf{m}_1 : m_1(\{\theta_1\}) = m_1(\{\theta_2\}) = \dots = m_1(\{\theta_n\}) = 1/n; \\ \mathbf{m}_2 : m_2(\{\theta_{n+1}\}) = m_2(\{\theta_{n+2}\}) = \dots = m_2(\{\theta_{2n}\}) = 1/n; \end{aligned}$$

In this example, we set  $n$  from 1 to 7, i.e., the size of FOD is from 2 to 14. We use  $d_J$ ,  $d_T$ ,  $d_C$ ,  $d_F$ ,  $d_{BI}^E$  and  $d_{BI}^C$  to calculate the distance between  $\mathbf{m}_1$  and  $\mathbf{m}_2$  given different values of  $n$ . The distance values are shown in Fig. 12. As we can see in Fig. 12, all the distance measures except for our proposed  $d_{BI}^E$  remain unchanged with the increase of  $n$ . Our proposed  $d_{BI}^E$  decreases with the increase of  $n$ , which intuitively makes sense. The reason is as follows. With the increase of  $n$  from  $k-1$  to  $k$ , the cardinality of the FOD, i.e.,  $|\Theta| = 2(k-1)$  also increases to  $2k$ . Then, the number of all possible ‘‘focal’’ elements<sup>10</sup> increases from  $2^{2(k-1)}$  to  $2^k$ .

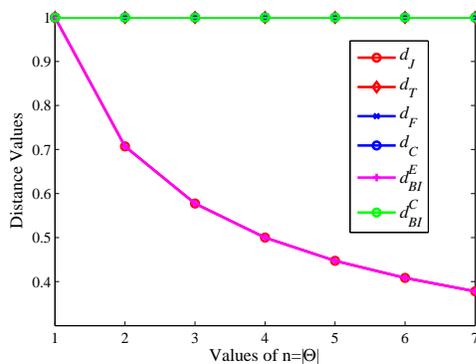


Figure 12. Distance between  $\mathbf{m}_1$  and  $\mathbf{m}_2$  for Example 13.

<sup>10</sup>Here ‘‘focal’’ elements refer to all the subsets of the FOD  $\Theta$ . They could have non-zero mass values or zero mass values.

Note that for each BBA, there are only  $n$  focal elements with non-zero mass assignment. When  $n = k - 1$ , for each BBA, there are  $2(k - 1)$  focal elements in total with non-zero value; when  $n = 2k$ , for each BBA, there are  $2(k - 1)$  focal elements in total with non-zero value. So, the number of focal elements with non-zero mass assignment increases from  $2(k - 1)$  to  $2k$ , i.e., only two more focal elements with non-zero mass assignment are added.

On the other hand, when  $n = k - 1$ , for each BBA, there are  $2^{2(k-1)} - (k - 1) - 1$  focal elements in total with zero mass assignment; when  $n = k$ , for each BBA, there are  $2^{2k} - k - 1$  focal elements in total with zero mass assignment. That is, with the increase of  $n$  from  $k - 1$  to  $k$ , there are  $2^{2k} - k - 1 - (2^{2(k-1)} - (k - 1) - 1) = 3 \times 4^{k-1} + 1$  more focal elements with zero mass values.

The common part (‘‘focal’’ elements with zero mass assignment) between  $\mathbf{m}_1$  and  $\mathbf{m}_2$  is significantly enlarged. At the same time, their different parts (those focal elements with non-zero values) only slightly increases of 2. Therefore, their distance should decrease. So, our proposed  $d_{BI}^E$  also behaves as expected in this case.

### H. Brief summary

According to above examples, our proposed  $d_{BI}^E$  behaves as expected in all the cases, in contrary to other measures compared.  $d_J$  also behaves well in many cases, however, in some special cases, it provides counter-intuitive behaviors. Our proposed  $d_{BI}^C$  behaves as expected in many cases, however, it is insensitive to the change of BBA due to the  $L_\infty$  norm used in its definition. Other measures like  $d_C$ ,  $d_F$ ,  $d_T$  are not strict distance metrics. They generate counter-intuitive behaviors in some cases, although they can be used to describe the dissimilarity between BOEs in particular cases.

Note that the results of the above examples can only show that our proposed distance measures behave as expected in those cases in the examples. Whether the rationality of our proposed measures has more generalized meaning needs further theoretical analysis besides the testing based on examples.

In the following part, simulation results based on random experiments are presented.

### I. Simulation

In this section, different measures are compared based on random simulations.

Relationship analyses are helpful for the joint use of multiple distance measures. Almost all the available distance measures have their own pros and cons. If one does not trust any single distance measure, one can use two distance measures together to construct a 2-D measure to describe the dissimilarity between two BOEs, e.g., Liu’s 2-D measure [40]. Then how to describe such a complementarity between members in a 2-D measure? As referred in Jousselme’s survey [22], a low correlation (close to 0) between two measures means that they quantify two distinct (and possibly complementary) aspects of the distance between two belief functions, while a

high correlation means that they are redundant. Hence, weakly correlated pairs of distances could be good candidates for 2-D measures.

The relationships between different measures are described using scatter plots and the correlation coefficient. The basic procedure of the simulations is as follows.

Let  $\mathcal{D}$  denote the set of distance measures used here, which includes  $d_J$ ,  $d_T$ ,  $d_C$ ,  $d_F$ ,  $d_{BI}^E$  and  $d_{BI}^C$ . Here, we calculate the correlation between different distance measures as follows.

- 1) Set the size of FOD to  $|\Theta|$  and generate  $N_s$  BBAs:  $\mathbf{m}^s$  ( $s = 1, \dots, N_s$ ) according to Algorithm 1 [39] in Table I.
- 2) Generate a reference BBA  $\mathbf{m}^r$  according to Algorithm 1.
- 3) Pick up a distance pair  $d_x$  and  $d_y$ , where  $d_x, d_y \in \mathcal{D}$  and calculate  $(d_x(\mathbf{m}^r, \mathbf{m}^s), d_y(\mathbf{m}^r, \mathbf{m}^s))$  for all  $s = 1, \dots, N_s$ .
- 4) Draw the scatter plot for  $(d_x(\mathbf{m}^r, \mathbf{m}^s), d_y(\mathbf{m}^r, \mathbf{m}^s))$  ( $s = 1, \dots, N_s$ ) to show the correlation between  $d_x$  and  $d_y$ .
- 5) Compute the correlation coefficient [22] for  $d_x$  and  $d_y$ :

$$CR(d_x, d_y) = \frac{\sum_{s=1}^{N_s} (d_x^s - \bar{d}_x) (d_y^s - \bar{d}_y)}{\sqrt{\sum_{s=1}^{N_s} (d_x^s - \bar{d}_x)^2} \sqrt{\sum_{s=1}^{N_s} (d_y^s - \bar{d}_y)^2}}, \quad (34)$$

where  $d_x^s$  denotes  $d_x(\mathbf{m}^r, \mathbf{m}^s)$ ,  $d_y^s$  denotes  $d_y(\mathbf{m}^r, \mathbf{m}^s)$ ,  $\bar{d}_x$  denotes the mean of  $d_x^s$ ,  $s = 1, \dots, N_s$ , and  $\bar{d}_y$  denotes the mean of  $d_y^s$ ,  $s = 1, \dots, N_s$ . For each pair  $d_x$  and  $d_y$  in  $\mathcal{D}$ , we calculate their correlation coefficient, to obtain a correlation matrix  $CR$ .

In simulations, we generate five types of BBAs:

- *Complete BBA*: A BBA with  $2^{|\Theta|} - 1$  focal elements with non-zero mass assignment.
- *Fixed length BBA*: A BBA with a fixed number of focal elements.
- *Simple support BBA*:  $m(A) = a, m(\Theta) = 1 - a$ , where  $A \subset \Theta$  and  $a \in [0, 1]$ .
- *Dichotomous BBA*:  $m(A) = a, m(\bar{A}) = b, m(\Theta) = 1 - a - b$ , where  $A \subset \Theta$ ,  $\bar{A}$  is the complementary set of  $A \in \Theta$ ,  $a, b \in [0, 1]$  and  $a + b \leq 1$ .
- *Consonant support BBA*: A BBA with nested focal elements, e.g.,  $\{\theta_1\}, \{\theta_1, \theta_2\}, \{\theta_1, \theta_2, \theta_3\}$ .

One can just make minor modifications to Algorithm 1 to randomly generate the above types of BBA.

**Case A:** Here we set  $|\Theta| = 8$ . Randomly generate 4000 complete BBAs  $\mathbf{m}^s$ ,  $s = 1, \dots, 4000$ . The reference BBA (complete)  $\mathbf{m}^r$  is also randomly generated. According to the above simulation steps, we can obtain the scatter plots between each pair of distance measures in  $\mathcal{D} = \{d_J, d_T, d_C, d_F, d_{BI}^E, d_{BI}^C\}$  as shown in Fig. 13, where their corresponding correlation coefficients are also provided for convenience.

As we can see in Fig. 13, our proposed  $d_{BI}^E$  and  $d_{BI}^C$  have high correlation with Jousselme's distance  $d_J$ , which is a strict distance metric and performs well in many cases as demonstrated in the previous subsection.

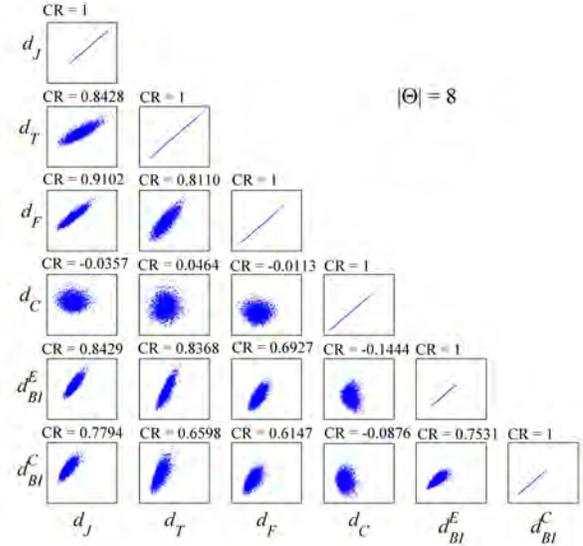


Figure 13. Scatter plots for  $|\Theta| = 8$  using complete BBAs.

$d_C$  always has low correlation with other measures, since it is actually a degree of conflict, which is different to the distance.

If a 2-D or 3-D measure to jointly use multiple distance measures is desired, we can refer to the scatter plots and corresponding CR values in Fig. 13. As aforementioned, the members in the 2-D measure should better have low correlation (close to 0), thus, they could be possibly complementary. As shown in Fig. 13, our proposed  $d_{BI}^E$  and  $d_{BI}^C$  have relatively low correlation with  $d_C$  and  $d_F$ , therefore,  $d_C$  and  $d_F$  are more proper to be selected to construct 2-D measures. The focus of this paper is not the 2-D measures. We mention 2-D measure just to show our motivation of the correlation analysis between different 1-D measures. If one is interested in the construction and applications of 2-D measures, one can refer to Liu's work [40], where  $d_T$  and  $d_C$  are used jointly as a 2-D measure.

As shown in the previous subsection,  $d_{BI}^E$  and  $d_J$  are two very appealing measures when compared with others, and they seem highly correlated to each other. Therefore, in the sequel, we will discuss the relationship between  $d_J$  and our proposed  $d_{BI}^E$  in details.

**Case B:** Although in Case A, the high correlation between  $d_{BI}^E$  and  $d_J$  has already been verified, with different FOD size  $|\Theta|$ , the correlation degree can be different. Here we use different FOD size  $|\Theta|$  to check whether the correlation between  $d_{BI}^E$  and  $d_J$  is greatly affected by  $|\Theta|$  or not, and to obtain the influence trend with the change of  $|\Theta|$ .

In this case, we set the size of the FOD to  $|\Theta| = 2, 3, 4, 5, 6, 7, 8$ , respectively. First, randomly generate 4000 complete BBAs, 4000 simple support BBAs, 4000 dichotomous BBAs and 4000 consonant support BBAs. Their corresponding reference BBAs (complete, simple support, dichotomous, consonant support)  $\mathbf{m}^r$ 's are also randomly generated.

Following the above steps and under different sizes of FOD, we can obtain the scatter plots between each pair of distance measures in  $\mathcal{D} = \{d_J, d_{BI}^E\}$  for the 4000 complete BBAs, 4000 simple support BBAs, 4000 dichotomous BBAs and 4000 consonant support BBAs, respectively, as shown in Fig. 14.

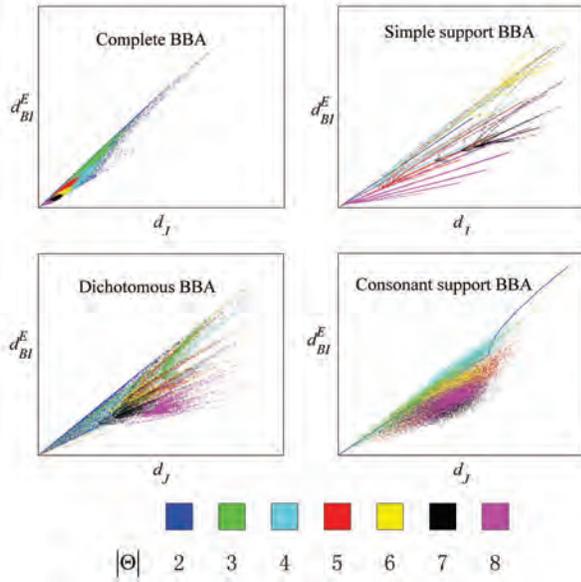


Figure 14. Scatter plots for  $|\Theta| = 2, \dots, 8$  using different types of BBAs.

With the increase of  $|\Theta|$ , the evolution of the correlation coefficient between  $d_J$  and  $d_{BI}^E$  for four different types of BBAs including complete, simple support, dichotomous, and consonant support are shown in Fig. 15.

As seen in Figs. 14 and 15, the increase of  $|\Theta|$  leads to the decrease of the correlation coefficient for all types of BBAs. No matter using which types of BBA,  $d_J$  and  $d_{BI}^E$  are highly correlated, although the correlation coefficient decreases with the increase of  $|\Theta|$ . As aforementioned, this to some extent shows the rationalities of our proposed new measure  $d_{BI}^E$ .

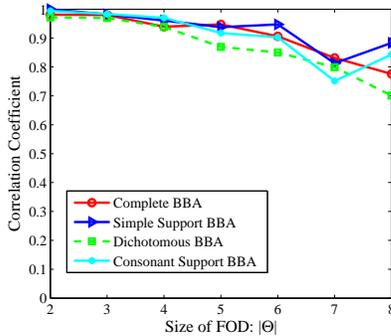


Figure 15. Evolution of the correlation coefficient between  $d_J$  and  $d_{BI}^E$  using different types of BBAs.

### J. Application of distance in BBA approximation evaluation

Here we provide an application of different distance measures of evidence in BBA approximations. The BBA approximation [23], [41], [42] aims to obtain a simpler BBA by removing some focal elements and thus to reduce the computational cost in the evidence combination and other operations in DST [1], [43]. A good BBA approximation should have little loss of information when simplifying the BBA. If the BBA obtained using an approximation is closer to the original BBA, such an approximation has less loss of information and thus, is more desired. Therefore, we can use the distance of evidence to evaluate BBA approximations.

Here three types of BBA approximations are compared including  $k-l-x$  [23], D1 [41] and summarization (Sum) [42]. Using  $k-l-x$ , the approximated BBA is obtained by

- 1) keeping no less than  $k$  focal elements;
- 2) keeping no more than  $l$  focal elements;
- 3) deleting the masses which are no larger than  $x$ .

Sum method [42] also keeps focal elements with the largest mass values as in  $k-l-x$ . The masses of removed focal elements are accumulated and assigned to their union set.

D1 method [41] is to keep some focal elements with the largest mass values in the original BBA and to re-assign the mass assignments of the other focal elements to those kept focal elements according to a well-designed criterion. See more details in related references [23], [41], [42].

$k-l-x$  has a coarse way of re-normalization, and Sum method re-assigns the masses of removed focal elements to their union set. D1 has a more subtle way to re-assign the mass, therefore, D1 should be a better method. Here we provide a simulation with distance of evidence as the evaluation criterion to check if the evaluation results agree with the analysis.

In our simulation,  $|\Theta| = 4$ . A complete BBA  $m$  (i.e., with  $2^4 - 1 = 15$  non-empty focal elements) can be randomly generated according to the Algorithm 1 in Table I. We use the distance of evidence ( $d_J, d_T, d_F, d_{BI}^E$  and  $d_{BI}^C$ , respectively) between the approximated BBA  $\hat{m}$  and the original one  $m$  in average as the performance evaluation criterion.

Our comparative analyses have 1000 Monte Carlo runs (i.e., totally 1000 complete BBAs are randomly generated). The number of remaining focal elements  $r$  for the approaches used here are set to from 14 down to 2 (decrease by 1). Then, different approximation results in each run can be obtained using the different approximations given a number  $r$ . The average (over 1000 runs) distance values between the original BBA  $m$  and the approximated BBA  $\hat{m}$  obtained using different approaches given different remaining focal elements number are shown in Fig. 16 (a)-(e).

Here the parameter in  $k-l-x$  is set as  $k = l = r$  and  $x = 0.5$ . As shown in Fig. 16 (a)-(e), using different distances, the distance values are different; however, the changing trends are the same, i.e., with the decrease of the number remaining focal elements, the distance value increases. This represents more

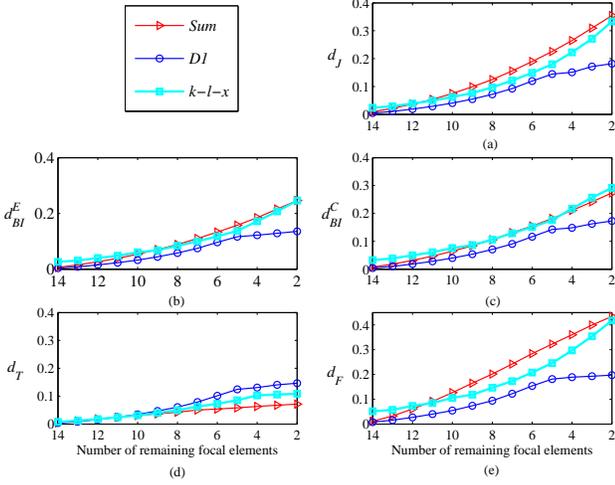


Figure 16. Evaluation of BBA approximations using different distances.

loss of information. Based on all the distances of evidence used here except for  $d_T$ , the BBA obtained by D1 is usually closer to the original BBA. It is experimentally shown that when using the distances of evidence including our new proposed ones, D1 is a better BBA approximation when compared with others. This is accordant to the analyses above, therefore, our proposed distances of evidence can be well used in performance evaluation in belief function related applications.

### K. Application of $d_{BI}^E$ in multiple criteria decision making

Here we provide a multiple criteria decision making (MCDM) application of our developed measure  $d_{BI}^E$ , which usually performs well in the previous examples and simulations.

Let's consider a selection problem in the car purchase. Four cars  $\{A_1, A_2, A_3, A_4\}$  are considered:

- $A_1$  = TOYOTA YARIS 69 VVT-i Tendance;
- $A_2$  = SUZUKI SWIFT MY15 1.2 VVT So'City;
- $A_3$  = VOLKSWAGEN POLO 1.0 60 Confortline;
- $A_4$  = OPEL CORSA 1.4 Turbo 100 ch Start/Stop Ed.;

Following criteria are for selecting the best car to purchase:

- $C_1$  is the price (in €);
- $C_2$  is fuel consumption (in L/km);
- $C_3$  is the CO<sub>2</sub> emission (in g/km);
- $C_4$  is the fuel tank volume (in L);
- $C_5$  is the trunk volume (in L);

From information extracted from car-makers technical characteristics available on the Internet<sup>11</sup>, we can build the score matrix  $\mathbf{S} = [S_{ij}]$  for the above four cars as

$$\mathbf{S} = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} & \begin{bmatrix} 15000 & 4.3 & 99 & 42 & 737 \\ 15290 & 5.0 & 116 & 42 & 892 \\ 15350 & 5.0 & 114 & 45 & 952 \\ 15490 & 5.3 & 123 & 45 & 1120 \end{bmatrix} \end{matrix}$$

When we use criteria  $C_1$ ,  $C_2$  and  $C_3$ , smaller is better. For criteria  $C_4$  and  $C_5$ , larger is better. We multiply values of columns  $C_1$ ,  $C_2$  and  $C_3$  by -1 to generate a modified score matrix  $\mathbf{S}'$  in order that the MCDM problem here is with homogeneous preference order ("larger is better") for each column:

$$\mathbf{S}' = \begin{matrix} & C_1 & C_2 & C_3 & C_4 & C_5 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} & \begin{bmatrix} -15000 & -4.3 & -99 & 42 & 737 \\ -15290 & -5.0 & -116 & 42 & 892 \\ -15350 & -5.0 & -114 & 45 & 952 \\ -15490 & -5.3 & -123 & 45 & 1120 \end{bmatrix} \end{matrix}$$

For simplicity, the importance  $imp(C_j)$  of each criteria  $C_j$  takes a value in  $\{1, 2, 3, 4, 5\}$ , where 1 means *the least important*, and 5 means *the most important*. Here,  $imp(C_1) = 5$ ,  $imp(C_2) = 4$ ,  $imp(C_3) = 4$ ,  $imp(C_4) = 1$  and  $imp(C_5) = 3$  are adopted, which means that the price (criteria  $C_1$ ) is the most important one and the volume of fuel tank (criteria  $C_4$ ) is the least important one. According to these importance values and after the normalization, we obtain the following vector of relative weights of criteria:  $\mathbf{w} = [\frac{5}{17}, \frac{4}{17}, \frac{4}{17}, \frac{1}{17}, \frac{3}{17}]$ .

We use the BF-TOPSIS (Belief Function based Technique for Order Preference by Similarity to Ideal Solution) approach [44] with our  $d_{BI}^E$  to solve the MCDM problem above.

First, from the score matrix  $\mathbf{S}'$ , generate BBAs<sup>12</sup>  $m_{i,j}(A_i)$ ,  $m_{i,j}(\bar{A}_i)$ , and  $m_{i,j}(A_i \cup \bar{A}_i)$  according to the BBA generation approach proposed in [44] as:

$$\begin{aligned} m_{1,1}(A_1) &= 0.9859, m_{1,1}(A_2 \cup A_3 \cup A_4) = 0.0047, \\ m_{1,1}(\Theta) &= 0.0094; m_{2,1}(A_2) = 1.0, \\ m_{2,1}(A_1 \cup A_3 \cup A_4) &= 0, m_{2,1}(\Theta) = 0; \\ m_{3,1}(A_3) &= 0.0022, m_{3,1}(A_1 \cup A_2 \cup A_4) = 0.9932, \\ m_{3,1}(\Theta) &= 0.0046; m_{4,1}(A_4) = 1.0, \\ m_{4,1}(A_1 \cup A_2 \cup A_3) &= 0, m_{4,1}(\Theta) = 0; \end{aligned}$$

$$\begin{aligned} m_{1,2}(A_1) &= 1.0, m_{1,2}(A_2 \cup A_3 \cup A_4) = 0, m_{1,2}(\Theta) = 0; \\ m_{2,2}(A_2) &= 0.1250, m_{2,2}(A_1 \cup A_3 \cup A_4) = 0.4375, \\ m_{2,2}(\Theta) &= 0.4375; \\ m_{3,2}(A_3) &= 0.1250, m_{3,2}(A_1 \cup A_2 \cup A_4) = 0.4375, \\ m_{3,2}(\Theta) &= 0.4375; \\ m_{4,2}(A_4) &= 1.0, m_{4,2}(A_1 \cup A_2 \cup A_3) = 0, m_{4,2}(\Theta) = 0; \end{aligned}$$

$$\begin{aligned} m_{1,3}(A_1) &= 1.0, m_{1,3}(A_2 \cup A_3 \cup A_4) = 0, m_{1,3}(\Theta) = 0; \\ m_{2,3}(A_2) &= 0.1250, m_{2,3}(A_1 \cup A_3 \cup A_4) = 0.4375, \\ m_{2,3}(\Theta) &= 0.4375; \\ m_{3,3}(A_3) &= 0.1964, m_{3,3}(A_1 \cup A_2 \cup A_4) = 0.3750, \\ m_{3,3}(\Theta) &= 0.4286; \\ m_{4,3}(A_4) &= 1.0, m_{4,3}(A_1 \cup A_2 \cup A_3) = 0, m_{4,3}(\Theta) = 0; \end{aligned}$$

$$\begin{aligned} m_{1,4}(A_1) &= 0, m_{1,4}(A_2 \cup A_3 \cup A_4) = 1, m_{1,4}(\Theta) = 0; \\ m_{2,4}(A_2) &= 0, m_{2,4}(A_1 \cup A_3 \cup A_4) = 1, m_{2,4}(\Theta) = 0; \\ m_{3,4}(A_3) &= 1.0, m_{3,4}(A_1 \cup A_2 \cup A_4) = 0, m_{3,4}(\Theta) = 0; \\ m_{4,4}(A_4) &= 1.0, m_{4,4}(A_1 \cup A_2 \cup A_3) = 0, m_{4,4}(\Theta) = 0; \end{aligned}$$

$$\begin{aligned} m_{1,5}(A_1) &= 0, m_{1,5}(A_2 \cup A_3 \cup A_4) = 1, m_{1,5}(\Theta) = 0; \\ m_{2,5}(A_2) &= 0.1990, m_{2,5}(A_1 \cup A_3 \cup A_4) = 0.3825, \\ m_{2,5}(\Theta) &= 0.4185; \\ m_{3,5}(A_3) &= 0.3530, m_{3,5}(A_1 \cup A_2 \cup A_4) = 0.2231, \\ m_{3,5}(\Theta) &= 0.4239; \\ m_{4,5}(A_4) &= 1.0, m_{4,5}(A_1 \cup A_2 \cup A_3) = 0, m_{4,5}(\Theta) = 0; \end{aligned}$$

Second, for each alternative  $A_i$ , compute the  $d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}})$  between  $\mathbf{m}_{i,j}$  and the best ideal BBA defined by  $\mathbf{m}_{i,j}^{\text{best}}(A_i) \triangleq$

<sup>12</sup> $i = 1, \dots, 4$  denotes the index of the alternative;  $j = 1, \dots, 5$  denotes the index of the criterion.

<sup>11</sup><http://www.choisir-sa-voiture.com>

1, and the distances  $d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}})$  between  $\mathbf{m}_{i,j}$  and the worst ideal BBA defined by  $\mathbf{m}_{i,j}^{\text{worst}}(\bar{A}_i) \triangleq 1$ . Then, two distance matrices<sup>13</sup> are obtained:

$$\mathbf{D}_{BI}^{\text{best}} = \begin{bmatrix} 0 & 0 & 0 & 0.8660 & 0.8660 \\ 0.6151 & 0.7032 & 0.7071 & 0.8660 & 0.6419 \\ 0.7100 & 0.7032 & 0.6430 & 0 & 0.5102 \\ 0.8660 & 0.8660 & 0.8660 & 0 & 0 \end{bmatrix},$$

and

$$\mathbf{D}_{BI}^{\text{worst}} = \begin{bmatrix} 0.8660 & 0.8660 & 0.8660 & 0 & 0 \\ 0.2804 & 0.2033 & 0.1938 & 0 & 0.2552 \\ 0.1885 & 0.2033 & 0.2555 & 0.8660 & 0.3819 \\ 0 & 0 & 0 & 0.8660 & 0.8660 \end{bmatrix}.$$

Here, the element  $D_{BI}^{\text{best}}(i, j) = d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}})$  and  $D_{BI}^{\text{worst}}(i, j) = d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}})$ .

Third, compute the weighted average of  $d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}})$  values with relative importance weighting factor  $w_j$  of criteria  $C_j$ . Similarly, compute the weighted average of  $d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}})$  values. More specifically, compute

$$d^{\text{best}}(A_i) \triangleq \sum_{j=1}^5 w_j \cdot d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{best}}), \quad (35)$$

$$d^{\text{worst}}(A_i) \triangleq \sum_{j=1}^5 w_j \cdot d_{BI}^E(\mathbf{m}_{i,j}, \mathbf{m}_{i,j}^{\text{worst}}). \quad (36)$$

The relative closeness of the alternative  $A_i$  with respect to ideal best solution  $A^{\text{best}}$  is then defined by

$$Cl(A_i, A^{\text{best}}) \triangleq \frac{d^{\text{worst}}(A_i)}{d^{\text{worst}}(A_i) + d^{\text{best}}(A_i)}. \quad (37)$$

Since  $d^{\text{best}}(A_i) \geq 0$  and  $d^{\text{worst}}(A_i) \geq 0$ , then  $Cl(A_i, A^{\text{best}}) \in [0, 1]$ . If  $d^{\text{best}}(A_i) = 0$ , it means that the alternative  $A_i$  coincides with the ideal best solution and thus  $Cl(A_i, A^{\text{best}}) = 1$  (the relative closeness of  $A_i$  with respect to  $A^{\text{best}}$  is maximal). Contrariwise, if  $d^{\text{worst}}(A_i) = 0$ , it means that the alternative  $A_i$  coincides with the ideal worst solution and thus  $Cl(A_i, A^{\text{best}}) = 0$  (the relative closeness of  $A_i$  with respect to  $A^{\text{best}}$  is minimal).

In the final, the set of alternatives is ranked according to the descending order of  $Cl(A_i, A^{\text{best}}) \in [0, 1]$ , where a larger  $Cl(A_i, A^{\text{best}})$  value means a better alternative (or a higher preference).

Based on the score matrix  $S'$  and importance of criteria,  $A_1$  tends to be the best car to buy, since the three most important criteria clearly take their best values for car  $A_1$ . When using the classical TOPSIS [45] method with the Euclidean distance, we obtain the preference order  $A_4 \succ A_1 \succ A_3 \succ A_2$ , where  $A_4$  is the best choice and  $A_2$  is the worst one. When we use the BF-TOPSIS method based on our proposed  $d_{BI}^E$ , we obtain a more satisfactory preference order  $A_1 \succ A_3 \succ A_2 \succ A_4$ .

As shown in this application example, our proposed distance measure can be well used in the multiple criterion decision making.  $d_{BI}^E$  has also been used successfully in other kind of applications related to risk management and for protecting housing areas against torrential floods in France [46], [47].

<sup>13</sup>One can also try to use other distance measures for belief functions as referred above. Here we only use  $d_{BI}^E$  for illustration.

## VI. CONCLUSIONS

Two novel distance measures of evidence have been proposed based on the distance measures between belief intervals. According to the comparisons between our proposed measures and the existing ones based on examples and simulations, it is shown that our proposed distances well describe the degree of closeness between different BOEs. Our results demonstrate that Euclidean distance based on belief intervals works better than the Chebyshev distance based on belief intervals.

Besides their good behaviors, the main interest of our proposed distances of evidence is that they have been established directly in the belief functions framework, contrary to most of other distance measures that switch from belief functions to probabilistic or fuzzy set framework, which leads to loss of information and bad behaviors in general.

Note that in this paper, many justifications or verifications of our new proposed distance measures are based on numerical examples and simulations. Numerical examples in belief functions related fields are usually designed according to the subjective intuitions, which lack objective criteria and the standard testing data. Furthermore, the results and conclusions only based on examples are usually incomplete. Therefore, more thorough justifications including theoretical analysis and more examples in special cases are needed to further examine our new measures. However, the theoretical evaluation or justification in belief functions related fields is still premature.

Therefore, our future work will focus on the theoretical and the objective evaluation and analysis of the belief functions related fields. We will try to establish the standard testing BBAs for the distance measures in the theory of belief functions. Our proposed distance measures will also be tested based on more experiments and simulations to find the possible counter-intuitive examples and analyze the reasons for the possible counter-intuitive behaviors. Our new distance measures will be applied to more belief functions related applications, e.g., the performance evaluations, for the further verification.

Furthermore, all the distance measures including ours are under the closed-world assumption. That is, when the mass assignment for the emptyset is positive, they cannot be used to measure the closeness between BOEs. Therefore, generalizing our new distance metrics to the open-world assumption is one of our future research directions.

## REFERENCES

- [1] G. Shafer, *A mathematical theory of evidence*. Princeton, NJ, USA: Princeton university press Princeton, 1976, vol. 1.
- [2] G. Lin, J. Liang, and Y. Qian, "An information fusion approach by combining multigranulation rough sets and evidence theory," *Information Sciences*, vol. 314, pp. 184–199, 2015.
- [3] Z. Liu, Q. Pan, and J. Dezert, "Evidential classifier for imprecise data based on belief functions," *Knowledge-Based Systems*, vol. 52, pp. 246–257, Nov. 2013.
- [4] Z. Liu, Q. Pan, J. Dezert, and G. Mercier, "Credal c-means clustering method based on belief functions," *Knowledge-Based Systems*, vol. 74, pp. 119–132, Jan. 2015.
- [5] D. Han, J. Dezert, J.-M. Tacnet, and C. Han, "A fuzzy-cautious owa approach with evidential reasoning," in *Proceedings of the 15th International Conference on Information Fusion (FUSION)*, Singapore, July 2012, pp. 278–285.

- [6] X. Xu, Z. Zhang, D.-L. Xu, and Y.-W. Chen, "Interval-valued evidence updating with reliability and sensitivity analysis for fault diagnosis." *Int. J. Computational Intelligence Systems*, vol. 9, no. 3, pp. 396–415, 2016. [Online]. Available: <http://dx.doi.org/10.1080/18756891.2016.1175808>
- [7] F. Smarandache and J. Dezert, *Advances and Applications of DSMT for Information Fusion-Collected Works-Volume 3*. American Research Press, 2009.
- [8] P. Wang, "A defect in Dempster-Shafer theory," in *Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence*, Seattle, USA, July 1994, pp. 560–566.
- [9] J. Dezert and A. Tchamova, "On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule," *International Journal of Intelligent Systems*, vol. 29, no. 3, pp. 223–252, 2014.
- [10] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [11] F. Smarandache and J. Dezert, *Advances and Applications of DSMT for Information Fusion-Collected Works-Volume 4*. American Research Press, 2015.
- [12] T. Denœux, "A neural network classifier based on Dempster-Shafer theory," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 30, no. 2, pp. 131–150, 2000.
- [13] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.
- [14] D. Han, J. Dezert, and C. Han, "New basic belief assignment approximations based on optimization," in *Proceedings of the 15th International Conference on Information Fusion*, Singapore, July 2012, pp. 286–293.
- [15] F. Cuzzolin, "Lp consonant approximations of belief functions," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 2, pp. 420–436, 2014.
- [16] D. Han, J. Dezert, and Y. Yang, "Two novel methods for BBA approximation based on focal element redundancy," in *Proceedings of the 18th International Conference on Information Fusion*, Washington DC, USA, July 2015, pp. 428–434.
- [17] Y. Yang and Y. Liu, "Iterative approximation of basic belief assignment based on distance of evidence," *PLoS ONE*, vol. 11, no. 2, pp. 1–27, Feb. 2016.
- [18] S. Hariz, Z. Elouedi, and K. Mellouli, "Clustering approach using belief function theory," in *Proceedings of the 12th International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, Varna, Bulgaria, Sept 2006, pp. 162–171.
- [19] J. Schubert, "Clustering decomposed belief functions using generalized weights of conflict," *International Journal of Approximate Reasoning*, vol. 48, no. 2, pp. 466–480, June 2008.
- [20] D. Mercier, B. Quost, and T. Denœux, "Refined modeling of sensor reliability in the belief function framework using contextual discounting," *Information Fusion*, vol. 9, no. 2, pp. 246–258, April 2008.
- [21] A. Martin, A.-L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Proceedings of the 11th International Conference on Information Fusion*. Cologne, Germany: ISIF, June 2008, pp. 1–8.
- [22] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, no. 2, pp. 118–145, 2012.
- [23] B. Tessem, "Approximations for efficient computation in the theory of evidence," *Artificial Intelligence*, vol. 61, no. 2, pp. 315–329, 1993.
- [24] D. Han, J. Dezert, C. Han, and Y. Yang, "New dissimilarity measures in evidence theory," in *Proceedings of the 14th International Conference on Information Fusion*, Chicago, USA, July 2011, pp. 1–7.
- [25] F. Cuzzolin, "A geometric approach to the theory of evidence," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, no. 4, pp. 522–534, 2008.
- [26] A. Irpino and R. Verde, "Dynamic clustering of interval data using a Wasserstein-based distance," *Pattern Recognition Letters*, vol. 29, no. 11, pp. 1648–1658, 2008.
- [27] D. Han, J. Dezert, and Y. Yang, "New distance measures of evidence based on belief intervals," in *Proceedings of the 3rd International Conference on Belief Functions*, Oxford, UK, Sept 2014, pp. 432–441.
- [28] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *International Journal of General Systems*, vol. 9, no. 4, pp. 249–260, 1983.
- [29] A.-L. Jousselme, C. Liu, D. Grenier, and É. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 36, no. 5, pp. 890–903, 2006.
- [30] D. Harmanec and G. J. Klir, "Measuring total uncertainty in Dempster-Shafer theory: A novel approach," *International Journal of General Systems*, vol. 22, no. 4, pp. 405–419, 1994.
- [31] Y. Yang and D. Han, "A new distance-based total uncertainty measure in the theory of belief functions," *Knowledge-Based Systems*, vol. 94, pp. 114–123, 2016.
- [32] L. A. Zadeh, "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination," *AI Magazine*, vol. 7, no. 2, pp. 85–90, 1986.
- [33] X. Deng, D. Han, J. Dezert, Y. Deng, and Y. Shyr, "Evidence combination from an evolutionary game theory perspective," *IEEE Transactions on Cybernetics*, vol. accepted, to appear, 2016.
- [34] P. Smets, "Decision making in the TBM: the necessity of the pignistic transformation," *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 133–147, 2005.
- [35] D. Han, J. Dezert, and Z. Duan, "Evaluation of probability transformations of belief functions for decision making," *IEEE Transactions on Systems Man and Cybernetics: Systems*, vol. 46, no. 1, pp. 93–108, 2016.
- [36] D. Han, Y. Deng, C. Han, and Y. Yang, "Some notes on betting commitment distance in evidence theory," *Science China - Information Sciences*, vol. 55, no. 3, pp. 558–565, 2012.
- [37] M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the Jaccard index matrix," *International Journal of Approximate Reasoning*, vol. 54, no. 5, pp. 615–626, 2013.
- [38] A. Antonucci, "An interval-valued dissimilarity measure for belief functions based on credal semantics," in *Proceedings of the 2nd International Conference on Belief Functions*, Compiègne, France, May 2012, pp. 37–44.
- [39] T. Burger and S. Destercke, "Random generation of mass functions: A short howto," in *Proceedings of the 2nd International Conference on Belief Functions*, Compiègne, France, May 2012, pp. 145–152.
- [40] W. Liu, "Analyzing the degree of conflict among belief functions," *Artificial Intelligence*, vol. 170, no. 11, pp. 909–924, 2006.
- [41] M. Bauer, "Approximations for decision making in the Dempster-Shafer theory of evidence," in *Proceedings of the 12th international conference on Uncertainty in artificial intelligence*. Portland, Oregon, USA: Morgan Kaufmann Publishers Inc., August 1-3 1996, pp. 73–80.
- [42] J. D. Lowrance, D. G. Thomas, and M. S. Thomas, "A framework for evidential-reasoning systems," in *Proceedings of the 5th National Conference on Artificial Intelligence, AAAI-86*. Philadelphia, Pennsylvania: AAAI, August 1986, pp. 896–903.
- [43] P. Smets, "Practical uses of belief functions," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, vol. 15. Stockholm, Sweden: Morgan Kaufmann Publishers Inc., 1999, pp. 612–621.
- [44] J. Dezert, D. Han, and H. Yin, "A new belief function based approach for multi-criteria decision-making support," in *Proceedings of the 19th International Conference on Information Fusion*, Heidelberg, Germany, July 2016, pp. 782–789.
- [45] C.-L. Hwang and K.-S. Yoon, *Multiple Attribute Decision Making: Methods and Applications*. Berlin Heidelberg: Springer-Verlag, 1981.
- [46] S. Carladous, J.-M. Tacnet, J. Dezert, and M. Batton-Hubert, "Evaluation of efficiency of torrential protective structures with new BF-TOPSIS methods," in *Proceedings of the 19th International Conference on Information Fusion*, Heidelberg, Germany, July 2016, pp. 2267–2274.
- [47] S. Carladous, J.-M. Tacnet, J. Dezert, D. Han, and M. Batton-Hubert, "Applying ER-MCDA and BF-TOPSIS to decide on effectiveness of torrent protection," in *Proceedings of the 4th International Conference on Belief Functions*, Prague, Czech, Sept 2016.

# Credibilistic Independence of Two Propositions

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, A. Tchamova, D. Han, *Credibilistic Independence of Two Propositions*, in Proc. of Int. Conf. on Information Fusion (Fusion 2018), Cambridge, UK, July 10–13, 2018, and reprinted with permission.

**Abstract**—In this paper the notion of (probabilistic) independence of two events defined classically in the theory of probability is extended in the theory of belief functions as the credibilistic independence of two propositions. This new notion of independence which is compatible with the probabilistic independence as soon as the belief function is Bayesian, is defined from Fagin-Halpern belief conditioning formulas drawn from Total Belief Theorem (TBT) when working in the framework of belief functions to model epistemic uncertainties. We give some illustrative examples of this notion at the end of the paper.

**Keywords:** credibilistic independence, belief functions, belief conditioning, total belief theorem.

## I. INTRODUCTION

In this paper the notion of (probabilistic) independence of two events defined classically in the theory of probability [1] is extended in the theory of belief functions [2]. We call it the *credibilistic independence* of two propositions to make a clear distinction between the origin of uncertainty related to events (i.e. the random or stochastic uncertainty), and in a more general context the origin of uncertainty of propositions (i.e. the epistemic uncertainty due to lack of knowledge). The epithet *credibilistic* refers to a credal system chosen for the codification of belief. In this work our credal system is the mathematical framework of belief functions.

Several works have been proposed in the past to define different notions of independences in imprecise probability framework and in the theory of belief functions. For example, Couso et al. [3] did propose several notions of independences illustrated by different combined Ellsberg's urns experiments. In 2000's Ben Yaghlane, Smets and Mellouli [4], [5] did explore the notion of independence and they define the *doxastic independence*. Their proposal is however essentially based on Dempster's rule of combination which is known problematic and incompatible with imprecise conditional probabilistic calculus as shown in [6]–[8]. More recently Jirousek and Vejnarova in [9] did propose a definition of conditional independence which is based on some complicate factorization principles of the joint basic belief assignment (BBA) into separate marginal spaces of the variables. All aforementioned works share two same basic principles for attempting to define the notion(s) of independence: 1) work on joint (Cartesian) product space, and 2) work with BBAs. These two fundamental principles yield to quite complicate definitions of independence(s) difficult to use by most engineers or researchers for their own applications or developments.

In this research work we adopt a radically different standpoint. We work with a BBA defined with respect to a single frame of discernment<sup>1</sup> (FoD), and we work directly with belief intervals induced by Fagin-Halpern conditioning rule [6], [7], rather than some factorization principles of joint BBA or extension principles of marginal BBAs. Our approach is constructive, easier than previous attempts to define independence, and consistent with the notion of probabilistic (or stochastic) independence of two events defined in the theory of probability. Our notion of credibilistic independence can be used easily to check if two propositions are credibilistically independent, or not, given a BBA. This new approach could be helpful for practitioners of belief functions. We do not have yet made more investigations for showing its usefulness for applications, but we expect it will generate some interest because this problem has been already explored by several researchers in the past based on different standpoints.

This paper is organized as follows. After a brief recall of basics of probability theory and belief functions in Sections II and III, we characterize mathematically the notion of *credibilistic independence* of two propositions in Section IV. Some basic illustrative examples are shown in Section V, with conclusions in Section VI.

## II. BASICS OF PROBABILITY THEORY

In probability theory [1], the elements  $\theta_i$  of the space  $\Theta$  are experimental outcomes. The subsets of  $\Theta$  are called events and the event  $\{\theta_i\}$  consisting of the single element  $\theta_i$  is an elementary event. The space  $\Theta$  is called the *sure event* and the empty set  $\emptyset$  is the *impossible event*. We assign to each event  $A$  a number  $P(A)$  in  $[0, 1]$ , called the probability of  $A$ , which satisfies the three Kolmogorov's axioms: 1)  $P(\emptyset) = 0$ ; 2)  $P(\Theta) = 1$ ; and 3) if  $A \cap B = \{\emptyset\}$ , then  $P(A \cup B) = P(A) + P(B)$ . The fundamental Total Probability Theorem (TPT), also called the law of total probability, see [1] states that for any event  $B$  and any partition  $\{A_1, A_2, \dots, A_k\}$  of the space  $\Theta$ , the following equality holds

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k). \quad (1)$$

<sup>1</sup>It can be any Cartesian product space in fact. The main point is that the (joint) BBA we work with is defined with respect to this space.

Starting from TPT formula (1) and assuming  $P(B) > 0$ , we get for any  $i \in \{1, \dots, k\}$  (after dividing each side of (1) by  $P(B)$  and rearranging terms) the equality

$$\frac{P(A_i \cap B)}{P(B)} = 1 - \sum_{\substack{j=1, \dots, k \\ j \neq i}} \frac{P(A_j \cap B)}{P(B)} = 1 - \frac{P(\bar{A}_i \cap B)}{P(B)}. \quad (2)$$

This equality allows us to define the conditional probability  $P(A_i|B)$  by<sup>2</sup>

$$P(A_i|B) \triangleq P(A_i \cap B)/P(B). \quad (3)$$

One can verify that the conditional probability (3) satisfies the three axioms of the Theory of Probability [1].

Similarly, by considering an event  $A_i$  of  $\Theta$  and the partition  $\{B, \bar{B}\}$  of  $\Theta$ , the formula  $P(A_i) = P(A_i \cap B) + P(A_i \cap \bar{B})$  applies, and by dividing it by  $P(A_i)$  (assuming  $P(A_i) > 0$ ), one gets

$$\frac{P(A_i \cap B)}{P(A_i)} = 1 - \frac{P(A_i \cap \bar{B})}{P(A_i)}. \quad (4)$$

This allows to define the reverse conditional probability  $P(B|A_i)$  by

$$P(B|A_i) \triangleq P(A_i \cap B)/P(A_i). \quad (5)$$

**Probabilistic Independence:** Two events  $A_i$  and  $B$  are said to be *probabilistically independent* (or P-independence for short) if and only if  $P(A_i|B) = P(A_i)$  and  $P(B|A_i) = P(B)$ . From conditioning formulas (3) and (5) and because conditional probabilities formulas  $P(A_i|B)$  and  $P(B|A_i)$  are mathematically defined only if  $P(B) > 0$  and  $P(A_i) > 0$ , one determines the condition of P-independence (which is well defined even if  $P(A_i) = 0$ , or  $P(B) = 0$ , or both) by the formula

$$P(A_i \cap B) = P(A_i)P(B). \quad (6)$$

### III. BASICS OF BELIEF FUNCTIONS

Based on Dempster's works [10], [11], Shafer did introduce Belief Functions (BF) to model the epistemic uncertainty and to reason under uncertainty in his Mathematical Theory of Evidence [2], also known as Dempster-Shafer Theory (DST). We consider a finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all exhaustive and exclusive elements of  $\Theta$  represent the set of the potential solutions of the problem under concern. The set of all subsets of  $\Theta$  is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . A basic belief assignment (BBA) associated with a given source of evidence is defined as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying the conditions  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is the mass of belief of subset  $A$  committed by the source of evidence (SoE). A focal element  $X$  of a BBA  $m(\cdot)$  is an element of  $2^\Theta$  such that  $m(X) > 0$ . Note that the empty set  $\emptyset$  is not a focal element of a BBA because  $m(\emptyset) = 0$  (closed-world assumption of Shafer's model for

the FoD). The set of all focal elements of  $m(\cdot)$  is denoted<sup>3</sup>  $\mathcal{F}_\Theta(m) \triangleq \{X \subseteq \Theta | m(X) > 0\} = \{X \in 2^\Theta | m(X) > 0\}$ . Belief and plausibility functions are defined by<sup>4</sup>

$$Bel(A) = \sum_{\substack{X \in 2^\Theta \\ X \subseteq A}} m(X) = \sum_{X \in \mathcal{F}_A(m)} m(X), \quad (7)$$

$$Pl(A) = \sum_{\substack{X \in 2^\Theta \\ X \cap A \neq \emptyset}} m(X) = 1 - Bel(\bar{A}), \quad (8)$$

where  $\bar{A} \triangleq \Theta - \{A\} = \{X | X \in \Theta \text{ and } X \notin A\}$ , is the complement of  $A$  in  $\Theta$  and the minus symbol denotes the set difference operator. The width  $U(A^*) = Pl(A) - Bel(A) = \sum_{X \in \mathcal{F}_{A^*}(m)} m(X)$  of the belief interval  $[Bel(A), Pl(A)]$  is called the *uncertainty on A* committed by the SoE.  $\mathcal{F}_{A^*}(m)$  is the set of focal elements of  $m(\cdot)$  not included in  $A$  and not included in  $\bar{A}$ , that is  $\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_A(m) - \mathcal{F}_{\bar{A}}(m)$  and  $U(A^*)$  represents the imprecision on the (subjective) probability of  $A$  granted by the SoE which provides the BBA  $m(\cdot)$ . When all elements of  $\mathcal{F}_\Theta(m)$  are only singletons,  $m(\cdot)$  is called a *Bayesian BBA* [2] and its corresponding  $Bel(\cdot)$  and  $Pl(\cdot)$  functions are homogeneous to a same (subjective) probability measure  $P(\cdot)$ , and in this case  $\mathcal{F}_{A^*}(m) = \emptyset$ . According to Shafer's Theorem 2.9, see [2] page 39 with its proof on page 51, the belief functions can be characterized without referencing to a BBA. The quantities  $m(\cdot)$  and  $Bel(\cdot)$  are one-to-one, and for any  $A \subseteq \Theta$  the BBA  $m(\cdot)$  is obtained from  $Bel(\cdot)$  by Möbius inverse formula (see [2], p.39)

$$m(A) = \sum_{B \subseteq A \subseteq \Theta} (-1)^{|A-B|} Bel(B). \quad (9)$$

Because for any partition  $\{A_1, \dots, A_k\}$  of the FoD  $\Theta$ , the equality  $\mathcal{F}_\Theta(m) = \mathcal{F}_{A_1}(m) \cup \dots \cup \mathcal{F}_{A_k}(m) \cup \mathcal{F}_{A^*}(m)$  with  $\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m)$  is valid, the following Total Belief Theorem (TBT) holds – see proof in the companion paper [7].

**Total Belief Theorem (TBT):** Let's consider a FoD  $\Theta$  with  $|\Theta| \geq 2$  elements and a BBA  $m(\cdot)$  defined on  $2^\Theta$  with the set of focal elements  $\mathcal{F}_\Theta(m)$ . For any chosen partition  $\{A_1, \dots, A_k\}$  of  $\Theta$  and for any  $B \subseteq \Theta$ , one has

$$Bel(B) = \sum_{i=1, \dots, k} Bel(A_i \cap B) + U(A^* \cap B), \quad (10)$$

where  $U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) \in [0, 1]$ .

By expressing  $Bel(\bar{B})$  using TBT and noting that  $Pl(B) = 1 - Bel(\bar{B})$ , one get the Total Plausibility Theorem (TPIT) [7], which states that for any partition  $\{A_1, \dots, A_k\}$  of  $\Theta$  and any  $B \subseteq \Theta$ , one has

$$Pl(B) = \sum_{i=1, \dots, k} Pl(\bar{A}_i \cup B) + 1 - k - U(A^* \cap \bar{B}), \quad (11)$$

<sup>3</sup>More generally, the set of all focal elements of  $m(\cdot)$  included in a subset  $A \subseteq \Theta$  is denoted  $\mathcal{F}_A(m)$ .

<sup>4</sup>By convention, a *sum of non existing terms* (if it occurs in formulas depending on the given BBA) is always set to zero.

<sup>2</sup>The notation  $\triangleq$  means *equal by definition*

where  $U(A^* \cap \bar{B}) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_{\bar{B}}(m)} m(X) \in [0, 1]$ .

In DST framework, Shafer [2] did propose to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD with Dempster's rule of combination (i.e. the normalized conjunctive rule). The justification and behavior of Dempster's rule have however been strongly disputed from both theoretical and practical standpoints as reported in [12]–[15]. Furthermore, Shafer did use also Dempster's rule to establish formulas for conditional belief and plausibility functions [2]. Unfortunately, Shafer's conditioning formulas are inconsistent with lower and upper bounds of imprecise conditional probability values as discussed in [6], [16], [18] – see also Ellsberg's urn example in [7]. That is why we do not recommend Shafer's conditioning and Dempster' rule in applications involving belief functions. This standpoint has been already shared by several authors before us, see by example [6], [8], [16], [19]–[21].

Recently in [7], we have proved that Fagin-Halpern conditional belief and plausibility formulas [6], [16], [17] can be directly obtained from TBT to define the conditional belief as the lower envelope (i.e. the infimum) of a family of conditional probability functions to make belief conditioning consistent with imprecise conditional probability calculus. In this paper we do not enter in details on the justification of Fagin-Halpern conditioning formulas but we just need to recall their expressions because they will be used in the next section to define the notion of credibilistic independence (or C-independence for short). Assuming  $Bel(B) > 0$ , Fagin and Halpern proposed the following conditional formulas (FH formulas for short)

$$Bel(A|B) = Bel(A \cap B) / (Bel(A \cap B) + Pl(\bar{A} \cap B)), \quad (12)$$

$$Pl(A|B) = Pl(A \cap B) / (Pl(A \cap B) + Bel(\bar{A} \cap B)). \quad (13)$$

Fagin and Halpern proved in [6] that  $Bel(A|B)$  given by (12) is a true belief function<sup>5</sup>. Later, Sundberg and Wagner in [20] (p. 268) did give a clearer proof also (not very easy to follow though). By switching notations and assuming  $Bel(A) > 0$ , the previous FH formulas yield

$$Bel(B|A) = Bel(A \cap B) / (Bel(A \cap B) + Pl(\bar{B} \cap A)), \quad (14)$$

$$Pl(B|A) = Pl(A \cap B) / (Pl(A \cap B) + Bel(\bar{B} \cap A)). \quad (15)$$

In [7], we did also generalize Bayes' Theorem for working in the framework of belief functions as follows.

**Generalized Bayes' Theorem (GBT):** For any partition  $\{A_1, \dots, A_k\}$  of a FoD  $\Theta$ , any belief function  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$ , and any subset  $B$  of  $\Theta$  with  $Bel(B) > 0$ , one has for  $i \in \{1, \dots, k\}$

$$Bel(A_i|B) = \frac{Bel(B|A_i)q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i)q(A_i, B) + U((\bar{A}_i \cap B)^*)}, \quad (16)$$

<sup>5</sup>satisfying the three conditions of Shafer's Theorem 2.9, see [2] page 39.

where

$$\begin{aligned} q(A_i, B) &\triangleq Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i), \\ U((\bar{B} \cap A_i)^*) &\triangleq Pl(\bar{B} \cap A_i) - Bel(\bar{B} \cap A_i), \\ U((\bar{A}_i \cap B)^*) &\triangleq Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B), \\ U(B^* \cap A_i) &\triangleq \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{A_i}(m)} m(X). \end{aligned}$$

Note that FH formulas are consistent with Bayes formula (i.e. conditional probability formula) when the underlying BBA  $m(\cdot)$  is Bayesian. Indeed if  $m(\cdot)$  is Bayesian, then  $Pl(A \cap B) = Bel(A \cap B) = P(A \cap B)$ ,  $Pl(\bar{A} \cap B) = Bel(\bar{A} \cap B) = P(\bar{A} \cap B)$  and  $Pl(\bar{B} \cap A) = Bel(\bar{B} \cap A) = P(\bar{B} \cap A)$  and FH formulas become

$$Bel(A|B) = Pl(A|B) = P(A \cap B) / P(B) = P(A|B), \quad (17)$$

$$Bel(B|A) = Pl(B|A) = P(A \cap B) / P(A) = P(B|A). \quad (18)$$

The advantage of FH formulas is their complete compatibility with the bounds of conditional probability calculus [20] and their theoretical constructive justification drawn from TBT.

#### IV. NOTION OF CREDIBILISTIC INDEPENDENCE

In this section we generalize in the belief functions framework the notion of probabilistic independence of two events  $A$  and  $B$  expressed by the condition  $P(A \cap B) = P(A)P(B)$ .

##### A. Definition of credibilistic independence

To define the credibilistic independence of two propositions  $A$  and  $B$ , we start from the FH belief conditioning formulas (12)–(15) and we impose the Credibilistic Independence Constraints (CIC) by analogy of what has been done in the framework of probabilistic framework. So, we require the conditions

$$Bel(A|B) = Bel(A), \quad (19)$$

$$Bel(B|A) = Bel(B), \quad (20)$$

$$Pl(A|B) = Pl(A), \quad (21)$$

$$Pl(B|A) = Pl(B), \quad (22)$$

which reflect the notion of independence of propositions  $A$  and  $B$ .

Working with conditional belief expressions, the formula (12) and the condition (19) yield

$$Bel(A)[Bel(A \cap B) + Pl(\bar{A} \cap B)] = Bel(A \cap B)$$

or equivalently

$$Bel(A)Pl(\bar{A} \cap B) = Bel(A \cap B)[1 - Bel(A)]$$

By noting that  $1 - Bel(A) = Pl(\bar{A})$  and dividing both sides of the previous equality by  $Pl(\bar{A})$  (assumed strictly positive), we get

$$Bel(A \cap B) = \frac{Bel(A)}{Pl(\bar{A})} Pl(\bar{A} \cap B) \quad (23)$$

Similarly, the formula (14) and the condition (20) yield

$$Bel(B)[Bel(A \cap B) + Pl(A \cap \bar{B})] = Bel(A \cap B),$$

or equivalently

$$Bel(B)Pl(A \cap \bar{B}) = Bel(A \cap B)[1 - Bel(B)].$$

By noting that  $1 - Bel(B) = Pl(\bar{B})$  and dividing both sides of the previous equality by  $Pl(\bar{B})$  (assumed strictly positive), we get

$$Bel(A \cap B) = \frac{Bel(B)}{Pl(\bar{B})}Pl(A \cap \bar{B}). \quad (24)$$

If CIC (19) and (20) are satisfied, then because of (23) and (24), one must have also the following equality satisfied

$$Bel(A \cap B) = \frac{Bel(A)}{Pl(\bar{A})}Pl(\bar{A} \cap B) = \frac{Bel(B)}{Pl(\bar{B})}Pl(A \cap \bar{B}).$$

This equality imposes the following condition to be satisfied

$$Bel(A)Pl(\bar{B})Pl(\bar{A} \cap B) = Pl(\bar{A})Bel(B)Pl(A \cap \bar{B}). \quad (25)$$

One sees that this equality is always satisfied if one has

$$Pl(A \cap \bar{B}) = Bel(A)Pl(\bar{B}), \quad (26)$$

$$Pl(\bar{A} \cap B) = Pl(\bar{A})Bel(B). \quad (27)$$

Working with conditional plausibility expressions, the formula (13) and the condition (21) yield

$$Pl(A)[Pl(A \cap B) + Bel(\bar{A} \cap B)] = Pl(A \cap B),$$

or equivalently

$$Pl(A \cap B) = \frac{Pl(A)}{Bel(\bar{A})}Bel(\bar{A} \cap B). \quad (28)$$

The formula (15) and the condition (22) yield

$$Pl(B)[Pl(A \cap B) + Bel(A \cap \bar{B})] = Pl(A \cap B),$$

or equivalently

$$Pl(A \cap B) = \frac{Pl(B)}{Bel(\bar{B})}Bel(A \cap \bar{B}). \quad (29)$$

If CIC (21) and (22) are satisfied, then because of (28) and (29), one must have also the following equality satisfied

$$Pl(A \cap B) = \frac{Pl(A)}{Bel(\bar{A})}Bel(\bar{A} \cap B) = \frac{Pl(B)}{Bel(\bar{B})}Bel(A \cap \bar{B}).$$

This equality imposes the following condition to be satisfied

$$Pl(A)Bel(\bar{B})Bel(\bar{A} \cap B) = Bel(\bar{A})Pl(B)Bel(A \cap \bar{B}). \quad (30)$$

One sees that this equality is always satisfied if one has

$$Bel(A \cap \bar{B}) = Pl(A)Bel(\bar{B}), \quad (31)$$

$$Bel(\bar{A} \cap B) = Bel(\bar{A})Pl(B). \quad (32)$$

In summary, the four CIC are satisfied whenever the two following conditions are satisfied for the two belief intervals  $[Bel(A \cap \bar{B}), Pl(A \cap \bar{B})]$  and  $[Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)]$ .

• **Condition  $C_1$ :**

$$[Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [Pl(A)Bel(\bar{B}), Bel(A)Pl(\bar{B})]. \quad (33)$$

• **Condition  $C_2$ :**

$$[Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [Bel(\bar{A})Pl(B), Pl(\bar{A})Bel(B)]. \quad (34)$$

The conditions  $C_1$  and  $C_2$  are in fact just necessary conditions but not sufficient conditions because one needs also to impose the coherence conditions  $C_3$  and  $C_4$  stating that right bound of any belief interval must always be greater (or equal) than its left bound. Hence the following inequalities (35) and (37) must also be satisfied.

• **Condition  $C_3$ :** The constraint  $Bel(A \cap \bar{B}) \leq Pl(A \cap \bar{B})$  and (33) impose to have

$$Pl(A)Bel(\bar{B}) \leq Bel(A)Pl(\bar{B}), \quad (35)$$

which is equivalent to the condition<sup>6</sup>

$$Pl(A) - Bel(A) \leq Pl(A)Pl(B) - Bel(A)Bel(B). \quad (36)$$

• **Condition  $C_4$ :** The constraint  $Bel(\bar{A} \cap B) \leq Pl(\bar{A} \cap B)$  and (34) impose to have

$$Bel(\bar{A})Pl(B) \leq Pl(\bar{A})Bel(B), \quad (37)$$

which is equivalent to the condition<sup>7</sup>

$$Pl(B) - Bel(B) \leq Pl(A)Pl(B) - Bel(A)Bel(B). \quad (38)$$

Thus, the conditions  $C_1$ ,  $C_2$ ,  $C_3$  and  $C_4$  characterize mathematically the notion of credibilistic independence (C-Indep) between two propositions  $A$  and  $B$  according to a given BBA. This allows us to establish the following theorem.

**C-Indep Theorem:** Consider a FoD  $\Theta$  and a BBA  $m(\cdot) : 2^\Theta \mapsto [0, 1]$  and  $A$  and  $B$  two subsets of  $\Theta$ . The two propositions  $A$  and  $B$  are said credibilistically independent if and only if

$$[Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [Pl(A)Bel(\bar{B}), Bel(A)Pl(\bar{B})],$$

$$[Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [Bel(\bar{A})Pl(B), Pl(\bar{A})Bel(B)],$$

and

$$Pl(A) - Bel(A) \leq Pl(A)Pl(B) - Bel(A)Bel(B),$$

$$Pl(B) - Bel(B) \leq Pl(A)Pl(B) - Bel(A)Bel(B),$$

where  $Bel(\cdot)$  and  $Pl(\cdot)$  are respectively the belief and plausibility functions related to the BBA  $m(\cdot)$ .

**Remark:** Fagin-Halpern formulas (12)–(15) are defined only if  $Bel(B) > 0$  and if  $Bel(A) > 0$ . This means that  $[Bel(A), Pl(A)] = ]a_1, a_2]$  is a left open interval (excluding  $a_1 = 0$  and with  $a_1 \leq a_2 \leq 1$ ) and  $[Bel(B), Pl(B)] = ]b_1, b_2]$  is also a left open interval (excluding  $b_1 = 0$  and with  $b_1 \leq b_2 \leq 1$ ). The credibilistic independence conditions of C-Indep Theorem can however be satisfied even<sup>8</sup> if  $Bel(A) = 0$ , or  $Bel(B) = 0$ , but in this case the Fagin-Halpern formulas yield 0/0 indeterminate form, which is perfectly normal.

<sup>6</sup>Substitute  $Bel(\bar{B})$  by  $1 - Pl(B)$ ,  $Pl(\bar{B})$  by  $1 - Bel(B)$  and rearrange terms.

<sup>7</sup>Substitute  $Bel(\bar{A})$  by  $1 - Pl(A)$ ,  $Pl(\bar{A})$  by  $1 - Bel(A)$  and rearrange terms.

<sup>8</sup>This is similar to probabilistic independence condition, where the condition  $P(A \cap B) = P(A)P(B)$  is valid even if  $P(A) = 0$ , or  $P(B) = 0$ , or if both equalities hold.

### B. Discussion

The propositions  $A$  and  $B$  can be credibilistically independent even if some of their lower or upper bounds equal respectively to zero or one as it will be shown in the next section. In this case, one can make a preliminary simple (pre-filtering) test to check the necessary condition that the left (lower) bound of belief interval must always be less (or equal) to right bound. For establishing such a test, it is worth noting that the following implications are true.

$$A \subseteq B \Rightarrow Bel(A) \leq Bel(B), \quad (39)$$

$$A \subseteq B \Rightarrow Pl(A) \leq Pl(B). \quad (40)$$

**Proof:** Indeed, if  $A \subseteq B$ , then  $B - A$  (the complement of  $A$  in  $B$ ) is also a subset of  $B$ . Since we have  $B = A \cup (B - A)$  and  $A \cap (B - A) = \emptyset$ , from the definition of  $Bel(\cdot)$  function, one can write

$$\begin{aligned} Bel(B) &= \sum_{X \subseteq B} m(X) \\ &= \sum_{X \subseteq A \cup (B-A)} m(X) \\ &= \sum_{X \subseteq A} m(X) + \sum_{X \subseteq B-A} m(X), \end{aligned}$$

which is obviously greater (or equal) to  $Bel(A) = \sum_{X \subseteq A} m(X)$ . Therefore (39) is true.

Because<sup>9</sup>  $A \subseteq B \Rightarrow \bar{B} \subseteq \bar{A}$  where  $\bar{A} \triangleq \Theta - A$  and  $\bar{B} \triangleq \Theta - B$  (the complements of  $A$  and of  $B$  in the FoD  $\Theta$ ), one always has  $Bel(\bar{B}) \leq Bel(\bar{A})$ . Hence,  $-Bel(\bar{A}) \leq -Bel(\bar{B})$ , and thus  $[Pl(A) = 1 - Bel(\bar{A})] \leq [Pl(B) = 1 - Bel(\bar{B})]$ . Therefore (40) is also true.

Because  $A \cap B$  is always included in  $A$  and in  $B$ , one always has  $Bel(A \cap B) \leq Bel(A)$  and  $Bel(A \cap B) \leq Bel(B)$ . For the same reason,  $Pl(A \cap B) \leq Pl(A)$  and  $Pl(A \cap B) \leq Pl(B)$ . Therefore the following inequalities always hold

$$Bel(A \cap B) \leq \min\{Bel(A), Bel(B)\}, \quad (41)$$

$$Pl(A \cap B) \leq \min\{Pl(A), Pl(B)\}. \quad (42)$$

Let's examine the bounds of the belief interval for the condition  $C_1$  given in (33), which is

$$[Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [Pl(A)Bel(\bar{B}), Bel(A)Pl(\bar{B})].$$

- Lower bound of belief interval: Because  $Bel(A \cap \bar{B}) \leq \min\{Bel(A), Bel(\bar{B})\}$  and  $Bel(A \cap \bar{B}) = Pl(A)Bel(\bar{B})$ , the following condition

$$Pl(A)Bel(\bar{B}) \leq \min\{Bel(A), Bel(\bar{B})\},$$

must be satisfied. In fact, because  $Pl(A)Bel(\bar{B}) \leq Bel(\bar{B})$  is always true because  $Pl(A) \in [0, 1]$ , the following coherence condition must hold

$$Pl(A)Bel(\bar{B}) \leq Bel(A), \quad (43)$$

<sup>9</sup>Letting  $x \in \bar{B}$  says that  $x$  does not belong to  $B$ , but the hypothesis  $A \subseteq B$  tells us that  $A$  is included in  $B$  and hence  $x$  does not belong to  $A$  as well, or in other words  $x \in \bar{A}$ . Therefore we have proven  $\bar{B} \subseteq \bar{A}$ .

or equivalently (because  $Bel(\bar{B}) = 1 - Pl(B)$ )

$$Pl(A) - Bel(A) \leq Pl(A)Pl(B). \quad (44)$$

Note that the constraint (43) is a bit less restrictive than the inequality (35) of condition  $C_3$ . This coherence constraint says that the uncertainty on  $A$  must be less than the product of plausibilities of  $A$  and of  $B$  if one wants to have equality for the lower bound of belief interval  $Bel(A \cap \bar{B}) = Pl(A)Bel(\bar{B})$  possible.

- Upper bound of belief interval: Because  $Pl(A \cap \bar{B}) \leq \min\{Pl(A), Pl(\bar{B})\}$  and  $Pl(A \cap \bar{B}) = Bel(A)Pl(\bar{B})$ , the following condition

$$Bel(A)Pl(\bar{B}) \leq \min\{Pl(A), Pl(\bar{B})\},$$

must be satisfied. In fact, because  $Bel(A)Pl(\bar{B}) \leq Pl(\bar{B})$  is always true because  $Bel(A) \in [0, 1]$ , the following coherence condition must hold

$$Bel(A)Pl(\bar{B}) \leq Pl(A). \quad (45)$$

Using the fact that  $Pl(\bar{B}) = 1 - Bel(B)$  in (45), and rearranging terms we get

$$Bel(A)(1 - Bel(B)) \leq Pl(A), \quad (46)$$

$$Bel(A) - Bel(A)Bel(B) \leq Pl(A), \quad (47)$$

$$-Bel(A)Bel(B) \leq Pl(A) - Bel(A). \quad (48)$$

As we see, the inequality (48) is always satisfied because  $Bel(A)$ ,  $Bel(B)$  and  $Pl(A)$  belong to  $[0, 1]$  and because  $Pl(A) \geq Bel(A)$ , so that  $-Bel(A)Bel(B) \leq 0$  whereas  $Pl(A) - Bel(A) \geq 0$ .

Thus, there is in fact no need for a coherence constraint for the upper bound of belief interval to allow the equality  $Pl(A \cap \bar{B}) = Bel(A)Pl(\bar{B})$  possible.

Let's examine the bounds of the belief interval for the condition  $C_2$  given in (34), which is

$$[Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [Bel(\bar{A})Pl(B), Pl(\bar{A})Bel(B)].$$

- Lower bound of belief interval: Because  $Bel(\bar{A} \cap B) \leq \min\{Bel(\bar{A}), Bel(B)\}$  and  $Bel(\bar{A} \cap B) = Bel(\bar{A})Pl(B)$ , the following condition

$$Bel(\bar{A})Pl(B) \leq \min\{Bel(\bar{A}), Bel(B)\},$$

must be satisfied. In fact, because  $Bel(\bar{A})Pl(B) \leq Bel(\bar{A})$  is always true because  $Pl(B) \in [0, 1]$ , the following coherence condition must hold

$$Bel(\bar{A})Pl(B) \leq Bel(B), \quad (49)$$

or equivalently (because  $Bel(\bar{A}) = 1 - Pl(A)$ )

$$Pl(B) - Bel(B) \leq Pl(A)Pl(B). \quad (50)$$

Note that the constraint (49) is a bit less restrictive than the inequality (37) of condition  $C_4$ . This coherence constraint says that the uncertainty on  $B$  must be less than

the product of plausibilities of  $A$  and of  $B$  if one wants to have equality for the lower bound of belief interval  $Bel(\bar{A} \cap B) = Bel(\bar{A})Pl(B)$  possible.

- Upper bound of belief interval: Because  $Pl(\bar{A} \cap B) \leq \min\{Pl(\bar{A}), Pl(B)\}$  and  $Pl(\bar{A} \cap B) = Pl(\bar{A})Bel(B)$ , the following condition

$$Pl(\bar{A})Bel(B) \leq \min\{Pl(\bar{A}), Pl(B)\},$$

must be satisfied. In fact, because  $Pl(\bar{A})Bel(B) \leq Pl(\bar{A})$  is always true because  $Bel(B) \in [0, 1]$ , the following coherence condition must hold

$$Pl(\bar{A})Bel(B) \leq Pl(B). \quad (51)$$

Using the fact that  $Pl(\bar{A}) = 1 - Bel(A)$  in (51), and rearranging terms we get

$$(1 - Bel(A))Bel(B) \leq Pl(B), \quad (52)$$

$$Bel(B) - Bel(A)Bel(B) \leq Pl(B), \quad (53)$$

$$-Bel(A)Bel(B) \leq Pl(B) - Bel(B). \quad (54)$$

As we see, the inequality (54) is always satisfied because  $Bel(A)$ ,  $Bel(B)$  and  $Pl(B)$  belong to  $[0, 1]$  and because  $Pl(B) \geq Bel(B)$ , so that  $-Bel(A)Bel(B) \leq 0$  whereas  $Pl(B) - Bel(B) \geq 0$ .

Thus, there is in fact no need for a coherence constraint for the upper bound of belief interval to allow the equality  $Pl(\bar{A} \cap B) = Pl(\bar{A})Bel(B)$  possible.

In summary, the conditions

$$Pl(A) - Bel(A) \leq Pl(A)Pl(B), \quad (55)$$

$$Pl(B) - Bel(B) \leq Pl(A)Pl(B), \quad (56)$$

are necessary for the coherence of belief interval bounds defined in the conditions  $C_1$  and  $C_2$ . They express the fact that the width of belief interval (i.e. the uncertainty) of the proposition  $A$  and  $B$  must be less than the product of their plausibilities. The conditions (55)–(56) are very convenient to test quickly the non credibilistic independence of  $A$  and  $B$ , because if at least one condition (55), or (56) (or both) is not satisfied, then we are sure that  $A$  and  $B$  cannot be credibilistically independent. If the inequalities (55)–(56) are satisfied, we need to check if the conditions  $C_1$ ,  $C_2$ ,  $C_3$  and  $C_4$  are also satisfied to declare the credibilistic independence of  $A$  and  $B$ .

### C. Special case: Bayesian belief functions

The notion of credibilistic independence defined in the previous section is a generalization of the notion of probabilistic independence. This can be justified (and verified) by examining what provides the conditions  $C_1$ ,  $C_2$ ,  $C_3$  and  $C_4$  in the limit case when the BBA  $m(\cdot)$  is Bayesian. In this case, belief function  $Bel(\cdot)$  and plausibility function  $Pl(\cdot)$  coincide with a probability measure  $P(\cdot)$ , which means that the conditions  $C_3$  and  $C_4$  characterized by formulas (36) and (38) are always satisfied because  $Pl(A) = Bel(A)$ ,

and  $Pl(B) = Bel(B)$ . Moreover, the conditions  $C_1$  and  $C_2$  become equalities between the following degenerate intervals

$$\begin{aligned} [P(A \cap \bar{B}), P(A \cap \bar{B})] &= [P(A)P(\bar{B}), P(A)P(\bar{B})], \\ [P(\bar{A} \cap B), P(\bar{A} \cap B)] &= [P(\bar{A})P(B), P(\bar{A})P(B)], \end{aligned}$$

or equivalently

$$\begin{aligned} P(A \cap \bar{B}) &= P(A)P(\bar{B}), \\ P(\bar{A} \cap B) &= P(\bar{A})P(B). \end{aligned}$$

These conditions are in fact equivalent to the probabilistic independence condition  $P(A \cap B) = P(A)P(B)$ . This can be shown from the TPT formulas  $P(A \cap \bar{B}) + P(A \cap B) = P(A)$  and  $P(A \cap B) + P(\bar{A} \cap B) = P(B)$  as follows.

- If  $P(A \cap \bar{B}) = P(A)P(\bar{B})$ , then  $P(A \cap \bar{B}) + P(A \cap B) = P(A)P(\bar{B}) + P(A \cap B) = P(A)$ , and thus  $P(A \cap B) = P(A)(1 - P(\bar{B})) = P(A)P(B)$ .
- If  $P(\bar{A} \cap B) = P(\bar{A})P(B)$ , then  $P(A \cap B) + P(\bar{A} \cap B) = P(A \cap B) + P(\bar{A})P(B) = P(B)$ , and thus  $P(A \cap B) = (1 - P(\bar{A}))P(B) = P(A)P(B)$ .

Therefore, one has proved that our notion of credibilistic independence derived from FH conditioning coincides with the notion of probabilistic independence as soon as the belief function under consideration is Bayesian.

## V. ILLUSTRATIVE EXAMPLES

For convenience (and not for significance), we give some simple examples illustrating the credibilistic independence between two propositions  $A$  and  $B$  with respect to some given basic belief assignments, so that the reader will be able to check by himself how to perform the derivations.

### A. Example 1 (Bayesian case)

Let consider the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$  and the (uniform) Bayesian BBA defined by  $m(\theta_i) = 1/6$  for  $i = 1, 2, \dots, 6$ . Consider the two propositions (subsets)  $A$  and  $B$  of  $\Theta$  defined as  $A \triangleq \theta_1 \cup \theta_2$  and  $B \triangleq \theta_2 \cup \theta_4 \cup \theta_6$ . In this case,  $\bar{A} = \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_6$  and  $\bar{B} = \theta_1 \cup \theta_3 \cup \theta_5$ . We have also  $A \cap \bar{B} = \theta_1$ , and  $\bar{A} \cap B = \theta_4 \cup \theta_6$ . Because  $m(\cdot)$  is a Bayesian BBA,  $Bel(X) = Pl(X) = P(X)$  for  $X \in 2^\Theta$ . Here one has

$$\begin{aligned} Bel(A) &= Bel(\theta_1 \cup \theta_2) = m(\theta_1) + m(\theta_2) = 1/3, \\ Pl(A) &= Pl(\theta_1 \cup \theta_2) = m(\theta_1) + m(\theta_2) = 1/3, \\ Bel(\bar{A}) &= Bel(\theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_6) = 1 - Pl(A) = 2/3, \\ Pl(\bar{A}) &= Pl(\theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_6) = 1 - Bel(A) = 2/3, \\ Bel(B) &= m(\theta_2) + m(\theta_4) + m(\theta_6) = 1/2, \\ Pl(B) &= m(\theta_2) + m(\theta_4) + m(\theta_6) = 1/2, \\ Bel(\bar{B}) &= Bel(\theta_1 \cup \theta_3 \cup \theta_5) = 1 - Pl(B) = 1/2, \\ Pl(\bar{B}) &= Pl(\theta_1 \cup \theta_3 \cup \theta_5) = 1 - Bel(B) = 1/2, \\ Bel(A \cap \bar{B}) &= m(\theta_1) = 1/6, \\ Pl(A \cap \bar{B}) &= m(\theta_1) = 1/6, \\ Bel(\bar{A} \cap B) &= Bel(\theta_4 \cup \theta_6) = m(\theta_4) + m(\theta_6) = 1/3, \\ Pl(\bar{A} \cap B) &= Pl(\theta_4 \cup \theta_6) = m(\theta_4) + m(\theta_6) = 1/3. \end{aligned}$$

Conditions  $C_1$  and  $C_2$  are satisfied because

$$C_1: \begin{cases} [Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [\frac{1}{6}, \frac{1}{6}], \\ [Pl(A)Bel(\bar{B}), Bel(A)Pl(\bar{B})] = [\frac{1}{3} \cdot \frac{1}{2}, \frac{1}{3} \cdot \frac{1}{2}] = [\frac{1}{6}, \frac{1}{6}], \end{cases}$$

$$C_2: \begin{cases} [Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [\frac{1}{3}, \frac{1}{3}], \\ [Bel(\bar{A})Pl(B), Pl(\bar{A})Bel(B)] = [\frac{2}{3} \cdot \frac{1}{2}, \frac{2}{3} \cdot \frac{1}{2}] = [\frac{1}{3}, \frac{1}{3}]. \end{cases}$$

The condition  $C_3 : Pl(A)Bel(\bar{B}) \leq Bel(A)Pl(\bar{B})$  is satisfied because  $Pl(A)Bel(\bar{B}) = \frac{1}{3} \cdot \frac{1}{2} = Bel(A)Pl(\bar{B})$ . The condition  $C_4 : Bel(\bar{A})Pl(B) \leq Pl(\bar{A})Bel(B)$  is also satisfied because  $Bel(\bar{A})Pl(B) = \frac{2}{3} \cdot \frac{1}{2} = Pl(\bar{A})Bel(B)$ .

Because the conditions  $C_1, C_2, C_3$  and  $C_4$  are satisfied, the propositions  $A$  and  $B$  are credibilistically independent. In fact, in this Bayesian case,  $A$  and  $B$  are also probabilistically independent because  $P(A \cap B) = P(\theta_2) = P(A)P(B)$ . Note that the coherence conditions (55) and (56) are of course satisfied because

$$[Pl(A) - Bel(A) = 0] \leq [Pl(A)Pl(B) = (1/3) \cdot (1/2)],$$

$$[Pl(B) - Bel(B) = 0] \leq [Pl(A)Pl(B) = (1/3) \cdot (1/2)].$$

### B. Example 2 (Non Bayesian case)

Let consider the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$  and the two propositions (subsets)  $A$  and  $B$  of  $\Theta$  defined as  $A \triangleq \theta_1 \cup \theta_2 \cup \theta_3$  and  $B \triangleq \theta_3 \cup \theta_4$ . In this case,  $\bar{A} = \theta_4 \cup \theta_5$  and  $\bar{B} = \theta_1 \cup \theta_2 \cup \theta_5$ . We have also  $A \cap \bar{B} = (\theta_1 \cup \theta_2 \cup \theta_3) \cap (\theta_1 \cup \theta_2 \cup \theta_5) = \theta_1 \cup \theta_2$  and  $\bar{A} \cap B = (\theta_4 \cup \theta_5) \cap (\theta_3 \cup \theta_4) = \theta_4$ . Suppose that the BBA  $m(\cdot)$  is simply defined as<sup>10</sup>

$$m(\theta_1) = 0.5, \quad m(\theta_3) = 0.1, \quad m(\theta_1 \cup \theta_3) = 0.4.$$

Based on the BBA  $m(\cdot)$ , the belief and plausibilities of propositions involved in the derivations are

$$[Bel(A), Pl(A)] = [1, 1], \quad [Bel(\bar{A}), Pl(\bar{A})] = [0, 0],$$

$$[Bel(B), Pl(B)] = [0.1, 0.5], \quad [Bel(\bar{B}), Pl(\bar{B})] = [0.5, 0.9],$$

$$[Bel(A \cap B), Pl(A \cap B)] = [0.1, 0.5],$$

$$[Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [0.5, 0.9],$$

$$[Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [0, 0].$$

The condition  $C_1$  is satisfied because

$$C_1: \begin{cases} [Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [0.5, 0.9], \\ [Pl(A)Bel(\bar{B}), Bel(A)Pl(\bar{B})] = [1 \cdot 0.5, 1 \cdot 0.9]. \end{cases}$$

The condition  $C_2$  is also satisfied because

$$C_2: \begin{cases} [Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [0, 0], \\ [Bel(\bar{A})Pl(B), Pl(\bar{A})Bel(B)] = [0 \cdot 0.5, 0 \cdot 0.1]. \end{cases}$$

The condition  $C_3$  given by  $Pl(A)Bel(\bar{B}) \leq Bel(A)Pl(\bar{B})$  is satisfied because  $Pl(A)Bel(\bar{B}) = 1 \cdot 0.5 = 0.5$  and  $Bel(A)Pl(\bar{B}) = 1 \cdot 0.9 = 0.9$ .

<sup>10</sup>All other elements of  $2^\Theta$  which are not focal elements of the BBA  $m(\cdot)$  receive a zero value.

The condition  $C_4$  given by  $Bel(\bar{A})Pl(B) \leq Pl(\bar{A})Bel(B)$  is satisfied because  $Bel(\bar{A})Pl(B) = 0 \cdot 0.5 = 0$  and  $Pl(\bar{A})Bel(B) = 0 \cdot 0.1 = 0$ .

Therefore the propositions  $A$  and  $B$  are credibilistically independent. One can easily verify using Fagin-Halpern formulas that  $[Bel(A|B), Pl(A|B)] = [Bel(A), Pl(A)]$  and  $[Bel(B|A), Pl(B|A)] = [Bel(B), Pl(B)]$ . Indeed, in applying (12) and (13) one gets

$$Bel(A|B) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\bar{A} \cap B)} = \frac{0.1}{0.1 + 0} = 1 = Bel(A),$$

$$Pl(A|B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\bar{A} \cap B)} = \frac{0.5}{0.5 + 0} = 1 = Pl(A),$$

and in applying (14) and (15), one gets

$$Bel(B|A) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(A \cap \bar{B})} = \frac{0.1}{0.1 + 0.9} = 0.1 = Bel(B),$$

$$Pl(B|A) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(A \cap \bar{B})} = \frac{0.5}{0.5 + 0.5} = 0.5 = Pl(B).$$

Note that the coherence conditions (55) and (56) are of course satisfied because

$$[Pl(A) - Bel(A) = 0] \leq [Pl(A)Pl(B) = 1 \cdot 0.5 = 0.5],$$

$$[Pl(B) - Bel(B) = 0.4] \leq [Pl(A)Pl(B) = 1 \cdot 0.5 = 0.5].$$

### C. Example 3 (Non Bayesian case)

Here we consider a more interesting example where the widths of belief intervals are not all restricted to zero. Consider the frame of discernment  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$  and the very simple BBA  $m(\cdot)$  defined by  $m(\theta_3) = 0.7$  and  $m(\theta_1 \cup \theta_3) = 0.3$ . We consider the propositions  $A \triangleq \theta_1 \cup \theta_3$  and  $B \triangleq \theta_2 \cup \theta_3$ . In this case, we have  $\bar{A} = \theta_2 \cup \theta_4$ ,  $\bar{B} = \theta_1 \cup \theta_4$ ,  $A \cap B = \theta_3$ ,  $A \cap \bar{B} = \theta_1$  and  $\bar{A} \cap B = \theta_2$ . Based on the BBA  $m(\cdot)$ , the belief and plausibilities of propositions involved in the derivations are

$$[Bel(A), Pl(A)] = [1, 1], \quad [Bel(\bar{A}), Pl(\bar{A})] = [0, 0],$$

$$[Bel(B), Pl(B)] = [0.7, 1], \quad [Bel(\bar{B}), Pl(\bar{B})] = [0, 0.3],$$

$$[Bel(A \cap B), Pl(A \cap B)] = [0.7, 1],$$

$$[Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [0, 0.3],$$

$$[Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [0, 0].$$

The condition  $C_1$  is satisfied because

$$C_1: \begin{cases} [Bel(A \cap \bar{B}), Pl(A \cap \bar{B})] = [0, 0.3], \\ [Pl(A)Bel(\bar{B}), Bel(A)Pl(\bar{B})] = [1 \cdot 0, 1 \cdot 0.3]. \end{cases}$$

The condition  $C_2$  is satisfied because

$$C_2: \begin{cases} [Bel(\bar{A} \cap B), Pl(\bar{A} \cap B)] = [0, 0], \\ [Bel(\bar{A})Pl(B), Pl(\bar{A})Bel(B)] = [0 \cdot 1, 0 \cdot 0.7]. \end{cases}$$

The condition  $C_3$  given by  $Pl(A)Bel(\bar{B}) \leq Bel(A)Pl(\bar{B})$  is satisfied because  $Pl(A)Bel(\bar{B}) = 1 \cdot 0 = 0$  and  $Bel(A)Pl(\bar{B}) = 1 \cdot 0.3 = 0.3$ .

The condition  $C_4$  given by  $Bel(\bar{A})Pl(B) \leq Pl(\bar{A})Bel(B)$  is satisfied because  $Bel(\bar{A})Pl(B) = 0 \cdot 1 = 0$  and  $Pl(\bar{A})Bel(B) = 0 \cdot 0.7 = 0$ .

Therefore the propositions  $A$  and  $B$  are credibilistically independent. One can easily verify using Fagin-Halpern formulas that  $[Bel(A|B), Pl(A|B)] = [Bel(A), Pl(A)]$  and  $[Bel(B|A), Pl(B|A)] = [Bel(B), Pl(B)]$ . Indeed, in applying (12) and (13) one gets

$$Bel(A|B) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\bar{A} \cap B)} = \frac{0.7}{0.7 + 0} = 1 = Bel(A),$$

$$Pl(A|B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\bar{A} \cap B)} = \frac{1}{1 + 0} = 1 = Pl(A),$$

and in applying (14) and (15), one gets

$$Bel(B|A) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(A \cap \bar{B})} = \frac{0.7}{0.7 + 0.3} = 0.7 = Bel(B),$$

$$Pl(B|A) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(A \cap \bar{B})} = \frac{1}{1 + 0} = 1 = Pl(B).$$

Note that the coherence conditions (55) and (56) are of course satisfied because

$$\begin{aligned} [Pl(A) - Bel(A) = 0] &\leq [Pl(A)Pl(B) = 1 \cdot 1 = 1], \\ [Pl(B) - Bel(B) = 0.3] &\leq [Pl(A)Pl(B) = 1 \cdot 1 = 1]. \end{aligned}$$

## VI. CONCLUSIONS

In this paper the notion of credibilistic independence of two propositions has been proposed in the framework of belief functions. It is a generalization of the notion of (probabilistic) independence of two events defined classically in the theory of probability. Our definition is totally consistent with the probabilistic independence when the basic belief assignment is Bayesian because it is based on Fagin-Halpern belief conditioning formulas (derived from Total Belief Theorem) which are consistent with imprecise conditional probability calculus. Simple examples of the notion of credibilistic independence have also been given to illustrate how to test easily the credibilistic independence of two propositions in practice from a given basic belief assignment.

## REFERENCES

- [1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw-Hill, pp. 37–38, 1984.
- [2] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Press, 1976.
- [3] I. Couso, S. Moral, *Examples of Independence for Imprecise Probabilities*, Proc. of 1st Int. Symp. on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June - 2 July 1999.
- [4] B. Ben Yaghlane, Ph. Smets, K. Mellouli, *Independence Concepts for Belief Functions*, in Proc. of IPMU, Vol. 1, pp 357–364, Madrid, 2000.

- [5] B. Ben Yaghlane, Ph. Smets, K. Mellouli, *Belief Function Independence: II. The Conditional Case*, IJAR Vol. 31, pp. 31–75, 2001.
- [6] R. Fagin, J.Y. Halpern, *A new approach to updating beliefs*, UAI Conf. Proc., 317-325, 1991.
- [7] J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Generalized Bayes Theorem*, Proc. of Fusion 2018, July 2018, Cambridge, UK.
- [8] J.N. Heendeni, K. Premaratne, M.N. Murthi, J. Uscinski, M. Scheutz, *A generalization of Bayesian Inference in the Dempster-Shafer Belief Theoretic Framework*, Proc. of Fusion 2016, Germany, July 2016.
- [9] R. Jirousek, J. Vejnárova, *Compositional models and conditional independence in evidence theory*, IJAR, Vol. 52, pp. 316–334, 2011.
- [10] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., (38):325–339, 1967.
- [11] A.P. Dempster, *A generalization of Bayesian inference*, J. of Royal Stat. Soc., (B30):205–247, 1968.
- [12] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Fusion 2012 Proc., Singapore, July 9–12, 2012.
- [13] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., (29):223-252, 2014 (with erratum in DSMT Book Vol. 5).
- [14] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's rule doesn't behave as Bayes rule with informative priors*, in Proc. of 2013 IEEE Int. Symposium on INnovations in Intelligent SysTems and Application (INISTA 2013), Albena, Bulgaria, June 19–21, 2013.
- [15] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's fusion rule is not a generalization of Bayes fusion rule*, in Proc. of Fusion 2013 Int. Conf. on Information Fusion, Istanbul, Turkey, July 9-12, 2013 (with erratum in DSMT Book Vol. 5).
- [16] J.Y. Halpern, R. Fagin, *Two views of belief: belief as generalized probability and belief as evidence*, Art. Intel., (54): 275–317, 1992.
- [17] J.Y. Halpern, *Reasoning about uncertainty*, MIT Press, 2003.
- [18] D. Dubois, T. Denœux, *Conditioning in Dempster-Shafer Theory: Prediction vs. Revision*, in Proc. of Belief 2012 Int. Conf., Compiègne, France, May 2012.
- [19] J.-Y. Jaffray, *Bayesian updating and belief functions*, IEEE Trans. on SMC, (22):1144–1152, 1992.
- [20] C. Sunberg, C. Wagner, *Generalized finite differences and Bayesian conditioning of Choquet capacities*, Adv. in Appl. Math., (13), 1992.
- [21] P.K. Black, *Is Shafer general Bayes?*, in proc. of 3rd UAI Workshop, USA, 1987.

# Correct Proof of Shafer's Belief Conditioning Formulas

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, A. Tchamova, D. Han, *Correct Proof of Shafer's Belief Conditioning Formulas*, in Proc. of Chinese Automation Congress (CAC 2018), Xi'an Jiaotong Univ., Xi'an, China, Nov. 30-December 2nd, 2018, and reprinted with permission.

**Abstract**—In his *Mathematical Theory of Evidence* published in 1976, Shafer did propose belief and plausibility conditioning formulas based on Dempster's rule of combination. It turns out that the proof given by Shafer for belief conditioning is incorrect and in this paper we present the correct proof of Shafer's belief conditioning formula.

**Keywords:** belief functions, Shafer's conditioning.

## I. INTRODUCTION

In his *Mathematical Theory of Evidence* published in 1976 [1], Glenn Shafer did propose belief and plausibility conditioning formulas based on Dempster's rule of combination. It turns out that the proof of Theorem 3.6 given by Shafer in [1] (p. 66) for belief conditioning is incorrect and we will explain why. In this paper we present the correct proof of Shafer's belief conditioning formulas. This paper must not be considered as a support for Shafer's belief conditioning approach because we recommend Fagin-Halpern conditioning approach [2] instead (see our paper [3] for justification). It is only a clarification of correct obtaining of Shafer's conditioning formulas, no less no more.

## II. BASICS OF BELIEF FUNCTIONS

Based on Dempster's works [4], [5], Shafer did introduce Belief Functions (BF) to model the epistemic uncertainty and to reason under uncertainty [1]. Shafer's theory of evidence is often called Dempster-Shafer Theory (DST) in the literature. We consider a finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \dots, \theta_n\}$ , with  $n > 1$ , and where all exhaustive and exclusive elements of  $\Theta$  represent the set of the potential solutions of the problem under concern. The set of all subsets of  $\Theta$  is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . A basic belief assignment (BBA) associated with a given source of evidence is defined as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying the conditions  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is the mass of belief of subset  $A$  committed by the source of evidence (SoE). A focal element  $X$  of a BBA  $m(\cdot)$  is an element of  $2^\Theta$  such that  $m(X) > 0$ . Note that the empty set  $\emptyset$  is not a focal element of a BBA because  $m(\emptyset) = 0$  (closed-world assumption of Shafer's model for the FoD). The set of all focal elements (i.e. the core) of  $m(\cdot)$  is denoted  $\mathcal{F}_\Theta(m) \triangleq \{X \subseteq \Theta | m(X) > 0\} = \{X \in 2^\Theta | m(X) > 0\}$ ,

and the set of focal elements of  $m(\cdot)$  included in  $A \subseteq \Theta$  is denoted  $\mathcal{F}_A(m) \triangleq \{X \in \mathcal{F}_\Theta(m) | X \cap A = X\}$ . Belief and plausibility functions are defined by<sup>1</sup>

$$\begin{aligned} Bel(A) &= \sum_{\substack{X \in 2^\Theta \\ X \subseteq A}} m(X) \\ &= \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \subseteq A}} m(X) = \sum_{X \in \mathcal{F}_A(m)} m(X), \end{aligned} \quad (1)$$

$$\begin{aligned} Pl(A) &= \sum_{\substack{X \in 2^\Theta \\ X \cap A \neq \emptyset}} m(X) \\ &= \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \cap A \neq \emptyset}} m(X) = 1 - Bel(\bar{A}). \end{aligned} \quad (2)$$

When all elements of  $\mathcal{F}_\Theta(m)$  are only singletons,  $m(\cdot)$  is called a *Bayesian BBA* [1] and its corresponding  $Bel(\cdot)$  and  $Pl(\cdot)$  functions are homogeneous to a same (subjective) probability measure  $P(\cdot)$ . The vacuous BBA representing a totally non informative source of evidence is characterized by the BBA  $m(\Theta) = 1$ . According to Shafer's Theorem 1 (see [1] page 39, with its proof on page 51), the belief functions can be characterized without referencing to a BBA. The quantities  $m(\cdot)$  and  $Bel(\cdot)$  are one-to-one, and the BBA  $m(\cdot)$  is obtained from  $Bel(\cdot)$  by Möbius inverse formula (see [1], p. 39).

In DST, Shafer [1] did propose to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD  $\Theta$  with Dempster's rule (i.e. the normalized conjunctive rule). Mathematically Dempster's rule of combination of  $s \geq 2$  BBAs is defined by  $m_{12\dots s}^{DS}(\emptyset) = 0$ , and for any  $X \neq \emptyset \in 2^\Theta$

$$\begin{aligned} m_{12\dots s}^{DS}(X) &= [m_1 \oplus \dots \oplus m_s](X) \\ &\triangleq m_{12\dots s}^{CR}(X) / (1 - m_{12\dots s}^{CR}(\emptyset)), \end{aligned} \quad (3)$$

where  $m_{12\dots s}^{CR}(X) \triangleq \sum_{\substack{X_1, \dots, X_s \in 2^\Theta \\ X_1 \cap X_2 \cap \dots \cap X_s = X}} \prod_{i=1}^s m_i(X_i)$  is the conjunctive rule (CR) of combination. The term  $m_{12\dots s}^{CR}(\emptyset)$

<sup>1</sup>By convention, a *sum of non existing terms* (if it occurs in formulas depending on the given BBA) is always set to zero.

reflects the amount of dissonance between the sources [6]. Dempster's rule is commutative and associative and preserves the neutrality of vacuous BBA in the fusion process. This rule has been disputed from both theoretical and practical standpoints, see [7]–[13] for discussions. In this paper we do not focus on Dempster's rule, but only on Shafer's belief conditioning formulas based on Dempster's rule.

### A. Shafer's conditioning formulas

In this section we present briefly Shafer's belief conditioning approach as proposed by Shafer in [1]. Suppose that the effect of a new evidence on the frame of discernment  $\Theta$  is to establish a particular subset  $B \subset \Theta$  with certainty. Then  $Bel_2$  defined by  $Bel_2(A) = 1$  if  $B \subset A$  and  $Bel_2(A) = 0$  if  $B \not\subset A$  will give a degree of belief one to the proposition corresponding to  $B$  and to every proposition implied by it [1], p.66. Shafer established the following important theorem<sup>2</sup> for conditional belief and plausibility.

**Theorem 3.6** [1], p. 67: Suppose  $Bel_2$  is defined by above two equations, and  $Bel_1$  is another belief function over  $\Theta$ . Then  $Bel_1$  and  $Bel_2$  are combinable if and only if  $Bel_1(\bar{B}) < 1$ . If  $Bel_1$  and  $Bel_2$  are combinable, let  $Bel_1(\cdot|B)$  denote  $Bel_1 \oplus Bel_2$ , and let  $Pl_1$  and  $Pl_1(\cdot|B)$  denote the upper probability functions for  $Bel_1$  and  $Bel_1 \oplus Bel_2$ , respectively. Then for all  $A \subset \Theta$ ,

$$Bel_1(A|B) = \frac{Bel_1(A \cup \bar{B}) - Bel_1(\bar{B})}{1 - Bel_1(\bar{B})}, \quad (4)$$

$$Pl_1(A|B) = \frac{Pl_1(A \cap B)}{Pl_1(B)}. \quad (5)$$

Shafer's proof of this theorem is in [1] (see pages 71–72), but we reproduce it here for convenience for a better identification of the mistake in this proof.

**Shafer's Proof of Theorem 3.6** (as given in [1]):  $Bel_1(\bar{B}) < 1$  if and only if  $B$  overlaps the core of  $Bel_1$ , and since  $B$  is the core of  $Bel_2$ , this is indeed equivalent to  $Bel_1$  being combinable with  $Bel_2$ . Denote the basic probability assignments of  $Bel_1$ ,  $Bel_2$  and  $Bel_1 \oplus Bel_2$  by  $m_1$ ,  $m_2$  and  $m$ . Since  $B$  is the only focal element of  $Bel_2$ , and  $m_2(B) = 1$ , Dempster's rule yields

$$m(A) = \frac{\sum_{A_i \cap B=A} m_1(A_i)}{1 - \sum_{A_i \cap B=\emptyset} m_1(A_i)} = \frac{\sum_{B \cap C=A} m_1(C)}{1 - Bel_1(\bar{B})}, \quad (6)$$

<sup>2</sup>In his theorem Shafer uses the notation  $P^*$  for upper probability instead of  $Pl$  used generally in the literature to denote the plausibility function.

and

$$Bel_1(A|B) = \sum_{D \subset A} m(D) = \frac{\sum_{\emptyset \neq D \subset A} \sum_{B \cap C=D} m_1(C)}{1 - Bel_1(\bar{B})} \quad (7)$$

$$= \frac{\sum_{\emptyset \neq B \cap C \subset A} m_1(C)}{1 - Bel_1(\bar{B})} \quad (8)$$

$$= \frac{\sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C)}{1 - Bel_1(\bar{B})} \quad (9)$$

$$= \frac{Bel_1(A \cup \bar{B}) - Bel_1(\bar{B})}{1 - Bel_1(\bar{B})}. \quad (10)$$

Hence

$$Pl_1(A|B) = 1 - Bel_1(\bar{A}|B) \quad (11)$$

$$= \frac{1 - Bel_1(\bar{B}) - Bel_1(\bar{A} \cup \bar{B}) + Bel_1(\bar{B})}{1 - Bel_1(\bar{B})} \quad (12)$$

$$= \frac{1 - Bel_1(\overline{A \cap B})}{1 - Bel_1(\bar{B})} = \frac{Pl_1(A \cap B)}{Pl_1(B)}. \quad \blacksquare \quad (13)$$

### III. WHY SHAFER'S PROOF IS INCORRECT

Although Shafer's formulas (4)–(5) are correct<sup>3</sup>, we show why Shafer's proof is incorrect. To obtain the final expression of  $Bel_1(A|B)$  given by (10), Shafer goes from (8) to (9) in the proof of Theorem 3.6. So, Shafer implicitly assumes that the following equality is valid

$$\sum_{\emptyset \neq B \cap C \subset A} m_1(C) = \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C). \quad (14)$$

In fact, (14) is wrong as shown in the next simple counterexample. Hence, Shafer's proof for  $Bel_1(A|B)$  is incorrect. This mistake casts doubts on the correctness of formulas in Theorem 3.6. However, we show in the next section that formulas given in Theorem 3.6 are in fact correct and we give in this paper their correct proofs. It is quite easy to verify that

<sup>3</sup>if one accepts Shafer's standpoint for belief conditioning based on Dempster's rule.

(9) is not equal to (10) because<sup>4</sup>

$$\begin{aligned}
 Bel_1(A \cup \bar{B}) &= \sum_{C \subset A \cup \bar{B}} m_1(C) \\
 &= \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C) + \sum_{\substack{C \subset A \cup \bar{B} \\ C \subset B}} m_1(C) \\
 &= \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C) + \sum_{C \subset (A \cup \bar{B}) \cap B} m_1(C) \\
 &= \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C) + \sum_{C \subset (A \cap B) \cup (\bar{B} \cap B)} m_1(C) \\
 &= \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C) + \sum_{C \subset (A \cap B) \cup \emptyset} m_1(C) \\
 &= \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C) + \sum_{C \subset (A \cap B)} m_1(C) \\
 &= \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C) + Bel_1(A \cap B).
 \end{aligned}$$

Therefore, the numerators of (9) and (10) are different in general because

$$\begin{aligned}
 \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C) &= Bel_1(A \cup \bar{B}) - Bel_1(A \cap B) \quad (15) \\
 &\neq Bel_1(A \cup \bar{B}) - Bel_1(\bar{B}).
 \end{aligned}$$

**Remark:** One may argue that there is just a small typo error in Shafer's book, and in fact the incorrect expression  $\sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C)$  in (14), must be replaced by  $\sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset \bar{B}}} m_1(C)$ . Even if one admits this possibility of typo error in Shafer's proof, it is not trivial to prove the (modified/corrected) equality

$$\sum_{\substack{C \\ \emptyset \neq B \cap C \subset A}} m_1(C) = \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset \bar{B}}} m_1(C), \quad (16)$$

to get the final Shafer's belief conditioning formula. That is why we provide a complete exact and detailed proof of Shafer's belief conditioning formula in section IV.

*A simple counter-example of Shafer's proof*

Consider the following FoD  $\Theta = \{\theta_1, \dots, \theta_7\}$  satisfying Shafer's model. We consider and denote the focal elements of  $m_1(\cdot)$  as follows  $A \triangleq \{\theta_2, \theta_3, \theta_4, \theta_5, \theta_7\} = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_7$ ,  $B \triangleq \{\theta_1, \theta_2, \theta_3, \theta_4\} = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$ ,  $C_1 \triangleq \{\theta_3, \theta_5, \theta_6\} = \theta_3 \cup \theta_5 \cup \theta_6$ ,  $C_2 \triangleq \{\theta_4, \theta_7\} = \theta_4 \cup \theta_7$ ,  $C_3 \triangleq \theta_2$ , and the BBA  $m_1(\cdot)$  defined on the FoD  $\Theta$  given by  $m_1(A) = 0.1$ ,  $m_1(B) = 0.1$ ,  $m_1(C_1) = 0.2$ ,  $m_1(C_2) = 0.3$  and  $m_1(C_3) = 0.3$ . We consider the subset  $B = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$  being the conditioning term, characterized by the BBA  $m_2(B) = 1$ ,

<sup>4</sup>The denominators of (9) and (10) being equal, we just need to verify if the numerators of (9) and (10) are equal, or not.

hence  $Bel_2(B) = 1$ . Note that  $\bar{B} = \Theta \setminus B = \{\theta_5, \theta_6, \theta_7\}$  and  $Bel_1(\bar{B}) = 0$  because there is no focal elements of  $m_1(\cdot)$  included in  $\bar{B} = \theta_5 \cup \theta_6 \cup \theta_7$ .

• Let us calculate at first the sum  $S_1 \triangleq \sum_{\substack{C \\ \emptyset \neq B \cap C \subset A}} m_1(C)$  involved in (8). All focal elements  $C$  of  $m_1(\cdot)$  such that  $\emptyset \neq B \cap C \subset A$  are the focal elements  $A$ ,  $C_1$ ,  $C_2$  and  $C_3$  because  $B \cap A = \theta_2 \cup \theta_3 \cup \theta_4 \neq \emptyset$  and  $\theta_2 \cup \theta_3 \cup \theta_4 \subset A$ ,  $B \cap C_1 = \theta_3 \neq \emptyset$  and  $\theta_3 \subset A$ ,  $B \cap C_2 = \theta_4 \neq \emptyset$  and  $\theta_4 \subset A$ ,  $B \cap C_3 = \theta_2 \neq \emptyset$  and  $\theta_2 \subset A$ . The focal element  $C = B$  of  $m_1(\cdot)$  is not involved in the sum  $S_1$  because if  $C = B$ , then  $B \cap C = B \cap B = B \not\subset A$ . Therefore, one gets

$$\begin{aligned}
 S_1 &= m_1(A) + m_1(C_1) + m_1(C_2) + m_1(C_3) \\
 &= 0.1 + 0.2 + 0.3 + 0.3 = 0.9.
 \end{aligned}$$

Hence, based on (8) which is the correct expression obtained from (7), one gets the correct value of Shafer's belief conditioning

$$Bel_1(A|B) = S_1 / (1 - Bel_1(\bar{B})) = 0.9 / (1 - 0) = 0.9.$$

• Let us calculate the sum  $S_2 \triangleq \sum_{\substack{C \subset A \cup \bar{B} \\ C \not\subset B}} m_1(C)$  involved in (9). First note that  $A \cup \bar{B} = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$  and the focal elements  $C$  of  $m_1(\cdot)$  such that  $C \subset (A \cup \bar{B})$  and  $C \not\subset B$  are the three focal elements  $A$ ,  $C_1$  and  $C_2$  because  $A \subset A \cup \bar{B}$  and  $A \not\subset B$ ,  $C_1 = \theta_3 \cup \theta_5 \cup \theta_6 \subset A \cup \bar{B}$  and  $C_1 \not\subset B$ ,  $C_2 = \theta_4 \cup \theta_7 \subset A \cup \bar{B}$  and  $C_2 \not\subset B$ . The focal element  $B = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$  of  $m_1(\cdot)$  is not included in  $A \cup \bar{B} = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$  because, in this example,  $B$  is not included in  $A$ , and of course because  $B \cap \bar{B} = \emptyset$ . The focal element  $C_3 = \theta_2$  of  $m_1(\cdot)$  is included in  $A \cup \bar{B} = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$  but  $C_3 = \theta_2$  is also included in  $B = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$ , so that the condition  $C_3 \not\subset B$  is not satisfied. Based on these remarks, one gets for  $S_2$

$$S_2 = m_1(A) + m_1(C_1) + m_1(C_2) = 0.1 + 0.2 + 0.3 = 0.6.$$

We can verify that the value of  $S_2$  corresponds to the value obtained with the correct formula (15), because  $Bel_1(A \cup \bar{B}) = m_1(A) + m_1(C_1) + m_1(C_2) + m_1(C_3) = 0.9$  and  $Bel_1(A \cap B) = m_1(C_3) = 0.3$  so that  $S_2 = Bel_1(A \cup \bar{B}) - Bel_1(A \cap B) = 0.9 - 0.3 = 0.6$ . Hence, based on (9), one would get an incorrect value of Shafer's belief conditioning

$$Bel_1(A|B) = S_2 / (1 - Bel_1(\bar{B})) = 0.6 / (1 - 0) = 0.6.$$

Clearly, this counter-example shows that  $S_1 \neq S_2$  and proves that the equality (14) is incorrect. This simple counter examples illustrates that the proof of Theorem 3.6 given by Shafer is incorrect.

IV. CORRECT PROOF OF FORMULAS OF THEOREM 3.6

Starting from Dempster's rule we have  $m(\emptyset) = 0$  and for all  $A \neq \emptyset \in 2^\Theta$ ,

$$m(A) = [m_1 \oplus m_2](A) = \frac{\sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = A}} m_1(X_1)m_2(X_2)}{1 - \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2)}. \quad (17)$$

Because in conditioning by  $B \neq \emptyset$ ,  $m_2(\cdot)$  is defined by  $m_2(X_2) = 1$  if  $X_2 = B$ , and  $m_2(X_2) = 0$  otherwise, the previous expression reduces for  $A \neq \emptyset$  to

$$m(A) = \frac{\sum_{\substack{X_1 \in 2^\emptyset \\ \emptyset \neq X_1 \cap B = A}} m_1(X_1)}{1 - \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B = \emptyset}} m_1(X_1)} = \frac{\sum_{\substack{X_1 \in 2^\emptyset \\ \emptyset \neq X_1 \cap B = A}} m_1(X_1)}{1 - Bel_1(\bar{B})}, \quad (18)$$

because  $Bel_1(\bar{B}) = \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \subseteq \bar{B}}} m_1(X_1) = \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B = \emptyset}} m_1(X_1)$ .

Using the definition of the belief function,  $Bel_1(A|B)$  for  $B \neq \emptyset$  is given by

$$\begin{aligned} Bel_1(A|B) &= \sum_{\substack{Y \in 2^\emptyset \\ Y \subseteq A}} m(Y) \\ &= \sum_{\substack{Y \in 2^\emptyset \\ Y \subseteq A}} \frac{\sum_{\substack{X_1 \in 2^\emptyset \\ \emptyset \neq X_1 \cap B = Y}} m_1(X_1)}{1 - Bel_1(\bar{B})} \\ &= \frac{\sum_{\substack{Y \in 2^\emptyset \\ Y \subseteq A}} \sum_{\substack{X_1 \in 2^\emptyset \\ \emptyset \neq X_1 \cap B = Y}} m_1(X_1)}{1 - Bel_1(\bar{B})} \\ &= \frac{\sum_{\substack{X_1 \in 2^\emptyset \\ \emptyset \neq X_1 \cap B \subseteq A}} m_1(X_1)}{1 - Bel_1(\bar{B})}. \end{aligned} \quad (19)$$

Note that equation (19) is the same as Shafer's equation (8) using slight modified notations<sup>5</sup> for better presentation in the sequel.

Because  $m_1(\cdot)$  is a normalized BBA, one has for all  $B \in 2^\emptyset$

$$\sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B = \emptyset}} m_1(X_1) + \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B \neq \emptyset}} m_1(X_1) = 1. \quad (20)$$

Also, for any  $A \in 2^\emptyset$  and in partitioning  $2^\emptyset$  in the subsets  $\{Y \in 2^\emptyset | Y \subseteq A\}$  and  $\{Y \in 2^\emptyset | Y \not\subseteq A\}$ , the following equality also always holds

$$\begin{aligned} &\sum_{\substack{Y \in 2^\emptyset \\ Y \subseteq A}} \left[ \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B \cap Y = \emptyset}} m_1(X_1) + \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B \cap Y \neq \emptyset}} m_1(X_1) \right] \\ &+ \sum_{\substack{Y \in 2^\emptyset \\ Y \not\subseteq A}} \left[ \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B \cap Y = \emptyset}} m_1(X_1) + \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B \cap Y \neq \emptyset}} m_1(X_1) \right] = 1. \end{aligned} \quad (21)$$

<sup>5</sup>We have also replaced symbol  $\subset$  by  $\subseteq$  for clarity.

This equality can be rewritten equivalently as

$$\begin{aligned} &\sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B = \emptyset) \subseteq A}} m_1(X_1) + \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \subseteq A}} m_1(X_1) \\ &+ \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B = \emptyset) \not\subseteq A}} m_1(X_1) + \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \not\subseteq A}} m_1(X_1) = 1. \end{aligned} \quad (22)$$

The second term of the left hand side of (22) corresponds to the numerator of  $Bel_1(A|B)$  given in (19). We can express it as

$$\begin{aligned} &\sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \subseteq A}} m_1(X_1) = 1 - \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B = \emptyset) \subseteq A}} m_1(X_1) \\ &- \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B = \emptyset) \not\subseteq A}} m_1(X_1) - \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \not\subseteq A}} m_1(X_1). \end{aligned}$$

Because

$$\sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B = \emptyset) \subseteq A}} m_1(X_1) + \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B = \emptyset) \not\subseteq A}} m_1(X_1) = \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B = \emptyset}} m_1(X_1),$$

one gets

$$\begin{aligned} &\sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \subseteq A}} m_1(X_1) = 1 - \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B = \emptyset}} m_1(X_1) \\ &- \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \not\subseteq A}} m_1(X_1) \\ &= 1 - Bel_1(\bar{B}) - Pl_1(\bar{A} \cap B). \end{aligned}$$

The last previous equality comes from the fact that

$$\begin{aligned} Bel_1(\bar{B}) &= \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \subseteq \bar{B}}} m_1(X_1) = \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B = \emptyset}} m_1(X_1), \\ Pl_1(\bar{A} \cap B) &= \sum_{\substack{X_1 \in 2^\emptyset \\ X_1 \cap B \cap \bar{A} \neq \emptyset}} m_1(X_1) = \sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \not\subseteq A}} m_1(X_1). \end{aligned}$$

Therefore, the numerator of  $Bel_1(A|B)$  given in (19) equals  $1 - Pl_1(\bar{A} \cap B) - Bel_1(\bar{B})$ . Because  $Pl_1(\bar{A} \cap B) = 1 - Bel_1(\bar{A} \cap B) = 1 - Bel_1(A \cup \bar{B})$ , one finally gets for the numerator of  $Bel_1(A|B)$

$$\sum_{\substack{X_1 \in 2^\emptyset \\ (X_1 \cap B \neq \emptyset) \subseteq A}} m_1(X_1) = Bel_1(A \cup \bar{B}) - Bel_1(\bar{B}), \quad (23)$$

and the final expression of  $Bel_1(A|B)$  is given by

$$Bel_1(A|B) = (Bel_1(A \cup \bar{B}) - Bel_1(\bar{B})) / (1 - Bel_1(\bar{B})). \quad (24)$$

This expression coincides with the final expression (10) given by Shafer in his flawed proof. The derivation of  $Pl_1(A|B)$  given in Shafer's proof is correct since we have proved that the expression of  $Bel_1(A|B)$  is correct.

## V. CONCLUSION

In this paper we have shown why the proof of belief conditioning formulas given by Shafer is wrong and we have illustrated this incorrectness with a simple counter-example. After the identification of the mistake in Shafer's proof, we have provided the correct proof of final expressions of Shafer's belief conditioning formulas. For readers interested in belief conditioning, we provide a solid justification against the belief conditioning method proposed by Shafer in our companion paper [3]. Our criticism of Shafer's conditioning approach is based on the Total Belief Theorem and Generalized Bayes' Theorem.

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.
- [2] R. Fagin, J.Y. Halpern, *A new approach to updating beliefs*, UAI Proc., 1991.
- [3] J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Generalized Bayes' Theorem*, in Proc. of Int. Conf. on Information Fusion (Fusion 2018), Cambridge, UK, July 2018.
- [4] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., Vol. 38, pp. 325–339, 1967.
- [5] A.P. Dempster, *A generalization of Bayesian inference*, J. Roy. Stat. Soc., B30, 1968.
- [6] D. Dubois, H. Prade, *A set-theoretic view of belief functions*, Int. J. General Systems, Vol. 12, pp. 193–226, 1976.
- [7] L.A. Zadeh, *A Mathematical Theory of Evidence (book review)*, AI Mag., Vol. 5, 1984.
- [8] J. Lemmer, *Confidence factors, empiricism and the Dempster-Shafer theory of evidence*, in Proc. of 1st UAI Conf., pp. 16–176, 1985.
- [9] P.K. Black, *Is Shafer general Bayes?*, in Proc. of 3rd UAI Workshop, USA, 1987.
- [10] P. Wang, *A defect in Dempster-Shafer theory*, in Proc. of UAI Conference, 1994.
- [11] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Int. Conf. on Information Fusion (Fusion 2012), Singapore, July 9–12, 2012.
- [12] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's fusion rule is not a generalization of Bayes fusion rule*, in Proc. of Int. Conf. on Information Fusion (Fusion 2013), Istanbul, Turkey, July 9–12, 2013.
- [13] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., Vol. 29, pp. 223–252, 2014.



# Total Belief Theorem and Generalized Bayes' Theorem

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Generalized Bayes' Theorem*, in Proc. of Int. Conf. on Information Fusion (Fusion 2018), Cambridge, UK, July 10–13, 2018 (Best Paper Awards), and reprinted with permission.

**Abstract**—This paper presents two new theoretical contributions for reasoning under uncertainty: 1) the Total Belief Theorem (TBT) which is a direct generalization of the Total Probability Theorem, and 2) the Generalized Bayes' Theorem drawn from TBT. A constructive justification of Fagin-Halpern belief conditioning formulas proposed in the nineties is also given. We also show how our new approach and formulas work through simple illustrative examples.

**Keywords:** Total Belief Theorem (TBT), Generalized Bayes' Theorem (GBT), belief functions.

## I. INTRODUCTION

This paper presents new theoretical results for reasoning under uncertainty with belief functions (BF) introduced by Shafer in [1] in Dempster-Shafer Theory (DST). The first important result is the Total Belief Theorem (TBT) which is a generalization of the Total Probability Theorem (TPT) for the belief functions framework. From TBT, one can provide a solid justification of Fagin-Halpern (FH) belief conditioning formulas [3]–[5] which are generalizations of the classical conditional probability formulas. These theoretical results allow us to establish rigorously the Generalized Bayes' Theorem (GBT). The belief conditioning problem is challenging, not new, and one of the two main methods usually adopted by users working with BF is : 1) Shafer's belief conditioning method based on Dempster's rule of combination [1], or 2) the belief conditioning method consistent with imprecise probability calculus bounds [2], [6], [7] based on the lower and upper probability interpretation of belief functions popularized by Fagin and Halpern [3]. In this paper we focus on the second approach of belief conditioning because Dempster's rule of combination presents serious problems as reported in [8]–[16]. Smets did also attempt to generalize Bayes' Theorem (BT) and did propose his own GBT [17] on the basis of conditional embedding, conjunctive merging and Shafer's conditioning. Unfortunately, Smets' approach remains doubtful as reported in [18]. Our new GBT establishment is obtained by a direct constructive manner from TBT. It does not need extra assumptions nor some underlying ad-hoc construction principles. Also, we prove that our TBT and GBT presented in this work are fully consistent with classical TPT and BT as soon as the belief functions are Bayesian.

This paper starts with a brief review of very basics of Probability Theory, including the Total Probability Theorem (TPT) and Bayes' Theorem (BT) in Section II because this helps to have a better understanding of the generalizations we propose. A brief review of belief functions is given in Section III, followed by classical Shafer's and Fagin-Halpern's belief conditioning methods respectively in Sections IV and V. In Section VI, we present the decomposition of the set of focal elements of any basic belief assignment (BBA) that allows us to establish formally the TBT and its generalization on Cartesian product space. The Section VII presents and justifies the new belief conditioning formulas drawn from TBT which are fully consistent with Fagin-Halpern conditioning formulas. This section also presents the generalization of Bayes' theorem in the framework of belief functions. We illustrate our new theoretical results with a quite simple GBT example in Section VIII to show how to make derivations of GBT and to prove that Shafer's conditioning results are inconsistent with GBT. Section IX concludes this paper.

## II. TOTAL PROBABILITY THEOREM & BAYES' FORMULA

### A. Total Probability Theorem

In probability theory, the elements  $\theta_i$  of the space  $\Theta$  are experimental outcomes. The subsets of  $\Theta$  are called events and the event  $\{\theta_i\}$  consisting of the single element  $\theta_i$  is an elementary event. The space  $\Theta$  is called the *sure event* and the empty set  $\emptyset$  is the *impossible event*. We assign to each event  $A$  a number  $P(A)$  in  $[0, 1]$ , called the probability of  $A$ , which satisfies the three Kolmogorov's conditions: 1)  $P(\emptyset) = 0$ ; 2)  $P(\Theta) = 1$ ; and 3) if  $A \cap B = \{\emptyset\}$ , then  $P(A \cup B) = P(A) + P(B)$ . These conditions are the axioms of the theory of probability [20]. The fundamental Theorem of the probability theory is the Total Probability Theorem (TPT), also called the law of total probability, see [20] which can be stated as follows.

**Total Probability Theorem (TPT):** Consider an event  $B$  and any partition<sup>1</sup>  $\{A_1, A_2, \dots, A_k\}$  of the space  $\Theta$ . Then

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k). \quad (1)$$

<sup>1</sup>A partition of  $\Theta$  is a collection of exclusive subsets of  $\Theta$  whose union equals  $\Theta$ .

### B. Conditional probability and Bayes' formula

Starting from TPT formula (1) and assuming  $P(B) > 0$ , we get for any  $i \in \{1, \dots, k\}$  after dividing each side of (1) by  $P(B)$  and rearranging terms the equality

$$\frac{P(A_i \cap B)}{P(B)} = 1 - \sum_{\substack{j=1, \dots, k \\ j \neq i}} \frac{P(A_j \cap B)}{P(B)} = 1 - \frac{P(\bar{A}_i \cap B)}{P(B)}, \quad (2)$$

which allows us to define the conditional probability  $P(A_i|B)$  by<sup>2</sup>

$$P(A_i|B) \triangleq P(A_i \cap B)/P(B). \quad (3)$$

Similarly, by considering an event  $A_i$  of  $\Theta$  and the partition  $\{B, \bar{B}\}$  of  $\Theta$ , the TPT formula  $P(A_i) = P(A_i \cap B) + P(A_i \cap \bar{B})$  applies, and by dividing it by  $P(A_i)$  (assuming  $P(A_i) > 0$ ), one gets

$$\frac{P(A_i \cap B)}{P(A_i)} = 1 - \frac{P(A_i \cap \bar{B})}{P(A_i)}, \quad (4)$$

which allows to define the conditional probability  $P(B|A_i)$  by

$$P(B|A_i) \triangleq P(A_i \cap B)/P(A_i). \quad (5)$$

From (3) and (5), one deduces the equality

$$P(A_i \cap B) = P(A_i|B)P(B) = P(B|A_i)P(A_i). \quad (6)$$

From equality (6) and assuming  $P(B) > 0$  and  $P(A_i) > 0$ , we get

$$P(A_i|B) = P(B|A_i)P(A_i)/P(B), \quad (7)$$

$$P(B|A_i) = P(A_i|B)P(B)/P(A_i). \quad (8)$$

Using (1) and noting that  $P(A_i \cap B) = P(B|A_i)P(A_i)$ , we get

$$P(B) = \sum_{i=1}^k P(B|A_i)P(A_i). \quad (9)$$

Substituting (9) in (7), we obtain Bayes' Theorem (BT) formula stated mathematically as the following equation

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^k P(B|A_i)P(A_i)}. \quad (10)$$

One can verify that the conditional probability defined by (3) satisfies the three axioms of the Theory of Probability [20].

Previously,  $A_i$  and  $B$  were events (subsets) of the same space  $\Theta$ . If  $A_i \subseteq \Theta_1$  and  $B \subseteq \Theta_2$  with  $\Theta_1 \neq \Theta_2$ , which corresponds to a so-called combined experiment [20], similar conditioning formulas can also be established by working in the Cartesian product space  $\Theta \triangleq \Theta_1 \times \Theta_2$  whose elementary elements are all the ordered pairs  $(x_p, y_q)$  with  $x_p \in \Theta_1$  and  $y_q \in \Theta_2$ . The two experiments are viewed as a single combined one whose outcomes are pairs  $(x_p, y_q)$ . In this space  $\Theta = \Theta_1 \times \Theta_2$ ,  $x_p$  is not an elementary element but a subset of  $n$  elements of  $\Theta$ , i.e.  $\{x_p\} = \{(x_p, y_1), \dots, (x_p, y_n)\}$ . Similarly,  $y_q$  is not an elementary element but a subset of  $m$  elements of  $\Theta$ , i.e.  $\{y_q\} = \{(x_1, y_q), \dots, (x_m, y_q)\}$ . If  $A_i \subseteq \Theta_1$  and  $B \subseteq \Theta_2$ , then  $A_i \times B = \{(x_p, y_q) | x_p \in A_i, y_q \in B\} \subseteq \Theta$ . If one forms  $A_i \times \Theta_2$  and  $\Theta_1 \times B$  one sees that  $A_i \times B = (A_i \times \Theta_2) \cap (\Theta_1 \times B) = (\Theta_1 \times B) \cap (A_i \times \Theta_2)$ . Because the event  $A_i \times \Theta_2$  occurs in the combined experiment if the event  $A_i$  of the experiment 1 occurs no matter what the outcome of experiment 2 is, one has  $P(A_i \times \Theta_2) = P_1(A_i)$  where  $P_1(A_i)$  is the probability of event  $A_i$  in the experiment 1. Similarly, the event  $\Theta_1 \times B$  occurs if  $B$  occurs in experiment 2 no matter what the outcome of experiment 1 is, so that  $P(\Theta_1 \times B) = P_2(B)$  where  $P_2(B)$  is the probability of event  $B$  in the experiment 2. Considering a partition  $\{A_1, A_2, \dots, A_k\}$  of  $\Theta_1$  and a subset (event)  $B \subseteq \Theta_2$ ,

<sup>2</sup>The notation  $\triangleq$  means *equal by definition*.

and based on set theory and property of Cartesian product, one can establish also TPT formula

$$P(\Theta_1 \times B) = \sum_{i=1, \dots, k} P((\Theta_1 \times B) \cap (A_i \times \Theta_2)),$$

and Bayes' formula

$$P(A_i \times \Theta_2 | \Theta_1 \times B) = \frac{P(\Theta_1 \times B | A_i \times \Theta_2) P(A_i \times \Theta_2)}{\sum_{i=1}^k P(\Theta_1 \times B | A_i \times \Theta_2) P(A_i \times \Theta_2)}.$$

That is why, for notation convenience (and notation abuse), we can just use classical formulas even when working with different sets of experimental outcomes  $\Theta_1$  and  $\Theta_2$ . One just has to keep in mind that in this case  $A_i$  must be understood as  $A_i \times \Theta_2$  and  $B$  as  $\Theta_1 \times B$ .

### III. BASICS OF BELIEF FUNCTIONS

Based on Dempster's works [2], [19], Shafer did introduce Belief Functions (BF) to model the epistemic uncertainty<sup>3</sup> and to reason under uncertainty [1]. We consider a finite discrete frame of discernement (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all exhaustive and exclusive elements of  $\Theta$  represent the set of the potential solutions of the problem under concern. The set of all subsets of  $\Theta$  is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . A basic belief assignment (BBA) associated with a given source of evidence is defined as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying the conditions  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is the mass of belief of subset  $A$  committed by the source of evidence (SoE). A focal element  $X$  of a BBA  $m(\cdot)$  is an element of  $2^\Theta$  such that  $m(X) > 0$ . Note that the empty set  $\emptyset$  is not a focal element of a BBA because  $m(\emptyset) = 0$  (closed-world assumption of Shafer's model for the FoD). The set of all focal elements of  $m(\cdot)$  is denoted

$$\mathcal{F}_\Theta(m) \triangleq \{X \subseteq \Theta | m(X) > 0\} = \{X \in 2^\Theta | m(X) > 0\}. \quad (11)$$

The set of focal elements of  $m(\cdot)$  included in  $A \subseteq \Theta$  is denoted

$$\mathcal{F}_A(m) \triangleq \{X \in \mathcal{F}_\Theta(m) | X \cap A = X\}. \quad (12)$$

Note that if  $A \subseteq B \subseteq \Theta$ , then  $\mathcal{F}_A(m) \subseteq \mathcal{F}_B(m)$ . Also,  $\forall A, B \subseteq \Theta$  one has  $\mathcal{F}_{A \cap B}(m) = \mathcal{F}_A(m) \cap \mathcal{F}_B(m)$ , but  $\mathcal{F}_{A \cup B}(m) \neq \mathcal{F}_A(m) \cup \mathcal{F}_B(m)$  in general. The set  $\mathcal{F}_\Theta(m)$  can always be partitioned as  $\{\mathcal{F}_A(m), \mathcal{F}_{\bar{A}}(m), \mathcal{F}_{A^*}(m)\}$  where<sup>4</sup>

$$\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_A(m) - \mathcal{F}_{\bar{A}}(m) \quad (13)$$

$$= \{X \in \mathcal{F}_\Theta(m) | X \cap A \neq \emptyset \text{ and } X \cap \bar{A} \neq \emptyset\}, \quad (14)$$

represents the set of focal elements of  $m(\cdot)$  which are not subsets of  $A$  and not subsets of  $\bar{A} \triangleq \Theta - \{A\} = \{X | X \in \Theta \text{ and } X \notin A\}$ , where  $\bar{A}$  is the complement of  $A$  in  $\Theta$  and the minus symbol denotes the set difference operator.

Belief and plausibility functions are defined by<sup>5</sup>

$$Bel(A) = \sum_{\substack{X \in 2^\Theta \\ X \subseteq A}} m(X) = \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \subseteq A}} m(X) = \sum_{X \in \mathcal{F}_A(m)} m(X), \quad (15)$$

$$Pl(A) = \sum_{\substack{X \in 2^\Theta \\ X \cap A \neq \emptyset}} m(X) = \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \cap A \neq \emptyset}} m(X) = 1 - Bel(\bar{A}). \quad (16)$$

The width  $U(A^*) = Pl(A) - Bel(A)$  of the belief interval  $[Bel(A), Pl(A)]$  is called the *uncertainty on A* committed by the

<sup>3</sup>Also called sometimes the cognitive uncertainty by some authors.

<sup>4</sup>For notation convenience, we use  $A^*$  to denote focal elements of  $m(\cdot)$  which are not in  $A$ , nor in  $\bar{A}$ .

<sup>5</sup>By convention, a *sum of non existing terms* (if it occurs in formulas depending on the given BBA) is always set to zero.

SoE. It represents the imprecision on the (subjective) probability of  $A$  granted by the SoE which provides the BBA  $m(\cdot)$ . The uncertainty  $U(A^*)$  can also be expressed directly as

$$U(A^*) = \sum_{X \in \mathcal{F}_{A^*}(m)} m(X). \quad (17)$$

It is worth noting that  $U(\bar{A}^*) = Pl(\bar{A}) - Bel(\bar{A}) = (1 - Bel(A)) - (1 - Pl(A)) = Pl(A) - Bel(A) = U(A^*)$ , or equivalently

$$U(\bar{A}^*) = \sum_{X \in \mathcal{F}_{\bar{A}^*}(m)} m(X), \quad (18)$$

where  $\mathcal{F}_{\bar{A}^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{A}}(m) - \mathcal{F}_A(m) = \mathcal{F}_{A^*}(m)$ .

When all elements of  $\mathcal{F}_\Theta(m)$  are only singletons,  $m(\cdot)$  is called a *Bayesian BBA* [1] and its corresponding  $Bel(\cdot)$  and  $Pl(\cdot)$  functions are homogeneous to a same (subjective) probability measure  $P(\cdot)$ . In this case  $\mathcal{F}_{A^*}(m) = \mathcal{F}_{\bar{A}^*}(m) = \emptyset$ . According to Shafer's Theorem 1 below, see [1] page 39 with its proof on page 51, the belief functions can be characterized without referencing to a BBA.

**Theorem 1:** If  $\Theta$  is a FoD, then a function  $Bel : 2^\Theta \mapsto [0, 1]$  is a belief function if and only if it satisfies the following conditions:

- B1) Belief in impossible event is zero, that is  $Bel(\emptyset) = 0$ .
- B2) Belief in the certain event is one, that is  $Bel(\Theta) = 1$ .
- B3) For every positive integer  $n$ , and for every collection  $A_1, \dots, A_n$  of subsets of  $\Theta$

$$Bel(A_1 \cup \dots \cup A_n) \geq \sum_{\substack{I \subseteq \{1, \dots, n\} \\ I \neq \emptyset}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i). \quad (19)$$

Quantities  $m(\cdot)$  and  $Bel(\cdot)$  are one-to-one, and for any  $A \subseteq \Theta$  the BBA  $m(\cdot)$  is obtained from  $Bel(\cdot)$  by Möbius inverse formula (see [1], p.39)

$$m(A) = \sum_{B \subseteq A \subseteq \Theta} (-1)^{|A-B|} Bel(B). \quad (20)$$

Shafer [1] did propose to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD with Dempster's rule (i.e. the normalized conjunctive rule). However Dempster's rule has been strongly disputed from both theoretical and practical standpoints as reported in [16], [21], [22]. In particular, the high (or even very low) conflict level between the sources can be totally ignored by Dempster's rule which is a very serious problem [15]. Also, Shafer's conditioning (based on Dempster's rule) is inconsistent with the probabilistic conditioning (see next section).

#### IV. SHAFER'S CONDITIONING

##### A. Shafer's conditioning formulas

Shafer's conditioning formulas are established in Theorem 3.6 p. 66 of [1] from Dempster's rule of combination of the original BBA  $m(\cdot)$  with the BBA  $m_B(B) = 1$  focused on  $B$ . We review them for convenience. For  $A, B \subseteq \Theta$  with  $Pl(B) > 0$ ,  $Bel(A|B)$  and  $Pl(A|B)$  are given by

$$Bel(A|B) = (Bel(A \cup \bar{B}) - Bel(\bar{B})) / (1 - Bel(\bar{B})), \quad (21)$$

$$Pl(A|B) = Pl(A \cap B) / Pl(B). \quad (22)$$

The expression (21) of  $Bel(A|B)$  is equivalent to

$$Bel(A|B) = (Pl(B) - Pl(B \cap \bar{A})) / Pl(B), \quad (23)$$

because one has always (from definition of belief functions)  $Pl(B) = 1 - Bel(\bar{B})$  and the numerator of (21) can be written as

$$Bel(A \cup \bar{B}) - Bel(\bar{B}) = Pl(B) - Pl(B \cap \bar{A}).$$

If  $A = \emptyset$ ,  $Bel(\emptyset|B) = Pl(\emptyset|B) = 0$ , and if  $A = \Theta$ ,  $Bel(\Theta|B) = Pl(\Theta|B) = 1$ . Also, if  $B = \Theta$ ,  $Bel(A|\Theta) = Bel(A)$

and  $Pl(A|\Theta) = Pl(A)$ . Note that if  $B = A$  in (22)–(23), we get  $Bel(A|A) = Pl(A|A) = 1$  which fits with the common sense.

In reversing the roles played by  $A$  and  $B$  and switching the notations in previous expressions, the following formulas also hold (assuming  $Pl(A) > 0$ )

$$Bel(B|A) = (Pl(A) - Pl(A \cap \bar{B})) / Pl(A), \quad (24)$$

$$Pl(B|A) = Pl(B \cap A) / Pl(A). \quad (25)$$

From (22) and (25), one deduces that

$$Pl(A \cap B) = Pl(A|B)Pl(B) = Pl(B|A)Pl(A).$$

Hence, the following formula applies for conditional plausibilities when  $Pl(B) > 0$

$$Pl(A|B) = Pl(B|A)Pl(A) / Pl(B). \quad (26)$$

Shafer's formula (25) is similar to conditional probabilities (3) when replacing plausibility by probability. So, at first glance it seems appealing. In the sequel we show why this is not the case.

##### B. Drawback of Shafer's conditioning

The main drawback of Shafer's conditioning is that the bounds of belief interval  $[Bel(A|B), Pl(A|B)]$  obtained by (21)–(22) are in general incompatible with lower and upper bounds of the conditional probability  $P(A|B)$ . This problem makes Shafer's conditioning based on Dempster's rule very disputable and cast doubts on pertinence (validity) of Shafer's conditioning results when used in applications. This serious problem has already been reported and addressed by several authors [3], [6], [7], [11] with some examples. To easily show this incompatibility of Shafer's conditioning with probability calculus we present briefly the famous Ellsberg urn example [23].

**Example 1 (Ellsberg urn):** We consider an urn with red (R) balls, black (B) and yellow (Y) balls. One knows that 1/3 of balls are red balls and 2/3 of balls are black and yellow balls. So the a priori information about the chance to pick a ball in the urn can be represented by a (parametric) probability mass function  $P(\cdot)$

$$P(R) = 1/3, \quad P(B) = 2/3 - x, \quad P(Y) = x,$$

where  $x$  is an unknown number/parameter in  $[0, 2/3]$ . Therefore,  $P(B)$  and  $P(Y)$  are unknown but their bounds are known. In fact, this problem can be seen as a problem of imprecise probabilities where  $P(R) + P(B) + P(Y) = 1$  with

$$P(R) \in [1/3, 1/3], \quad P(B) \in [0, 2/3], \quad P(Y) \in [0, 2/3].$$

Now let's suppose that someone picks a ball at random in the urn and tell us that the color of the ball is not black, i.e. the event  $\bar{B} = R \cup Y$  has occurred. How do we must revise (update) our prior probabilities with this new information? The correct answer to this question is obtained by computing the conditional probabilities  $P(R|\bar{B})$ ,  $P(B|\bar{B})$  and  $P(Y|\bar{B})$  and by analyzing their bounds. This is done using the fact that  $P(\bar{B}) = P(R \cup Y) = P(R) + P(Y) - P(R \cap Y) = P(R) + P(Y) = (1/3) + x$ . Indeed,  $P(R \cap Y) = 0$  because the events  $R$  and  $Y$  are mutually exclusive. So, we get

$$P(R|\bar{B}) = \frac{P(R \cap (R \cup Y))}{P(R \cup Y)} = \frac{P(R)}{(1/3) + x} = \frac{1/3}{(1/3) + x},$$

$$P(B|\bar{B}) = \frac{P(B \cap (R \cup Y))}{P(R \cup Y)} = \frac{P(\emptyset)}{(1/3) + x} = \frac{0}{(1/3) + x},$$

$$P(Y|\bar{B}) = \frac{P(Y \cap (R \cup Y))}{P(R \cup Y)} = \frac{P(Y)}{(1/3) + x} = \frac{x}{(1/3) + x}.$$

If  $x = 0$ , then  $P(R|\bar{B}) = 1$  and  $P(Y|\bar{B}) = 0$ . If  $x = 2/3$ , then  $P(R|\bar{B}) = 1/3$  and  $P(Y|\bar{B}) = 2/3$ . Therefore after conditioning we get

$$P(R|\bar{B}) \in [1/3, 1], \quad P(B|\bar{B}) \in [0, 0], \quad P(Y|\bar{B}) \in [0, 2/3].$$

Let's examine what we get with Shafer's conditioning. The problem is modeled using the *a priori* BBA  $m(\cdot)$  defined on the FoD  $\Theta = \{R, B, Y\}$  with  $m(R) = 1/3$  and  $m(B \cup Y) = 2/3$  which gives the belief intervals  $[Bel(R), Pl(R)] = [1/3, 1/3]$ ,  $[Bel(B), Pl(B)] = [0, 2/3]$  and  $[Bel(Y), Pl(Y)] = [0, 2/3]$ . With Shafer's conditioning formulas and noting that  $Pl(R) = 1/3$ ,  $Pl(B) = 2/3$ ,  $Pl(Y) = 2/3$ , and  $Pl(R \cup Y) = 1$ , we get

$$\begin{aligned} Bel(R|\bar{B}) &= \frac{Pl(R \cup Y) - Pl((R \cup Y) \cap (B \cup Y))}{Pl(R \cup Y)} = \frac{1 - Pl(Y)}{1} = 1/3, \\ Bel(B|\bar{B}) &= \frac{Pl(R \cup Y) - Pl((R \cup Y) \cap (R \cup Y))}{Pl(R \cup Y)} = \frac{1 - Pl(R \cup Y)}{1} = 0, \\ Bel(Y|\bar{B}) &= \frac{Pl(R \cup Y) - Pl((R \cup Y) \cap (R \cup B))}{Pl(R \cup Y)} = \frac{1 - Pl(R)}{1} = 2/3, \\ Pl(R|\bar{B}) &= \frac{Pl(R \cap (R \cup Y))}{Pl(R \cup Y)} = \frac{Pl(R)}{Pl(R \cup Y)} = 1/3, \\ Pl(B|\bar{B}) &= \frac{Pl(B \cap (R \cup Y))}{Pl(R \cup Y)} = \frac{Pl(\emptyset)}{1} = 0, \\ Pl(Y|\bar{B}) &= \frac{Pl(Y \cap (R \cup Y))}{Pl(R \cup Y)} = \frac{Pl(Y)}{Pl(R \cup Y)} = 2/3. \end{aligned}$$

Hence with Shafer's conditioning we get results incompatible with the real bounds of conditional probabilities because

$$\begin{aligned} [Bel(R|\bar{B}), Pl(R|\bar{B})] &= [1/3, 1/3] \neq [1/3, 1], \\ [Bel(B|\bar{B}), Pl(B|\bar{B})] &= [0, 0], \\ [Bel(Y|\bar{B}), Pl(Y|\bar{B})] &= [2/3, 2/3] \neq [0, 2/3]. \end{aligned}$$

## V. FAGIN-HALPERN CONDITIONING

Fagin and Halpern (FH) proposed in [3], [4] to define the conditional belief as the lower envelope (i.e. the infimum) of a family of conditional probability functions to make belief conditioning consistent with imprecise conditional probability calculus.

### A. Fagin-Halpern conditioning formulas

Assuming  $Bel(B) > 0$ , Fagin and Halpern proposed the following conditional formulas (FH formulas for short)

$$Bel(A|B) = Bel(A \cap B) / (Bel(A \cap B) + Pl(\bar{A} \cap B)), \quad (27)$$

$$Pl(A|B) = Pl(A \cap B) / (Pl(A \cap B) + Bel(\bar{A} \cap B)). \quad (28)$$

They prove in [3] that  $Bel(A|B)$  given by (27) satisfies the three conditions of Theorem 1 and so FH belief conditioning is an appealing solution for BF conditioning. However, it is quite obscure how Fagin and Halpern did obtain (construct) FH formulas. A justification has been given by Sundberg and Wagner in [7] (p. 268) but it is not very easy to follow. In this paper, we justify clearly and directly the establishment of FH formulas from the simple and direct consequence of the Total Belief Theorem (TBT).

Similarly, by switching notations and assuming  $Bel(A) > 0$ , the previous FH formulas can be rewritten as

$$Bel(B|A) = Bel(A \cap B) / (Bel(A \cap B) + Pl(\bar{B} \cap A)), \quad (29)$$

$$Pl(B|A) = Pl(A \cap B) / (Pl(A \cap B) + Bel(\bar{B} \cap A)). \quad (30)$$

As we see, FH formulas are also consistent with Bayes' formula when the underlying BBA  $m(\cdot)$  is Bayesian. Indeed if  $m(\cdot)$  is Bayesian, then  $Pl(A \cap B) = Bel(A \cap B) = P(A \cap B)$ ,  $Pl(\bar{A} \cap B) = Bel(\bar{A} \cap B) = P(\bar{A} \cap B)$  and  $Pl(\bar{B} \cap A) = Bel(\bar{B} \cap A) = P(\bar{B} \cap A)$  and FH formulas become equivalent to

$$Bel(A|B) = Pl(A|B) = P(A \cap B) / (P(A \cap B) + P(\bar{A} \cap B)). \quad (31)$$

Thanks to TPT formula (1), the denominator involved in these formula is  $P(A \cap B) + P(\bar{A} \cap B) = P(B)$ , therefore

$$Bel(A|B) = Pl(A|B) = P(A \cap B) / P(B) = P(A|B). \quad (32)$$

Similarly, one can also easily verify that

$$Bel(B|A) = Pl(B|A) = P(A \cap B) / P(A) = P(B|A). \quad (33)$$

### B. Advantage of Fagin-Halpern conditioning

The advantage of FH conditioning is its complete compatibility with the conditional probability calculus [7], [25]. We show what provides FH conditioning in the previous Ellsberg urn example.

**Ellsberg urn example revisited:** Applying FH conditioning formulas with the conditioning event  $\bar{B} = R \cup Y$  we obtain

$$\begin{aligned} Bel(R|\bar{B}) &= \frac{Bel(R \cap (R \cup Y))}{Bel(R \cap (R \cup Y)) + Pl((B \cup Y) \cap (R \cup Y))} \\ &= \frac{Bel(R)}{Bel(R) + Pl(Y)} = \frac{1/3}{(1/3) + (2/3)} = 1/3, \\ Pl(R|\bar{B}) &= \frac{Pl(R \cap (R \cup Y))}{Bel((B \cup Y) \cap (R \cup Y)) + Pl(R \cap (R \cup Y))} \\ &= \frac{Pl(R)}{Bel(Y) + Pl(R)} = \frac{1/3}{0 + (1/3)} = 1. \end{aligned}$$

Similarly, we can verify that  $Bel(B|\bar{B}) = 0$ ,  $Pl(B|\bar{B}) = 0$ ,  $Bel(Y|\bar{B}) = 0$  and  $Pl(Y|\bar{B}) = 2/3$ . Therefore with these conditioning formulas, we get the correct bounds of the imprecise conditional probabilities

$$\begin{aligned} [Bel(R|\bar{B}), Pl(R|\bar{B})] &= [1/3, 1], \\ [Bel(B|\bar{B}), Pl(B|\bar{B})] &= [0, 0], \\ [Bel(Y|\bar{B}), Pl(Y|\bar{B})] &= [0, 2/3]. \end{aligned}$$

One can also verify that  $Bel(\emptyset|\bar{B}) = 0$ ,  $Bel(R \cup B|\bar{B}) = 1/3$ ,  $Bel(R \cup Y|\bar{B}) = 1$ ,  $Bel(B \cup Y|\bar{B}) = 0$  and  $Bel(R \cup B \cup Y|\bar{B}) = 1$ . Applying Möbius inverse formula (20) with  $Bel(\cdot|\bar{B})$ , one gets the conditional BBA  $m(R|\bar{B}) = 1/3$  and  $m(R \cup Y|\bar{B}) = 2/3$ , whereas with Shafer's conditioning one gets  $m(R|\bar{B}) = 1/3$  and  $m(Y|\bar{B}) = 2/3$ . One sees that with Shafer's conditioning, because  $(B \cup Y) \cap (R \cup Y) \neq \emptyset$  the mass  $m(B \cup Y) = 2/3$  is entirely transferred (optimistically) to the most specific focal element  $Y$  included in  $\bar{B} = R \cup Y$ . With FH conditioning, the mass  $m(B \cup Y) = 2/3$  is entirely transferred (pessimistically, or cautiously) to the least specific focal element  $R \cup Y$  included in  $\bar{B} = R \cup Y$ .

## VI. TOTAL BELIEF THEOREM (TBT)

In this section, we extend TPT theorem to BF and we establish the Total Belief Theorem (TBT) based on a decomposition of  $\mathcal{F}_\Theta(m)$ .

### A. Decomposition of $\mathcal{F}_\Theta(m)$

Let us consider a FoD  $\Theta = \{\theta_1, \dots, \theta_{|\Theta|}\}$  with  $|\Theta| > 1$  elements, and a BBA  $m(\cdot)$  defined on  $2^\Theta$  with a given set of focal elements  $\mathcal{F}_\Theta(m)$ . Considering any partition  $\{A_1, A_2, \dots, A_k\}$  of the FoD  $\Theta$ , then  $\mathcal{F}_\Theta(m)$  can be obtained by the union of following subsets

$$\mathcal{F}_\Theta(m) = \mathcal{F}_{A_1}(m) \cup \dots \cup \mathcal{F}_{A_k}(m) \cup \mathcal{F}_{A^*}(m). \quad (34)$$

where  $\mathcal{F}_{A_i}(m)$  ( $i = 1, \dots, k$ ) is the set of focal elements of  $m(\cdot)$  included in  $A_i$ , and  $\mathcal{F}_{A^*}(m)$  is the set of focal elements of  $m(\cdot)$  which are not included in  $A_i$ ,  $i = 1, \dots, k$ . We use the notation  $A^*$  for representing the entity characterized by the focal set  $\mathcal{F}_{A^*}(m)$  mathematically defined by

$$\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m). \quad (35)$$

The entity  $A^*$  has in general no explicit form and it is used only for notation convenience and conciseness. Because  $A_i$  for  $i = 1, \dots, k$  are mutually exclusive (disjoint), the sets  $\mathcal{F}_{A_i}(m)$  are also mutually exclusive and therefore  $\cap_{i=1, \dots, k} (\mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m)) = \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m)$  because all possible intersections of focal sets including  $\mathcal{F}_{A_i}(m) \cap \mathcal{F}_{A_j}(m)$  for  $j \neq i$  equal the empty set. Hence  $\mathcal{F}_{A^*}(m)$  can also be expressed as

$$\mathcal{F}_{A^*}(m) = \cap_{i=1, \dots, k} \bar{\mathcal{F}}_{A_i}(m), \quad (36)$$

where  $\bar{\mathcal{F}}_{A_i}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m) = \mathcal{F}_{\bar{A}_i}(m) + \mathcal{F}_{A_i^*}(m)$  because when partitioning  $\Theta$  as  $\{A_i, \bar{A}_i\}$  one has  $\mathcal{F}_{A_i^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m) - \mathcal{F}_{\bar{A}_i}(m)$ .

**Example 2:** Consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$  and a BBA  $m(\cdot)$  defined on  $2^\Theta$ , with set of focal elements  $\mathcal{F}_\Theta(m) = \{X_1, X_2, \dots, X_8\}$  chosen as follows:  $X_1 = \theta_1$ ,  $X_2 = \theta_1 \cup \theta_2$ ,  $X_3 = \theta_2 \cup \theta_3$ ,  $X_4 = \theta_3 \cup \theta_4$ ,  $X_5 = \theta_4$ ,  $X_6 = \theta_4 \cup \theta_5$ ,  $X_7 = \theta_1 \cup \theta_3 \cup \theta_5$ , and  $X_8 = \theta_5$ . Consider also the partition  $\{A_1, A_2, A_3\}$  of  $\Theta$  with  $A_1 = \{\theta_1, \theta_2\}$ ,  $A_2 = \{\theta_3, \theta_4\}$  and  $A_3 = \{\theta_5\}$ . Therefore,

$$\begin{aligned} \mathcal{F}_{A_1}(m) &= \{X_1, X_2\} = \{\theta_1, \theta_1 \cup \theta_2\}, \\ \mathcal{F}_{A_2}(m) &= \{X_4, X_5\} = \{\theta_3 \cup \theta_4, \theta_4\}, \\ \mathcal{F}_{A_3}(m) &= \{X_8\} = \{\theta_5\}, \\ \mathcal{F}_{A^*}(m) &= \{X_1, \dots, X_8\} - \{X_1, X_2\} - \{X_4, X_5\} - \{X_8\}, \\ &= \{X_3, X_6, X_7\} = \{\theta_2 \cup \theta_3, \theta_4 \cup \theta_5, \theta_1 \cup \theta_3 \cup \theta_5\}, \\ \bar{\mathcal{F}}_{A_1}(m) &= \mathcal{F}_\Theta(m) - \{X_1, X_2\} = \{X_3, X_4, X_5, X_6, X_7, X_8\}, \\ \bar{\mathcal{F}}_{A_2}(m) &= \mathcal{F}_\Theta(m) - \{X_4, X_5\} = \{X_1, X_2, X_3, X_6, X_7, X_8\}, \\ \bar{\mathcal{F}}_{A_3}(m) &= \mathcal{F}_\Theta(m) - \{X_8\} = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}. \end{aligned}$$

Applying (36), one gets

$$\bar{\mathcal{F}}_{A_1}(m) \cap \bar{\mathcal{F}}_{A_2}(m) \cap \bar{\mathcal{F}}_{A_3}(m) = \{X_3, X_6, X_7\} = \mathcal{F}_{A^*}(m).$$

### B. Total Belief Theorem (TBT)

Based on the previous decomposition of  $\mathcal{F}_\Theta(m)$  according to any partition  $\{A_1, \dots, A_k\}$  of the FoD  $\Theta$ , the following TBT holds.

**Total Belief Theorem (TBT):** Let's consider a FoD  $\Theta$  with  $|\Theta| \geq 2$  elements and a BBA  $m(\cdot)$  defined on  $2^\Theta$  with the set of focal elements  $\mathcal{F}_\Theta(m)$ . For any chosen partition  $\{A_1, \dots, A_k\}$  of  $\Theta$  and for any  $B \subseteq \Theta$ , one has

$$Bel(B) = \sum_{i=1, \dots, k} Bel(A_i \cap B) + U(A^* \cap B), \quad (37)$$

where  $\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m)$  and

$$U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X). \quad (38)$$

**Proof of TBT:** See appendix.

$A^*$  is a shorthand notation for the entity associated to the set of focal elements  $\mathcal{F}_{A^*}(m)$  of the BBA  $m(\cdot)$  involved in the summation (38) of  $U(A^* \cap B)$ . From (38), one sees that  $U(A^* \cap B) \in [0, 1]$ . If one applies TBT with  $B = \Theta$ , we get for any chosen partition  $\{A_1, \dots, A_k\}$  of  $\Theta$ ,  $\sum_{i=1, \dots, k} Bel(A_i) + U(A^*) = 1$  where  $U(A^*) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m)} m(X)$ . This equality corresponds to TPT if  $U(A^*) = 0$  (i.e. there is no uncertainty on the value of probabilities of  $A_i$ ,  $i = 1, \dots, k$ ). Note that if  $B = \Theta$  and if the FoD  $\Theta$  is simply partitioned as  $\{A \triangleq A_1, \bar{A} \triangleq A_2\}$ , then  $U(A^* \cap B) = U(A^* \cap \Theta) = U(A^*) = Pl(A) - Bel(A) = Pl(\bar{A}) - Bel(\bar{A})$ .

**Corollary 1 of TBT:** If  $m(\cdot)$  is Bayesian, then TBT is consistent with the Total Probability Theorem (TPT) because  $U(A^* \cap B) = 0$  and  $Bel(\cdot)$  is homogeneous to a probability measure.

In expressing  $Bel(\bar{B})$  with TBT and noting that  $Pl(B) = 1 - Bel(\bar{B})$ , one can also easily establish the following (not so elegant) Total Plausibility Theorem (TPIT).

**Total Plausibility Theorem (TPIT):** For any partition  $\{A_1, \dots, A_k\}$  of  $\Theta$  and any  $B \subseteq \Theta$ , one has

$$Pl(B) = \sum_{i=1, \dots, k} Pl(\bar{A}_i \cup B) + 1 - k - U(A^* \cap \bar{B}). \quad (39)$$

### C. Example for TBT

Consider the FoD  $\Theta = \{\theta_i, i = 1, \dots, 7\}$  and  $\mathcal{F}_\Theta(m) = \{X_1, X_2, \dots, X_9\}$  of a BBA  $m(\cdot)$  defined over  $2^\Theta$  as in Table I. Consider also the partition  $\{A_1, A_2, A_3\}$  of  $\Theta$  with  $A_1 \triangleq \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$ ,  $A_2 \triangleq \theta_2 \cup \theta_5$  and  $A_3 \triangleq \theta_6$  and the subset  $B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$  of  $\Theta$ . The Table II summarizes the belief values of different subsets of  $\Theta$  which are needed to apply TBT.

Focal element $X$	BBA $m(X)$
$X_1 = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_7$	$m(X_1) = 0.01$
$X_2 = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$	$m(X_2) = 0.02$
$X_3 = \theta_3 \cup \theta_5 \cup \theta_6$	$m(X_3) = 0.03$
$X_4 = \theta_4 \cup \theta_7$	$m(X_4) = 0.04$
$X_5 = \theta_2$	$m(X_5) = 0.20$
$X_6 = \theta_6 \cup \theta_7$	$m(X_6) = 0.30$
$X_7 = \theta_2 \cup \theta_3 \cup \theta_7$	$m(X_7) = 0.20$
$X_8 = \theta_1 \cup \theta_4 \cup \theta_6$	$m(X_8) = 0.15$
$X_9 = \theta_6$	$m(X_9) = 0.05$

Table I  
FOCAL ELEMENTS AND THEIR MASSES.

Subsets of $\Theta$	$Bel(\cdot)$
$B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$	$Bel(B) = 0.39$
$A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$	$Bel(A_1) = 0.04$
$A_2 = \theta_2 \cup \theta_5$	$Bel(A_2) = 0.20$
$A_3 = \theta_6$	$Bel(A_3) = 0.05$
$A_1 \cap B = \theta_4 \cup \theta_7$	$Bel(A_1 \cap B) = 0.04$
$A_2 \cap B = \theta_5$	$Bel(A_2 \cap B) = 0$
$A_3 \cap B = \theta_6$	$Bel(A_3 \cap B) = 0.05$

Table II  
BELIEF VALUES USED FOR THE DERIVATIONS.

In this example, one has

$$\begin{aligned} \mathcal{F}_B(m) &= \{X_4, X_6, X_9\} \text{ and } \mathcal{F}_{\bar{B}}(m) = \{X_5\}, \\ \mathcal{F}_{A_1}(m) &= \{X_4\} \text{ and } \mathcal{F}_{\bar{A}_1}(m) = \{X_5, X_9\}, \\ \mathcal{F}_{A_2}(m) &= \{X_5\} \text{ and } \mathcal{F}_{\bar{A}_2}(m) = \{X_4, X_6, X_8, X_9\}, \\ \mathcal{F}_{A_3}(m) &= \{X_9\} \text{ and } \mathcal{F}_{\bar{A}_3}(m) = \{X_1, X_2, X_4, X_5, X_7\}, \\ \mathcal{F}_{A^*}(m) &= \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \mathcal{F}_{A_2}(m) - \mathcal{F}_{A_3}(m), \\ &= \{X_1, X_2, X_3, X_6, X_7, X_8\}. \end{aligned}$$

Therefore, one has

$$U(A^* \cap B) = \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) = m(X_6) = 0.30.$$

In applying TBT formula (37), one can easily verify

$$\begin{aligned} Bel(B) &= \sum_{i=1, \dots, 3} Bel(B \cap A_i) + U(A^* \cap B) \\ &= 0.04 + 0 + 0.05 + 0.30 = 0.39. \end{aligned}$$

### D. Generalization of TBT

As explained in Section II-B, we have to work in Cartesian product space  $\Theta = \Theta_1 \times \Theta_2$  if the partition  $\{A_1, \dots, A_k\}$  is related to a given FoD  $\Theta_1$  and  $B$  is a subset of an other FoD  $\Theta_2$ . Because  $\{A_1, \dots, A_k\}$  is a partition of  $\Theta_1$ , then  $\{A_1 \times \Theta_2, \dots, A_k \times \Theta_2\}$  defines a partition of  $\Theta = \Theta_1 \times \Theta_2$  and because  $\Theta_1 \times B = \bigcup_{i=1, \dots, k} ((\Theta_1 \times B) \cap (A_i \times \Theta_2))$ , one can always apply TBT in the Cartesian space  $\Theta$ . More precisely, one has

$$Bel(\Theta_1 \times B) = \sum_{i=1, \dots, k} Bel(A_i \times B) + U(A^* \times B), \quad (40)$$

and where  $U(A^* \times B) \triangleq U((A^* \times \Theta_2) \cap (\Theta_1 \times B))$ .

This formula can be used if and only if one knows the joint BBA  $m(\cdot)$  (or equivalently the joint belief) defined over the powerset of the Cartesian space  $\Theta = \Theta_1 \times \Theta_2$ .

## VII. CONDITIONAL BELIEF FUNCTIONS AND GBT

Before justifying FH conditioning from TBT and presenting the Generalized Bayes' Theorem for BF, we establish a useful lemma.

**Lemma 1:** Consider a FoD  $\Theta$  with a given BBA  $m(\cdot)$  defined over  $\Theta$ , for partition  $\{A_i, \bar{A}_i\}$  of  $\Theta$  and any  $B \subseteq \Theta$ , one always has

$$0 \leq U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \leq 1, \quad (41)$$

where  $U((\bar{A}_i \cap B)^*) = \sum_{X \in \mathcal{F}_{(\bar{A}_i \cap B)^*}(m)} m(X)$  and  $U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X)$ .

**Proof of Lemma 1:** See appendix.

## A. Conditional belief and plausibility

We consider a partition  $\{A_i, \bar{A}_i\}$  of the FoD  $\Theta$  and a subset  $B$  of  $\Theta$ . Using TBT, one has

$$Bel(B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) + U(A^* \cap B). \quad (42)$$

Hence

$$Bel(B) - U(A^* \cap B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B). \quad (43)$$

Moreover, since one has (by definition)

$$U((\bar{A}_i \cap B)^*) = Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B), \quad (44)$$

from the equality (44), one gets

$$Bel(\bar{A}_i \cap B) = Pl(\bar{A}_i \cap B) - U((\bar{A}_i \cap B)^*). \quad (45)$$

Putting the expression of  $Bel(\bar{A}_i \cap B)$  above into (43) and rearranging terms, one gets

$$Bel(B) + \Delta(U) = Bel(A_i \cap B) + Pl(\bar{A}_i \cap B), \quad (46)$$

where  $\Delta(U) \triangleq U((\bar{A}_i \cap B)^*) - U(A^* \cap B)$ , and  $\Delta(U) \in [0, 1]$  because of Lemma 1.

Assuming  $Bel(B) > 0$ , and dividing left and right sides of the equality (46) by  $Bel(B) + \Delta(U)$ , one gets

$$1 = \frac{Bel(A_i \cap B)}{Bel(B) + \Delta(U)} + \frac{Pl(\bar{A}_i \cap B)}{Bel(B) + \Delta(U)}. \quad (47)$$

Hence, the equality (47) suggests to define the conditional belief  $Bel(A_i|B)$  and  $Pl(A_i|B)$  as follows

$$Bel(A_i|B) \triangleq Bel(A_i \cap B) / (Bel(B) + \Delta(U)), \quad (48)$$

$$Pl(\bar{A}_i|B) \triangleq Pl(\bar{A}_i \cap B) / (Bel(B) + \Delta(U)). \quad (49)$$

Using equality (46), the previous conditioning formulas can be rewritten more concisely as

$$Bel(A_i|B) = Bel(A_i \cap B) / (Bel(A_i \cap B) + Pl(\bar{A}_i \cap B)), \quad (50)$$

$$Pl(\bar{A}_i|B) = Pl(\bar{A}_i \cap B) / (Bel(A_i \cap B) + Pl(\bar{A}_i \cap B)). \quad (51)$$

Replacing  $\bar{A}_i$  by  $A_i$  in notations of formulas (49)–(51) we get<sup>6</sup> the following expressions for conditional plausibility  $Pl(A_i|B)$

$$Pl(A_i|B) \triangleq \frac{Pl(A_i \cap B)}{Bel(B) + U((A_i \cap B)^*) - U(A^* \cap B)}, \quad (52)$$

$$Pl(A_i|B) = \frac{Pl(A_i \cap B)}{Bel(\bar{A}_i \cap B) + Pl(A_i \cap B)}. \quad (53)$$

Formulas (50) and (53) coincide with FH formulas [4] originally proposed from a very good intuition. In this work, we derive them only from TBT by a direct constructive manner. Note that  $Bel(A_i|B)$  given in (48) satisfies  $Bel(\emptyset|B) = 0$ ,  $Bel(\Theta|B) = 1$ ,

<sup>6</sup>It is worth to note that one has always  $U(A^* \cap B) = \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) = U(\bar{A}^* \cap B)$  because  $\mathcal{F}_{A^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{A}_i}(m) - \mathcal{F}_{\bar{A}_i}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{A}_i}(m) - \mathcal{F}_{A_i}(m) = \mathcal{F}_{\bar{A}^*}(m)$ .

and  $Bel(A_i|B) \in [0, 1]$  conditions. To prove that  $Bel(A_i|B)$  defined by (50) is a belief function one must also prove that it is an  $n$ -monotone ( $n \geq 2$ ) Choquet's capacity [24] on the finite set  $\Theta$ , or equivalently that the condition B3 of Theorem 1 holds for  $Bel(\cdot|B)$ . The proof of B3 is difficult, but three different proofs have been already given by Fagin and Halpern [3], Jaffray [6], and Sundberg and Wagner [7], the latter one being the clearest of fashion.

## B. Generalization of Bayes' Theorem

Starting from (48) with  $\Delta(U) \triangleq U((\bar{A}_i \cap B)^*) - U(A^* \cap B)$  and replacing  $Bel(B)$  by the expression (37) of TBT, we get

$$Bel(A_i|B) = \frac{Bel(A_i \cap B)}{\sum_{i=1, \dots, k} Bel(A_i \cap B) + U((\bar{A}_i \cap B)^*)}. \quad (54)$$

Similarly, in assuming  $Bel(A_i) > 0$ , Fagin-Halpern expression of  $Bel(B|A_i)$  given by

$$Bel(B|A_i) = \frac{Bel(B \cap A_i)}{Bel(B \cap A_i) + Pl(\bar{B} \cap A_i)}, \quad (55)$$

is equivalent to the formula

$$Bel(B|A_i) = \frac{Bel(B \cap A_i)}{Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i)}, \quad (56)$$

where

$$U((\bar{B} \cap A_i)^*) \triangleq Pl(\bar{B} \cap A_i) - Bel(\bar{B} \cap A_i) \quad (57)$$

$$= \sum_{X \in \mathcal{F}_{(\bar{B} \cap A_i)^*}(m)} m(X), \quad (58)$$

with  $\mathcal{F}_{(\bar{B} \cap A_i)^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{B} \cap A_i}(m) - \mathcal{F}_{B \cup \bar{A}_i}(m)$ , and where

$$U(B^* \cap A_i) \triangleq \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{A_i}(m)} m(X), \quad (59)$$

with  $\mathcal{F}_{B^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_B(m) - \mathcal{F}_{\bar{B}}(m)$ .

From (56), one obtains

$$Bel(A_i \cap B) = Bel(B|A_i)[Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i)].$$

Replacing the above expression of  $Bel(A_i \cap B)$  into the formula (54), we obtain the formula

$$Bel(A_i|B) = \frac{Bel(B|A_i)q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i)q(A_i, B) + U((\bar{A}_i \cap B)^*)}, \quad (60)$$

where the factor  $q(A_i, B)$  introduced here for notation conciseness is defined by

$$q(A_i, B) \triangleq Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i). \quad (61)$$

This allows to establish the Generalized Bayes' Theorem (GBT).

**Generalized Bayes' Theorem (GBT):** For any partition  $\{A_1, \dots, A_k\}$  of a FoD  $\Theta$ , any belief function  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$ , and any subset  $B$  of  $\Theta$  with  $Bel(B) > 0$ , then one has for  $i \in \{1, \dots, k\}$

$$Bel(A_i|B) = \frac{Bel(B|A_i)q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i)q(A_i, B) + U((\bar{A}_i \cap B)^*)}, \quad (62)$$

$$U((\bar{A}_i \cap B)^*) \triangleq Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B),$$

where  $U((\bar{A}_i \cap B)^*) \triangleq \sum_{X \in \mathcal{F}_{(\bar{A}_i \cap B)^*}(m)} m(X) = Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B)$ , and where the factor  $q(A_i, B)$  is defined by (61).

**Lemma 2:** GBT reduces to BT if  $Bel(\cdot)$  is a Bayesian BF.

**Proof:** See appendix.

**Remark:** When  $A_i \subseteq \Theta_1$  and  $B \subseteq \Theta_2$  with  $\Theta_1 \neq \Theta_2$ , we must work in the Cartesian product space  $\Theta = \Theta_1 \times \Theta_2$  and the GBT formula is similar to (62) in replacing  $A_i$  by  $A_i \times \Theta_2$ , and  $B$  by  $\Theta_1 \times B$ . The application of GBT formula is not easy in general because it requires the knowledge of joint BBA  $m(\cdot)$  defined over  $2^{\Theta_1 \times \Theta_2}$  which is rarely known in practice. If the joint BBA  $m(\cdot)$  can be expressed (or approximated) as a function of two marginal BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  (assumed to be known) defined respectively over  $\Theta_1$  and  $\Theta_2$ , then GBT formula should become tractable.

**VIII. ILLUSTRATIVE EXAMPLE OF GBT**

Consider  $\Theta = \{\theta_i, i = 1, \dots, 7\}$ ,  $\mathcal{F}_\Theta(m) = \{X_1, X_2, \dots, X_9\}$  and  $m(\cdot)$  given in Table III.

Focal element $X$	BBA $m(X)$
$X_1 = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_7$	$m(X_1) = 0.01$
$X_2 = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$	$m(X_2) = 0.02$
$X_3 = \theta_3 \cup \theta_5 \cup \theta_6$	$m(X_3) = 0.03$
$X_4 = \theta_4 \cup \theta_7$	$m(X_4) = 0.04$
$X_5 = \theta_2$	$m(X_5) = 0.20$
$X_6 = \theta_6 \cup \theta_7$	$m(X_6) = 0.30$
$X_7 = \theta_2 \cup \theta_3 \cup \theta_7$	$m(X_7) = 0.20$
$X_8 = \theta_1 \cup \theta_4 \cup \theta_6$	$m(X_8) = 0.15$
$X_9 = \theta_6$	$m(X_9) = 0.05$

Table III  
FOCAL ELEMENTS AND THEIR MASSES.

Consider the partition  $\{A_1, A_2, A_3\}$  of  $\Theta$  with  $A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$ ,  $A_2 = \theta_2 \cup \theta_5$  and  $A_3 = \theta_6$ , and the subset  $B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$  of  $\Theta$  having belief  $Bel(B) = m(X_4) + m(X_6) + m(X_9) = 0.39$ . Table IV summarizes the BF values which are needed in the derivations.

Subsets $X$ of $\Theta$	$Bel(X)$	$Pl(X)$
$X = B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$	$Bel(B) = 0.39$	$Pl(B) = 0.80$
$X = A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$	$Bel(A_1) = 0.04$	$Pl(A_1) = 0.75$
$X = A_2 = \theta_2 \cup \theta_5$	$Bel(A_2) = 0.20$	$Pl(A_2) = 0.46$
$X = A_3 = \theta_6$	$Bel(A_3) = 0.05$	$Pl(A_3) = 0.53$
$X = A_1 \cap B = \theta_4 \cup \theta_7$	$Bel(X) = 0.04$	$Pl(X) = 0.72$
$X = A_2 \cap B = \theta_5$	$Bel(X) = 0$	$Pl(X) = 0.04$
$X = A_3 \cap B = \theta_6$	$Bel(X) = 0.05$	$Pl(X) = 0.53$
$X = \bar{A}_1 \cap B = \theta_5 \cup \theta_6$	$Bel(X) = 0.05$	$Pl(X) = 0.54$
$X = \bar{A}_2 \cap B = \theta_4 \cup \theta_6 \cup \theta_7$	$Bel(X) = 0.39$	$Pl(X) = 0.80$
$X = \bar{A}_3 \cap B = \theta_4 \cup \theta_5 \cup \theta_7$	$Bel(X) = 0.04$	$Pl(X) = 0.75$
$X = A_1 \cap \bar{B} = \theta_1 \cup \theta_3$	$Bel(X) = 0$	$Pl(X) = 0.41$
$X = A_2 \cap \bar{B} = \theta_2$	$Bel(X) = 0.20$	$Pl(X) = 0.43$
$X = A_3 \cap \bar{B} = \emptyset$	$Bel(X) = 0$	$Pl(X) = 0$

Table IV  
BELIEF AND PLAUSIBILITY VALUES USED FOR THE DERIVATIONS.

• **Results with Fagin-Halpern conditioning formulas**

Using (50) and (55) and the fact that  $Pl(A_i|B) = 1 - Bel(\bar{A}_i|B)$  and  $Pl(B|A_i) = 1 - Bel(\bar{B}|A_i)$ , we get the values of Tables V–VI.

Subsets of $\Theta$	$Bel(A_i B)$	$Pl(A_i B)$
$A_1$	$Bel(A_1 B) \approx 0.0690$	$Pl(A_1 B) \approx 0.9351$
$A_2$	$Bel(A_2 B) = 0$	$Pl(A_2 B) \approx 0.0930$
$A_3$	$Bel(A_3 B) \approx 0.0625$	$Pl(A_3 B) \approx 0.9298$

Table V  
 $Bel(A_i|B)$  AND  $Pl(A_i|B)$  WITH FAGIN-HALPERN CONDITIONING.

Subsets of $\Theta$	$Bel(B A_i)$	$Pl(B A_i)$
$A_1$	$Bel(B A_1) \approx 0.0889$	$Pl(B A_1) = 1$
$A_2$	$Bel(B A_2) = 0$	$Pl(B A_2) \approx 0.1667$
$A_3$	$Bel(B A_3) = 1$	$Pl(B A_3) = 1$

Table VI  
 $Bel(B|A_i)$  AND  $Pl(B|A_i)$  WITH FAGIN-HALPERN CONDITIONING.

To verify GBT, one calculates  $Bel(A_i)$ ,  $U((\bar{B} \cap A_i)^*)$  and  $U(B^* \cap A_i)$  for getting  $q(A_i, B)$ , and  $U((\bar{A}_i \cap B)^*)$ . These values are given in Table VII.  $q(A_1, B) = 0.45$  is calculated by  $q(A_1, B) \triangleq Bel(A_1) + U((\bar{B} \cap A_1)^*) - U(B^* \cap A_1) = 0.45$  because  $Bel(A_1) = 0.04$ ,  $U((\bar{B} \cap A_1)^*) = Pl(\bar{B} \cap A_1) - Bel(\bar{B} \cap A_1) = 0.41$  and  $U(B^* \cap A_1) = \sum_{X \in \mathcal{F}_{A_1}(m) | X \in \mathcal{F}_{B^*}(m)} m(X) = 0$ .  $U((\bar{A}_1 \cap B)^*) = 0.49$  is calculated by  $U((\bar{A}_1 \cap B)^*) = Pl(\bar{A}_1 \cap B) - Bel(\bar{A}_1 \cap B) = 0.54 - 0.05 = 0.49$ , and other values of Table VII are calculated similarly.

Subsets of $\Theta$	$q(A_i, B)$	$U((\bar{A}_i \cap B)^*)$
$A_1$	0.45	0.49
$A_2$	0.43	0.41
$A_3$	0.05	0.71

Table VII  
VALUES OF  $q(A_i, B)$  AND  $U((\bar{A}_i \cap B)^*)$  FOR GBT FORMULA.

One verifies that GBT formula (62) works because we retrieve correct values obtained with FH formula. Indeed, one has

$$\begin{aligned}
 Bel(A_1|B) &= \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)} \\
 &\approx \frac{0.0889 \cdot 0.45}{(0.0889 \cdot 0.45) + (0 \cdot 0.43) + (1 \cdot 0.05) + 0.49} \\
 &\approx 0.0690.
 \end{aligned}$$

Similarly, one can easily verify that one obtains  $Bel(A_2|B) = 0$  and  $Bel(A_3|B) \approx 0.0625$  with GBT.

• **Results with Shafer’s conditioning formulas**

With formulas (22)–(23), we get the values of Tables VIII–IX.

Subsets of $\Theta$	$Bel(A_i B)$	$Pl(A_i B)$
$A_1$	$Bel(A_1 B) = 0.3250$	$Pl(A_1 B) = 0.9000$
$A_2$	$Bel(A_2 B) = 0$	$Pl(A_2 B) = 0.0500$
$A_3$	$Bel(A_3 B) = 0.0625$	$Pl(A_3 B) = 0.6625$

Table VIII  
 $Bel(A_i|B)$  AND  $Pl(A_i|B)$  WITH SHAFER’S CONDITIONING.

Subsets of $\Theta$	$Bel(B A_i)$	$Pl(B A_i)$
$A_1$	$Bel(B A_1) \approx 0.4533$	$Pl(B A_1) \approx 0.9600$
$A_2$	$Bel(B A_2) \approx 0.0652$	$Pl(B A_2) \approx 0.0870$
$A_3$	$Bel(B A_3) = 1$	$Pl(B A_3) = 1$

Table IX  
 $Bel(B|A_i)$  AND  $Pl(B|A_i)$  WITH SHAFER’S CONDITIONING.

One sees that the conditional values are not coherent since they do not verify GBT because we obtain in this example

$$\begin{aligned}
 Bel(A_1|B) &= 0.3250 \text{ (using (23))} \\
 &\neq \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)} \\
 &\approx \frac{0.4533 \cdot 0.45}{(0.4533 \cdot 0.45) + (0.0652 \cdot 0.43) + (1 \cdot 0.05) + 0.49} \\
 &\approx 0.2642.
 \end{aligned}$$

Similarly, one can show that  $Bel(A_2|B) = 0$  (using (23))  $\neq 0.0405$  (using GBT) and  $Bel(A_3|B) = 0.0625$  (using (23))  $\neq 0.0504$  (using GBT). Hence, Ellsberg urn example and this example show clearly that Dempster’s rule of combination used by Shafer to establish his belief conditioning formulas does not provide coherent and satisfactory results since they are inconsistent with lower and upper bounds of imprecise conditional probabilities and they do not satisfy GBT established directly by a constructive manner without ad-hoc assumption.

## IX. CONCLUSION

This paper has presented new important results: the Total Belief Theorem (TBT), the justification of Fagin-Halpern conditioning from TBT, and the Generalized Bayes' Theorem (GBT). Our theoretical results allowed us to establish rigorously the Generalized Bayes' Theorem by a direct constructive manner from TBT. It does not need extra assumptions nor some underlying ad-hoc construction principles. Also, we prove that our TBT and GBT are fully consistent with classical TPT and Bayes Theorem as soon as the belief functions are Bayesian. That way this achievement could be an excellent ground for working in belief function framework. From Ellsberg's urn example and an illustrative example we have shown that Shafer's conditioning based on Dempster's rule provides results inconsistent with lower and upper bounds of imprecise conditional probabilities, and inconsistent with GBT. These new results should allow to reconcile practitioners of Bayesian reasoning with those of evidential reasoning.

## APPENDIX

## A. Proof of TBT

$$\begin{aligned}
 Bel(B) &= \sum_{X \in \mathcal{F}_\Theta(m) | X \subseteq B} m(X) \\
 &= \sum_{X \in \mathcal{F}_{A_1}(m) | X \in \mathcal{F}_B(m)} m(X) + \dots \\
 &\quad + \sum_{X \in \mathcal{F}_{A_k}(m) | X \in \mathcal{F}_B(m)} m(X) \\
 &\quad + \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) \\
 &= Bel(A_1 \cap B) + \dots + Bel(A_k \cap B) \\
 &\quad + \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) \\
 &= \sum_{i=1, \dots, k} Bel(A_i \cap B) + U(A^* \cap B),
 \end{aligned}$$

where  $U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X)$ .

## B. Proof of Lemma 1

For notation convenience, we denote

$$\begin{aligned}
 \Delta(U) &\triangleq U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \\
 &= [Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B)] \\
 &\quad - [Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) - Bel(B)] \\
 &= Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) + Bel(B) \\
 &\quad - Bel(A_i \cap B) - Bel(\bar{A}_i \cap B).
 \end{aligned}$$

To prove that  $\Delta(U) \geq 0$ , one needs to prove equivalently that  $Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) + Bel(B) \geq Bel(A_i \cap B) + Bel(\bar{A}_i \cap B)$ . Using TBT, one has  $Bel(B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) + U(A^* \cap B)$ , and replacing expression of  $Bel(B)$  in the previous inequality, one must verify if the following equality is satisfied

$$\begin{aligned}
 Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) + Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) + U(A^* \cap B) \\
 \geq Bel(A_i \cap B) + Bel(\bar{A}_i \cap B).
 \end{aligned}$$

After simplification, we have to check if inequality below holds

$$Pl(\bar{A}_i \cap B) + U(A^* \cap B) \geq Bel(\bar{A}_i \cap B).$$

Because  $Pl(\bar{A}_i \cap B) = Bel(\bar{A}_i \cap B) + U((\bar{A}_i \cap B)^*)$ , one has to check if  $Bel(\bar{A}_i \cap B) + U((\bar{A}_i \cap B)^*) + U(A^* \cap B) \geq Bel(\bar{A}_i \cap B)$ . After simplification (omitting both  $Bel(\bar{A}_i \cap B)$  in left and right

side of the previous inequality), one just has to prove the inequality  $U((\bar{A}_i \cap B)^*) + U(A^* \cap B) \geq 0$  in order to prove that  $\Delta(U) \geq 0$ . Because  $U((\bar{A}_i \cap B)^*) \in [0, 1]$  and  $U(A^* \cap B) \in [0, 1]$ , the previous inequality always holds which proves that  $U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \geq 0$ . Moreover because  $U(A^* \cap B) \in [0, 1]$ , then  $-U(A^* \cap B) \in [-1, 0]$ , and because  $U((\bar{A}_i \cap B)^*) \in [0, 1]$  one deduces that  $\Delta(U) = U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \leq 1$ .

## C. Proof of Lemma 2

If  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$  is a Bayesian belief function, then all focal elements of its corresponding BBA  $m(\cdot)$  are singletons of  $2^\Theta$ . In this case  $Bel(\cdot)$  and  $Pl(\cdot)$  functions coincide and therefore one has  $U((\bar{A}_i \cap B)^*) = Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) = 0$  and  $U((\bar{B} \cap A_i)^*) = Pl(\bar{B} \cap A_i) - Bel(\bar{B} \cap A_i) = 0$ . Any focal element (singleton) of  $m(\cdot)$  is either a subset of  $B$  or a subset of  $\bar{B}$  of the FoD  $\Theta$ . Therefore,  $\mathcal{F}_{B^*}(m) = \emptyset$ , which implies  $U(B^* \cap A_i) = 0$ , so that  $q(A_i, B) = Bel(A_i)$ . The GBT formula (62) with in this case  $q(A_i, B) = Bel(A_i)$  and  $U((\bar{A}_i \cap B)^*) = 0$  reduces to the formula  $Bel(A_i | B) = Bel(B | A_i) Bel(A_i) / \sum_{i=1}^k Bel(B | A_i) Bel(A_i)$ . This coincides with formula (10) since  $Bel(\cdot)$  (being a Bayesian belief function) is homogeneous to a probability measure  $P(\cdot)$ . This completes the proof that GBT formula is consistent with Bayes' Theorem formula when the Belief function is Bayesian.

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Press, 1976.
- [2] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., (38): 325–339, 1967.
- [3] R. Fagin, J.Y. Halpern, *A new approach to updating beliefs*, UAI Conf. Proc., pp. 317–325, 1991.
- [4] J.Y. Halpern, R. Fagin, *Two views of belief: belief as generalized probability and belief as evidence*, Art. Intel., (54):275–317, 1992.
- [5] J.Y. Halpern, *Reasoning about uncertainty*, MIT Press, 2003.
- [6] J.-Y. Jaffray, *Bayesian updating and belief functions*, IEEE Trans. on SMC, (22):1144–1152, 1992.
- [7] C. Sunberg, C. Wagner, *Generalized finite differences and Bayesian conditioning of Choquet capacities*, Adv. in Appl. Math., (13), 1992.
- [8] P. Diaconis, S.L. Zabell, *Updating subjective probability*, J. of Amer. Stat. Soc., (77):822–830, 1982.
- [9] L.A. Zadeh, *A Mathematical Theory of Evidence (book review)*, AI Magazine, (5):81–83, 1984.
- [10] J. Lemmer, *Confidence factors, empiricism and the Dempster-Shafer theory of evidence*, Proc. of 1st UAI Conf., pp. 16–176, 1985.
- [11] P.K. Black, *Is Shafer general Bayes?*, 3rd UAI Workshop, USA, 1987.
- [12] P. Wang, *A defect in Dempster-Shafer theory*, UAI Conf. Proc., 1994.
- [13] A. Gelman, *The boxer, the wrestler, and the coin flip: a paradox of robust Bayesian inference and belief functions*, Amer. Stat. (60):146–150, 2006.
- [14] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan 1988.
- [15] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Fusion 2012 Proc., Singapore, July 9–12, 2012.
- [16] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., (29):223–252, 2014 (erratum in this DSMT Book Vol. 5).
- [17] P. Smets, *Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem*, IJAR, Vol. 9, No. 1, pp. 1–35, 1993.
- [18] D. Dubois, T. Denoeux, *Conditioning in Dempster-Shafer Theory: Prediction vs. Revision*, Belief 2012 Conf. Proc., Compiègne, May 2012.
- [19] A.P. Dempster, *A generalization of Bayesian inference*, J. of Royal Stat. Soc., (B30):205–247, 1968.
- [20] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. New York: McGraw-Hill, pp. 37–38, 1984.
- [21] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's rule doesn't behave as Bayes rule with informative priors*, Proc. of 2013 IEEE INISTA Int. Conf., Albena, Bulgaria, June 2013.
- [22] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's fusion rule is not a generalization of Bayes fusion rule*, Proc. of Fusion 2013.
- [23] D. Ellsberg, *Risk, ambiguity, and the Savage axioms*, Quart. J. Econ. (75):643–669, 1961.
- [24] G. Choquet, *Theory of capacities*, Ann. Inst. Fourier, (5):131–295, 1953.
- [25] J.N. Heendani et al., *A generalization of Bayesian inference in the Dempster-Shafer belief theoretic framework*, Proc. of Fusion 2016.

# Total Belief Theorem and Conditional Belief Functions

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Conditional Belief Functions*, Int. Journal of Intelligent Systems, pp. 1–27, July 2018, and reprinted with permission.

**Abstract**—In this paper new theoretical results for reasoning with belief functions are obtained and discussed. After a judicious decomposition of the set of focal elements of a belief function, we establish the Total Belief Theorem (TBT) which is the direct generalization of the Total Probability Theorem when working in the framework of belief functions. The TBT is also generalized for dealing with different frames of discernments thanks to Cartesian product space. From TBT, we can derive and define formally the expressions of conditional belief functions which are consistent with the bounds of imprecise conditional probability. This work provides a direct establishment and solid justification of Fagin-Halpern belief conditioning formulas. The well-known Bayes' Theorem of Probability Theory is then generalized in the framework of belief functions and we illustrate it with an example at the end of this paper.

**Keywords:** Total Belief Theorem (TBT), conditional belief functions.

## I. INTRODUCTION

In this paper, we present new theoretical results for reasoning with belief functions (BF) introduced by Shafer in [1], known as Dempster-Shafer Theory (DST) in the literature. The first result is the establishment of the Total Belief Theorem (TBT) which can be interpreted as a generalization of the Total Probability Theorem (TPT) for the belief functions framework. TBT is essential for formally establishing conditional belief functions in a constructive manner whose expressions are consistent with original Dempster's idea (through eq. (4.8) in [2]), rediscovered independently and popularized by Fagin-Halpern in [3], [4]. TBT also allows us to present a new formulation of Generalized Bayes' Theorem (GBT).

Several methods have been proposed in the literature to address the belief conditioning problem. They essentially can be separated in two different approaches: 1) Shafer's belief conditioning method based on Dempster's rule of combination [1], and 2) the belief conditioning method consistent with imprecise probability calculus bounds [2], [3], [5]–[8] based on the lower and upper probability interpretation of belief functions.

Although Shafer's belief functions offer an appealing mathematical framework for modeling epistemic uncertainty, their use and the validity of the results obtained in the applications are very controversial both for uncertain information fusion as well as for belief conditioning mainly due to Shafer's

choice of Dempster's rule of combination as a pillar for combining evidences represented by belief functions and for conditioning. These well known problems of DST have already been reported and discussed by many experts in the fields over the last decades, see for example [9]–[23]. That is why in this paper we focus on the second approach of belief conditioning based on the lower and upper probability interpretation of BF.

It is worth noting that Smets in nineties [25] did propose a preliminary version of GBT to generalize Bayes' Theorem (BT) to belief functions but Smets' GBT is based on conditional embedding, conjunctive merging and Shafer's conditioning which make it quite complicate to apply and whose results have been cast in doubt in [26]. Here we propose a simpler and direct constructive manner to derive a new version of GBT without need of extra assumptions of some underlying ad-hoc principles as done by Smets. Of course, we prove that our TBT and GBT presented in this work are fully consistent with classical TPT and BT as soon as the belief functions are restricted to Bayesian belief functions (i.e. classical probability measures).

This paper is organized as follows. After a brief recall of basics of belief functions in Section II and Total Probability Theorem in Section III, we present probability conditioning and Bayes' theorem in Section IV followed by classical Shafer's and Fagin-Halpern's belief conditioning methods respectively in Sections V and VI. In Section VII, we present the decomposition of the set of focal elements of any basic belief assignment that allows us to establish formally the Total Belief Theorem and its generalization on Cartesian product space. The Section VIII presents and justifies the new belief conditioning formulas drawn from TBT which are fully consistent with Fagin-Halpern conditioning formulas. Section IX presents the generalization of Bayes' theorem in the framework of belief functions obtained from TBT. We illustrate our new theoretical results with a quite simple GBT example in Section X to show how to make derivations of GBT and to prove that Shafer's conditioning results are inconsistent with GBT. Section XI concludes this paper.

## II. BASICS OF BELIEF FUNCTIONS

Belief functions (BF) have been introduced by Shafer in [1] to model epistemic uncertainty based on preliminary works

done by Dempster [2], [27]. Shafer's Theory of Belief Functions is also referred as Dempster-Shafer Theory (DST) in the literature. We assume that the answer<sup>1</sup> of the problem under concern belongs to a known (or given) finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all elements of  $\Theta$  are exhaustive and exclusive<sup>2</sup>. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . A basic belief assignment (BBA) associated with a given source of evidence is defined as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying the conditions  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is called the mass of  $A$  committed by the source of evidence. Belief and plausibility functions are respectively defined by

$$Bel(A) = \sum_{\substack{X \in 2^\Theta \\ X \subseteq A}} m(X), \quad (1)$$

$$Pl(A) = \sum_{\substack{X \in 2^\Theta \\ X \cap A \neq \emptyset}} m(X) = 1 - Bel(\bar{A}). \quad (2)$$

where<sup>3</sup>  $\bar{A} \triangleq \Theta - \{A\} = \{X | X \in \Theta \text{ and } X \not\subseteq A\}$ , i.e.  $\bar{A}$  is the complement of  $A$  in  $\Theta$ . The notation  $\triangleq$  means *equal by definition*. The width  $Pl(A) - Bel(A)$  of the belief interval  $[Bel(A), Pl(A)]$  is usually called the *uncertainty on A* committed by the source of evidence, and will be denoted<sup>4</sup> by  $U(A^*)$ . It represents in fact the imprecision on the probability of  $A$  granted by the source of evidence, which provides the BBA  $m(\cdot)$ .

A focal element  $X$  of a BBA  $m(\cdot)$  is an element of  $2^\Theta$  such that  $m(X) > 0$ . Note that the empty set  $\emptyset$  is not a focal element of a BBA because  $m(\emptyset) = 0$  (close-world assumption of Shafer's model for the FoD). The set of all focal elements of  $m(\cdot)$  is denoted

$$\begin{aligned} \mathcal{F}_\Theta(m) &= \{X \subseteq \Theta | m(X) > 0\} \\ &= \{X \in 2^\Theta | m(X) > 0\}. \end{aligned} \quad (3)$$

Because  $m(\emptyset) = 0$ , one always has  $1 \leq |\mathcal{F}_\Theta(m)| \leq 2^{|\Theta|} - 1$ . The set of focal elements of  $m(\cdot)$  included in a subset  $A$  of  $\Theta$  is denoted

$$\mathcal{F}_A(m) \triangleq \{X \subseteq A \subseteq \Theta | m(X) > 0\} \quad (4)$$

$$= \{X \in \mathcal{F}_\Theta(m) | X \cap A = X\}. \quad (5)$$

Note that if  $A \subseteq B \subseteq \Theta$ , then  $\mathcal{F}_A(m) \subseteq \mathcal{F}_B(m)$ , and one always has<sup>5</sup>  $\mathcal{F}_A(m) \cap \mathcal{F}_B(m) = \mathcal{F}_{A \cap B}(m)$  for any subsets  $A$  and  $B$  of  $\Theta$ , but  $\mathcal{F}_{A \cup B}(m) \neq \mathcal{F}_A(m) \cup \mathcal{F}_B(m)$  in general<sup>6</sup>.

<sup>1</sup>i.e. the solution, or the decision to take.

<sup>2</sup>This is so-called Shafer's model of FoD [28].

<sup>3</sup>Here the minus symbol denotes the set difference operator [29], [30].

<sup>4</sup>In the literature it is usually denoted by  $U(A)$ . Here we use a new notation  $U(A^*)$  which is not anecdotic. This new notation reveals its importance for the consistency of notations used in formulas we give in this paper.

<sup>5</sup>Proof:  $\mathcal{F}_A(m) \cap \mathcal{F}_B(m) = \{X \in \mathcal{F}_\Theta(m) | (X \cap A) \cap (X \cap B) = X\} = \{X \in \mathcal{F}_\Theta(m) | X \cap (A \cap B) = X\} = \mathcal{F}_{A \cap B}(m)$ .

<sup>6</sup>For example, consider the focal elements given in the example of section X. One has  $A_1 \cup \bar{B} = \{\theta_1, \theta_3, \theta_4, \theta_7\} \cup \{\theta_1, \theta_2, \theta_3\} = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_7\}$  and therefore  $\mathcal{F}_{A_1 \cup \bar{B}} = \{X_2, X_4, X_5, X_7\}$ , but  $\mathcal{F}_{A_1} = \{X_4\}$  and  $\mathcal{F}_{\bar{B}} = \{X_5\}$ , so that  $\mathcal{F}_{A_1} \cup \mathcal{F}_{\bar{B}} = \{X_4, X_5\} \neq \mathcal{F}_{A_1 \cup \bar{B}}$ .

By definition, all elements of  $2^\Theta$  not in  $\mathcal{F}_\Theta(m)$  have a zero mass value, and therefore the definition of  $Bel(A)$  and  $Pl(A)$  given in (1)–(2) can also be expressed<sup>7</sup>

$$Bel(A) = \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \subseteq A}} m(X) = \sum_{X \in \mathcal{F}_A(m)} m(X), \quad (6)$$

$$Pl(A) = \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \cap A \neq \emptyset}} m(X) = 1 - Bel(\bar{A}). \quad (7)$$

The set of focal elements  $\mathcal{F}_\Theta(m)$  of the BBA  $m(\cdot)$  can always been partitioned as  $\{\mathcal{F}_A(m), \mathcal{F}_{\bar{A}}(m), \mathcal{F}_{A^*}(m)\}$  where

$$\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_A(m) - \mathcal{F}_{\bar{A}}(m) \quad (8)$$

$$= \{X \in \mathcal{F}_\Theta(m) | X \cap A \neq \emptyset \text{ and } X \cap \bar{A} \neq \emptyset\}, \quad (9)$$

represents the set of focal elements of  $m(\cdot)$  which are not subsets of  $A$  and not subsets of  $\bar{A} = \Theta - \{A\}$ .

The uncertainty  $U(A^*)$  can also be expressed directly as

$$U(A^*) = \sum_{X \in \mathcal{F}_{A^*}(m)} m(X). \quad (10)$$

It is worth noting that  $U(\bar{A}^*) = Pl(\bar{A}) - Bel(\bar{A}) = (1 - Bel(A)) - (1 - Pl(A)) = Pl(A) - Bel(A) = U(A^*)$ , or equivalently

$$U(\bar{A}^*) = \sum_{X \in \mathcal{F}_{\bar{A}^*}(m)} m(X), \quad (11)$$

where  $\mathcal{F}_{\bar{A}^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{A}}(m) - \mathcal{F}_A(m) = \mathcal{F}_{A^*}(m)$ .

When all elements of  $\mathcal{F}_\Theta(m)$  are only singletons,  $m(\cdot)$  is called a *Bayesian BBA* [1] and its corresponding  $Bel(\cdot)$  and  $Pl(\cdot)$  functions are homogeneous to a same (subjective) probability measure  $P(\cdot)$ .

The class of belief functions can be characterized without explicitly referencing to a BBA, see Shafers' theorem in [1] page 39, with its proof on page 51. More precisely, a mapping  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$  is a belief function if and only if  $Bel(\emptyset) = 0$ ,  $Bel(\Theta) = 1$  and for every positive integer  $n$  and every collection  $A_1, \dots, A_n$  of subsets of  $\Theta$

$$Bel(A_1 \cup \dots \cup A_n) \geq \sum_{\substack{I \subseteq \{1, \dots, n\} \\ I \neq \emptyset}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i). \quad (12)$$

There is a one-to-one relationship between a BBA  $m(\cdot)$  and its corresponding belief function  $Bel(\cdot)$ . The BBA  $m(\cdot)$  that produces a given belief function is unique and is obtained for any  $A \subseteq \Theta$  by the following Möbius inverse formula (see [1], p.39)

$$m(A) = \sum_{B \subseteq A \subseteq \Theta} (-1)^{|A-B|} Bel(B). \quad (13)$$

<sup>7</sup>More precisely, we should write  $Bel(A) = 0 + \sum_{X \in \mathcal{F}_\Theta(m)} m(X)$  to get a well defined value even there is no  $X \in \mathcal{F}_\Theta(m)$  such that  $X \subseteq A$ . For notation convenience, this zero additional term (as well other zero terms in formulas (10)–(11), (42), etc) will be omitted in the sequel being understood that *a sum of non existing terms is always equal to zero*.

In DST framework, Shafer [1] did propose to combine  $s \geq 2$  distinct sources of evidence represented by BBAs  $m_1(\cdot), \dots, m_s(\cdot)$  over the same FoD with Dempster's rule (i.e. the normalized conjunctive rule). Discussions on the justification of Dempster's rule with examples can be found in [21]–[23].

### III. TOTAL PROBABILITY THEOREM (TPT)

We recall briefly the Total Probability Theorem because we will present its extension in Belief function framework. In probability theory, the elements  $\theta_i$  of the space  $\Theta$  are experimental outcomes. The subsets of  $\Theta$  are called events and the event  $\{\theta_i\}$  consisting of the single element  $\theta_i$  is an elementary event. The space  $\Theta$  is called the *sure event* and the empty set  $\emptyset$  is the *impossible event*. We assign to each event  $A$  a number  $P(A)$  in  $[0, 1]$ , called the probability of  $A$ , which satisfies the three Kolmogorov's conditions: 1)  $P(\emptyset) = 0$ ; 2)  $P(\Theta) = 1$ ; and 3) if  $A \cap B = \{\emptyset\}$ , then  $P(A \cup B) = P(A) + P(B)$ . These conditions are the axioms of the theory of probability [30], [31]. The fundamental Theorem of the probability theory is the following Total Probability Theorem (TPT), also called a the law of total probability, see [31] and Theorem 1B of [32].

**Total Probability Theorem (TPT):** Consider an event  $B$  and any partition<sup>8</sup>  $\{A_i, i = 1, \dots, k\}$  of the space  $\Theta$ , then

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k). \quad (14)$$

### IV. CONDITIONAL PROBABILITY AND BAYES' FORMULA

Starting from TPT formula (14) and assuming  $P(B) > 0$ , we get for any  $i \in \{1, \dots, k\}$  after dividing each side of (14) by  $P(B)$  and rearranging terms the equality

$$\frac{P(A_i \cap B)}{P(B)} = 1 - \sum_{\substack{j=1, \dots, k \\ j \neq i}} \frac{P(A_j \cap B)}{P(B)} = 1 - \frac{P(\bar{A}_i \cap B)}{P(B)}, \quad (15)$$

which allows us to define the conditional probability  $P(A_i|B)$  by<sup>9</sup>

$$P(A_i|B) \triangleq P(A_i \cap B)/P(B). \quad (16)$$

Similarly, by considering an event  $A_i$  of  $\Theta$  and the partition  $\{B, \bar{B}\}$  of  $\Theta$ , the TPT formula  $P(A_i) = P(A_i \cap B) + P(A_i \cap \bar{B})$  applies, and by dividing it by  $P(A_i)$  (assuming  $P(A_i) > 0$ ), we get

$$P(A_i \cap B)/P(A_i) = 1 - P(A_i \cap \bar{B})/P(A_i), \quad (17)$$

which allows us to define also the conditional probability  $P(B|A_i)$  by

$$P(B|A_i) \triangleq P(A_i \cap B)/P(A_i). \quad (18)$$

From (16) and (18), one deduces the equality

$$P(A_i \cap B) = P(A_i|B)P(B) = P(B|A_i)P(A_i). \quad (19)$$

From (19) and assuming  $P(B) > 0$  we get  $P(A_i|B) = P(B|A_i)P(A_i)/P(B)$ , and assuming  $P(A_i) > 0$  we get  $P(B|A_i) = P(A_i|B)P(B)/P(A_i)$ .

<sup>8</sup>A partition of a set  $\Theta$  is a collection of mutually exclusive subsets of  $\Theta$  whose union equals  $\Theta$ .

<sup>9</sup>In probability theory, the notation  $P(A_i, B) \equiv P(A_i \cap B)$  is also used to represent the probability of the joint occurrence (intersection) of events  $A_i$  and  $B$ .

Using TPT formula (14) and noting that  $P(A_i \cap B) = P(B|A_i)P(A_i)$ , we get

$$P(B) = \sum_{i=1}^k P(B|A_i)P(A_i). \quad (20)$$

Substituting (20) in  $P(A_i|B) = P(B|A_i)P(A_i)/P(B)$ , we get the well-known Bayes' Theorem formula (BTF)

$$P(A_i|B) = P(B|A_i)P(A_i) / \sum_{i=1}^k P(B|A_i)P(A_i). \quad (21)$$

It can be easily verified that the conditional probability defined by (16) verifies the three axioms of the Theory of probability [31]: 1)  $P(\emptyset|B) = 0$ , 2)  $P(\Theta|B) = 1$  and 3) if  $A_1 \cap A_2 = \emptyset$ , then  $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$ .

In the previous presentation,  $A_i$  ( $i = 1, \dots, k$ ) and  $B$  are events (subsets) of the same space  $\Theta$ . How to proceed to compute  $P(A_i|B)$  if the events  $A_i$  ( $i = 1, \dots, k$ ) and  $B$  are subsets of different spaces, say if  $A_i \subseteq \Theta_1 = \{x_1, \dots, x_m\} = \{x_p, p = 1, 2, \dots, m\}$  ( $i = 1, \dots, k$ ), and if  $B \subseteq \Theta_2 = \{y_1, \dots, y_n\} = \{y_q, q = 1, 2, \dots, n\}$  with  $\Theta_1 \neq \Theta_2$ ? Such situation corresponds to a so-called combined experiment [31]. In fact, one can prove that similar conditioning formulas can also be established. For this, we need to work with the Cartesian product space  $\Theta \triangleq \Theta_1 \times \Theta_2$  whose elementary elements are all the ordered pairs  $(x_p, y_q)$  with  $x_p \in \Theta_1$  and  $y_q \in \Theta_2$ . The two experiments are viewed as a single combined one whose outcomes are pairs  $(x_p, y_q)$ . In this space  $\Theta = \Theta_1 \times \Theta_2$ ,  $x_p$  is not an elementary element but a subset of  $n$  elements of  $\Theta$ , i.e.  $\{x_p\} = \{(x_p, y_1), \dots, (x_p, y_n)\}$ . Similarly,  $y_q$  is not an elementary element but a subset of  $m$  elements of  $\Theta$ , i.e.  $\{y_q\} = \{(x_1, y_q), \dots, (x_m, y_q)\}$ . If  $A_i \subseteq \Theta_1$  and  $B \subseteq \Theta_2$ , then  $A_i \times B = \{(x_p, y_q) | x_p \in A_i, y_q \in B\} \subseteq \Theta$ . If one forms  $A_i \times \Theta_2$  and  $\Theta_1 \times B$  one sees that  $A_i \times B = (A_i \times \Theta_2) \cap (\Theta_1 \times B) = (\Theta_1 \times B) \cap (A_i \times \Theta_2)$ . Because the event  $A_i \times \Theta_2$  occurs in the combined experiment if the event  $A_i$  of the experiment 1 occurs no matter what the outcome of experiment 2, one has  $P(A_i \times \Theta_2) = P_1(A_i)$  where  $P_1(A_i)$  is the probability of event  $A_i$  in the experiment 1. Similarly, the event  $\Theta_1 \times B$  occurs if  $B$  occurs in experiment 2 no matter what the outcome of experiment 1, so that  $P(\Theta_1 \times B) = P_2(B)$  where  $P_2(B)$  is the probability of event  $B$  in the experiment 2. One considers a partition  $\{A_1, A_2, \dots, A_k\}$  of  $\Theta_1$  and a subset (event)  $B \subseteq \Theta_2$ . Based on set theory and property of Cartesian product, one has

$$\begin{aligned} \Theta_1 \times B &= (\Theta_1 \times B) \cap (\Theta_1 \times \Theta_2) \\ &= (\Theta_1 \times B) \cap ((A_1 \cup A_2 \cup \dots \cup A_k) \times \Theta_2) \\ &= (\Theta_1 \times B) \cap ((A_1 \times \Theta_2) \cup \dots \cup (A_k \times \Theta_2)) \\ &= \cup_i ((\Theta_1 \times B) \cap (A_i \times \Theta_2)). \end{aligned}$$

The elements  $A_i \times \Theta_2$ ,  $i = 1, \dots, k$  being disjoint<sup>10</sup>, one has the following TPT formula

$$\begin{aligned} P(\Theta_1 \times B) &= P(\cup_i ((\Theta_1 \times B) \cap (A_i \times \Theta_2))) \\ &= P((\Theta_1 \times B) \cap (A_1 \times \Theta_2)) + \\ &\quad \dots + P((\Theta_1 \times B) \cap (A_k \times \Theta_2)). \end{aligned} \quad (22)$$

After dividing each side of formula (22) by  $P(\Theta_1 \times B)$  (assumed positive) and rearranging terms, we get

$$\begin{aligned} \frac{P((A_i \times \Theta_2) \cap (\Theta_1 \times B))}{P(\Theta_1 \times B)} &= \\ 1 - \sum_{\substack{j=1, \dots, k \\ j \neq i}} \frac{P((A_j \times \Theta_2) \cap (\Theta_1 \times B))}{P(\Theta_1 \times B)}. \end{aligned} \quad (23)$$

<sup>10</sup>because  $A_i$  are disjoint since  $\{A_1, \dots, A_k\}$  is a partition of  $\Theta_1$ .

Formula (23) suggests naturally to define the conditional probability  $P(A_i \times \Theta_2 | \Theta_1 \times B)$  by

$$P(A_i \times \Theta_2 | \Theta_1 \times B) \triangleq P(A_i \times B) / P(\Theta_1 \times B). \quad (24)$$

Using same reasoning as before and working on Cartesian product space  $\Theta = \Theta_1 \times \Theta_2$ , one can also prove<sup>11</sup> that if  $P(A_i \times \Theta_2) > 0$  one can define

$$P(\Theta_1 \times B | A_i \times \Theta_2) = P(A_i \times B) / P(A_i \times \Theta_2). \quad (25)$$

From (24) and (25), one deduces the equality

$$P(A_i \times \Theta_2 | \Theta_1 \times B) P(\Theta_1 \times B) = P(\Theta_1 \times B | A_i \times \Theta_2) P(A_i \times \Theta_2). \quad (26)$$

From equality (26) and assuming  $P(\Theta_1 \times B) > 0$ , we get

$$P(A_i \times \Theta_2 | \Theta_1 \times B) = \frac{P(\Theta_1 \times B | A_i \times \Theta_2) P(A_i \times \Theta_2)}{P(\Theta_1 \times B)}. \quad (27)$$

From equality (26) and assuming  $P(A_i \times \Theta_2) > 0$  we get

$$P(\Theta_1 \times B | A_i \times \Theta_2) = \frac{P(A_i \times \Theta_2 | \Theta_1 \times B) P(\Theta_1 \times B)}{P(A_i \times \Theta_2)}. \quad (28)$$

Using TPT formula (22) and formula(25), we get  $P(\Theta_1 \times B) = \sum_{i=1}^k P(\Theta_1 \times B | A_i \times \Theta_2) P(A_i \times \Theta_2)$ . Putting this expression in (27), we obtain the Bayes' Theorem formula (BTF) when  $A \subseteq \Theta_1$  and  $B \subseteq \Theta_2$  and  $\Theta_1 \neq \Theta_2$ , which is written as

$$P(A_i \times \Theta_2 | \Theta_1 \times B) = \frac{P(\Theta_1 \times B | A_i \times \Theta_2) P(A_i \times \Theta_2)}{\sum_{i=1}^k P(\Theta_1 \times B | A_i \times \Theta_2) P(A_i \times \Theta_2)}. \quad (29)$$

For notation convenience, we can use classical formulas when working with different sets of experimental outcomes  $\Theta_1$  and  $\Theta_2$  with keeping in mind that in this case  $A_i$  must be understood as  $A_i \times \Theta_2$  and  $B$  as  $\Theta_1 \times B$ .

## V. SHAFER'S CONDITIONING

In the belief functions framework, Shafer did propose formulas to calculate conditional belief functions  $Bel(A|B)$  and  $Pl(A|B)$ . Shafer's formulas have been obtained from the conditional BBA  $m(\cdot|B)$  obtained from Dempster's rule of combination of the original BBA  $m(\cdot)$  with the BBA  $m_B(B) = 1$  focused on  $B$  under the condition that  $Bel(\bar{B}) < 1$ , or equivalently<sup>12</sup> under the condition that  $Pl(B) > 0$ . Shafer's conditioning formulas for belief and plausibility functions were established by Shafer in Theorem 3.6 p. 66 of [1]. For  $A, B \subseteq \Theta$  with  $Pl(B) > 0$ ,  $Bel(A|B)$  and  $Pl(A|B)$  are given by

$$Bel(A|B) = (Bel(A \cup \bar{B}) - Bel(\bar{B})) / (1 - Bel(\bar{B})), \quad (30)$$

$$Pl(A|B) = Pl(A \cap B) / Pl(B). \quad (31)$$

The expression (30) of  $Bel(A|B)$  is equivalent to

$$Bel(A|B) = (Pl(B) - Pl(B \cap \bar{A})) / Pl(B), \quad (32)$$

because one has always (from definition of belief functions)  $Pl(B) = 1 - Bel(\bar{B})$ , and the numerator of (30) can be written as

$$\begin{aligned} Bel(A \cup \bar{B}) - Bel(\bar{B}) &= (1 - Bel(\bar{B})) - (1 - Bel(A \cup \bar{B})) \\ &= Pl(B) - Pl(\overline{A \cup \bar{B}}) \\ &= Pl(B) - Pl(B \cap \bar{A}). \end{aligned}$$

<sup>11</sup>The proof is left to the reader due to space limitation restraint.

<sup>12</sup>Indeed, if  $Bel(\bar{B}) < 1$  then  $Pl(B) = 1 - Bel(\bar{B})$  is greater than zero.

Using (30)–(31) and taking  $A = \emptyset$ , we get  $Bel(\emptyset|B) = Pl(\emptyset|B) = 0$ , and taking  $A = \Theta$  we get  $Bel(\Theta|B) = Pl(\Theta|B) = 1$ . Also in taking  $B = \Theta$  we get  $Bel(A|\Theta) = Bel(A)$  and  $Pl(A|\Theta) = Pl(A)$ . Note that taking  $B = A$  in (31)–(32), we obtain  $Bel(A|A) = Pl(A|A) = 1$  which fits with the common sense.

In reversing the roles played by  $A$  and  $B$  and switching the notations in previous expressions, the following formulas also hold (assuming  $Pl(A) > 0$ )

$$Bel(B|A) = (Pl(A) - Pl(A \cap \bar{B})) / Pl(A), \quad (33)$$

$$Pl(B|A) = Pl(B \cap A) / Pl(A). \quad (34)$$

From (31) and (34), one deduces  $Pl(A \cap B) = Pl(A|B)Pl(B) = Pl(B|A)Pl(A)$ . Hence, the following formula applies for conditional plausibilities when  $Pl(B) > 0$

$$Pl(A|B) = Pl(B|A)Pl(A) / Pl(B). \quad (35)$$

Note that this formula for conditional plausibilities is similar to the expression for conditional probabilities given in (16) when replacing plausibilities by probabilities.

The main drawback of Shafer's conditioning is its incompatibility with probability calculus when working with imprecise probabilities. More precisely, the bounds of belief interval defined by  $[Bel(A|B), Pl(A|B)]$  obtained by (30)–(31) are in general<sup>13</sup> incompatible with lower and upper bounds of the conditional probability  $P(A|B)$ . This problem makes Shafer's conditioning very disputable and cast serious doubts on pertinence (validity) of Shafer's conditioning results when used in applications, which is a direct consequence of the validity of Dempster's rule reported in [3], [9]–[23], [33], [34]. Shafer's conditioning problem has already been reported and addressed by several authors [3], [6], [7], [14], [24] in the past with some examples. To easily show this incompatibility of Shafer's conditioning with probability calculus we present briefly the famous Ellsberg's urn example [35].

**Example 1 (Ellsberg's urn):** We consider an urn with red (R), black (B) and yellow (Y) balls. The a priori information one has on the repartition of the balls in the urn is the following: 1/3 of balls are red balls and 2/3 or balls are black and yellow balls. We don't know precisely the percentage of black balls, nor the percentage of yellow balls. So the a priori information about the chance to pick a ball in the urn can be represented by a (parametric) probability mass function  $P(\cdot)$  with  $P(R) = 1/3$ ,  $P(B) = 2/3 - x$ ,  $P(Y) = x$ , where  $x$  is an unknown number/parameter in  $[0, 2/3]$ ,  $P(R)$  is the probability to pick at random a red ball in the urn,  $P(B)$  is the probability to pick at random a black ball in the urn, and  $P(Y)$  is the probability to pick at random a yellow ball in the urn. Of course because  $x$  is unknown but bounded,  $P(B)$  and  $P(Y)$  are unknown but their bounds are known. In fact, this problem can be seen as a problem of imprecise probabilities where  $P(R) \in [1/3, 1/3]$ ,  $P(B) \in [0, 2/3]$ ,  $P(Y) \in [0, 2/3]$  and with the constraint  $P(R) + P(B) + P(Y) = 1$ . Now let's suppose that someone picks a ball at random in the urn and tell us that the color of the ball is not black, i.e. the event  $\bar{B} = R \cup Y$  has occurred. How do we must revise (update) our prior probabilities with this new information? The correct answer to this question is obtained by computing the conditional probabilities  $P(R|\bar{B})$ ,  $P(B|\bar{B})$  and  $P(Y|\bar{B})$  and by analyzing their bounds. This is done as follows using the fact that  $P(\bar{B}) = P(R \cup Y) = P(R) + P(Y) - P(R \cap Y) =$

<sup>13</sup>but if the BBA  $m(\cdot)$  is Bayesian.

$P(R) + P(Y) = (1/3) + x$ . Indeed,  $P(R \cap Y) = 0$  because the events  $R$  and  $Y$  are mutually exclusive. So, we get

$$\begin{aligned} P(R|\bar{B}) &= P(R \cap (R \cup Y))/P(R \cup Y) \\ &= P(R)/((1/3) + x) = (1/3)/((1/3) + x), \\ P(B|\bar{B}) &= P(B \cap (R \cup Y))/P(R \cup Y) \\ &= P(\emptyset)/((1/3) + x) = 0/((1/3) + x), \\ P(Y|\bar{B}) &= P(Y \cap (R \cup Y))/P(R \cup Y) \\ &= P(Y)/((1/3) + x) = x/((1/3) + x). \end{aligned}$$

If  $x = 0$ , then  $P(R|\bar{B}) = 1$  and  $P(Y|\bar{B}) = 0$ . If  $x = 2/3$ , then  $P(R|\bar{B}) = 1/3$  and  $P(Y|\bar{B}) = 2/3$ . Therefore after conditioning by  $\bar{B} = R \cup Y$  we get as bounds of conditional probabilities values the following intervals  $P(R|\bar{B}) \in [1/3, 1]$ ,  $P(B|\bar{B}) \in [0, 0]$ ,  $P(Y|\bar{B}) \in [0, 2/3]$  with the constraint  $P(R|\bar{B}) + P(B|\bar{B}) + P(Y|\bar{B}) = 1$ .

Let's examine what we get using Shafer's conditioning approach. For this, the problem is modeled directly in the belief function framework using the *a priori* BBA  $m(\cdot)$  defined on the FoD  $\Theta = \{R, B, Y\}$  with  $m(R) = 1/3$ ,  $m(B \cup Y) = 2/3$  which corresponds to the following *a priori* belief intervals  $[Bel(R), Pl(R)] = [1/3, 1/3]$ ,  $[Bel(B), Pl(B)] = [0, 2/3]$ ,  $[Bel(Y), Pl(Y)] = [0, 2/3]$ .

With Shafer's conditioning formulas and noting that  $Pl(R) = 1/3$ ,  $Pl(B) = 2/3$ ,  $Pl(Y) = 2/3$ , and  $Pl(R \cup Y) = 1$ , we get incompatible results with the real bounds of conditional probabilities because

$$\begin{aligned} [Bel(R|\bar{B}), Pl(R|\bar{B})] &= [1/3, 1/3] \text{ (by Shafer)} \\ &\neq [1/3, 1] \text{ (correct bounds),} \\ [Bel(B|\bar{B}), Pl(B|\bar{B})] &= [0, 0] \text{ (by Shafer)} \\ &= [0, 0] \text{ (correct bounds),} \\ [Bel(Y|\bar{B}), Pl(Y|\bar{B})] &= [2/3, 2/3] \text{ (by Shafer)} \\ &\neq [0, 2/3] \text{ (correct bounds).} \end{aligned}$$

To overcome this problem, Fagin and Halpern did propose a more efficient conditioning approach which is, by construction, always consistent with conditional probability bounds. It is presented in the next section.

## VI. FAGIN-HALPERN CONDITIONING

Fagin and Halpern (FH) proposed in [3], [4] to define the conditional belief as the lower envelope (i.e. the infimum) of a family of conditional probability functions to make belief conditioning consistent with imprecise conditional probability calculus. Assuming  $Bel(B) > 0$ , Fagin and Halpern proposed the following conditional formulas (FH formulas for short)

$$Bel(A|B) = Bel(A \cap B)/(Bel(A \cap B) + Pl(\bar{A} \cap B)), \quad (36)$$

$$Pl(A|B) = Pl(A \cap B)/(Pl(A \cap B) + Bel(\bar{A} \cap B)). \quad (37)$$

Fagin and Halpern did prove in [3] with long derivations and great effort that the conditional belief  $Bel(A|B)$  given by (36) satisfies also the three conditions for defining a true belief function according to Shafer's theorem in [1], p. 39. Therefore, the formula (36) is also a good candidate and serious alternative for conditioning belief functions. However, it is quite mysterious how Fagin and Halpern did obtain (construct) these close-form expressions. According to the authors, these expressions were rather established from a very good intuition. A better justification has been given by Sundberg and Wagner in [7] (p. 268) but it is still not so clear in our opinion. In this paper, we justify clearly and directly the establishment of FH formulas from the simple and direct consequence of the Total Belief Theorem (TBT) which is one of the main contributions of our work. From FH conditioning formulas (36)-(37), we can verify that the common sense results are also obtained, that is  $Bel(\emptyset|B) =$

$Pl(\emptyset|B) = 0$ ,  $Bel(\Theta|B) = Pl(\Theta|B) = 1$ ,  $Bel(A|\Theta) = Bel(A)$ ,  $Pl(A|\Theta) = Pl(A)$ , and  $Bel(A|A) = Pl(A|A) = 1$ .

FH conditioning formulas are consistent with Bayes conditioning formulas when the underlying BBA  $m(\cdot)$  is Bayesian. Indeed if  $m(\cdot)$  is Bayesian, then  $Pl(A \cap B) = Bel(A \cap B) = P(A \cap B)$ ,  $Pl(\bar{A} \cap B) = Bel(\bar{A} \cap B) = P(\bar{A} \cap B)$  and  $Pl(\bar{B} \cap A) = Bel(\bar{B} \cap A) = P(\bar{B} \cap A)$  so that the FH formulas become equivalent to  $Bel(A|B) = P(A \cap B)/(P(A \cap B) + P(\bar{A} \cap B))$  and  $Pl(A|B) = P(A \cap B)/(P(A \cap B) + P(\bar{A} \cap B))$ . Thanks to total probability theorem (TPT) formula (14), the denominator involved in these formula is  $P(A \cap B) + P(\bar{A} \cap B) = P(B)$ , therefore  $Bel(A|B) = Pl(A|B) = P(A \cap B)/P(B) = P(A|B)$ .

Similarly, one can also easily verify that  $Bel(B|A) = Pl(B|A) = P(A \cap B)/P(A) = P(B|A)$ . The advantage of FH conditioning is its complete compatibility with the conditional probability calculus [7]. Let us show what provides FH conditioning in the previous Ellsberg's urn example.

**Ellsberg's urn example revisited:** Let's see the result obtained by formulas (63) and (65) for Ellsberg's urn example. Applying formulas (63) and (65) with the conditioning event  $\bar{B} = R \cup Y$  we obtain

$$\begin{aligned} Bel(R|\bar{B}) &= \frac{Bel(R \cap (R \cup Y))}{Bel(R \cap (R \cup Y)) + Pl((B \cup Y) \cap (R \cup Y))} \\ &= \frac{1/3}{(1/3) + (2/3)} = 1/3, \\ Pl(R|\bar{B}) &= \frac{Pl(R \cap (R \cup Y))}{Bel((B \cup Y) \cap (R \cup Y)) + Pl(R \cap (R \cup Y))} \\ &= \frac{1/3}{0 + (1/3)} = 1, \\ Bel(B|\bar{B}) &= \frac{Bel(B \cap (R \cup Y))}{Bel(B \cap (R \cup Y)) + Pl((R \cup Y) \cap (R \cup Y))} \\ &= \frac{0}{0 + 1} = 0, \\ Pl(B|\bar{B}) &= \frac{Pl(B \cap (R \cup Y))}{Bel((R \cup Y) \cap (R \cup Y)) + Pl(B \cap (R \cup Y))} \\ &= \frac{0}{(1/3) + 0} = 0, \\ Bel(Y|\bar{B}) &= \frac{Bel(Y \cap (R \cup Y))}{Bel(Y \cap (R \cup Y)) + Pl((R \cup B) \cap (R \cup Y))} \\ &= \frac{0}{0 + (1/3)} = 0, \\ Pl(Y|\bar{B}) &= \frac{Pl(Y \cap (R \cup Y))}{Bel((R \cup B) \cap (R \cup Y)) + Pl(Y \cap (R \cup Y))} \\ &= \frac{2/3}{(1/3) + (2/3)} = 2/3. \end{aligned}$$

Hence with FH conditioning formulas, we get the correct conditional probability bounds

$$\begin{aligned} [Bel(R|\bar{B}), Pl(R|\bar{B})] &= [1/3, 1] \text{ (by Fagin-Halpern)} \\ &= [1/3, 1] \text{ (correct bounds),} \\ [Bel(B|\bar{B}), Pl(B|\bar{B})] &= [0, 0] \text{ (by Fagin-Halpern)} \\ &= [0, 0] \text{ (correct bounds),} \\ [Bel(Y|\bar{B}), Pl(Y|\bar{B})] &= [0, 2/3] \text{ (by Fagin-Halpern)} \\ &= [0, 2/3] \text{ (correct bounds).} \end{aligned}$$

We can also verify that  $Bel(\emptyset|\bar{B}) = 0$ ,  $Bel(R \cup B|\bar{B}) = 1/3$ ,  $Bel(R \cup Y|\bar{B}) = 1$ ,  $Bel(B \cup Y|\bar{B}) = 0$  and  $Bel(R \cup B \cup Y|\bar{B}) = 1$ . Applying Möbius inverse formula (13) with this conditional belief function  $Bel(\cdot|\bar{B})$ , we get the conditional mass of belief given by  $m(R|\bar{B}) = 1/3$  and  $m(R \cup Y|\bar{B}) = 2/3$  and all other mass

values are equal to zero, whereas with Shafer's approach based on Dempster's rule of combination we get  $m(R|\bar{B}) = 1/3$  and  $m(Y|\bar{B}) = 2/3$ . We see the difference between Shafer's and FH conditioning approaches. With Shafer's conditioning approach, because  $(B \cup Y) \cap (R \cup Y) \neq \emptyset$  the mass  $m(B \cup Y) = 2/3$  is entirely transferred (optimistically) to the most specific focal element  $Y$  included in  $\bar{B} = R \cup Y$ . With the FH conditioning method the mass  $m(B \cup Y) = 2/3$  is entirely transferred (pessimistically, or cautiously) to the least specific focal element  $R \cup Y$  included in  $\bar{B} = R \cup Y$ .

## VII. TOTAL BELIEF THEOREM (TBT)

In this section, we extend TPT theorem to belief and plausibility functions and we establish the Total Belief Theorem (TBT). Before this, we need to explain how the set of focal elements of a given BBA  $m(\cdot)$  must be decomposed because it is the basis of the establishment of TBT.

### A. Decomposition of the set of focal elements $\mathcal{F}_\Theta(m)$

Let us consider a FoD  $\Theta = \{\theta_1, \dots, \theta_{|\Theta|}\}$  with  $|\Theta| > 1$  elements, and a BBA  $m(\cdot)$  defined on  $2^\Theta$  with a given set of focal elements  $\mathcal{F}_\Theta(m)$ . Consider any partition  $\{A_1, A_2, \dots, A_k\}$  of the FoD  $\Theta$ , then one can always decompose  $\mathcal{F}_\Theta(m)$  as the union of following subsets

$$\mathcal{F}_\Theta(m) = \mathcal{F}_{A_1}(m) \cup \dots \cup \mathcal{F}_{A_k}(m) \cup \mathcal{F}_{A^*}(m). \quad (38)$$

where  $\mathcal{F}_{A_i}(m)$  ( $i = 1, \dots, k$ ) is the set of focal elements of  $m(\cdot)$  included in  $A_i$ , and  $\mathcal{F}_{A^*}(m)$  is the set of focal elements of  $m(\cdot)$  which are not included in  $A_i$ ,  $i = 1, \dots, k$ . We use the notation  $A^*$  for representing the entity characterized by the focal set  $\mathcal{F}_{A^*}(m)$  mathematically defined by

$$\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m). \quad (39)$$

The entity  $A^*$  has in general no explicit form and it is used only for notation convenience to make presentation of formulas more concise in the sequel. Because  $A_i$  for  $i = 1, \dots, k$  are mutually exclusive (disjoint), the sets  $\mathcal{F}_{A_i}(m)$  are also mutually exclusive and therefore  $\cap_{i=1, \dots, k} (\mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m)) = \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m)$  because all possible intersections of focal sets including  $\mathcal{F}_{A_i}(m) \cap \mathcal{F}_{A_j}(m)$  for  $j \neq i$  equal the empty set. Hence  $\mathcal{F}_{A^*}(m)$  can also be expressed as

$$\mathcal{F}_{A^*}(m) = \cap_{i=1, \dots, k} \bar{\mathcal{F}}_{A_i}(m), \quad (40)$$

where  $\bar{\mathcal{F}}_{A_i}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m) = \mathcal{F}_{\bar{A}_i}(m) + \mathcal{F}_{A^*}(m)$  represents the set of focal elements of  $m(\cdot)$  which are not subsets of  $A_i$ .

**Example 2:** Consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$  and a BBA  $m(\cdot)$  defined on  $2^\Theta$ , with set of focal elements  $\mathcal{F}_\Theta(m) = \{X_1, X_2, \dots, X_8\}$  with  $X_1 = \theta_1$ ,  $X_2 = \theta_1 \cup \theta_2$ ,  $X_3 = \theta_2 \cup \theta_3$ ,  $X_4 = \theta_3 \cup \theta_4$ ,  $X_5 = \theta_4$ ,  $X_6 = \theta_4 \cup \theta_5$ ,  $X_7 = \theta_1 \cup \theta_3 \cup \theta_5$  and  $X_8 = \theta_5$ . Consider the partition  $\{A_1, A_2, A_3\}$  of  $\Theta$  with  $A_1 = \{\theta_1, \theta_2\}$ ,  $A_2 = \{\theta_3, \theta_4\}$  and  $A_3 = \{\theta_5\}$ . In this example, one has

$$\begin{aligned} \mathcal{F}_{A_1}(m) &= \{X_1, X_2\} = \{\theta_1, \theta_1 \cup \theta_2\}, \\ \mathcal{F}_{A_2}(m) &= \{X_4, X_5\} = \{\theta_3 \cup \theta_4, \theta_4\}, \\ \mathcal{F}_{A_3}(m) &= \{X_8\} = \{\theta_5\}, \\ \mathcal{F}_{A^*}(m) &= \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \mathcal{F}_{A_2}(m) - \mathcal{F}_{A_3}(m) \\ &= \{X_3, X_6, X_7\} \\ &= \{\theta_2 \cup \theta_3, \theta_4 \cup \theta_5, \theta_1 \cup \theta_3 \cup \theta_5\}. \end{aligned}$$

One sees that

$$\begin{aligned} \bar{\mathcal{F}}_{A_1}(m) &= \mathcal{F}_\Theta(m) - \{X_1, X_2\} = \{X_3, X_4, X_5, X_6, X_7, X_8\}, \\ \bar{\mathcal{F}}_{A_2}(m) &= \mathcal{F}_\Theta(m) - \{X_4, X_5\} = \{X_1, X_2, X_3, X_6, X_7, X_8\}, \\ \bar{\mathcal{F}}_{A_3}(m) &= \mathcal{F}_\Theta(m) - \{X_8\} = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7\}. \end{aligned}$$

and applying (40), we get

$$\bar{\mathcal{F}}_{A_1}(m) \cap \bar{\mathcal{F}}_{A_2}(m) \cap \bar{\mathcal{F}}_{A_3}(m) = \{X_3, X_6, X_7\} = \mathcal{F}_{A^*}(m).$$

**Example 3:** Consider  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$  and a BBA  $m(\cdot)$  defined on  $2^\Theta$ , with the degenerate set of focal elements with only one focal element as follows  $\mathcal{F}_\Theta(m) = \{X_1 = \Theta\}$  corresponding to the vacuous BBA. Consider the partition  $\{A_1, A_2, A_3\}$  of  $\Theta$  where  $A_1 \triangleq \{x_3, x_5\}$ ,  $A_2 \triangleq \{x_2\}$  and  $A_3 \triangleq \{x_1, x_4\}$ . Then, we get  $\mathcal{F}_{A_1}(m) = \emptyset$ ,  $\mathcal{F}_{A_2}(m) = \emptyset$ ,  $\mathcal{F}_{A_3}(m) = \emptyset$  and  $\mathcal{F}_{A^*}(m) = \{X_1\} - \emptyset - \emptyset - \emptyset = \Theta$ . Note that,  $\bar{\mathcal{F}}_{A_1}(m) = \bar{\mathcal{F}}_{A_2}(m) = \bar{\mathcal{F}}_{A_3}(m) = \Theta$ , and therefore  $\bar{\mathcal{F}}_{A_1}(m) \cap \bar{\mathcal{F}}_{A_2}(m) \cap \bar{\mathcal{F}}_{A_3}(m) = \Theta = \mathcal{F}_{A^*}(m)$ , and of course  $\mathcal{F}_\Theta(m) = \mathcal{F}_{A_1}(m) \cup \mathcal{F}_{A_2}(m) \cup \mathcal{F}_{A_3}(m) \cup \mathcal{F}_{A^*}(m) = \emptyset \cup \emptyset \cup \emptyset \cup \Theta = \Theta$ .

### B. Total Belief Theorem (TBT)

Based on the previous decomposition of the set of focal elements  $\mathcal{F}_\Theta(m)$  according to any given partition  $\{A_1, \dots, A_k\}$  of the FoD  $\Theta$ , the following Total Belief Theorem (TBT) is established.

**Total Belief Theorem (TBT):** Let's consider a frame of discernment  $\Theta$  with  $|\Theta| \geq 2$  elements and a BBA  $m(\cdot)$  defined on  $2^\Theta$  with the set of focal elements  $\mathcal{F}_\Theta(m)$ . For any chosen partition  $\{A_1, \dots, A_k\}$  of  $\Theta$  and for any  $B \subseteq \Theta$ , one has

$$Bel(B) = \sum_{i=1, \dots, k} Bel(A_i \cap B) + U(A^* \cap B), \quad (41)$$

where  $\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m)$  and

$$U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X). \quad (42)$$

**Proof of TBT:** See appendix.

$A^*$  is a shorthand notation for the entity associated to the set of focal elements  $\mathcal{F}_{A^*}(m)$  of the BBA  $m(\cdot)$  involved in the summation (42) of  $U(A^* \cap B)$ . From the formula (42), one sees that  $U(A^* \cap B) \in [0, 1]$ . Note that if  $B = \Theta$  and if the FoD  $\Theta$  is simply partitioned as  $\{A \triangleq A_1, \bar{A} \triangleq A_2\}$ , then  $U(A^* \cap B) = U(A^* \cap \Theta) = U(A^*) = Pl(A) - Bel(A) = Pl(\bar{A}) - Bel(\bar{A})$ .

If one applies TBT with  $B = \Theta$ , we get for any chosen partition  $\{A_1, \dots, A_k\}$  of  $\Theta$

$$\sum_{i=1, \dots, k} Bel(A_i) + U(A^*) = 1, \quad (43)$$

where  $U(A^*) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m)} m(X)$ . This equality corresponds to TPT if  $U(A^*) = 0$  (i.e. there is no uncertainty on the value of probabilities of  $A_i$ ,  $i = 1, \dots, k$ ).

**Corollary of TBT:** If  $m(\cdot)$  is Bayesian, then TBT is consistent with the Total Probability Theorem (TPT).

**Proof:** See appendix.

From TBT one can establish the following (not so elegant) Total Plausibility Theorem (TPIT).

**Total Plausibility Theorem (TPIT):** For any BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$ , and for any partition  $\{A_1, \dots, A_k\}$  of  $\Theta$ , one has for any  $B \subseteq \Theta$

$$Pl(B) = \sum_{i=1, \dots, k} Pl(\bar{A}_i \cup B) + 1 - k - U(A^* \cap \bar{B}). \quad (44)$$

**Proof:** See appendix.

**Example 4:** Consider the FoD  $\Theta = \{\theta_1, \dots, \theta_7\}$  and the set of focal elements  $\mathcal{F}_\Theta(m) = \{X_1, X_2, \dots, X_9\}$  of a BBA  $m(\cdot)$  defined over  $2^\Theta$  given in Table I.

Table I  
FOCAL ELEMENTS AND THEIR MASSES.

Focal element $X$	BBA $m(X)$
$X_1 = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_7$	$m(X_1) = 0.01$
$X_2 = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$	$m(X_2) = 0.02$
$X_3 = \theta_3 \cup \theta_5 \cup \theta_6$	$m(X_3) = 0.03$
$X_4 = \theta_4 \cup \theta_7$	$m(X_4) = 0.04$
$X_5 = \theta_2$	$m(X_5) = 0.20$
$X_6 = \theta_6 \cup \theta_7$	$m(X_6) = 0.30$
$X_7 = \theta_2 \cup \theta_3 \cup \theta_7$	$m(X_7) = 0.20$
$X_8 = \theta_1 \cup \theta_4 \cup \theta_6$	$m(X_8) = 0.15$
$X_9 = \theta_6$	$m(X_9) = 0.05$

Let's consider the partition  $\{A_1, A_2, A_3\}$  of  $\Theta$  with  $A_1 \triangleq \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$ ,  $A_2 \triangleq \theta_2 \cup \theta_5$  and  $A_3 \triangleq \theta_6$ , and the subset  $B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$  of  $\Theta$  having positive belief  $Bel(B) = m(X_4) + m(X_6) + m(X_9) = 0.39$ . Table II summarizes the belief values of different subsets of  $\Theta$  which are needed in the derivations to apply TBT.

Table II  
BELIEF AND PLAUSIBILITY VALUES USED FOR THE DERIVATIONS.

Subsets of $\Theta$	$Bel(\cdot)$
$B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$	$Bel(B) = 0.39$
$A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$	$Bel(A_1) = 0.04$
$A_2 = \theta_2 \cup \theta_5$	$Bel(A_2) = 0.20$
$A_3 = \theta_6$	$Bel(A_3) = 0.05$
$A_1 \cap B = \theta_4 \cup \theta_7$	$Bel(A_1 \cap B) = 0.04$
$A_2 \cap B = \theta_5$	$Bel(A_2 \cap B) = 0$
$A_3 \cap B = \theta_6$	$Bel(A_3 \cap B) = 0.05$

In this example, one has

$$\begin{aligned} \mathcal{F}_B(m) &= \{X_4, X_6, X_9\} \text{ and } \mathcal{F}_{\bar{B}}(m) = \{X_5\}, \\ \mathcal{F}_{A_1}(m) &= \{X_4\} \text{ and } \mathcal{F}_{\bar{A}_1}(m) = \{X_5, X_9\}, \\ \mathcal{F}_{A_2}(m) &= \{X_5\} \text{ and } \mathcal{F}_{\bar{A}_2}(m) = \{X_4, X_6, X_8, X_9\}, \\ \mathcal{F}_{A_3}(m) &= \{X_9\} \text{ and } \mathcal{F}_{\bar{A}_3}(m) = \{X_1, X_2, X_4, X_5, X_7\}, \\ \mathcal{F}_{A^*}(m) &= \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \mathcal{F}_{A_2}(m) - \mathcal{F}_{A_3}(m) \\ &= \{X_1, X_2, X_3, X_6, X_7, X_8\}. \end{aligned}$$

Therefore,

$$U(A^* \cap B) = \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X) = m(X_6) = 0.30.$$

In applying TBT formula (41), one can easily verify that

$$\begin{aligned} Bel(B) &= Bel(B \cap A_1) + Bel(B \cap A_2) + Bel(B \cap A_3) \\ &\quad + U(A^* \cap B) \\ &= 0.04 + 0 + 0.05 + 0.30 = 0.39. \end{aligned}$$

### C. Special case : A partition with only two elements

If we consider any simple partition  $\{A, \bar{A}\}$  of the FoD  $\Theta$  and any  $B$  subset of  $\Theta$ , then the TBT and TPIT formulas (41) and (44) reduce to<sup>14</sup>

$$Bel(B) = Bel(A \cap B) + Bel(\bar{A} \cap B) + U(A^* \cap B), \quad (45)$$

$$Pl(B) = Pl(\bar{A} \cup B) + Pl(A \cup B) - 1 - U(A^* \cap \bar{B}). \quad (46)$$

<sup>14</sup>Take  $k = 2$ , and set  $A \triangleq A_1$  and  $\bar{A} \triangleq A_2$  in (41) and (44).

**Remark:** If the BBA  $m(\cdot)$  is Bayesian then  $U(A^* \cap B) = 0$  and  $U(A^* \cap \bar{B}) = 0$ . Therefore the previous formulas reduce to

$$Bel(B) = Bel(A \cap B) + Bel(\bar{A} \cap B), \quad (47)$$

$$Pl(B) = Pl(\bar{A} \cup B) + Pl(A \cup B) - 1. \quad (48)$$

$m(\cdot)$  being a Bayesian BBA,  $Bel(\cdot)$  and  $Pl(\cdot)$  are homogeneous to a same (possibly subjective) probability measure  $P(\cdot)$ . Therefore, the previous equalities can be rewritten as

$$P(B) = P(A \cap B) + P(\bar{A} \cap B), \quad (49)$$

$$P(B) = P(\bar{A} \cup B) + P(A \cup B) - 1. \quad (50)$$

The formula (49) is valid because  $\{A, \bar{A}\}$  is a partition of  $\Theta$  and because of TPT theorem. The formula (50) is nothing but a dual form of TPT formula. It is also valid because

$$\begin{aligned} P(\bar{A} \cup B) + P(A \cup B) - 1 &= P(\bar{A}) + P(B) - P(\bar{A} \cap B) \\ &\quad + P(A) + P(B) - P(A \cap B) - 1 \\ &= (P(\bar{A}) + P(A) - 1) + 2P(B) \\ &\quad - (P(\bar{A} \cap B) + P(A \cap B)) \\ &= 0 + 2P(B) - P(B) = P(B) \end{aligned}$$

### D. Generalization of TBT

Previously, the TBT formula was established when the partition  $\{A_1, \dots, A_k\}$  was related to a given FoD  $\Theta$  and  $B$  was a subset of the same FoD  $\Theta$ . We can generalize TBT in considering  $\{A_i, \dots, A_k\}$  as any partition of a FoD  $\Theta_1 = \{x_1, \dots, x_m\} = \{x_p, p = 1, 2, \dots, m\}$ , and  $B$  as being a subset of another FoD  $\Theta_2 = \{y_1, \dots, y_n\} = \{y_q, q = 1, 2, \dots, n\}$  with  $\Theta_1 \neq \Theta_2$ . For this, we need to work within the Cartesian product space  $\Theta \triangleq \Theta_1 \times \Theta_2$ . In the space  $\Theta = \Theta_1 \times \Theta_2$ ,  $x_p$  is not an elementary element but a subset of  $n$  elements of  $\Theta$ , i.e.  $\{x_p\} = \{(x_p, y_1), \dots, (x_p, y_n)\}$ . Similarly,  $y_q$  is not an elementary element but a subset of  $m$  elements of  $\Theta$ , i.e.  $\{y_q\} = \{(x_1, y_q), \dots, (x_m, y_q)\}$ . If  $A_i \subseteq \Theta_1$  and  $B \subseteq \Theta_2$ , then  $A_i \times B = \{(x_p, y_q) | x_p \in A_i; y_q \in B\} \subseteq \Theta$ . Because  $\{A_1, \dots, A_k\}$  is a partition of  $\Theta_1$ , then  $\{A_1 \times \Theta_2, \dots, A_k \times \Theta_2\}$  defines a partition of  $\Theta = \Theta_1 \times \Theta_2$ . Because  $\Theta_1 \times B = \bigcup_{i=1, \dots, k} ((\Theta_1 \times B) \cap (A_i \times \Theta_2))$ , we can apply TBT in the Cartesian space  $\Theta$ . More precisely,

$$\begin{aligned} Bel(\Theta_1 \times B) &= Bel(\bigcup_i ((\Theta_1 \times B) \cap (A_i \times \Theta_2))) \\ &= Bel((\Theta_1 \times B) \cap (A_1 \times \Theta_2)) + \\ &\quad \dots + Bel((\Theta_1 \times B) \cap (A_k \times \Theta_2)) \\ &\quad + U((A^* \times \Theta_2) \cap (\Theta_1 \times B)), \end{aligned}$$

where the quantity  $U((A^* \times \Theta_2) \cap (\Theta_1 \times B))$  is now defined by

$$U((A^* \times \Theta_2) \cap (\Theta_1 \times B)) \triangleq \sum_{X \in \mathcal{F}_{A^* \times \Theta_2}(m) | X \in \mathcal{F}_{\Theta_1 \times B}(m)} m(X). \quad (51)$$

The previous TBT formula when working in the Cartesian space  $\Theta = \Theta_1 \times \Theta_2$  can be written more concisely as

$$Bel(\Theta_1 \times B) = \sum_{i=1, \dots, k} Bel(A_i \times B) + U(A^* \times B), \quad (52)$$

because  $(\Theta_1 \times B) \cap (A_i \times \Theta_2) = (A_i \times \Theta_2) \cap (\Theta_1 \times B) = A_i \times B$ , and by notation convention  $U(A^* \times B) = U((A^* \times \Theta_2) \cap (\Theta_1 \times B))$ .

Note that the formula (52) can be used if and only if one knows the joint BBA  $m(\cdot)$  (or equivalently the joint belief) defined over the powerset of the Cartesian space  $\Theta = \Theta_1 \times \Theta_2$ .

## VIII. CONDITIONAL BELIEF FUNCTIONS BASED ON TBT

In this section we show how FaginHalpern belief conditioning formulas can be established directly from TBT. This result is important because it provides a solid construction of FH formulas and it justifies its use for applications where belief conditioning is necessary. For deriving FH formulas from TBT we consider a partition  $\{A_i, \bar{A}_i\}$  of the FoD  $\Theta$  and a subset  $B$  of  $\Theta$ . Using TBT, one has

$$Bel(B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) + U(A^* \cap B), \quad (53)$$

where  $\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_i}(m) - \mathcal{F}_{\bar{A}_i}(m)$  and

$$U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X). \quad (54)$$

Hence

$$Bel(B) - U(A^* \cap B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B). \quad (55)$$

At this stage, one may be tempted to divide right and left side of previous equality by  $Bel(B) - U(A^* \cap B)$  (assuming its positiveness) to get

$$1 = \frac{Bel(A_i \cap B)}{Bel(B) - U(A^* \cap B)} + \frac{Bel(\bar{A}_i \cap B)}{Bel(B) - U(A^* \cap B)},$$

which would suggest to define  $Bel(A_i|B)$  by taking

$$Bel(A_i|B) = Bel(A_i \cap B) / (Bel(B) - U(A^* \cap B)). \quad (56)$$

Unfortunately, it can be seen from Ellsberg's urn example that the conditional belief defined by (56) is inconsistent with bounds of imprecise conditional probabilities. Therefore, we need to go one step beyond in the calculus for defining consistent conditional belief and plausibility functions. Because by definition  $U((\bar{A}_i \cap B)^*) \triangleq Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B)$ , we have

$$Bel(\bar{A}_i \cap B) = Pl(\bar{A}_i \cap B) - U((\bar{A}_i \cap B)^*). \quad (57)$$

Putting this expression of  $Bel(\bar{A}_i \cap B)$  into (55) and rearranging terms, we get

$$Bel(B) + \Delta(U) = Bel(A_i \cap B) + Pl(\bar{A}_i \cap B), \quad (58)$$

with  $\Delta(U) \triangleq U((\bar{A}_i \cap B)^*) - U(A^* \cap B)$  and  $\Delta(U) \in [0, 1]$  (see proof in appendix).

Assuming  $Bel(B) > 0$ , and dividing each side of (58) by  $Bel(B) + \Delta(U)$ , we get

$$1 = \frac{Bel(A_i \cap B)}{Bel(B) + \Delta(U)} + \frac{Pl(\bar{A}_i \cap B)}{Bel(B) + \Delta(U)}, \quad (59)$$

or equivalently

$$\frac{Bel(A_i \cap B)}{Bel(B) + \Delta(U)} = 1 - \frac{Pl(\bar{A}_i \cap B)}{Bel(B) + \Delta(U)}. \quad (60)$$

Because the general relationship  $Bel(X) = 1 - Pl(\bar{X})$  between the belief and the plausibility must always be satisfied for any  $X \subseteq \Theta$ , the equality (60) allows to define the conditional belief  $Bel(A_i|B)$  and  $Pl(\bar{A}_i|B)$  by taking

$$Bel(A_i|B) \triangleq \frac{Bel(A_i \cap B)}{Bel(B) + \Delta(U)}, \quad (61)$$

$$Pl(\bar{A}_i|B) \triangleq \frac{Pl(\bar{A}_i \cap B)}{Bel(B) + \Delta(U)}. \quad (62)$$

Using equality (58), the previous conditioning formulas can be rewritten equivalently as

$$Bel(A_i|B) = \frac{Bel(A_i \cap B)}{Bel(A_i \cap B) + Pl(\bar{A}_i \cap B)}, \quad (63)$$

$$Pl(\bar{A}_i|B) = \frac{Pl(\bar{A}_i \cap B)}{Bel(A_i \cap B) + Pl(\bar{A}_i \cap B)}. \quad (64)$$

In replacing  $\bar{A}_i$  by  $A_i$  in notations of formulas (62)–(64) we get<sup>15</sup> the conditional plausibility  $Pl(A_i|B)$  as

$$\begin{aligned} Pl(A_i|B) &\triangleq \frac{Pl(A_i \cap B)}{Bel(B) + U((A_i \cap B)^*) - U(A^* \cap B)} \\ &= \frac{Pl(A_i \cap B)}{Bel(\bar{A}_i \cap B) + Pl(A_i \cap B)}. \end{aligned} \quad (65)$$

Formulas (63) and (65) coincide with Fagin-Halpern formulas [4] which were originally proposed from essentially a very good intuition. In this work, we have derived Fagin-Halpern formulas only from TBT using the proper decomposition of the set of focal elements of the a priori BBA. Note that the definition of  $Bel(A_i|B)$  given in (61) satisfies the conditions  $Bel(\emptyset|B) = 0$ ,  $Bel(\Theta|B) = 1$ , and  $Bel(A_i|B) \in [0, 1]$ . To prove that  $Bel(A_i|B)$  defined by (63) is a belief function one must prove that it is also an  $n$ -monotone ( $n \geq 2$ ) Choquet's capacity [36] on the finite set  $\Theta$ , or equivalently that the following inequality holds for any  $B \subseteq \Theta$  with  $Bel(B) > 0$  and for any collection  $A_1, \dots, A_n$  of subsets of  $\Theta$

$$Bel(A_1 \cup \dots \cup A_n|B) \geq \sum_{\substack{I \subseteq \{1, \dots, n\} \\ I \neq \emptyset}} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i|B).$$

The proof of this inequality is complicated. However, three very different proofs have already been given by Fagin and Halpern [3], Jaffray [6], and Sundberg and Wagner [7], the latter one being the clearest of fashion.

## IX. GENERALIZATION OF BAYES' THEOREM

In this section and thanks to the previous results, we generalize Bayes' Theorem (BT) in the framework of belief functions. Assuming  $Bel(B) > 0$ , we have shown that Fagin-Halpern expression of  $Bel(A_i|B)$  given by

$$Bel(A_i|B) = \frac{Bel(A_i \cap B)}{Bel(A_i \cap B) + Pl(\bar{A}_i \cap B)} \quad (66)$$

is equal to the formula (61), i.e.

$$Bel(A_i|B) = \frac{Bel(A_i \cap B)}{Bel(B) + U((\bar{A}_i \cap B)^*) - U(A^* \cap B)}. \quad (67)$$

In replacing  $Bel(B)$  by the expression (41) of TBT we get

$$Bel(A_i|B) = \frac{Bel(A_i \cap B)}{\sum_{i=1, \dots, k} Bel(A_i \cap B) + U((\bar{A}_i \cap B)^*)}. \quad (68)$$

Assuming  $Bel(A_i) > 0$ , Fagin-Halpern expression of  $Bel(B|A_i)$  given by

$$Bel(B|A_i) = \frac{Bel(B \cap A_i)}{Bel(B \cap A_i) + Pl(\bar{B} \cap A_i)} \quad (69)$$

is equal to

$$Bel(B|A_i) = \frac{Bel(B \cap A_i)}{Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i)}, \quad (70)$$

where

$$U((\bar{B} \cap A_i)^*) \triangleq Pl(\bar{B} \cap A_i) - Bel(\bar{B} \cap A_i) \quad (71)$$

$$= \sum_{X \in \mathcal{F}_{(\bar{B} \cap A_i)^*}(m)} m(X), \quad (72)$$

<sup>15</sup>It is worth to note that one has always  $U(A^*) = U(A_i^*) = U(\bar{A}_i^*)$  when partitioning  $\Theta$  as  $\{A_i, \bar{A}_i\}$  because  $U(A_i^*) = Pl(A_i) - Bel(A_i) = (1 - Bel(\bar{A}_i)) - (1 - Pl(\bar{A}_i)) = Pl(\bar{A}_i) - Bel(\bar{A}_i) = U(\bar{A}_i^*)$ .

with  $\mathcal{F}_{(\bar{B} \cap A_i)^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_{\bar{B} \cap A_i}(m) - \mathcal{F}_{B \cup \bar{A}_i}(m)$ , and where

$$U(B^* \cap A_i) \triangleq \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{A_i}(m)} m(X), \quad (73)$$

with  $\mathcal{F}_{B^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_B(m) - \mathcal{F}_{\bar{B}}(m)$ .

From (70), one obtains

$$Bel(A_i \cap B) = Bel(B|A_i)[Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i)].$$

By replacing the above expression of  $Bel(A_i \cap B)$  into (68), we obtain

$$Bel(A_i|B) = \frac{Bel(B|A_i)q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i)q(A_i, B) + U((\bar{A}_i \cap B)^*)}, \quad (74)$$

where the factor  $q(A_i, B)$  introduced here for notation conciseness is defined by

$$q(A_i, B) \triangleq Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i). \quad (75)$$

This result allows us to establish the following Generalized Bayes' Theorem (GBT).

**Generalized Bayes' Theorem (GBT):** For any partition  $\{A_i, i = 1, \dots, k\}$  of a FoD  $\Theta$ , any belief function  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$ , and any subset  $B$  of  $\Theta$  with  $Bel(B) > 0$ , then one has

$$Bel(A_i|B) = \frac{Bel(B|A_i)q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i)q(A_i, B) + U((\bar{A}_i \cap B)^*)}, \quad (76)$$

where

$$\begin{aligned} U((\bar{A}_i \cap B)^*) &\triangleq \sum_{X \in \mathcal{F}_{(\bar{A}_i \cap B)^*}(m)} m(X) \\ &= Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B), \end{aligned}$$

and

$$q(A_i, B) = Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i).$$

**Lemma:** GBT reduces to Bayes' Theorem if  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$  is a Bayesian belief function.

**Proof:** See appendix.

When  $A_i \subseteq \Theta_1$  and  $B \subseteq \Theta_2$  with  $\Theta_1 \neq \Theta_2$ , we must work in the Cartesian product space  $\Theta = \Theta_1 \times \Theta_2$  and the GBT formula is similar to (76) in replacing  $A_i$  by  $A_i \times \Theta_2$ ,  $B$  by  $\Theta_1 \times B$ , and where

$$\begin{aligned} U((\bar{A}_i \cap B)^*) &\triangleq \sum_{X \in \mathcal{F}_{((\bar{A}_i \times \Theta_2) \cap (\Theta_1 \times B))^*}(m)} m(X) \\ &= Pl((\bar{A}_i \times \Theta_2) \cap (\Theta_1 \times B)) \\ &\quad - Bel((\bar{A}_i \times \Theta_2) \cap (\Theta_1 \times B)), \quad (77) \end{aligned}$$

and where the factor  $q(A_i, B)$  must be replaced by

$$q(A_i \times \Theta_2, \Theta_1 \times B) \triangleq Bel(A_i \times \Theta_2) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i), \quad (78)$$

with

$$\begin{aligned} U((\bar{B} \cap A_i)^*) &\triangleq \sum_{X \in \mathcal{F}_{(\Theta_1 \times \bar{B}) \cap (A_i \times \Theta_2)^*}(m)} m(X) \\ &= Pl((\Theta_1 \times \bar{B}) \cap (A_i \times \Theta_2)) \\ &\quad - Bel((\Theta_1 \times \bar{B}) \cap (A_i \times \Theta_2)), \quad (79) \end{aligned}$$

$$U(B^* \cap A_i) \triangleq \sum_{X \in \mathcal{F}_{\Theta_1 \times B^*}(m) | X \in \mathcal{F}_{A_i \times \Theta_2}(m)} m(X), \quad (80)$$

and  $\mathcal{F}_{\Theta_1 \times B^*}(m) = \mathcal{F}_{\Theta_1 \times \Theta_2}(m) - \mathcal{F}_{\Theta_1 \times B}(m) - \mathcal{F}_{\Theta_1 \times \bar{B}}(m)$ .

In the formulas (77)–(80),  $X$  is an elementary element of the Cartesian space  $\Theta = \Theta_1 \times \Theta_2$ , and  $m(X)$  is the (joint) BBA value of  $X$  defined on the power set of Cartesian product space.

The application of GBT formula when working with  $A_i \subseteq \Theta_1$  and  $B \subseteq \Theta_2$  with  $\Theta_1 \neq \Theta_2$  is not easy in general because it requires the knowledge of joint BBA  $m(\cdot)$  defined over  $2^{\Theta_1 \times \Theta_2}$  which is rarely known in practice. If the joint BBA  $m(\cdot)$  can be expressed (or approximated) as a function of two marginal BBAs  $m^{\Theta_1}(\cdot)$  and  $m^{\Theta_2}(\cdot)$  (assumed to be known) defined respectively over  $\Theta_1$  and  $\Theta_2$ , then GBT formula should become tractable.

## X. ILLUSTRATIVE EXAMPLE OF GBT

In this section, we provide a complete quite simple illustrative example to show how belief conditioning formulas work and how to apply GBT.

Let us consider the FoD  $\Theta = \{\theta_i, i = 1, \dots, 7\}$  and the set of focal elements  $\mathcal{F}_\Theta(m) = \{X_1, X_2, \dots, X_9\}$  of a BBA  $m(\cdot)$  defined over  $2^\Theta$  given in Table III. Let's consider the partition  $\{A_1, A_2, A_3\}$  of  $\Theta$  with  $A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$ ,  $A_2 = \theta_2 \cup \theta_5$  and  $A_3 = \theta_6$ , and let consider the subset  $B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$  of  $\Theta$  having positive belief  $Bel(B) = m(X_4) + m(X_6) + m(X_9) = 0.39$ . Table IV summarizes the belief and plausibility values of different subsets of  $\Theta$  which are needed in the derivations.

Table III  
FOCAL ELEMENTS AND THEIR MASSES.

Focal element $X$	BBA $m(X)$
$X_1 = \theta_2 \cup \theta_3 \cup \theta_4 \cup \theta_5 \cup \theta_7$	$m(X_1) = 0.01$
$X_2 = \theta_1 \cup \theta_2 \cup \theta_3 \cup \theta_4$	$m(X_2) = 0.02$
$X_3 = \theta_3 \cup \theta_5 \cup \theta_6$	$m(X_3) = 0.03$
$X_4 = \theta_4 \cup \theta_7$	$m(X_4) = 0.04$
$X_5 = \theta_2$	$m(X_5) = 0.20$
$X_6 = \theta_6 \cup \theta_7$	$m(X_6) = 0.30$
$X_7 = \theta_2 \cup \theta_3 \cup \theta_7$	$m(X_7) = 0.20$
$X_8 = \theta_1 \cup \theta_4 \cup \theta_6$	$m(X_8) = 0.15$
$X_9 = \theta_6$	$m(X_9) = 0.05$

Table IV  
BELIEF AND PLAUSIBILITY VALUES USED FOR THE DERIVATIONS.

Subsets of $\Theta$	$Bel(\cdot)$	$Pl(\cdot)$
$B = \theta_4 \cup \theta_5 \cup \theta_6 \cup \theta_7$	$Bel(B) = 0.39$	$Pl(B) = 0.80$
$A_1 = \theta_1 \cup \theta_3 \cup \theta_4 \cup \theta_7$	$Bel(A_1) = 0.04$	$Pl(A_1) = 0.75$
$A_2 = \theta_2 \cup \theta_5$	$Bel(A_2) = 0.20$	$Pl(A_2) = 0.46$
$A_3 = \theta_6$	$Bel(A_3) = 0.05$	$Pl(A_3) = 0.53$
$A_1 \cap B = \theta_4 \cup \theta_7$	$Bel(A_1 \cap B) = 0.04$	$Pl(A_1 \cap B) = 0.72$
$A_2 \cap B = \theta_5$	$Bel(A_2 \cap B) = 0$	$Pl(A_2 \cap B) = 0.04$
$A_3 \cap B = \theta_6$	$Bel(A_3 \cap B) = 0.05$	$Pl(A_3 \cap B) = 0.53$
$\bar{A}_1 \cap B = \theta_5 \cup \theta_6$	$Bel(\bar{A}_1 \cap B) = 0.05$	$Pl(\bar{A}_1 \cap B) = 0.54$
$\bar{A}_2 \cap B = \theta_4 \cup \theta_6 \cup \theta_7$	$Bel(\bar{A}_2 \cap B) = 0.39$	$Pl(\bar{A}_2 \cap B) = 0.80$
$\bar{A}_3 \cap B = \theta_4 \cup \theta_5 \cup \theta_7$	$Bel(\bar{A}_3 \cap B) = 0.04$	$Pl(\bar{A}_3 \cap B) = 0.75$
$A_1 \cap \bar{B} = \theta_1 \cup \theta_3$	$Bel(A_1 \cap \bar{B}) = 0$	$Pl(A_1 \cap \bar{B}) = 0.41$
$A_2 \cap \bar{B} = \theta_2$	$Bel(A_2 \cap \bar{B}) = 0.20$	$Pl(A_2 \cap \bar{B}) = 0.43$
$A_3 \cap \bar{B} = \emptyset$	$Bel(A_3 \cap \bar{B}) = 0$	$Pl(A_3 \cap \bar{B}) = 0$

In this example, one has

$$\mathcal{F}_B(m) = \{X_4, X_6, X_9\},$$

$$\mathcal{F}_{\bar{B}}(m) = \{X_5\},$$

$$\begin{aligned} \mathcal{F}_{B^*}(m) &= \mathcal{F}_\Theta(m) - \mathcal{F}_B(m) - \mathcal{F}_{\bar{B}}(m) \\ &= \{X_1, X_2, X_3, X_7, X_8\}, \end{aligned}$$

$$\begin{aligned} \mathcal{F}_{A_1}(m) &= \{X_4\}, \\ \mathcal{F}_{\bar{A}_1}(m) &= \{X_5, X_9\}, \\ \mathcal{F}_{A_2}(m) &= \{X_5\}, \\ \mathcal{F}_{\bar{A}_2}(m) &= \{X_4, X_6, X_8, X_9\}, \\ \mathcal{F}_{A_3}(m) &= \{X_9\}, \\ \mathcal{F}_{\bar{A}_3}(m) &= \{X_1, X_2, X_4, X_5, X_7\}, \\ \mathcal{F}_{A^*}(m) &= \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \mathcal{F}_{A_2}(m) - \mathcal{F}_{A_3}(m) \\ &= \{X_1, X_2, X_3, X_6, X_7, X_8\}. \end{aligned}$$

• **Results with Fagin-Halpern conditioning formulas**

Using Fagin-Halpern conditioning formulas (63) and (69) and the fact that  $Pl(A_i|B) = 1 - Bel(\bar{A}_i|B)$  and  $Pl(B|A_i) = 1 - Bel(\bar{B}|A_i)$ , we obtain in this example the conditional belief and plausibility values given in Tables V-VI

Table V  
 $Bel(A_i|B)$  AND  $Pl(A_i|B)$  WITH FAGIN-HALPERN CONDITIONING.

Subsets of $\Theta$	$Bel(A_i B)$	$Pl(A_i B)$
$A_1$	$Bel(A_1 B) \approx 0.0690$	$Pl(A_1 B) \approx 0.9351$
$A_2$	$Bel(A_2 B) = 0$	$Pl(A_2 B) \approx 0.0930$
$A_3$	$Bel(A_3 B) \approx 0.0625$	$Pl(A_3 B) \approx 0.9298$

Table VI  
 $Bel(B|A_i)$  AND  $Pl(B|A_i)$  WITH FAGIN-HALPERN CONDITIONING.

Subsets of $\Theta$	$Bel(B A_i)$	$Pl(B A_i)$
$A_1$	$Bel(B A_1) \approx 0.0889$	$Pl(B A_1) = 1$
$A_2$	$Bel(B A_2) = 0$	$Pl(B A_2) \approx 0.1667$
$A_3$	$Bel(B A_3) = 1$	$Pl(B A_3) = 1$

To apply and verify GBT on this example, one needs to compute  $Bel(A_i)$ ,  $U((\bar{B} \cap A_i)^*)$  and  $U(B^* \cap A_i)$  to calculate  $q(A_i, B)$  factors and also  $U((\bar{A}_i \cap B)^*)$  because they enter in GBT formula (76). These values are listed in Table VII for convenience.

Table VII  
VALUES OF  $q(A_i, B)$  AND  $U((\bar{A}_i \cap B)^*)$  FOR GBT FORMULA.

Subsets of $\Theta$	$q(A_i, B)$	$U((\bar{A}_i \cap B)^*)$
$A_1$	0.45	0.49
$A_2$	0.43	0.41
$A_3$	0.05	0.71

The value  $q(A_1, B) = 0.45$  appearing in Table VII has been calculated as follows

$$q(A_1, B) \triangleq Bel(A_1) + U((\bar{B} \cap A_1)^*) - U(B^* \cap A_1) = 0.45,$$

because

$$\begin{aligned} Bel(A_1) &= 0.04, \\ U((\bar{B} \cap A_1)^*) &= Pl(\bar{B} \cap A_1) - Bel(\bar{B} \cap A_1) = 0.41, \\ U(B^* \cap A_1) &= \sum_{X \in \mathcal{F}_{A_1}(m) | X \in \mathcal{F}_{B^*}(m)} m(X) = 0. \end{aligned}$$

The value  $U((\bar{A}_1 \cap B)^*) = 0.49$  appearing in Table VII is calculated as follows

$$\begin{aligned} U((\bar{A}_1 \cap B)^*) &= Pl(\bar{A}_1 \cap B) - Bel(\bar{A}_1 \cap B) \\ &= 0.54 - 0.05 = 0.49. \end{aligned}$$

Other values of Table VII are calculated similarly.

One verifies that GBT formula (76) works because we retrieve correct values obtained with FH formula, given in Table V. Indeed, one has

$$\begin{aligned} Bel(A_1|B) &= \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)} \\ &\approx \frac{0.0889 \cdot 0.45}{(0.0889 \cdot 0.45) + (0 \cdot 0.43) + (1 \cdot 0.05) + 0.49} \\ &\approx 0.0690, \end{aligned}$$

$$\begin{aligned} Bel(A_2|B) &= \frac{Bel(B|A_2)q(A_2, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_2 \cap B)^*)} \\ &\approx \frac{0 \cdot 0.43}{(0.0889 \cdot 0.45) + (0 \cdot 0.43) + (1 \cdot 0.05) + 0.41} \\ &= 0, \end{aligned}$$

$$\begin{aligned} Bel(A_3|B) &= \frac{Bel(B|A_3)q(A_3, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_3 \cap B)^*)} \\ &\approx \frac{1 \cdot 0.05}{(0.0889 \cdot 0.45) + (0 \cdot 0.43) + (1 \cdot 0.05) + 0.71} \\ &\approx 0.0625. \end{aligned}$$

• **Results with Shafer's conditioning formulas**

Using Shafer's conditioning formulas (31) and (32), we obtain in this example the conditional belief and plausibility values given in Table VIII and IX.

Table VIII  
 $Bel(A_i|B)$  AND  $Pl(A_i|B)$  WITH SHAFER'S CONDITIONING.

Subsets of $\Theta$	$Bel(A_i B)$	$Pl(A_i B)$
$A_1$	$Bel(A_1 B) = 0.3250$	$Pl(A_1 B) = 0.9000$
$A_2$	$Bel(A_2 B) = 0$	$Pl(A_2 B) = 0.0500$
$A_3$	$Bel(A_3 B) = 0.0625$	$Pl(A_3 B) = 0.6625$

Table IX  
 $Bel(B|A_i)$  AND  $Pl(B|A_i)$  WITH SHAFER'S CONDITIONING.

Subsets of $\Theta$	$Bel(B A_i)$	$Pl(B A_i)$
$A_1$	$Bel(B A_1) \approx 0.4533$	$Pl(B A_1) \approx 0.9600$
$A_2$	$Bel(B A_2) \approx 0.0652$	$Pl(B A_2) \approx 0.0870$
$A_3$	$Bel(B A_3) = 1$	$Pl(B A_3) = 1$

As shown in the previous Ellsberg's urn example, one knows that Shafer's belief conditioning formulas are inconsistent with lower and upper bounds of imprecise conditional probabilities, and with this example one shows that Shafer's belief conditioning is also incompatible with GBT formula (76). We emphasize that GBT has been established by a constructive manner from TBT using a direct and relatively simple calculus<sup>16</sup> without need of rule of combination

<sup>16</sup>assuming  $Bel(B)$  and  $Bel(A_i)$  being positive to have well defined expressions as it is for this example.

of basic belief assignments. When using Shafer's belief conditioning formulas, one sees that the conditional values are not coherent since they do not verify GBT because we obtain in this example

$$\begin{aligned} Bel(A_1|B) &= 0.3250 \text{ (from the results in Table VIII using eq. (32))} \\ &\neq \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)} \\ &\approx \frac{0.4533 \cdot 0.45}{(0.4533 \cdot 0.45) + (0.0652 \cdot 0.43) + (1 \cdot 0.05) + 0.49} \\ &\approx 0.2642, \end{aligned}$$

$$\begin{aligned} Bel(A_2|B) &= 0 \text{ (from the results in Table VIII using eq. (32))} \\ &\neq \frac{Bel(B|A_2)q(A_2, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_2 \cap B)^*)} \\ &\approx \frac{0.0652 \cdot 0.43}{(0.4533 \cdot 0.45) + (0.0652 \cdot 0.43) + (1 \cdot 0.05) + 0.41} \\ &\approx 0.0405, \end{aligned}$$

$$\begin{aligned} Bel(A_3|B) &= 0.0625 \text{ (from the results in Table VIII using eq. (32))} \\ &\neq \frac{Bel(B|A_3)q(A_3, B)}{\sum_{i=1}^3 Bel(B|A_i)q(A_i, B) + U((\bar{A}_3 \cap B)^*)} \\ &\approx \frac{1 \cdot 0.05}{(0.4533 \cdot 0.45) + (0.0652 \cdot 0.43) + (1 \cdot 0.05) + 0.71} \\ &\approx 0.0504. \end{aligned}$$

Ellsberg's urn example and this example show clearly that Dempster's rule of combination used by Shafer to establish his belief and conditioning formulas does not provide coherent and satisfactory results since they are inconsistent with lower and upper bounds of imprecise conditional probabilities, and they do not satisfy GBT also.

## XI. CONCLUSION

In this paper new important results for reasoning with belief functions were obtained and discussed. The Total Belief Theorem (TBT) was established from a simple decomposition of the set of focal elements of any basic belief assignment. TBT is a generalization of Total Probability Theorem for belief functions, and based on it we are able to derive conditional belief and conditional plausibility functions that coincide with Fagin-Halpern conditioning formulas which are coherent with lower and upper bounds of imprecise conditional probability. Hence, this work provides a solid justification of the establishment of formulas presented by Fagin and Halpern. The TBT has been generalized for dealing with different frames of discernments as well thanks to the Cartesian product space. Also as a direct consequence of TBT, we have presented a generalization of the well-known Bayes' Theorem for the framework of belief functions called the Generalized Bayesian Theorem (GBT). We have proved that TBT and GBT reduce to TPT and BT respectively as soon as we work with Bayesian belief function because in this case the Bayesian belief function is homogeneous to a probability measure. On the base of Ellsberg's urn example and an illustrative example we have shown that Dempster's rule of combination used by Shafer to establish his belief and conditioning formulas does not provide coherent and satisfactory results because they are inconsistent with lower and upper bounds of imprecise conditional probabilities and because they do not satisfy GBT also. These new theoretical results should (we hope) reconcile the Bayesian reasoning practioners with evidential reasoning practioners and bring new foundations for reasoning with uncertainty thanks to belief functions.

## XII. ACKNOWLEDGMENT

The reported work is partially supported by the National Natural Science Foundation of China (No. 61573275, No. 61671370), Postdoctoral Science Research Foundation of Shaanxi Province (No.

2016BSHEDZZ46), Postdoctoral Science Foundation of China (No. 2016M592790), and Fundamental Research Funds for the Central Universities (No. xjj2016066)

## APPENDIX

### Proof of TBT

From the basic definition of  $Bel(B)$  one has for any  $B \subseteq \Theta$ ,  $Bel(B) = \sum_{X \in \mathcal{F}_\Theta(m) | X \subseteq B} m(X)$ . Because the set of focal elements  $\mathcal{F}_\Theta(m)$  can always be decomposed as the union  $\mathcal{F}_\Theta(m) = \mathcal{F}_{A_1}(m) \cup \dots \cup \mathcal{F}_{A_k}(m) \cup \mathcal{F}_{A^*}(m)$ , then one can always decompose the previous sum as follows

$$\begin{aligned} Bel(B) &= \sum_{X \in \mathcal{F}_\Theta(m) | X \subseteq B} m(X) \\ &= \sum_{X \in \mathcal{F}_{A_1}(m) | X \subseteq B(m)} m(X) + \dots \\ &\quad + \sum_{X \in \mathcal{F}_{A_k}(m) | X \subseteq B(m)} m(X) \\ &\quad + \sum_{X \in \mathcal{F}_{A^*}(m) | X \subseteq B(m)} m(X) \\ &= Bel(A_1 \cap B) + \dots + Bel(A_k \cap B) \\ &\quad + \sum_{X \in \mathcal{F}_{A^*}(m) | X \subseteq B(m)} m(X) \\ &= \sum_{i=1, \dots, k} Bel(A_i \cap B) + U(A^* \cap B), \end{aligned}$$

where  $U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \subseteq B(m)} m(X)$ , which completes the proof of TBT.

### Proof of the corollary of TBT

If  $m(\cdot)$  is Bayesian then any focal element  $X$  of  $\mathcal{F}_\Theta(m)$  is a singleton of  $2^\Theta$  which either belongs to  $A_i$ , or to  $\bar{A}_i$  (but it cannot belong to both). Therefore,  $\mathcal{F}_\Theta(m) = \mathcal{F}_{A_1}(m) \cup \dots \cup \mathcal{F}_{A_k}(m)$  and  $\mathcal{F}_{A^*}(m) = \emptyset$ . TBT formula applies with<sup>17</sup>  $U(A^* \cap B) = \sum_{X \in \mathcal{F}_{A^*}(m) | X \subseteq B(m)} m(X) = \sum_{X \in \emptyset | X \subseteq B(m)} m(X) = 0$  and thanks to TBT one has in this case for any partition  $\{A_1, \dots, A_k\}$  of  $\Theta$  and any subset  $B$  of  $\Theta$  the following equality satisfied

$$Bel(B) = \sum_{i=1, \dots, k} Bel(A_i \cap B). \quad (81)$$

When  $m(\cdot)$  is Bayesian, its corresponding belief function  $Bel(\cdot)$  is homogeneous to a probability measure  $P(\cdot)$  [1], and therefore the previous equality is consistent with TPT formula (14), which completes the proof of the corollary of the TBT.

<sup>17</sup>We recall that if a summation has no term then its value is set to zero.

### Proof of TPII

From equality  $Pl(B) = 1 - Bel(\bar{B})$  and TBT, one has

$$\begin{aligned}
 Pl(B) &= 1 - Bel(\bar{B}) \\
 &= 1 - \sum_{i=1, \dots, k} Bel(A_i \cap \bar{B}) - U(A^* \cap \bar{B}) \\
 &= 1 - \sum_{i=1, \dots, k} (Bel(A_i \cap \bar{B}) + 1 - 1) - U(A^* \cap \bar{B}) \\
 &= 1 - \sum_{i=1, \dots, k} (-1 + Bel(A_i \cap \bar{B}) + 1) - U(A^* \cap \bar{B}) \\
 &= 1 - \sum_{i=1, \dots, k} (-(1 - Bel(A_i \cap \bar{B})) + 1) - U(A^* \cap \bar{B}) \\
 &= 1 + \sum_{i=1, \dots, k} Pl(\overline{A_i \cap \bar{B}}) - k - U(A^* \cap \bar{B}) \\
 &= \sum_{i=1, \dots, k} Pl(\bar{A}_i \cup B) + 1 - k - U(A^* \cap \bar{B}),
 \end{aligned}$$

which completes the proof of TPII.

### Proof that $\Delta(U) \in [0, 1]$

$$\begin{aligned}
 \Delta(U) &\triangleq U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \\
 &= [Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B)] \\
 &\quad - [Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) - Bel(B)] \\
 &= Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) \\
 &\quad + Bel(B) - Bel(A_i \cap B) - Bel(\bar{A}_i \cap B).
 \end{aligned}$$

To prove that  $\Delta(U) \geq 0$ , one must prove equivalently that

$$\begin{aligned}
 Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) + Bel(B) \geq \\
 Bel(A_i \cap B) + Bel(\bar{A}_i \cap B). \quad (82)
 \end{aligned}$$

Using TBT, one has

$$Bel(B) = Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) + U(A^* \cap B).$$

Replacing expression of  $Bel(B)$  in inequality (82), one must verify if the following equality is satisfied

$$\begin{aligned}
 Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) + Bel(A_i \cap B) + Bel(\bar{A}_i \cap B) + U(A^* \cap B) \\
 \geq Bel(A_i \cap B) + Bel(\bar{A}_i \cap B).
 \end{aligned}$$

After simplification, we have to prove that the following inequality holds

$$Pl(\bar{A}_i \cap B) + U(A^* \cap B) \geq Bel(\bar{A}_i \cap B).$$

Because  $Pl(\bar{A}_i \cap B) = Bel(\bar{A}_i \cap B) + U((\bar{A}_i \cap B)^*)$ , one has to verify if the following inequality holds

$$Bel(\bar{A}_i \cap B) + U((\bar{A}_i \cap B)^*) + U(A^* \cap B) \geq Bel(\bar{A}_i \cap B).$$

After simplification (omitting both  $Bel(\bar{A}_i \cap B)$  in left and right side of the previous inequality), one has to prove that the inequality below is satisfied to prove that  $\Delta(U) \geq 0$

$$U((\bar{A}_i \cap B)^*) + U(A^* \cap B) \geq 0.$$

Because  $U((\bar{A}_i \cap B)^*) \in [0, 1]$  and  $U(A^* \cap B) \in [0, 1]$ , the previous inequality always holds which proves that  $U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \geq 0$ . Moreover because  $U(A^* \cap B) \in [0, 1]$ , then  $-U(A^* \cap B) \in [-1, 0]$ . Because  $U((\bar{A}_i \cap B)^*) \in [0, 1]$ , one deduces that  $U((\bar{A}_i \cap B)^*) - U(A^* \cap B) \leq 1$ . This completes the proof.

### Proof of Lemma

If  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$  is a Bayesian belief function, then all focal elements of its corresponding BBA  $m(\cdot)$  are singletons of  $2^\Theta$ . In this case  $Bel(\cdot)$  and  $Pl(\cdot)$  functions coincide and therefore one has  $U((\bar{A}_i \cap B)^*) = Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B) = 0$  and  $U((\bar{B} \cap A_i)^*) = Pl(\bar{B} \cap A_i) - Bel(\bar{B} \cap A_i) = 0$ . Any focal element (singleton) of  $m(\cdot)$  is either a subset of  $B$  or a subset of  $\bar{B}$  of the FoD  $\Theta$ . Therefore,  $\mathcal{F}_{B^*}(m) = \emptyset$ , which implies  $U(B^* \cap A_i) = 0$ , so that  $q(A_i, B) = Bel(A_i)$ . The GBT formula (76) with in this case  $q(A_i, B) = Bel(A_i)$  and  $U((\bar{A}_i \cap B)^*) = 0$  reduces to formula

$$Bel(A_i|B) = \frac{Bel(B|A_i)Bel(A_i)}{\sum_{i=1}^k Bel(B|A_i)Bel(A_i)},$$

which coincides with formula (21) because  $Bel(\cdot)$  (being a Bayesian belief function) is homogeneous to a probability measure  $P(\cdot)$ . This completes the proof that GBT formula is consistent with Bayesian Theorem formula when the Belief function is Bayesian.

### REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [2] A.P. Dempster, *Upper and lower probabilities induced by a multi-valued mapping*, Ann. of Math. Stat., Vol. 38, pp. 325-339, 1967.
- [3] R. Fagin, J.Y. Halpern, *A new approach to updating beliefs*, in Proc. Conf. on Uncertainty in Artificial Intelligence (UAI), pp. 317-325, 1991.
- [4] J.Y. Halpern, R. Fagin, *Two views of belief: belief as generalized probability and belief as evidence*, Artificial Intelligence, Vol. 54, pp. 275-317, 1992.
- [5] L. De Campos, T. Lamata, S. Moral, *The concept of conditional fuzzy measure*, Int. J. of Intel. Systems, Vol. 5, pp. 237-246, 1990.
- [6] J.-Y. Jaffray, *Bayesian updating and belief functions*, IEEE Trans on SMC, Vol. 22(5), pp. 1144-1152, Sept.-Oct. 1992.
- [7] C. Sunberg, C. Wagner, *Generalized finite differences and Bayesian conditioning of Choquet capacities*, Advances in Applied Mathematics, Vol. 13: pp. 262-272, 1992.
- [8] P. Walley, *Measures of uncertainty in expert systems*, Artificial Intelligence, Vol. 83, pp. 1-58, 1996.
- [9] J. Aitchison, *Discussion on Professor Dempster's paper*, Journal of the Royal Statistical Society, Serie B, Vol. 30, pp. 234-237, 1968.
- [10] P. Diaconis, *Review of "A Mathematical Theory of Evidence"*, Journal of the American Statistical Society, Vol. 73 (363), pp. 677-678, 1978.
- [11] P. Diaconis, S.L. Zabell, *Updating subjective probability*, Journal of the American Statistical Society, Vol. 77(380), pp. 822-830, 1982.
- [12] L.A. Zadeh, *Mathematical Theory of Evidence (book review)*, AI Magazine, Vol. 5(3), pp. 81-83, 1984.
- [13] J. Lemmer, *Confidence factors, empiricism and the Dempster-Shafer theory of evidence*, in Proc. of 1st Conf. on UAI, pp. 160-176, 1985.
- [14] P.K. Black, *Is Shafer general Bayes?*, in Proc. of 3rd Workshop on Uncertainty in Artificial Intelligence, Seattle, WA, USA, pp. 2-9, 1987.
- [15] D. Hunter, *Dempster-Shafer vs. probabilistic logic*, in Proc. 3rd AAAI Uncertainty in Artificial Intelligence Workshop, pp. 22-29, 1987.
- [16] F. Voorbraak, *On the justification of Dempster's rule of combination*, Utrecht Univ., Netherlands, Logic Group Preprint Series, No. 42, 1988.
- [17] P. Wang, *A defect in Dempster-Shafer theory*, in Proc. of 10th Conf. on Uncertainty in AI, pp. 560-566, 1994.
- [18] A. Gelman, *The boxer, the wrestler, and the coin flip: a paradox of robust Bayesian inference and belief functions*, American Statistician, Vol. 60(2), pp. 146-150, 2006.
- [19] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, 1988.
- [20] J. Pearl, *Reasoning with belief functions: a critical assessment*, Technical Report R-136, UCLA, 1989.
- [21] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Fusion 2012, Singapore, July 9-12, 2012.
- [22] A. Tchamova, J. Dezert, *On the behavior of Dempster's Rule of combination and the foundations of Dempster-Shafer theory*, in Proc. of 6th IEEE Int. Conf. on Int. Syst., Sofia, Bulgaria, Sept. 6-8, 2012.

- [23] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., Vol. 29(3), pp. 223–252, March 2014 (with erratum in DSMT Book Vol. 5).
- [24] J.N. Heendeni, K. Premaratne, M.N. Murthi, J. Uscinski, M. Scheutz, *A generalization of Bayesian inference in the Dempster-Shafer belief theoretic framework*, in Proc. of Fusion 2016 Int. Conf. on Information Fusion, Heidelberg, Germany, July 2016.
- [25] P. Smets, *Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem*, Int. J. of Approximate Reasoning, Vol. 9(1), pp. 1–35, 1993.
- [26] D. Dubois, T. Denceux, *Conditioning in Dempster-Shafer Theory: Prediction vs. Revision*, in Proc. of Belief 2012 Int. Conf., Compiègne, France, May 2012; (Belief Functions: Theory & Appl., AISC 164, Springer, 385–392).
- [27] A.P. Dempster, *A generalization of Bayesian inference*, J. of Royal Sta. Soc., B30, pp. 205–247, 1968.
- [28] F. Smarandache, J. Dezert J.(Editors), *Advances and applications of DSMT for information fusion*, Vol. 1–4, American Research Press, 2004–2015. <https://www.onera.fr/fr/staff/jean-dezert/references>
- [29] P.R. Halmos, *Naive Set Theory*, Springer-Verlag New York, 1974.
- [30] X.R. Li, *Probability, random signals, and statistics*, CRC Press, 1999.
- [31] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd Ed. New York: McGraw-Hill, pp. 37–38, 1984.
- [32] G. Grimmett, D. Welsh, *Probability: An Introduction*, Oxford Science Publications, 1986.
- [33] J. Dezert, A. Tchamova, D. Han, J.-M.-Tacnet, *Why Dempster's rule doesn't behave as Bayes rule with informative priors*, Proc. of 2013 IEEE Int. Symp. on INnovations in Intelligent SysTems and Application (INISTA 2013), Albena, Bulgaria, June 19–21, 2013.
- [34] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's fusion rule is not a generalization of Bayes fusion rule*, Fusion 2013 Proc., Istanbul, July 9–12, 2013 (with erratum in DSMT Book Vol. 5).
- [35] D. Ellsberg, *Risk, ambiguity, and the Savage axioms*, Quart. J. Econ., Vol. 75, pp. 643–669, 1961.
- [36] G. Choquet, *Theory of capacities*, Ann. Inst. Fourier, Vol. 5, pp. 131–295, 1953.



# A Simplified Formulation of Generalized Bayes' Theorem

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>, Thanuka Wickramaratne<sup>d</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

<sup>d</sup>ECE Dept., Univ. of Massachusetts, Lowell, MA, USA.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn, thanuka@uml.edu

Originally published as: J. Dezert, A. Tchamova, D. Han, T. Wickramaratne, *A Simplified Formulation of Generalized Bayes' Theorem*, in Proc. of Int. Conf. on Information Fusion (Fusion 2019), Ottawa, Canada, July 2–5, 2019, and reprinted with permission.

**Abstract**—In this paper we present a simple formulation of the Generalized Bayes' Theorem (GBT) which extends Bayes' theorem in the framework of belief functions. We also present the condition under which this new formulation is valid. We illustrate our theoretical results with simple examples.

**Keywords:** Generalized Bayes' Theorem (GBT), Simplified GBT (SGBT), Total Belief Theorem (TBT), belief functions.

## I. INTRODUCTION

Based on Dempster's works [1], [2], Shafer did introduce Belief Functions (BF) in 1976 to model the epistemic uncertainty<sup>1</sup> and to reason under uncertainty [3] which is referred as Dempster-Shafer Theory (DST) in the literature. Belief functions are mathematically well defined and they are very appealing from the theoretical standpoint because of their good ability to model uncertainty interpreted as imprecise probability measures in Dempster's original works.

From the end of 1970's the DST has however been cast in doubts because Dempster's rule of combination of Basic Belief Assignments (BBAs) yields counter intuitive results not only in high conflicting situations but also in low conflicting cases as well [4]–[6], and Shafer's conditioning formulas based on Dempster's rule are not consistent with conditional probability calculus [7], [8]. Discussions on the validity of DST can be found, for instance, in [4], [5], [9]–[13]. These two major concerns make DST quite risky for applications involving randomness and epistemic uncertainties and it should be replaced by better techniques to reason under uncertainty with belief functions.

In 2018 we did establish in [8], [14] two new important general results for reasoning with belief functions: the Total Belief Theorem (TBT), and the Generalized Bayes' Theorem (GBT). TBT and GBT generalize the well-known Total Probability Theorem (TPT) and Bayes' Theorem (BT) of the Probability Theory (PT). Thanks to these new theorems we have now in hands a generalized Bayesian inference mechanism for working with imprecise probability measures in the belief functions framework. Similarly to the probability theory requiring a good estimation of pdf (or pmf) involved in Bayes' formula to make a good inference, the major difficulty for applying GBT

is the knowledge (or good estimation) of all<sup>2</sup> BBAs required in GBT. For a given size of frame of discernment, GBT requires more computations than Bayes formula (if we would prefer to work with probabilities) because we need to work with BBAs defined on the powerset of the frame of discernment.

The general formulation of GBT presented in details in [8] is not easy to apply and that is why we present in this paper a simpler and more convenient formulation of GBT providing elegant and useful mathematical expressions. The obtention of these new formulas of GBT are established from a dichotomous partitioning of the frame of discernment.

This paper is organized as follows. After a short reminder of basics of belief functions in Section II and their constructions based on Dempster's multi-valued mapping, we present briefly the Total Belief Theorem and Fagin-Halpern conditioning in Section III, and the Generalized Bayes Theorem in Section IV. In section V we establish the Simplified GBT (SGBT) drawn from GBT for working with a dichotomous partitioning of the frame of discernment. Section VI presents and discusses two examples of SGBT results. Section VII concludes this paper.

## II. BASICS OF BELIEF FUNCTIONS

### A. Basic belief assignment

We consider a finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and where all exhaustive and exclusive elements of  $\Theta$  represent the set of the potential solutions of the problem under concern. The set of all subsets of  $\Theta$  (including the empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . A Basic Belief Assignment (BBA) associated with a given source of evidence is defined as the mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying the conditions  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The quantity  $m(A)$  is the mass of belief for subset  $A$  committed by the Source of Evidence (SoE).

### B. Focal elements

A focal element  $X$  of a BBA  $m(\cdot)$  is an element of  $2^\Theta$  such that  $m(X) > 0$ . Note that the empty set  $\emptyset$  is not a focal element of a BBA because  $m(\emptyset) = 0$  (closed-world

<sup>1</sup>Also called sometimes the cognitive uncertainty by some authors.

<sup>2</sup>possibly joint BBAs if we work on Cartesian product spaces [8].

assumption of Shafer's model for the FoD). The set of all focal elements of  $m(\cdot)$  is denoted

$$\begin{aligned}\mathcal{F}_\Theta(m) &\triangleq \{X \subseteq \Theta | m(X) > 0\} \\ &= \{X \in 2^\Theta | m(X) > 0\}.\end{aligned}\quad (1)$$

The set of focal elements of  $m(\cdot)$  included in  $A \subseteq \Theta$  is denoted, where  $\triangleq$  means *equal by definition*, by

$$\mathcal{F}_A(m) \triangleq \{X \in \mathcal{F}_\Theta(m) | X \cap A = X\}.\quad (2)$$

$\mathcal{F}_\Theta(m)$  can be partitioned as  $\{\mathcal{F}_A(m), \mathcal{F}_{\bar{A}}(m), \mathcal{F}_{A^*}(m)\}$  with

$$\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_A(m) - \mathcal{F}_{\bar{A}}(m),\quad (3)$$

which represents the set of focal elements of  $m(\cdot)$  which are not subsets of  $A$ , and not subsets of the complement of  $A$  in  $\Theta$  which is  $\bar{A} \triangleq \Theta - \{A\}$ . The minus symbol in  $\Theta - \{A\}$  denotes the set difference operator.

### C. Belief, plausibility and uncertainty

Belief and plausibility functions are defined as<sup>3</sup>

$$\begin{aligned}Bel(A) &\triangleq \sum_{\substack{X \in 2^\Theta \\ X \subseteq A}} m(X) \\ &= \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \subseteq A}} m(X) \\ &= \sum_{X \in \mathcal{F}_A(m)} m(X),\end{aligned}\quad (4)$$

$$\begin{aligned}Pl(A) &\triangleq \sum_{\substack{X \in 2^\Theta \\ X \cap A \neq \emptyset}} m(X) \\ &= \sum_{\substack{X \in \mathcal{F}_\Theta(m) \\ X \cap A \neq \emptyset}} m(X) \\ &= 1 - \sum_{X \in \mathcal{F}_{\bar{A}}(m)} m(X) \\ &= 1 - Bel(\bar{A}).\end{aligned}\quad (5)$$

The length of the belief interval  $[Bel(A), Pl(A)]$  is usually called by abuse of terminology the *uncertainty on A* committed by the SoE. In fact it represents the imprecision on the (possibly subjective) probability of  $A$  granted by the SoE which provides the BBA  $m(\cdot)$ . We denote it  $U(A^*)$ , and it is defined as

$$U(A^*) \triangleq Pl(A) - Bel(A) = \sum_{X \in \mathcal{F}_{A^*}(m)} m(X).\quad (6)$$

If all the elements of  $\mathcal{F}_\Theta(m)$  are singletons,  $m(\cdot)$  is called a *Bayesian BBA* [3], and its corresponding  $Bel(\cdot)$  and  $Pl(\cdot)$  functions are homogeneous to a same (subjective) probability measure  $P(\cdot)$ . In this case  $\mathcal{F}_{A^*}(m) = \mathcal{F}_{\bar{A}^*}(m) = \emptyset$ . Shafer

<sup>3</sup>By convention, a *sum of non existing terms* (if it occurs in formulas depending on the given BBA) is always set to zero.

did prove in [3] (p.39) that  $m(\cdot)$ ,  $Bel(\cdot)$  and  $Pl(\cdot)$  are one-to-one, and for any  $A \subseteq \Theta$ ,  $m(\cdot)$  is obtained from  $Bel(\cdot)$  by Möbius inverse formula

$$m(A) = \sum_{B \subseteq A \subseteq \Theta} (-1)^{|A-B|} Bel(B).\quad (7)$$

### D. Interpretation and construction of belief functions

In original Dempster's works [1] belief  $Bel(A)$  and plausibility  $Pl(A)$  are interpreted as lower and upper bounds of an unknown probability  $P(A)$ , and so  $Bel(A) \leq P(A) \leq Pl(A)$ . The construction of  $m(A)$ ,  $Bel(A)$  and  $Pl(A)$  are mathematically well defined from an underlying random variable with a known probability measure and a given multi-valued mapping as follows:

- Consider a random variable  $x$  with its set of possible values in  $\mathcal{X} = \{x_1, \dots, x_m\}$  with known probabilities  $p_j = P(x = x_j)$ ,  $j = 1, \dots, m$ ;
- Consider a FoD  $\Theta = \{\theta_1, \dots, \theta_n\}$  for the variable  $\theta$  under concern;
- Consider/learn a multi-valued mapping  $\Gamma : \mathcal{X} \mapsto 2^\Theta$  such that if  $x = x_i$  then  $\theta \in A$ , so that  $A = \Gamma(x_i) \in 2^\Theta$ ;
- The belief (lower proba) and plausibility (upper proba) that  $\theta \in A$  are given by [1]

$$\begin{aligned}P_*(A) &= Bel(A) = Bel(\theta \in A) \\ &= P(\{x \in \mathcal{X} | \Gamma(x) \neq \emptyset, \Gamma(x) \subseteq A\}),\end{aligned}\quad (8)$$

$$\begin{aligned}P^*(A) &= Pl(A) = Pl(\theta \in A) \\ &= P(\{x \in \mathcal{X} | \Gamma(x) \cap A \neq \emptyset\}).\end{aligned}\quad (9)$$

The mass of belief that  $\theta$  belongs to  $A$  is given by

$$m(A) = P(\{x \in \mathcal{X} | \Gamma(x) \neq \emptyset, \Gamma(x) = A\}).\quad (10)$$

**Example for multi-valued mapping:** Paul has been killed and Police asks a witness  $W$ : Who did you see killing Paul? Witness answer is Mary. To estimate the confidence of this testimony report one has to consider if this witness  $W$  is more or less precise when he is reliable, or if he is not reliable. So the state of  $W$  can belong to  $\mathcal{X} = \{x_1, x_2, x_3\}$  where  $x_1$  means  $W$  is precise,  $x_2$  means  $W$  is approximate, and  $x_3$  means  $W$  is not reliable. We suppose that the a priori probabilities of the state of  $W$  are  $P(x_1) = 0.3$ ,  $P(x_2) = 0.1$  and  $P(x_3) = 0.6$ . As FoD  $\Theta$ , we consider a set of three suspects that includes the unknown killer

$$\Theta = \{\theta_1 = \text{Mary}, \theta_2 = \text{Peter}, \theta_3 = \text{John}\}.$$

If we define the multivalued mapping  $\Gamma(\cdot)$  as follows

$$\begin{aligned}\Gamma(x_1 = W \text{ is precise}) &= \theta_1, \\ \Gamma(x_2 = W \text{ is approximate}) &= \{\theta_1, \theta_2\}, \\ \Gamma(x_3 = W \text{ is not reliable}) &= \{\theta_1, \theta_2, \theta_3\} = \Theta.\end{aligned}$$

$\Gamma(x_1 = W \text{ is precise}) = \theta_1$  means that if  $W$  is precise then Mary has killed Paul.  $\Gamma(x_2 = W \text{ is approximate}) = \{\theta_1, \theta_2\}$  means that if  $W$  is less precise then Mary or Peter have killed Paul.  $\Gamma(x_3 = W \text{ is not reliable}) = \Theta$  means that if  $W$  is not reliable then we have no useful information about the killer.

Applying formulas (8) and (9), one gets

$$\begin{aligned}
 Bel(\emptyset) &= P(\{x|\Gamma(x) \subseteq \emptyset\}) = P(\emptyset) = 0 \\
 &= 1 - Pl(\Theta), \\
 Bel(\theta_1) &= P(\{x|\Gamma(x) \subseteq \theta_1\}) = P(x_1) = 0.3 \\
 &= 1 - Pl(\theta_2 \cup \theta_3), \\
 Bel(\theta_2) &= P(\{x|\Gamma(x) \subseteq \theta_2\}) = 0 \\
 &= 1 - Pl(\theta_1 \cup \theta_3), \\
 Bel(\theta_3) &= P(\{x|\Gamma(x) \subseteq \theta_3\}) = 0 \\
 &= 1 - Pl(\theta_1 \cup \theta_2), \\
 Bel(\theta_1 \cup \theta_2) &= P(\{x|\Gamma(x) \subseteq \theta_1 \cup \theta_2\}) = P(\{x_1, x_2\}) \\
 &= P(x_1) + P(x_2) = 0.4 \\
 &= 1 - Pl(\theta_3), \\
 Bel(\theta_1 \cup \theta_3) &= P(\{x|\Gamma(x) \subseteq \theta_1 \cup \theta_3\}) = P(x_1) = 0.3 \\
 &= 1 - Pl(\theta_2), \\
 Bel(\theta_2 \cup \theta_3) &= P(\{x|\Gamma(x) \subseteq \theta_2 \cup \theta_3\}) = 0 \\
 &= 1 - Pl(\theta_1), \\
 Bel(\Theta) &= P(\{x|\Gamma(x) \subseteq \Theta\}) = P(\{x_1, x_2, x_3\}) \\
 &= P(x_1) + P(x_2) + P(x_3) = 1 \\
 &= 1 - Pl(\emptyset),
 \end{aligned}$$

and

$$\begin{aligned}
 Pl(\emptyset) &= P(\{x|\Gamma(x) \cap \emptyset \neq \emptyset\}) = P(\emptyset) = 0 \\
 &= 1 - Bel(\Theta), \\
 Pl(\theta_1) &= P(x|\Gamma(x) \cap \theta_1 \neq \emptyset) = P(\{x_1, x_2, x_3\}) \\
 &= P(x_1) + P(x_2) + P(x_3) = 1 \\
 &= 1 - Bel(\theta_2 \cup \theta_3), \\
 Pl(\theta_2) &= P(\{x|\Gamma(x) \cap \theta_2 \neq \emptyset\}) = P(\{x_2, x_3\}) \\
 &= P(x_2) + P(x_3) = 0.7 \\
 &= 1 - Bel(\theta_1 \cup \theta_3), \\
 Pl(\theta_3) &= P(\{x|\Gamma(x) \cap \theta_3 \neq \emptyset\}) \\
 &= P(x_3) = 0.6 \\
 &= 1 - Bel(\theta_1 \cup \theta_2), \\
 Pl(\theta_1 \cup \theta_2) &= P(\{x|\Gamma(x) \cap (\theta_1 \cup \theta_2) \neq \emptyset\}) = P(\{x_1, x_2, x_3\}) \\
 &= P(x_1) + P(x_2) + P(x_3) = 1 \\
 &= 1 - Bel(\theta_3), \\
 Pl(\theta_1 \cup \theta_3) &= P(\{x|\Gamma(x) \cap (\theta_1 \cup \theta_3) \neq \emptyset\}) = P(\{x_1, x_2, x_3\}) \\
 &= P(x_1) + P(x_2) + P(x_3) = 1 \\
 &= 1 - Bel(\theta_2), \\
 Pl(\theta_2 \cup \theta_3) &= P(\{x|\Gamma(x) \cap (\theta_2 \cup \theta_3) \neq \emptyset\}) = P(\{x_2, x_3\}) \\
 &= P(x_2) + P(x_3) = 0.7 \\
 &= 1 - Bel(\theta_1), \\
 Pl(\Theta) &= P(\{x|\Gamma(x) \cap (\theta_1 \cup \theta_2 \cup \theta_3) \neq \emptyset\}) = P(\{x_1, x_2, x_3\}) \\
 &= P(x_1) + P(x_2) + P(x_3) = 1 \\
 &= 1 - Bel(\emptyset).
 \end{aligned}$$

In applying formula (10), one gets finally the BBA

$$\begin{aligned}
 m(\emptyset) &= P(\{x|\Gamma(x) = \emptyset\}) = P(\emptyset) = 0, \\
 m(\theta_1) &= P(\{x|\Gamma(x) = \theta_1\}) = P(x_1) = 0.3, \\
 m(\theta_2) &= P(\{x|\Gamma(x) = \theta_2\}) = 0, \\
 m(\theta_3) &= P(\{x|\Gamma(x) = \theta_3\}) = 0,
 \end{aligned}$$

$$\begin{aligned}
 m(\theta_1 \cup \theta_2) &= P(\{x|\Gamma(x) = \theta_1 \cup \theta_2\}) = P(x_2) = 0.1, \\
 m(\theta_1 \cup \theta_3) &= P(\{x|\Gamma(x) = \theta_1 \cup \theta_3\}) = 0, \\
 m(\theta_2 \cup \theta_3) &= P(\{x|\Gamma(x) = \theta_2 \cup \theta_3\}) = 0, \\
 m(\Theta) &= P(\{x|\Gamma(x) = \Theta\}) = P(x_3) = 0.6.
 \end{aligned}$$

Some authors have proposed different interpretations of belief functions to escape the probabilistic framework introduced by Dempster to save DST of its inherent contradiction mainly due to the choice of Dempster's rule of combination and Shafer's conditioning approach based on Dempster's rule. The most important attempt has been done in 1990's by Smets in [15] with his axiomatic Transferable Belief Model (TBM). It however remains disputable because of the ambiguous (or inconsistent/double) interpretation of the empty set.

In this paper we adopt the original Dempster's interpretation and construction of belief functions because it is mathematically well defined, clear and consistent.

### III. TBT AND FAGIN-HALPERN CONDITIONING

#### A. Total Belief Theorem

In [8], we have generalized the Total Probability Theorem (TPT) [16] for working with belief functions and we proved the following simple and important theorem.

**Total Belief Theorem (TBT):** Let's consider a FoD  $\Theta$  with  $|\Theta| \geq 2$  elements and a BBA  $m(\cdot)$  defined on  $2^\Theta$  with the set of focal elements  $\mathcal{F}_\Theta(m)$ . For any chosen partition  $\{A_1, \dots, A_k\}$  of  $\Theta$  and for any  $B \subseteq \Theta$ , one has

$$Bel(B) = \sum_{i=1, \dots, k} Bel(A_i \cap B) + U(A^* \cap B), \quad (11)$$

where

$$U(A^* \cap B) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m) | X \in \mathcal{F}_B(m)} m(X), \quad (12)$$

and  $\mathcal{F}_{A^*}(m) \triangleq \mathcal{F}_\Theta(m) - \mathcal{F}_{A_1}(m) - \dots - \mathcal{F}_{A_k}(m)$ .

**Proof of TBT:** see [8], with example.

From (12), one sees that  $U(A^* \cap B) \in [0, 1]$ . If one applies TBT with  $B = \Theta$ , we get  $\sum_{i=1, \dots, k} Bel(A_i) + U(A^*) = 1$  where  $U(A^*) \triangleq \sum_{X \in \mathcal{F}_{A^*}(m)} m(X)$ . This equality corresponds to TPT if  $U(A^*) = 0$  (i.e. there is no imprecision on the value of probabilities of  $A_i$ ,  $i = 1, \dots, k$ ).

In spite of its apparent simplicity the TBT is very important because it provides a strong theoretical justification of Fagin-Halpern (FH) belief and plausibility conditioning formulas [7], [17] proposed in 1990's as a very serious alternative to Shafer's conditioning formulas. Indeed, it can be easily proved with a simple counter-example (e.g. Ellsberg's urn example - see [8]) that conditioning formulas established by Shafer from Dempster's rule of combination are not consistent with bounds of the conditional probabilities. The main advantage of FH conditioning formulas is that they provide exact bounds of imprecise conditional probability and they coincide exactly

with the conditional probability when the belief functions involved in FH formulas are Bayesian.

### B. Fagin-Halpern belief conditioning formulas

In [8] we have proved that the TBT justifies the following FH conditioning formulas (assuming  $Bel(B) > 0$ )

$$Bel(A|B) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\bar{A} \cap B)}, \quad (13)$$

$$Pl(A|B) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\bar{A} \cap B)}. \quad (14)$$

Fagin and Halpern in [7] proved that  $Bel(\cdot|B)$  is a true belief function and so FH belief conditioning is an appealing solution for belief and plausibility conditioning. A proof that FH formulas are belief functions has been also given by Sundberg and Wagner in [18]. Hence TBT provides a complete justification of FH formulas which offers a full compatibility with the conditional probability calculus [18], [19].

Similarly, by interchanging notations  $A$  and  $B$  and assuming  $Bel(A) > 0$ , the previous FH formulas can be expressed as

$$Bel(B|A) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(B \cap \bar{A})}, \quad (15)$$

$$Pl(B|A) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(B \cap \bar{A})}. \quad (16)$$

When  $m(\cdot)$  is Bayesian  $Bel(\cdot) = Pl(\cdot) = P(\cdot)$ , and so  $Pl(A \cap B) = Bel(A \cap B) = P(A \cap B)$ ,  $Pl(\bar{A} \cap B) = Bel(\bar{A} \cap B) = P(\bar{A} \cap B)$  and  $Pl(\bar{B} \cap A) = Bel(\bar{B} \cap A) = P(\bar{B} \cap A)$ . FH formulas above reduce to

$$Bel(A|B) = Pl(A|B) = \frac{P(A \cap B)}{P(A \cap B) + P(\bar{A} \cap B)}.$$

From TPT [16])  $P(A \cap B) + P(\bar{A} \cap B) = P(B)$ , thus

$$Bel(A|B) = Pl(A|B) = P(A \cap B)/P(B) = P(A|B). \quad (17)$$

Similarly, one can also easily verify that

$$Bel(B|A) = Pl(B|A) = P(A \cap B)/P(A) = P(B|A). \quad (18)$$

Hence from (17) and (18) one obtains the well-known equality

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A). \quad (19)$$

### IV. GENERALIZED BAYES' THEOREM

In [8] we did also establish from TBT the following Generalized Bayes' Theorem (GBT) and lemma.

**Generalized Bayes' Theorem (GBT):** For any partition  $\{A_1, \dots, A_k\}$  of a FoD  $\Theta$ , any belief function  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$ , and any subset  $B$  of  $\Theta$  with  $Bel(B) > 0$ , one has for  $i \in \{1, \dots, k\}$

$$Bel(A_i|B) = \frac{Bel(B|A_i)q(A_i, B)}{\sum_{i=1}^k Bel(B|A_i)q(A_i, B) + U((\bar{A}_i \cap B)^*)}, \quad (20)$$

where

$$q(A_i, B) \triangleq Bel(A_i) + U((\bar{B} \cap A_i)^*) - U(B^* \cap A_i), \quad (21)$$

with

$$U((\bar{B} \cap A_i)^*) \triangleq Pl(\bar{B} \cap A_i) - Bel(\bar{B} \cap A_i), \quad (22)$$

$$U(B^* \cap A_i) \triangleq \sum_{X \in \mathcal{F}_{B^*} \setminus \{X \in \mathcal{F}_{A_i}(m)\}} m(X), \quad (23)$$

and where

$$U((\bar{A}_i \cap B)^*) \triangleq Pl(\bar{A}_i \cap B) - Bel(\bar{A}_i \cap B). \quad (24)$$

**Lemma 1:** GBT degenerates to Bayes' theorem formula if  $Bel(\cdot)$  is a Bayesian BF, that is

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^k P(B|A_i)P(A_i)}. \quad (25)$$

### V. SIMPLIFIED FORMULATION OF GBT

In this section we establish a simplified formulation of GBT which will be denoted SGBT for short in the sequel. Because the GBT formula (20) is not very easy to use and quite difficult to compute in applications, we propose a more useful simplified formulation of GBT which is drawn from (20) when considering only a simple dichotomous partitioning of the frame of discernment  $\Theta$ . More precisely we consider a partition  $\{A, \bar{A}\}$  of  $\Theta$  with  $A \subseteq \Theta$  and  $\bar{A}$  is the complement of  $A$  in  $\Theta$ , that is  $\bar{A} = \Theta - \{A\}$ . We establish the following theorem which is the main contribution of this paper.

**Simplified Generalized Bayes' Theorem (SGBT):** For any partition  $\{A, \bar{A}\}$  of a FoD  $\Theta$ , any belief function  $Bel(\cdot) : 2^\Theta \mapsto [0, 1]$ , and any subset  $B$  of  $\Theta$ , one has

- If  $Pl(A \cap \bar{B}) > 0$  (Condition  $C_1$ )

$$Bel(A|B) = \frac{Bel(B|A)Pl(A \cap \bar{B})}{Bel(B|A)Pl(A \cap \bar{B}) + Pl(\bar{B}|A)Pl(\bar{A} \cap B)}, \quad (26)$$

- If  $Bel(A \cap \bar{B}) > 0$  (Condition  $C_2$ )

$$Pl(A|B) = \frac{Pl(B|A)Bel(A \cap \bar{B})}{Pl(B|A)Bel(A \cap \bar{B}) + Bel(\bar{B}|A)Bel(\bar{A} \cap B)}, \quad (27)$$

and if the denominators involved in formulas (26) and (27) are strictly positive.

Note that if condition  $C_2$  is satisfied then the condition  $C_1$  is also satisfied, but not necessarily the converse.

**Proof of SGBT:** From GBT formula (20), we replace the terms by their expressions to obtain SGBT formulas (26)–(27). For notation convenience, we denote  $A_1 \triangleq A$  and  $A_2 \triangleq \bar{A}$ . Hence the GBT formula reduces to

$$Bel(A|B) = \frac{Num}{Den} = \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^2 Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)},$$

where

$$\begin{aligned} Num &\triangleq Bel(B|A_1)q(A_1, B), \\ Den &\triangleq Bel(B|A_1)q(A_1, B) \\ &\quad + Bel(B|A_2)q(A_2, B) + U((\bar{A}_1 \cap B)^*), \end{aligned}$$

and

$$\begin{aligned} q(A_1, B) &= Bel(A_1) + U((\bar{B} \cap A_1)^*) - U(B^* \cap A_1) \\ &= Bel(A_1) + \underbrace{Pl(\bar{B} \cap A_1) - Bel(\bar{B} \cap A_1)}_{U((\bar{B} \cap A_1)^*)} - U(B^* \cap A_1). \end{aligned}$$

Because  $\mathcal{F}_{B^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_B(m) - \mathcal{F}_{\bar{B}}(m)$  one has

$$\begin{aligned} U(B^* \cap A_1) &= \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{A_1}(m)} m(X) \\ &= \sum_{X \in \mathcal{F}_\Theta(m) - \mathcal{F}_B(m) - \mathcal{F}_{\bar{B}}(m) | X \in \mathcal{F}_{A_1}(m)} m(X) \\ &= \sum_{X \in \mathcal{F}_\Theta(m) | X \in \mathcal{F}_{A_1}(m)} m(X) \\ &\quad - \sum_{X \in \mathcal{F}_B(m) | X \in \mathcal{F}_{A_1}(m)} m(X) \\ &\quad - \sum_{X \in \mathcal{F}_{\bar{B}}(m) | X \in \mathcal{F}_{A_1}(m)} m(X) \\ &= Bel(A_1) - Bel(A_1 \cap B) - Bel(A_1 \cap \bar{B}). \end{aligned}$$

Therefore

$$\begin{aligned} q(A_1, B) &= Bel(A_1) + \underbrace{Pl(\bar{B} \cap A_1) - Bel(\bar{B} \cap A_1)}_{U((\bar{B} \cap A_1)^*)} \\ &\quad - \underbrace{[Bel(A_1) - Bel(A_1 \cap B) - Bel(A_1 \cap \bar{B})]}_{U(B^* \cap A_1)} \\ &= Pl(A_1 \cap \bar{B}) + Bel(A_1 \cap B). \end{aligned}$$

Similarly, one has

$$\begin{aligned} q(A_2, B) &= Bel(A_2) + U((\bar{B} \cap A_2)^*) - U(B^* \cap A_2) \\ &= Bel(A_2) + \underbrace{Pl(\bar{B} \cap A_2) - Bel(\bar{B} \cap A_2)}_{U((\bar{B} \cap A_2)^*)} \\ &\quad - \underbrace{[Bel(A_2) - Bel(A_2 \cap B) - Bel(A_2 \cap \bar{B})]}_{U(B^* \cap A_2)} \\ &= Pl(A_2 \cap \bar{B}) + Bel(A_2 \cap B). \end{aligned}$$

The value  $U((\bar{A}_1 \cap B)^*)$  is given by

$$U((\bar{A}_1 \cap B)^*) = Pl(\bar{A}_1 \cap B) - Bel(\bar{A}_1 \cap B).$$

Therefore the numerator and denominator of  $Bel(A|B)$  are

$$\begin{aligned} Num &\triangleq Bel(B|A_1)q(A_1, B) \\ &= Bel(B|A_1)[Pl(A_1 \cap \bar{B}) + Bel(A_1 \cap B)] \\ &= Bel(B|A)[Pl(A \cap \bar{B}) + Bel(A \cap B)], \end{aligned}$$

$$\begin{aligned} Den &\triangleq Bel(B|A_1)q(A_1, B) + Bel(B|A_2)q(A_2, B) \\ &\quad + U((\bar{A}_1 \cap B)^*) \\ &= Bel(B|A_1)[Pl(A_1 \cap \bar{B}) + Bel(A_1 \cap B)] \\ &\quad + Bel(B|A_2)[Pl(A_2 \cap \bar{B}) + Bel(A_2 \cap B)] \\ &\quad + [Pl(\bar{A}_1 \cap B) - Bel(\bar{A}_1 \cap B)] \\ &= Bel(B|A)[Pl(A \cap \bar{B}) + Bel(A \cap B)] \\ &\quad + Bel(B|\bar{A})[Pl(\bar{A} \cap \bar{B}) + Bel(\bar{A} \cap B)] \\ &\quad + [Pl(\bar{A} \cap B) - Bel(\bar{A} \cap B)]. \end{aligned}$$

Because  $Bel(B|A) = Bel(A \cap B) / [Bel(A \cap B) + Pl(A \cap \bar{B})]$  and  $Bel(B|\bar{A}) = Bel(\bar{A} \cap B) / [Bel(\bar{A} \cap B) + Pl(\bar{A} \cap \bar{B})]$  based on FH formulas, after basic algebra one can verify that  $Num = Bel(A \cap B)$  and  $Den = Bel(A \cap B) + Pl(\bar{A} \cap B)$ .

Because  $Bel(B|\bar{A}) = Bel(\bar{A} \cap B) / [Bel(\bar{A} \cap B) + Pl(\bar{A} \cap \bar{B})]$ , the term  $Bel(B|\bar{A})[Pl(\bar{A} \cap \bar{B}) + Bel(\bar{A} \cap B)]$  involved in  $Den$  equals  $Bel(\bar{A} \cap B)$ . Hence the expression of  $Den$  reduces to

$$\begin{aligned} Den &= Bel(B|A)[Pl(A \cap \bar{B}) + Bel(A \cap B)] \\ &\quad + \underbrace{Bel(B|\bar{A})[Pl(\bar{A} \cap \bar{B}) + Bel(\bar{A} \cap B)]}_{Bel(\bar{A} \cap B)} \\ &\quad + [Pl(\bar{A} \cap B) - Bel(\bar{A} \cap B)] \\ &= Bel(B|A)[Pl(A \cap \bar{B}) + Bel(A \cap B)] + Pl(\bar{A} \cap B). \end{aligned}$$

If  $Pl(\bar{B}|A) = Pl(A \cap \bar{B}) / [Pl(A \cap \bar{B}) + Bel(A \cap B)] > 0$  and if we multiply the expressions of  $Num$  and  $Den$  by  $Pl(\bar{B}|A)$  one gets

$$\begin{aligned} Bel(A|B) &= \frac{Num}{Den} = \frac{Num \cdot Pl(\bar{B}|A)}{Den \cdot Pl(\bar{B}|A)} \\ &= \frac{Num \cdot \frac{Pl(A \cap \bar{B})}{Pl(A \cap \bar{B}) + Bel(A \cap B)}}{Den \cdot \frac{Pl(A \cap \bar{B})}{Pl(A \cap \bar{B}) + Bel(A \cap B)}} \\ &= \frac{Bel(B|A)Pl(A \cap \bar{B})}{Bel(B|A)Pl(A \cap \bar{B}) + Pl(\bar{B}|A)Pl(\bar{A} \cap B)}, \end{aligned}$$

which corresponds exactly to the SGBT formula (26).

The SGBT formula (27) can also be obtained similarly from GBT by expressing at first  $Bel(\bar{A}|B) = Bel(A_2|B)$  as

$$\begin{aligned} Bel(\bar{A}|B) &= \frac{Num'}{Den'} = \frac{Bel(B|A_2)q(A_2, B)}{\sum_{i=1}^2 Bel(B|A_i)q(A_i, B) + U((\bar{A}_2 \cap B)^*)} \\ &= \frac{Bel(B|A_2)q(A_2, B)}{Bel(B|A_1)q(A_1, B) + Bel(B|A_2)q(A_2, B) + U((\bar{A}_2 \cap B)^*)}, \end{aligned}$$

where<sup>4</sup>

$$\begin{aligned} Num' &\triangleq Bel(B|A_2)q(A_2, B) \\ &= Bel(B|\bar{A})[Pl(\bar{A} \cap \bar{B}) + Bel(\bar{A} \cap B)] \\ &= Bel(\bar{A} \cap B), \\ Den' &\triangleq Bel(B|A_1)q(A_1, B) + Bel(B|A_2)q(A_2, B) \\ &\quad + U((\bar{A}_2 \cap B)^*) \\ &= Bel(B|A)[Pl(A \cap \bar{B}) + Bel(A \cap B)] \\ &\quad + Bel(B|\bar{A})[Pl(\bar{A} \cap \bar{B}) + Bel(\bar{A} \cap B)] \\ &\quad + [Pl(A \cap B) - Bel(A \cap B)] \\ &= Bel(\bar{A} \cap B) + Pl(A \cap B). \end{aligned}$$

<sup>4</sup>Here  $U((\bar{A}_2 \cap B)^*) = Pl(\bar{A}_2 \cap B) - Bel(\bar{A}_2 \cap B) = Pl(A \cap B) - Bel(A \cap B)$  because  $\bar{A}_2 = \bar{A} = A$ .

If  $Bel(\bar{B}|A) = Bel(A \cap \bar{B}) / [Bel(A \cap \bar{B}) + Pl(A \cap B)] > 0$  and if we multiply  $Num'$  and  $Den'$  by  $Bel(\bar{B}|A)$  one gets<sup>5</sup>

$$\begin{aligned} Bel(\bar{A}|B) &= \frac{Num'}{Den'} = \frac{Num' \cdot Bel(\bar{B}|A)}{Den' \cdot Bel(\bar{B}|A)} \\ &= \frac{Bel(\bar{A} \cap B) Bel(\bar{B}|A)}{Bel(\bar{A} \cap B) Bel(\bar{B}|A) + Pl(A \cap B) Bel(\bar{B}|A)} \\ &= \frac{Bel(\bar{A} \cap B) Bel(\bar{B}|A)}{Bel(\bar{A} \cap B) Bel(\bar{B}|A) + Pl(B|A) Bel(A \cap \bar{B})}. \end{aligned}$$

Hence

$$\begin{aligned} Pl(A|B) &= 1 - Bel(\bar{A}|B) \\ &= \frac{Pl(B|A) Bel(A \cap \bar{B})}{Pl(B|A) Bel(A \cap \bar{B}) + Bel(\bar{B}|A) Bel(\bar{A} \cap B)}, \end{aligned}$$

which corresponds to SGBT formula (27).

Therefore, one has proved that expression (26) can be obtained from GBT if  $Pl(A \cap \bar{B}) > 0$ , and expression (27) can be obtained from GBT if  $Bel(A \cap \bar{B}) > 0$ . This completes the proof of SGBT.

**Lemma 2:** SGBT formulas (26) and (27) coincide with conditional probability formula  $P(A|B) = P(B|A)P(A)/P(B) = P(A \cap B)/P(B)$  if the belief function is Bayesian.

**Proof:** Replacing  $Bel(\cdot)$  and  $Pl(\cdot)$  by  $P(\cdot)$  in (26) and (27) we get  $P(A|B) = \frac{P(B|A)P(A \cap \bar{B})}{P(B|A)P(A \cap \bar{B}) + P(\bar{B}|A)P(A \cap B)} = \frac{P(B|A)P(A \cap \bar{B})}{P(B|A)P(A)} = \frac{P(A \cap \bar{B})P(A \cap \bar{B}) + \frac{P(A \cap \bar{B})}{P(A)}P(A \cap B)}{P(A \cap \bar{B}) + P(A \cap B)}$  because  $P(A \cap B) + P(\bar{A} \cap B) = P(B)$ . This completes the proof of lemma 2.

In appendix we also prove that  $Bel(A|B) \leq Pl(A|B)$  when using SGBT formulas (26) and (27).

## VI. EXAMPLES

In this section we give two simple interesting examples of application of SGBT. Example 1 shows that GBT and SGBT works fine because conditions  $C_1$  and  $C_2$  are satisfied, whereas the example 2 shows that GBT works fine but SGBT doesn't work because of violation of condition  $C_1$ .

### A. Example 1

We consider  $\Theta = \{x_1, x_2, x_3, x_4\}$  and the BBA chosen as follows  $m(x_1) = 0.05$ ,  $m(x_2) = 0.03$ ,  $m(x_1 \cup x_2) = 0.02$ ,  $m(x_3) = 0.04$ ,  $m(x_4) = 0.06$ ,  $m(x_3 \cup x_4) = 0.10$ ,  $m(x_2 \cup x_3) = 0.30$  and  $m(x_1 \cup x_2 \cup x_3 \cup x_4) = m(\Theta) = 0.40$ . We also consider the partition  $\Theta = \{A = \{x_1, x_2\}, \bar{A} = \{x_3, x_4\}\}$  and the subset  $B = \{x_2, x_3\}$ . Hence one has

$$\Theta = \underbrace{\{x_1, x_2\}}_A \cup \underbrace{\{x_3, x_4\}}_{\bar{A}}.$$

with  $A = \{x_1, x_2\} = x_1 \cup x_2$ ,  $\bar{A} = \{x_3, x_4\} = x_3 \cup x_4$ ,  $B = \{x_2, x_3\} = x_2 \cup x_3$ , and  $\bar{B} = \{x_1, x_4\} = x_1 \cup x_4$ .

<sup>5</sup>From FH formulas  $Pl(A \cap B) Bel(\bar{B}|A) = Pl(B|A) Bel(A \cap \bar{B})$ .

The set of focal elements in this example is

$$\mathcal{F}_\Theta(m) = \{x_1, x_2, x_1 \cup x_2, x_3, x_4, x_3 \cup x_4, x_2 \cup x_3, x_1 \cup x_2 \cup x_3 \cup x_4\}.$$

The sets of focal elements included in  $A$  and in  $\bar{A}$  are  $\mathcal{F}_A(m) = \{x_1, x_2, x_1 \cup x_2\}$ , and  $\mathcal{F}_{\bar{A}}(m) = \{x_3, x_4, x_3 \cup x_4\}$ , and one has  $\mathcal{F}_{A^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_A(m) - \mathcal{F}_{\bar{A}}(m) = \{x_2 \cup x_3, x_1 \cup x_2 \cup x_3 \cup x_4\}$ . The sets of focal elements included in  $B$  and in  $\bar{B}$  are  $\mathcal{F}_B(m) = \{x_2, x_3, x_2 \cup x_3\}$ ,  $\mathcal{F}_{\bar{B}}(m) = \{x_1, x_4\}$ , and one has  $\mathcal{F}_{B^*}(m) = \mathcal{F}_\Theta(m) - \mathcal{F}_B(m) - \mathcal{F}_{\bar{B}}(m) = \{x_1 \cup x_2, x_3 \cup x_4, x_1 \cup x_2 \cup x_3 \cup x_4\}$ . From the BBA  $m(\cdot)$  we get the following belief and plausibility values listed in Table I which are useful for making derivations of FH, GBT and SGBT formulas.

Subsets of $\Theta$	$Bel(\cdot)$	$Pl(\cdot)$
$A = x_1 \cup x_2$	$Bel(A) = 0.10$	$Pl(A) = 0.80$
$\bar{A} = x_3 \cup x_4$	$Bel(\bar{A}) = 0.20$	$Pl(\bar{A}) = 0.90$
$B = x_2 \cup x_3$	$Bel(B) = 0.37$	$Pl(B) = 0.89$
$\bar{B} = x_1 \cup x_4$	$Bel(\bar{B}) = 0.11$	$Pl(\bar{B}) = 0.63$
$A \cap B = x_2$	$Bel(A \cap B) = 0.03$	$Pl(A \cap B) = 0.75$
$A \cap \bar{B} = x_1$	$Bel(A \cap \bar{B}) = 0.05$	$Pl(A \cap \bar{B}) = 0.47$
$\bar{A} \cap B = x_3$	$Bel(\bar{A} \cap B) = 0.04$	$Pl(\bar{A} \cap B) = 0.84$
$\bar{A} \cap \bar{B} = x_4$	$Bel(\bar{A} \cap \bar{B}) = 0.06$	$Pl(\bar{A} \cap \bar{B}) = 0.56$

Table I

BELIEF AND PLAUSIBILITY VALUES USED FOR THE DERIVATIONS.

- Application of FH formulas: with (13)-(14) one gets

$$\begin{aligned} Bel(A|B) &= \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\bar{A} \cap B)} = \frac{0.03}{0.03 + 0.84} \\ &\approx 0.03448275, \\ Pl(A|B) &= \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\bar{A} \cap B)} = \frac{0.75}{0.75 + 0.04} \\ &\approx 0.94936708. \end{aligned}$$

With FH formulas (15)-(16), one gets

$$\begin{aligned} Bel(B|A) &= \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\bar{B} \cap A)} = \frac{0.03}{0.03 + 0.47} \\ &= 0.06, \\ Pl(B|A) &= \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(\bar{B} \cap A)} = \frac{0.75}{0.75 + 0.05} \\ &= 0.9375. \end{aligned}$$

- Application of SGBT formulas: with (26) and (27) one gets<sup>6</sup>

$$\begin{aligned} Bel(A|B) &= \frac{Bel(B|A) Pl(A \cap \bar{B})}{Bel(B|A) Pl(A \cap \bar{B}) + Pl(\bar{B}|A) Pl(\bar{A} \cap B)} \\ &= \frac{Bel(B|A) Pl(A \cap \bar{B})}{Bel(B|A) Pl(A \cap \bar{B}) + [1 - Bel(B|A)] Pl(\bar{A} \cap B)} \\ &= \frac{0.06 \cdot 0.47}{0.06 \cdot 0.47 + [1 - 0.06] \cdot 0.84} = \frac{0.0282}{0.0282 + 0.7896} \\ &\approx 0.03448275, \end{aligned}$$

<sup>6</sup>It is worth noting that conditions  $C_1$  and  $C_2$  are satisfied in this example because  $Pl(A \cap \bar{B}) = 0.47$  and  $Bel(A \cap \bar{B}) = 0.05$ .

$$\begin{aligned}
 Pl(A|B) &= \frac{Pl(B|A)Bel(A \cap \bar{B})}{Bel(\bar{B}|A)Bel(\bar{A} \cap B) + Pl(B|A)Bel(A \cap \bar{B})} \\
 &= \frac{Pl(B|A)Bel(A \cap \bar{B})}{[1 - Pl(B|A)]Bel(\bar{A} \cap B) + Pl(B|A)Bel(A \cap \bar{B})} \\
 &= \frac{0.9375 \cdot 0.05}{[1 - 0.9375]0.04 + 0.9375 \cdot 0.05} \\
 &= \frac{0.046875}{0.0025 + 0.046875} \approx 0.94936708.
 \end{aligned}$$

• Application of GBT formula (20): we denote  $A_1 = A = x_1 \cup x_2$  and  $A_2 = \bar{A} = x_3 \cup x_4$ . Here GBT formula (20) becomes

$$Bel(A|B) = \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^2 Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)},$$

where  $Bel(B|A_1)$  and  $Bel(B|A_2)$  terms are given by

$$\begin{aligned}
 Bel(B|A_1) &\equiv Bel(B|A) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(\bar{B} \cap A)} \\
 &= \frac{0.03}{0.03 + 0.47} = 0.06, \\
 Bel(B|A_2) &\equiv Bel(B|\bar{A}) = \frac{Bel(\bar{A} \cap B)}{Bel(\bar{A} \cap B) + Pl(\bar{B} \cap \bar{A})} \\
 &= \frac{0.04}{0.04 + 0.56} \approx 0.06666667.
 \end{aligned}$$

The terms  $q(A_1, B)$  and  $q(A_2, B)$  are given by

$$\begin{aligned}
 q(A_1, B) &= Bel(A_1) + U((\bar{B} \cap A_1)^*) - U(B^* \cap A_1) \\
 &= 0.10 + 0.42 - 0.02 = 0.50, \\
 q(A_2, B) &= Bel(A_2) + U((\bar{B} \cap A_2)^*) - U(B^* \cap A_2) \\
 &= 0.20 + 0.50 - 0.10 = 0.60,
 \end{aligned}$$

because

$$\begin{aligned}
 U((\bar{B} \cap A_1)^*) &= Pl(\bar{B} \cap A_1) - Bel(\bar{B} \cap A_1) \\
 &= Pl(\bar{B} \cap A) - Bel(\bar{B} \cap A) \\
 &= 0.47 - 0.05 = 0.42, \\
 U(B^* \cap A_1) &= \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{A_1}(m)} m(X) \\
 &= \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_A(m)} m(X) \\
 &= m(x_1 \cup x_2) = 0.02,
 \end{aligned}$$

and

$$\begin{aligned}
 U((\bar{B} \cap A_2)^*) &= Pl(\bar{B} \cap A_2) - Bel(\bar{B} \cap A_2) \\
 &= Pl(\bar{B} \cap \bar{A}) - Bel(\bar{B} \cap \bar{A}) \\
 &= 0.56 - 0.06 = 0.50, \\
 U(B^* \cap A_2) &= \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{A_2}(m)} m(X) \\
 &= \sum_{X \in \mathcal{F}_{B^*}(m) | X \in \mathcal{F}_{\bar{A}}(m)} m(X) \\
 &= m(x_3 \cup x_4) = 0.10.
 \end{aligned}$$

The value  $U((\bar{A}_1 \cap B)^*)$  involved in the denominator of  $Bel(A|B)$  expression is given by

$$\begin{aligned}
 U((\bar{A}_1 \cap B)^*) &= Pl(\bar{A} \cap B) - Bel(\bar{A} \cap B) \\
 &= 0.84 - 0.04 = 0.80.
 \end{aligned}$$

Replacing all these values in GBT formula of  $Bel(A|B)$  we get

$$\begin{aligned}
 Bel(A|B) &\equiv Bel(A_1|B) \\
 &= \frac{Bel(B|A_1)q(A_1, B)}{\sum_{i=1}^2 Bel(B|A_i)q(A_i, B) + U((\bar{A}_1 \cap B)^*)} \\
 &\approx \frac{0.06 \cdot 0.50}{0.06 \cdot 0.50 + 0.06666667 \cdot 0.60 + 0.80} \\
 &\approx \frac{0.03}{0.870000002} \approx 0.03448275.
 \end{aligned}$$

As shown,  $Bel(A|B)$  calculated by GBT and by SGBT formulas are consistent with the value calculated directly from FH formulas. For calculating  $Pl(A|B)$ , we calculate at first  $Bel(\bar{A}|B) = Bel(A_2|B)$  and then  $Pl(A|B) = 1 - Bel(\bar{A}|B)$ . Applying GBT formula for calculating  $Bel(A_2|B)$ , one has

$$\begin{aligned}
 Bel(\bar{A}|B) &= Bel(A_2|B) \\
 &= \frac{Bel(B|A_2)q(A_2, B)}{\sum_{i=1}^2 Bel(B|A_i)q(A_i, B) + U((\bar{A}_2 \cap B)^*)}.
 \end{aligned}$$

The values of  $Bel(B|A_i)$ ,  $q(A_i, B)$  for  $i = 1, 2$  have been calculated previously and  $U((\bar{A}_2 \cap B)^*)$  is given by

$$\begin{aligned}
 U((\bar{A}_2 \cap B)^*) &= Pl(\bar{A}_2 \cap B) - Bel(\bar{A}_2 \cap B) \\
 &= Pl(A \cap B) - Bel(A \cap B) \\
 &= 0.75 - 0.03 = 0.72.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 Bel(\bar{A}|B) &\equiv Bel(A_2|B) \\
 &= \frac{Bel(B|A_2)q(A_2, B)}{\sum_{i=1}^2 Bel(B|A_i)q(A_i, B) + U((\bar{A}_2 \cap B)^*)} \\
 &\approx \frac{0.06666667 \cdot 0.60}{0.06 \cdot 0.50 + 0.06666667 \cdot 0.60 + 0.72} \\
 &\approx \frac{0.040000002}{0.790000002} \approx 0.05063292.
 \end{aligned}$$

and finally we get

$$Pl(A|B) = 1 - Bel(\bar{A}|B) \approx 0.94936708.$$

From this very simple example we have verified that FH formulas, GBT formula and simplified GBT formula are all consistent because the conditions  $C_1$  and  $C_2$  are satisfied.

## B. Example 2

We consider the example of [8] (Section VIII). We verify that SGBT formula (26) works because  $Bel(B|A_1) = 0.0889$ ,  $Pl(A_1 \cap \bar{B}) = 0.41$ ,  $Pl(\bar{B}|A_1) = 1 - Bel(B|A_1) = 1 - 0.0889 = 0.9111$  and  $Pl(\bar{A}_1 \cap B) = 0.54$  so that

$$\begin{aligned} Bel(A_1|B) &= \frac{Bel(B|A_1)Pl(A_1 \cap \bar{B})}{Bel(B|A_1)Pl(A_1 \cap \bar{B}) + Pl(\bar{B}|A_1)Pl(\bar{A}_1 \cap B)} \\ &= \frac{0.0889 \cdot 0.41}{0.0889 \cdot 0.41 + 0.9111 \cdot 0.54} = 0.0690. \end{aligned}$$

which is the same value of what we get by applying directly FH formula, or GBT formula (20). The SGBT formula (26) works because the condition  $C_1$  (i.e.  $Pl(A_1 \cap \bar{B}) = 0.41 > 0$ ) is satisfied. Similarly, using (26), one has for  $Bel(A_2|B)$

$$\begin{aligned} Bel(A_2|B) &= \frac{Bel(B|A_2)Pl(A_2 \cap \bar{B})}{Bel(B|A_2)Pl(A_2 \cap \bar{B}) + Pl(\bar{B}|A_2)Pl(A_2 \cap B)} \\ &= \frac{0 \cdot 0.43}{0 \cdot 0.43 + 1 \cdot 0.80} = 0, \end{aligned}$$

which is the same value of what we get by applying directly FH formula, or GBT formula (20). Here SGBT formula (26) works because the condition  $C_1$  (i.e.  $Pl(A_2 \cap \bar{B}) = 0.43 > 0$ ) is satisfied.

For the value  $Bel(A_3|B) = 0.0625$  computed by FH conditioning formula, or by GBT formula (20) things are different because when applying SGBT formula (26) we get 0/0 indetermination. Indeed,

$$\begin{aligned} Bel(A_3|B) &= \frac{Bel(B|A_3)Pl(A_3 \cap \bar{B})}{Bel(B|A_3)Pl(A_3 \cap \bar{B}) + Pl(\bar{B}|A_3)Pl(\bar{A}_3 \cap B)} \\ &= \frac{1 \cdot 0}{1 \cdot 0 + 0 \cdot 0.75} = \frac{0}{0}. \end{aligned}$$

So one sees that SGBT formula (26) does not work for computing  $Bel(A_3|B)$  in this case because the condition  $C_1$  (i.e.  $Pl(A_3 \cap \bar{B}) > 0$ ) is not satisfied which is normal. In this case the correct value  $Bel(A_3|B) = 0.0625$  must be calculated by GBT or FH formulas.

Therefore in practice a special attention must always be paid to conditions  $C_1$  and  $C_2$  before applying SGBT formulas, and in case of violation of one of these conditions, one needs to work back directly with FH or GBT formulas.

## VII. CONCLUSION

The main contribution of this paper is the derivation of a simplified formulation of Generalized Bayes' Theorem, called SGBT, which extends Bayesian Theorem in the frame of belief functions. The simplification is imposed from the fact that the general formulation of GBT is not easy to apply in real world applications. It is drawn from GBT for working with a dichotomous partitioning of the frame of discernment. The conditions under which this new formulation is valid are presented. The theoretical results obtained are illustrated with simple theoretical examples. The challenging question of application of GBT and SGBT to solve real-world problems is under investigation.

### A. Proof that $Bel(A|B) \leq Pl(A|B)$ from SGBT formula

To prove that  $Bel(A|B) \leq Pl(A|B)$  from SGBT formulas (26)-(27) one needs to prove the following inequality

$$\frac{Bel(B|A)Pl(A \cap \bar{B})}{Bel(B|A)Pl(A \cap \bar{B}) + Pl(\bar{B}|A)Pl(\bar{A} \cap B)} \leq \frac{Pl(B|A)Bel(A \cap \bar{B})}{Bel(\bar{B}|A)Bel(\bar{A} \cap B) + Pl(B|A)Bel(A \cap \bar{B})}.$$

After basic algebraic manipulations on the previous inequality, one has to prove if  $R_1 \leq R_2 \cdot R_3 \cdot R_4$ . where, for the notation convenience,  $R_1 = Bel(B|A)/Pl(B|A)$ ,  $R_2 = Bel(A \cap \bar{B})/Pl(A \cap \bar{B})$ ,  $R_3 = Pl(\bar{B}|A)/Bel(\bar{B}|A)$  and  $R_4 = Pl(\bar{A} \cap B)/Bel(\bar{A} \cap B)$ . Our proof is done by contradiction as follows.

Let us assume that  $R_2 \cdot R_3 \cdot R_4 < R_1$  is valid, that is

$$\underbrace{\frac{Bel(A \cap \bar{B})}{Pl(A \cap \bar{B})}}_{R_2} \cdot \underbrace{\frac{Pl(\bar{B}|A)}{Bel(\bar{B}|A)}}_{R_3} \cdot \underbrace{\frac{Pl(\bar{A} \cap B)}{Bel(\bar{A} \cap B)}}_{R_4} < \underbrace{\frac{Bel(B|A)}{Pl(B|A)}}_{R_1} \quad (28)$$

Because  $R_2 \leq 1$ , one has necessarily  $R_2 \cdot R_3 \cdot R_4 \leq R_3 \cdot R_4$ , so we must have (if our assumption is valid)  $R_3 \cdot R_4 < R_1$ , that is

$$\underbrace{\frac{1 - Bel(B|A)}{1 - Pl(B|A)}}_{R_3} \cdot \underbrace{\frac{Pl(\bar{A} \cap B)}{Bel(\bar{A} \cap B)}}_{R_4} < \underbrace{\frac{Bel(B|A)}{Pl(B|A)}}_{R_1}, \quad (29)$$

or equivalently

$$[1 - Bel(B|A)]Pl(B|A) < Bel(B|A)[1 - Pl(B|A)] \underbrace{\frac{Bel(\bar{A} \cap B)}{Pl(\bar{A} \cap B)}}_{1/R_4}.$$

Because  $Bel(\bar{A} \cap B)/Pl(\bar{A} \cap B) \leq 1$  then

$$Bel(B|A)[1 - Pl(B|A)] \underbrace{\frac{Bel(\bar{A} \cap B)}{Pl(\bar{A} \cap B)}}_{1/R_4} \leq Bel(B|A)[1 - Pl(B|A)].$$

So we must have (if our assumption is valid)

$$[1 - Bel(B|A)]Pl(B|A) < Bel(B|A)[1 - Pl(B|A)], \quad (30)$$

which is (after rearranging terms) equivalent to have the inequality  $Pl(B|A) < Bel(B|A)$  satisfied. However, from Fagin-Halpern definitions of conditional belief function and properties of belief functions the previous inequality  $Pl(B|A) < Bel(B|A)$  is never satisfied. Therefore our assumption  $R_2 \cdot R_3 \cdot R_4 < R_1$  is not valid and one has necessarily  $R_1 \leq R_2 \cdot R_3 \cdot R_4$ , which completes the proof that  $Bel(A|B) \leq Pl(A|B)$  when  $Bel(A|B)$  and  $Pl(A|B)$  are calculated by the SGBT formulas (26) and (27).

## REFERENCES

- [1] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, Ann. of Math. Stat., Vol. 38, pp. 325–339, 1967.
- [2] A.P. Dempster, *A generalization of Bayesian inference*, J. of Royal Stat. Soc., Vol. B30, pp. 205–247, 1968.
- [3] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Press, 1976.
- [4] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, in Proc. of Fusion 2012 Conf., Singapore, 9–12 July 2012.
- [5] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. J. of Intell. Syst., Vol. 29, pp. 223–252, 2014 (with erratum in DSMT Book - Vol. 5).
- [6] F. Smarandache, J. Dezert, A. Tchamova, *Examples where Dempster's rule is insensitive to the conflict level between the sources of evidence*, Octogon Math. Magazine, Vol. 25, No. 2, pp. 284–290, October 2017.
- [7] R. Fagin, J.Y. Halpern, *A new approach to updating beliefs*, in Proc. of 6th UAI Conf., pp. 347–374, Cambridge, MA, USA, 27–29 July 1990 .
- [8] J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Generalized Bayes' Theorem*, in Proc. of Fusion 2018 Conf., Cambridge, UK, 10–13 July 2018.
- [9] L. Zadeh, *A Mathematical Theory of Evidence (book review)*, AI Magazine, Vol. 5, pp. 81–83, 1984.
- [10] J. Lemmer, *Confidence factors, empiricism and the Dempster-Shafer theory of evidence*, Machine Intelligence and Pattern Recognition, Vol. 4, pp. 117–125, 1986.
- [11] P.K. Black, *Is Shafer general Bayes?*, In Proc. of 3rd UAI Conf., Seattle, WA, USA, 1987.
- [12] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publ. Inc., San Mateo, CA, USA, 1988.
- [13] P. Wang, *A defect in Dempster-Shafer theory*, in Proc of 10th UAI Conf., pp. 560–566, Univ. of Washington, Seattle, WA, USA, 29–31 July 1994.
- [14] J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Conditional Belief Functions*, Int. Journal of Intelligent Systems, Vol. 33, No. 12, pp. 2314–2340, December 2018.
- [15] P. Smets, R. Kennes, *The Transferable Belief Model*, Artificial Intelligence, Vol. 66, No. 2, pp.191–234, 1994.
- [16] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd edition, McGraw-Hill, New York, pp. 37–38, 1984.
- [17] J.Y. Halpern, R. Fagin, *Two views of belief: belief as generalized probability and belief as evidence*, Artificial Intelligence, Vol. 54, pp. 275–317, 1992.
- [18] C. Sunberg, C. Wagner, *Generalized finite differences and Bayesian conditioning of Choquet capacities*, Advances in Applied Mathematics, Vol. 13, No. 3, pp. 262–272, September 1992.
- [19] J.N Heendeni et al., *A generalization of Bayesian inference in the Dempster-Shafer belief theoretic framework*, in Proc. of Fusion 2016, Heidelberg, Germany, 5–8 July 2016.



# A New Adaptive Switching Median Filter for Impulse Noise Reduction with Pre-detection Based on Evidential Reasoning

Zhe Zhang<sup>a</sup>, Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>

<sup>a</sup>MOE KLINNS Lab, Institute of Integrated Automation,  
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>c</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China.

Emails: zhezhang@stu.xjtu.edu.cn, deqhan@xjtu.edu.cn, jean.dezert@onera.fr, jiafeiyy@mail.xjtu.edu.cn

Originally published as: Z. Zhang, D. Han, J. Dezert, Y. Yang, *A New Adaptive Switching Median Filter for Impulse Noise Reduction with Pre-Detection Based on Evidential Reasoning*, Signal Processing, Vol. 147, pp. 173–198, 2018, and reprinted with permission.

**Abstract**—Image denoising is a fundamental problem in image processing. The switching filtering is a popular approach to reduce the impulse noise. It faces two challenges including the impulse noise detection and filter design. The traditional detection methods based on single criterion or multiple criteria encounter uncertainty problems and produce many miss-detections and false alarms, especially when the image is severely corrupted. In this paper, the uncertainties encountered in the impulse noise detection are addressed using the theory of belief functions, and a multi-criteria detection strategy for the impulse noise based on evidential reasoning is proposed. Based on the pre-detection, an adaptive median filter is designed, which adaptively determines the size of the filtering window according to the estimated global noise density and the degree of local corruption. Experimental results and related analyses show that our proposed image denoising method for the impulse noise has superior performance compared with several state-of-the-art denoising methods.

**Keywords:** image denoising, impulse noise, multi-criteria detection, evidential reasoning, adaptive median filtering.

## I. INTRODUCTION

Digital images can be corrupted by various types of noise during the image acquisition and transmission. The impulse noise is one of the most common types, which is encountered in cases with quick transients, e.g., faulty switching during imaging [1]. The intensity of a pixel corrupted by the impulse noise tends to be much higher or lower than those of its uncorrupted neighbors. The impulse noise dramatically influences the image quality and makes images unsuitable for subsequent human understanding or image processing such as the edge detection [2], segmentation [3], object recognition [4], image analysis [5] and image understanding [6].

Till now, the impulse noise reduction problem has not been well solved and has attracted extensive research interests. The median filtering is the most popular approaches for the impulse noise reduction. The standard median (SM) filter [7] replaces the target pixel's intensity by the median of intensities of its neighbors. Various modifications of the SM filter have been proposed, such as the weighted median (WM) filter [8] and the center weighted median (CWM) filter [9]. However, all these filters apply the median operations to each pixel

ignoring whether the target pixel is corrupted or not. This might destroy the details contributed from uncorrupted pixels and lead to image quality degradation. To deal with this problem, switching median filters [10] were proposed, which introduce the noise detection prior to the filtering. Since only the corrupted pixels will be filtered and the uncorrupted pixels remain intact, more details can be preserved and better filtering performance can be achieved if the pre-detection result is accurate enough.

In recent years, sparse representation (SR) [11] is widely used in image denoising [12], [13], [14], especially for Gaussian noise. For the impulse noise, the noise detector is incorporated into SR model and the weighted dictionary learning method was proposed for impulse noise denoising [15], [16], [17]. Both median filtering and SR based method face the challenge of noise detector designing.

There have emerged two major criteria for the impulse noise detection including the extreme property and discontinuity property. Some detectors only use a single criterion, which may involve some uncertainty problems. For example, the boundary discriminative noise detection (BDND) [18] and the efficient improvements on the BDND (IBDND) [19] use the criterion of extreme property. Both algorithms label a pixel as the noise if it is assigned to the low-intensity range or high-intensity range according to the histogram distribution in a local window centered at that given pixel. However, these detectors easily lead to false alarms since not all the pixels with low-intensity or high-intensity are noise. There are other detectors that only use the criterion of discontinuity property. Such detectors can be found in the adaptive impulse detection using center-weighted median (ACWM) filter [20], directional weighted median (DWM) filter [21], contrast enhancement-based (CEF) filter [22], adaptive switching median (ASWM) filter [23], weighted couple sparse representation (WCSR) model [24] and the denoising framework combining the detection mechanism based on the robust outlyingness ratio with the NL-means (ROR-NLM) [25]. They label a pixel as the noise if its similarity with its neighbors is lower than a preset threshold.

However, when the noise density is high, the impulse noise pixels might not show the discontinuity property since there are too many noise pixels in their neighbors. Since pixels with extreme or discontinuity property may not always be the noise, the detectors based on a single criterion will involve uncertainty problems and tend to yield incorrect detection results. Both criteria have their own rationalities; however, they are one-sided. It should be better to jointly use them when detecting the impulse noise. Therefore, some approaches using the above two criteria jointly have been proposed, e.g., the detector in noise adaptive soft-switching median (NASM) filter [26] and the detector based on the cloud model (CM) [27]. These two-step detection methods first recognize the suspected noise pixels using the extreme criterion, and then distinguish the noise pixels from the suspected noise pixels using the discontinuity criterion. However, they can easily produce miss-detections when some noise pixels are not detected as the suspected noise in the first step. Therefore, the two-step type joint use is not preferred.

To deal with the uncertainties encountered in the impulse noise detection and avoid the drawbacks of the two-step-type joint use of detection criteria, in this paper, a new detection approach for the impulse noise is proposed, which uses the two criteria simultaneously based on the theory of belief functions [28]. In our detection approach, the extreme property is described using the interval data distance between the target pixel's intensity and the intensity range of the whole noisy image (expressed as an interval number). The discontinuity property is described using the rank-ordered absolute differences (ROAD) statistic [29]. The uncertainty problem encountered in the impulse noise detection, e.g., pixels with extreme or discontinuity property may not always be the noise, are modeled by belief functions and are further handled through the evidence combination.

The impulse noise detector implementation is the main work of this paper. Based on the detection result, an adaptive median filter is designed, which adaptively determines the size of filtering window according to the estimated global noise density and local corrupted degree. Experimental results show that our proposed adaptive switching median filter with pre-detection based on evidential reasoning (ASMF-DBER) has superior performance compared with several state-of-the-air switch median filters and the SR based method.

## II. BASIS OF IMPULSE NOISE AND UNCERTAINTY

### PROBLEMS ENCOUNTERED IN IMPULSE NOISE DETECTION

#### A. Impulse noise model

When an image is corrupted by the impulse noise, some pixels are changed and their intensities are extremely high or extremely low. We use the same impulse noise model as used in BDND [18]. Assume that the noise pixels take values in two fixed sets  $S_1 = \{0, 1, \dots, \alpha\}$  and  $S_2 = \{255 - \alpha, 255 - (\alpha - 1), \dots, 255\}$  for an 8-bit monochrome image. Let  $s_{i,j}$  and  $x_{i,j}$  be the pixels' intensities at location  $(i, j)$  in the original and noisy images, respectively. Let  $n_{i,j}$  be the noise which is independent of  $s_{i,j}$  and corresponds to a

random value uniformly distributed in the set  $S_1$  and  $S_2$ . Let  $p$  denote the probability that a pixel is corrupted. The probability mass function (pmf) [30] of  $x_{i,j}$  is given by

$$P(x_{i,j}) = \begin{cases} p, & \text{for } x_{i,j} = n_{i,j}, \\ 1 - p, & \text{for } x_{i,j} = s_{i,j}. \end{cases} \quad (1)$$

Specially, if  $\alpha = 0$ , the intensities of noise pixels can only take the two extreme values 0 or 255. This type of impulse noise is also called the salt-and-pepper noise. Since  $n_{i,j}$  is independent of  $s_{i,j}$ , it is possible that  $n_{i,j} = s_{i,j}$ . This kind of pixel should be regarded as uncorrupted.

#### B. Uncertainties encountered in impulse noise detection

The impulse noise has two properties:

- 1) *Extreme property*: The intensity of an impulse noise pixel is usually an extreme value (0 or 255) or close to an extreme value.
- 2) *Discontinuity property*: The intensity of an impulse noise pixel tends to be much higher or lower than those of its neighbors.

These two properties are often used as detection criteria for the impulse noise. Some detectors only use one of the criteria:

- 1) Detectors based on the criterion of extreme property: These detectors label a pixel as the noise, if it is assigned to the low-intensity range or high-intensity range according to the histogram distribution in a local window centered at that pixel, e.g., BDND [18] and IBDND [19] detectors.
- 2) Detectors based on the criterion of discontinuity property: These detectors label a pixel as the noise, if its dissimilarity with its neighbors is larger than a preset threshold, such as ACWM [20], DWM [21], CEF [22], ASWM [23], ROR-NLM [25] and WCSR [24].

However, such single criterion based detectors may involve the following uncertainty problems:

- 1) Uncertainty in extreme criterion: Some signal pixels may also be detected as the noise, since their intensities are very close to extreme values, e.g., some edge pixels and texture pixels. Moreover, in some bright or dark area, the intensity range of signal pixels may overlap with that of the impulse noise pixels. Therefore, when using the extreme criterion alone, it is uncertain to judge those signal pixels with extreme property to be the impulse noise or not.
- 2) Uncertainty in discontinuity criterion: The discontinuity property of the impulse noise pixels becomes weaker with the increase of noise density since there are many noise pixels in their neighbors. At the same time, some signal pixels may show discontinuity. Therefore, with only the discontinuity criterion, it is uncertain to judge a pixel to be the impulse noise or not.

Due to these uncertainties, the single criterion based detectors are to some extent one-sided and tend to yield incorrect detection results. Hence, it should be better to jointly use the two criteria to implement a more comprehensive detection.

Some two-step detection methods, like NASM [26] and CM [27], jointly use these two criteria in two consecutive steps. They first recognize suspected noise pixels according to the extreme criterion, and then distinguish noise pixels from suspected noise pixels according to the discontinuity criterion, as illustrated in Fig. 1. In the first step, only using the extreme criterion, some noise pixels may not be detected as the suspected noise and therefore are miss-detected straightly. These miss-detected pixels will not undergo the filtering so that these two-step methods can easily lead to poor noise-reduction capabilities. Therefore, when detecting the impulse noise, it should be better to use these two criteria simultaneously, but not in two consecutive steps.

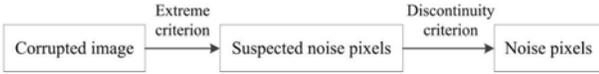


Figure 1. The two-step detection method.

To deal with these uncertainties encountered in the single criterion based detections and to implement a more comprehensive detection by using these two criteria simultaneously, we propose an evidential reasoning based impulse noise detection approach thanks to the ability of belief functions to model uncertainty and for reasoning under uncertainty. The theory of belief functions [28] are briefly recalled first below.

### III. IMPULSE NOISE DETECTION BASED ON EVIDENTIAL REASONING

#### A. Basic of evidence theory

The theory of belief functions, also called Dempster-Shafer evidence theory (DST) [28], is a theoretical framework for uncertainty modeling and reasoning.

In DST, elements in the frame of discernment (FOD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_l\}$  are mutually exclusive and exhaustive. The power set  $2^\Theta$  of the FOD  $\Theta$  is the set of all subsets of  $\Theta$ . Define a function  $m$  from  $2^\Theta$  to  $[0, 1]$  as a basic belief assignment (BBA, also called a mass function) satisfying

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad m(\emptyset) = 0. \quad (2)$$

$m(A)$  depicts the evidence support to the proposition  $A$ . If  $m(A) > 0$ ,  $A$  is called a focal element.

The plausibility function ( $Pl$ ) and belief function ( $Bel$ ) are defined as:

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B), \quad (3)$$

$$Bel(A) = \sum_{B \subseteq A} m(B). \quad (4)$$

The belief interval [28], [31]  $[Bel(A), Pl(A)]$  represents the imprecision of the support to the proposition  $A$ .

Dempster's rule of combination [28], which is used for combining two distinct pieces of evidence, is defined as

$$(m_1 \oplus m_2)(A) = \begin{cases} 0, & A = \emptyset, \\ \frac{1}{1-K} \sum_{B \cap C = A} m_1(B)m_2(C), & A \neq \emptyset, \end{cases} \quad (5)$$

where  $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$  represents the total conflict or contradictory mass assignments.

For a probabilistic decision-making, Smets defined the pig-nistic probability transformation [32] to transform a BBA into a probability measure BetP:

$$\text{BetP}(\theta_i) \triangleq \sum_{\theta_i \in A} \frac{m(A)}{|A|}, \quad \forall \theta_i \in \Theta, \quad (6)$$

where  $|A|$  denotes the cardinality of  $A$ . The decision is made by choosing the element in FOD which has the highest BetP value. Note that there are still other probability transformations of BBA, see [33] for details.

#### B. Evidential modeling for uncertainties and fusion based detection

To deal with the uncertainties encountered in the impulse noise detection, we propose a detection method based on evidential reasoning, which uses the extreme criterion and discontinuity criterion simultaneously. The flow chart of the detection algorithm is illustrated in Fig. 2.

Here we propose two methods for uncertainties modeling. One proposed method models the uncertainties of extreme criterion and discontinuity criterion with two BBAs, respectively (denoted by method I). The other proposed method treats this impulse noise detection with two criteria as a multi-criteria decision making problem, and uses cautious ordered weighted averaging with evidential reasoning (COWA-ER) method [34] to generate BBAs (denoted by method II). In both methods, we use the distance of interval numbers to describe the extreme property and use the rank-ordered absolute differences (ROAD) statistic [29] to describe the discontinuity property.

##### B-1 Evidential modeling method I and fusion based detection

###### 1) Evidential modeling for the uncertainty in extreme criterion

According to the extreme property of the impulse noise, the intensity of an impulse noise pixel must be an extreme value or close to an extreme value.

Since all of the pixels' intensities in an image are within a range ( $[0, 255]$  for an 8-bit image), the intensity information of an image can be represented by an interval number. An interval number  $\tilde{a}$  in  $\mathbb{R}$  is a set of real numbers that lie between two real numbers, i.e.,  $\tilde{a} = [a_1, a_2] = \{x | a_1 \leq x \leq a_2\}$ ,  $a_1, a_2 \in \mathbb{R}$  and  $a_1 \leq a_2$ . The intensity information of an image can be expressed as an interval number  $\tilde{I} = [I_{\min}, I_{\max}]$ , where  $I_{\min}$  and  $I_{\max}$  denote the minimum and maximum intensities of the image, respectively. Furthermore, a single pixel's intensity  $x$  can also be viewed as an interval number  $[x, x]$ , whose upper bound and lower bound are equal.

The distance of interval numbers is a measure of dissimilarity between interval numbers. Here we use it to describe the closeness between a pixel's intensity and the extreme values. Various types of distance for interval numbers [35] have been proposed. Here, we use the following strict distance metric.

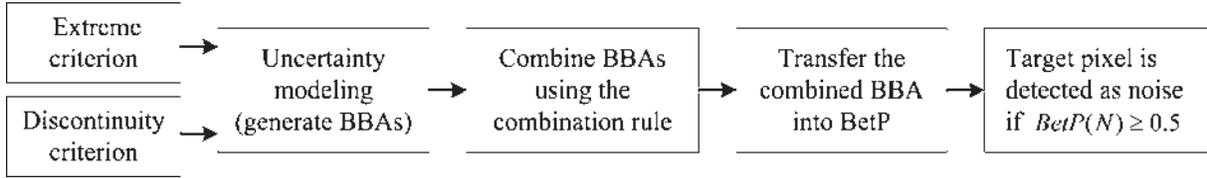


Figure 2. Noise detection algorithm based on evidential reasoning.

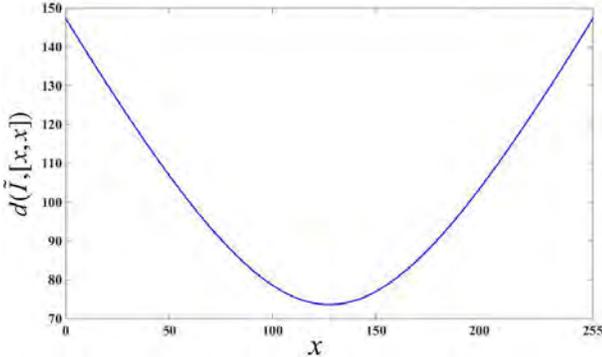
Given two interval numbers  $\tilde{a} = [a_1, a_2]$  and  $\tilde{b} = [b_1, b_2]$ , the distance between  $\tilde{a}$  and  $\tilde{b}$  is defined as:

$$d(\tilde{a}, \tilde{b}) = \sqrt{\left[\frac{(a_1+a_2)-(b_1+b_2)}{2}\right]^2 + \frac{1}{3}\left[\frac{(a_2-a_1)-(b_2-b_1)}{2}\right]^2}. \quad (7)$$

According to (7), the distance between  $\tilde{I}$  and  $[x, x]$ , which describes the closeness between a pixel's intensity  $x$  and the extreme values,  $I_{\min}$  and  $I_{\max}$ , is:

$$d(\tilde{I}, [x, x]) = \sqrt{\left(\frac{I_{\min}+I_{\max}}{2} - x\right)^2 + \frac{(I_{\max}-I_{\min})^2}{12}}. \quad (8)$$

Here,  $x$  takes values in the interval  $[I_{\min}, I_{\max}]$ . An illustration of  $d(\tilde{I}, [x, x])$  is shown in Fig. 3, where the intensity range of the image is set as  $[0, 255]$ . When  $x$  takes the median of  $[0, 255]$ , i.e., 127 or 128,  $d(\tilde{I}, [x, x])$  reaches the minimum value. The closer between  $x$  and the extreme value, e.g., 0 or 255, the higher the value of  $d(\tilde{I}, [x, x])$ . Thus,  $d(\tilde{I}, [x, x])$  can be used to describe the extreme property.


 Figure 3. An illustration of  $d(\tilde{I}, [x, x])$ .

Suppose that the only possible type of the noise existing in the given image is the impulse noise. The pixel whose intensity is far from the extreme values must be the signal, but the pixel whose intensity is close to an extreme value may not be the noise. For example, in some bright or dark area, the intensity range of signal pixels may overlap with that of the impulse noise pixels. Here, we use the belief function to describe this uncertainty.

We set a detection window with a size of  $w_D \times w_D$  centered at the given pixel at  $(i, j)$ :

$$W_D(i, j) = \left\{x_{i-s, j-t} \mid -\frac{(w_D-1)}{2} \leq s, t \leq \frac{(w_D-1)}{2}\right\}, \quad (9)$$

where  $x_{i-s, j-t}$  is the intensity of the pixel at  $(i-s, j-t)$ .

We focus on two distances in  $W_D$ :

- 1)  $d_c$  denotes the distance between the center pixel's intensity and the interval number  $\tilde{I}$ , where  $\tilde{I}$  expresses the intensity range of the pixels in the whole image.
- 2)  $d_0$  denotes the minimum distance in  $W_D$  between a pixel's intensity and  $\tilde{I}$ , where the pixel is the one whose intensity is the farthest one in  $W_D$  from the extreme values. Thus, this pixel is most unlikely to be the noise in  $W_D$  according to the extreme criterion.

We also focus on another two distances in the whole image:

- 1)  $d_{\text{ext}}$  denotes distance between  $\tilde{I}$  and the extreme value:  $I_{\min}$  or  $I_{\max}$ . It is the maximum distance in the image between a pixel's intensity and  $\tilde{I}$ . If a pixel's intensity is close to the extreme value, its distance to  $\tilde{I}$  is close to  $d_{\text{ext}}$ .
- 2)  $d_{\text{med}}$  denotes distance between  $\tilde{I}$  and the median of  $\tilde{I}$ . It is the minimum possible distance in the image between a pixel's intensity and  $\tilde{I}$ . If a pixel's intensity is much far from the extreme values, its distance to  $\tilde{I}$  is close to  $d_{\text{med}}$ .

Finally, we construct a BBA  $m_1$  using the above distances to model the uncertainty of whether the center pixel is corrupted by the impulse noise or not according to the extreme criterion:

$$\begin{cases} m_1(N) = \frac{d_c - d_0}{d_{\text{ext}} - d_0 + \varepsilon}, \\ m_1(S) = 1 - \frac{d_c - d_{\text{med}}}{d_{\text{ext}} - d_{\text{med}}}, \\ m_1(\Theta) = 1 - m_1(N) - m_1(S). \end{cases} \quad (10)$$

Here, the FOD  $\Theta = \{N, S\}$ , where  $N$  denotes the noise and  $S$  denotes the signal. The parameter  $\varepsilon$  is a small positive real number to avoid  $m_1(N)$  to be 1, when the intensity of the center pixel is an extreme value. It means that a pixel with an extreme value should not be absolutely recognized as the noise because it might be the signal actually. Furthermore, Dempster's rule of combination in (5) has the problem of one ballot veto when one BBA is assigned 1 on one singleton  $\theta_i$  ( $\theta_i \in \Theta$ ), while 0 on other singletons. That is, if  $m_1(N) = 1$ , i.e.,  $m_1(S) = 0$ , no matter what  $m_2$  is, the combined BBA has  $m(S) = 0$ , which indicates the center pixel cannot be the signal.

Since  $d_{\text{ext}} \geq d_c \geq d_0 \geq d_{\text{med}}$ ,  $0 \leq (d_c - d_0)/(d_{\text{ext}} - d_0 + \varepsilon) < 1$  and  $0 \leq (d_c - d_{\text{med}})/(d_{\text{ext}} - d_{\text{med}}) \leq 1$ . That is,  $0 \leq m_1(N) < 1$  and  $0 \leq m_1(S) \leq 1$ . Besides, as  $d_{\text{med}} \leq d_0$ ,

$$m_1(N) \leq \frac{d_c - d_0}{d_{\text{ext}} - d_0} \leq \frac{d_c - d_{\text{med}}}{d_{\text{ext}} - d_{\text{med}}},$$

2	100	103	121	133
145	120	114	115	126
139	132	$x_{i,j}$	8	249
145	254	131	124	127
129	150	147	144	130

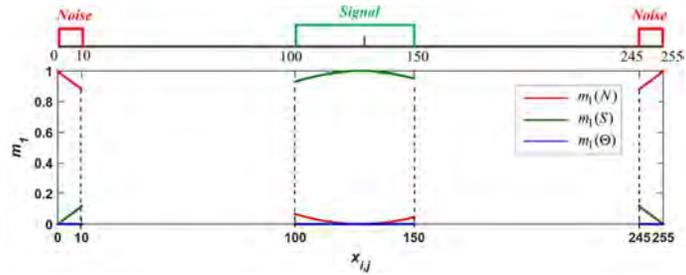


Figure 4. Case *a* and the corresponding  $m_1$  of the center pixel. (a) A detection window for case *a*. (b) Corresponding  $m_1$  when  $x$  takes different values.

230	10	234	240	246
238	236	253	239	241
248	243	$x_{i,j}$	246	249
5	250	233	236	240
255	247	242	239	1

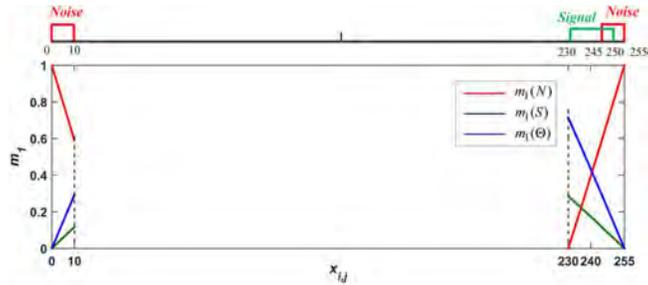


Figure 5. Case *b* and the corresponding  $m_1$  of the center pixel. (a) A detection window for case *b*. (b) Corresponding  $m_1$  when  $x$  takes different values.

which means  $m_1(N) \leq 1 - m_1(S)$  so that  $m_1(N) + m_1(S) \leq 1$ . Therefore,  $m_1$  satisfies the constraint in (2), and  $m_1$  is a legitimate BBA.

According to  $m_1$ , the center pixel will have a large value of  $m_1(N)$  only when its intensity is close to the extreme value, and at the same time far from the intensity being the closest to the median of  $\tilde{I}$  in  $W_D$ . The center pixel will have a large value of  $m_1(S)$  when its intensity is close to the median of  $\tilde{I}$ .

Here, we consider two different cases about the detection window and the corresponding BBA  $m_1$  of the center pixel when  $\tilde{I} = [0, 255]$ ,  $\varepsilon = 0.1$  and  $\alpha = 10$ , i.e., noise pixels take values in the sets of  $S_1 = \{0, 1, \dots, 10\}$  and  $S_2 = \{245, 246, \dots, 255\}$ .

- 1) For the most common case, the intensities of signal pixels in a detection window  $W_D$  are far from the extreme values, as the example shown in Fig. 4(a), where the intensity range of signal pixels is  $[100, 150]$ . The pixel with the intensity of 127 is the most unlikely to be the noise in  $W_D$ , since 127 is the farthest intensity in  $W_D$  from the extreme values. If the center pixel's intensity  $x_{i,j}$  is in the range of  $[0, 10]$  or  $[245, 255]$ , it is close to the extreme value, and at the same time far from 127. Thus, the center pixel is assigned a large value of  $m_1(N)$ . If the center pixel's intensity is in the range of  $[100, 150]$ , it is close to the median of  $\tilde{I}$ . Thus, it is assigned a large value of  $m_1(S)$ .
- 2) In some cases, all of the signal pixels' intensities in a detection window are close to an extreme value, as the example shown in Fig. 5(a), where the intensity range of signal pixels is  $[230, 250]$ . The pixel with the intensity of 230 is the one that is most unlikely to be the noise in  $W_D$ , since 230 is the farthest intensity in  $W_D$  from

the extreme values. If the center pixel's intensity  $x_{i,j}$  is in the range of  $[0, 10]$  or  $[245, 255]$ , it is close to the extreme value and far from 230. Thus, the center pixel is assigned a large value of  $m_1(N)$ . If the center pixel's intensity  $x_{i,j}$  is in the range of  $[230, 240]$ , the center pixel is assigned a small value of  $m_1(N)$  since its intensity is close to 230. At the same time, it is assigned a small value of  $m_1(S)$  since  $x_{i,j}$  is far from the median of  $\tilde{I}$ . Therefore,  $m_1(\Theta)$  is large, which indicates it is hard to say whether the center pixel is the noise or signal. In summary, in this case, it is hard to get a crisp description of the beliefs for the corresponding decisions according to the extreme criterion, since the intensity range of signal pixels overlaps with that of the noise pixels. However, our BBA  $m_1$  keeps the large uncertainty (large  $m_1(\Theta)$ ), which is helpful to avoid the arbitrary detection decision.

From the above we can see that when using the extreme criterion for noise detection, our evidential method uses the BBA to describe the beliefs of the corresponding detection decisions, which does not make a hard decision like two-steps methods but keeps the uncertainty. This is more cautious and can reduce the information loss for the final fusion-based detection.

#### 2) Evidential modeling for the uncertainty in discontinuity criterion

According to the discontinuity property of the impulse noise, the intensity of an impulse noise pixel tends to be much higher or lower than the intensities of its neighbors. For some signal pixels, e.g., edge pixels and signal pixels in the bright or dark area, they are easily detected as the noise

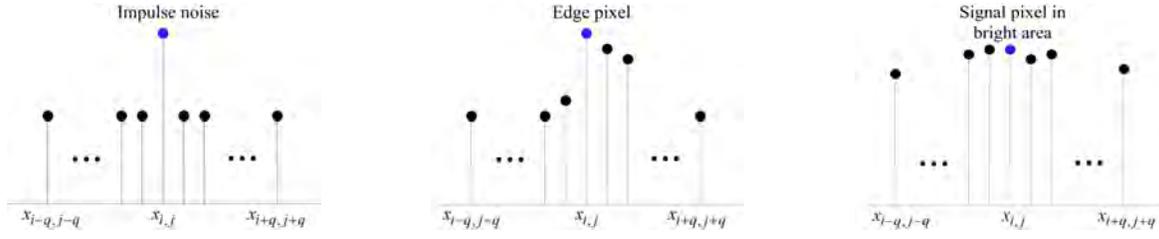


Figure 6. One-dimensional illustration of the differences between some signal pixels and the impulse noise. (a) Intensities of the impulse noise and its neighbors. (b) Intensities of an edge pixel and its neighbors. (c) Intensities of signal pixels in the bright area.

according to the extreme criterion. However, they have some differences from the noise pixel according to the discontinuity criterion. The intensity of an edge pixel is higher or lower than only a portion of the intensities of its neighbors. The intensity of a signal pixel in the bright or dark area is similar to the intensities of its neighbors.

For a pixel at  $(i, j)$ , we consider its neighbors' intensity information in the same detection window  $W_D(i, j) = \{x_{i-s,j-t} - (w_D - 1)/2 \leq s, t \leq (w_D - 1)/2\}$  as used in modeling the uncertainty in the extreme criterion. Those differences between the signal, edge, and the bright area pixels reflected in the discontinuity property are illustrated in Fig. 6. Here for simplification, one-dimensional expressions of the pixels' intensities in the detection windows are used, where  $q = (w_D - 1)/2$ .

We use the rank-ordered absolute differences (ROAD) statistic [29] to describe such differences reflected in the discontinuity property. Define  $diff(x_{i,j}, x_{i-s,j-t}) = x_{i,j} - x_{i-s,j-t}$  as the absolute difference of the intensities between the center pixel at  $(i, j)$  and its neighbor at  $(i-s, j-t)$ , where  $x_{i-s,j-t} \in W_D(i, j)$ . If the size of  $W_D(i, j)$  is  $M = w_D \times w_D$ , there will be  $M - 1$  neighbors in the window, and therefore the amount of  $diff(x_{i,j}, \cdot)$  is  $M - 1$ . These  $diff(x_{i,j}, \cdot)$  can describe the dissimilarity between the center pixel and its neighbors. To further analyze this dissimilarity, sort these  $M - 1$   $diff(x_{i,j}, \cdot)$  values in the ascending order and denote  $r_g(x_{i,j})$  as the  $g$ th smallest  $diff(x_{i,j}, \cdot)$ . Finally, calculate the sum of the first  $n$  smallest  $diff(x_{i,j}, \cdot)$  as

$$ROAD_n(x_{i,j}) = \sum_{g=1}^n r_g(x_{i,j}), \quad (11)$$

where  $2 \leq n \leq M - 1$ .

If the center pixel is the noise,  $diff(x_{i,j}, \cdot)$  is small when its neighbor is a noise pixel whose intensity is close to the same extreme value as the center pixel.

If the center pixel is the signal without extreme property,  $diff(x_{i,j}, \cdot)$  is large when its neighbor is the noise.

If the center pixel is the signal with extreme property,  $diff(x_{i,j}, \cdot)$  is large when its neighbor is a noise whose intensity is close to the other extreme value.

Therefore, when the noise density is low, the impulse noise has large value of  $ROAD_{M-1}(x_{i,j})$  as well as the sum of its smallest  $n$   $diff(x_{i,j}, \cdot)$  values, i.e.,  $ROAD_n(x_{i,j})$ . The

signal pixel has small value of  $ROAD_{M-1}(x_{i,j})$  as well as  $ROAD_n(x_{i,j})$ .

With the increase of the noise density, the quantity of impulse noise pixels increases in a detection window. If the center pixel is the noise, the amount of small  $diff(x_{i,j}, \cdot)$  will increase since there are more noise neighbors having the similar intensities with the center pixel. At the same time, the amount of very large  $diff(x_{i,j}, \cdot)$  also increases since there are more noise neighbors having the intensities close to the other extreme value. Thus,  $ROAD_n(x_{i,j})$  becomes smaller but  $ROAD_{M-1}(x_{i,j})$  has no significant change.

If the center pixel is the signal, the amount of large  $diff(x_{i,j}, \cdot)$  will increase since there are more noise neighbors. Thus,  $ROAD_{M-1}(x_{i,j})$  becomes larger but  $ROAD_n(x_{i,j})$  has no significant change.

In summary,  $ROAD_n(x_{i,j})$  is large only when the center pixel is the noise and the noise density is small;  $ROAD_{M-1}(x_{i,j})$  is small only when the center pixel is the signal and the noise density is small. With the increasing of the noise density, the differences between the signal and the noise reflected in  $ROAD_n(x_{i,j})$  and  $ROAD_{M-1}(x_{i,j})$  narrow. This means that the discontinuity property of the impulse noise pixels becomes weaker with the increase of the noise density. It is unreasonable to use the discontinuity criterion to make hard decisions for detection. Therefore, we construct a BBA  $m_2$  to describe the beliefs of the corresponding detection decisions according to the discontinuity criterion.

As we have discussed above, only the noise pixel can have large  $ROAD_n(x_{i,j})$  and only the signal pixel can have small  $ROAD_{M-1}(x_{i,j})$ . Thus, for a given center pixel, the larger value of  $ROAD_n(x_{i,j})$  it has, the larger belief it should be assigned to being detected as the noise; the smaller value of  $ROAD_{M-1}(x_{i,j})$  it has, the larger belief should be assigned to being detected as the signal. For a center pixel with the intensity of  $x_{i,j}$ , its  $m_2$  is constructed as follow. We take  $n = (M - 1)/2$ , which means that we focus on the first half small  $diff(x_{i,j}, \cdot)$  when considering the belief of that a pixel should be detected as the noise.

$$\begin{cases} m_2(N) = \frac{ROAD_{\frac{(M-1)}{2}}(x_{i,j})}{\frac{(M-1)}{2} \times (I_{\max} - I_{\min})}, \\ m_2(S) = 1 - \frac{ROAD_{M-1}(x_{i,j})}{(M-1) \times (I_{\max} - I_{\min})}, \\ m_2(\Theta) = 1 - m_2(N) - m_2(S). \end{cases} \quad (12)$$

Here,  $I_{\max}$  and  $I_{\min}$  denote the maximum and minimum intensities of the whole image, respectively.

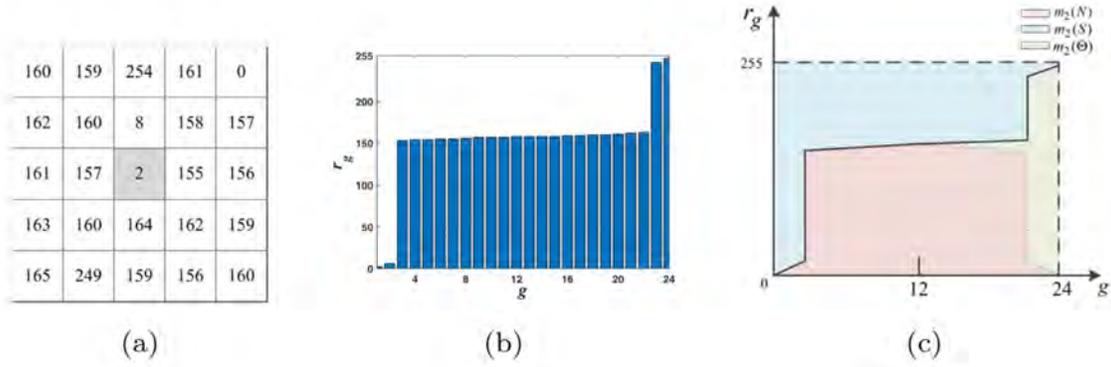


Figure 7. Visual representation of  $m_2$ . (a) A detection window. (b) The illustration of  $r_g(2)$ . (c) Graphical representation of belief assignments.

Since  $r_g(x_{i,j}) \leq I_{\max} - I_{\min}$ ,  $\text{ROAD}_{\frac{(M-1)}{2}}(x_{i,j}) \leq \frac{(M-1)}{2} \times (I_{\max} - I_{\min})$  and  $\text{ROAD}_{M-1}(x_{i,j}) \leq (M-1) \times (I_{\max} - I_{\min})$ . That is  $0 \leq m_2(N) \leq 1$  and  $0 \leq m_2(S) \leq 1$ . Besides, since

$$\begin{aligned} m_2(N) + m_2(S) &= 1 - \left( \frac{\text{ROAD}_{M-1}(x_{i,j})}{(M-1) \times (I_{\max} - I_{\min})} - \frac{\text{ROAD}_{\frac{(M-1)}{2}}(x_{i,j})}{\frac{(M-1)}{2} \times (I_{\max} - I_{\min})} \right) \\ &= 1 - \frac{\sum_{g=\frac{(M+1)}{2}}^{M-1} r_g(x_{i,j}) - \text{ROAD}_{\frac{(M-1)}{2}}(x_{i,j})}{(M-1) \times (I_{\max} - I_{\min})}, \end{aligned}$$

and  $\sum_{g=\frac{(M+1)}{2}}^{M-1} r_g(x_{i,j}) \geq \text{ROAD}_{\frac{(M-1)}{2}}(x_{i,j})$ , there exists  $m_2(N) + m_2(S) \leq 1$ . Thus,  $m_2$  satisfies the constraint of BBA in (2) and  $m_2$  is a legitimate BBA.

For a given pixel, mass values in the BBA  $m_2$ :  $m_2(N)$ ,  $m_2(S)$  and  $m_2(\Theta)$  can be represented as the areas of regions as shown in Fig. 7.

Fig. 7(a) illustrates an example of a detection window. The intensity of the center pixel  $x_{i,j} = 2$ . We suppose that the largest intensity difference  $I_{\max} - I_{\min}$  in the image is 255. Since the size of the window  $M$  is 25, we can get 24  $\text{dif}(x_{i,j}, \cdot)$  values. The ascending ordered  $\text{dif}(x_{i,j}, \cdot)$ , i.e.,  $r_g(x_{i,j})$ , ( $g = 1, 2, \dots, 24$ ), are expressed as the histogram in Fig. 7(b). We specify the area of the rectangular region in Fig. 7(b) with the vertex points:  $(0, 0)$ ,  $(24, 0)$ ,  $(0, 255)$  and  $(24, 255)$  as 1. It means that we represent the value of  $(M-1) \times (I_{\max} - I_{\min})$  in (12) using a region with an area of 1. Thus, the value of  $\text{ROAD}_{M-1}(x_{i,j}) / [(M-1) \times (I_{\max} - I_{\min})]$  can be represented by the region determined by  $r_g(x_{i,j})$ , ( $g = 1, 2, \dots, 24$ ) in Fig. 7(b). That is, the value of  $m_2(S)$  in (12), i.e.,  $1 - \text{ROAD}_{M-1}(x_{i,j}) / [(M-1) \times (I_{\max} - I_{\min})]$  can be represented as the blue area in Fig. 7(c). Similarly, the value of  $\text{ROAD}_{\frac{(M-1)}{2}}(x_{i,j}) / [(M-1) \times (I_{\max} - I_{\min})]$  can be represented by the region determined by  $r_g(x_{i,j})$ , ( $g = 1, 2, \dots, 12$ ) in Fig. 7(b). That is, the value of  $m_2(N)$  in (12), i.e.,  $2 \times \text{ROAD}_{\frac{(M-1)}{2}}(x_{i,j}) / [(M-1) \times (I_{\max} - I_{\min})]$ , can be represented as the pink area<sup>1</sup> in Fig. 7(c). Thus, the value of  $m_2(\Theta)$  is represented as the remanent area, i.e. the green area in Fig. 7(c).

<sup>1</sup>Since the value of  $m_2(N)$  is the double of  $\text{ROAD}_{\frac{(M-1)}{2}}(x_{i,j}) / [(M-1) \times (I_{\max} - I_{\min})]$ ,  $m_2(N)$  can be represented as the double of the region determined by  $r_g(x_{i,j})$ , ( $g = 1, 2, \dots, 12$ ).

Fig. 8 shows the graphical representations of  $m_2$  for different kinds of center pixels with different noise density levels (25%, 50% and 75%). Center pixels include the impulse noise pixel (the first column in Fig. 8), the signal pixels with extreme property, such as the edge pixel (the second column in Fig. 8), the signal pixels in the bright or dark area (the third column in Fig. 8), and the common signal pixels with no extreme property (the last column in Fig. 8). Here, we assume that the largest intensity differences  $I_{\max} - I_{\min}$  for the whole image in all cases are 255.

In Fig. 8, the signal pixels (from the second column to the last column) have large values of  $m_2(S)$  indicating that for signal pixels, large beliefs are assigned to being detected as the signal under all noise density levels.

For the impulse noise pixel, when the noise density  $\leq 50\%$  (Fig. 8(a) or Fig. 8(e)), it has large value of  $m_2(N)$  indicating that large belief is assigned to being detected as the noise. However, when the noise density is larger than 50% (Fig. 8(i)), the impulse noise pixel has small value of  $m_2(N)$  indicating that only small belief is assigned to being detected as the noise. But at the same time, its value of  $m_2(\Theta)$  is large indicating that the uncertainty degree of discontinuity criterion is large when the image is corrupted severely.

Our evidential method uses the BBA to describe the beliefs of the corresponding detection decisions according to the discontinuity criterion. We do not make the hard decision directly but keep the uncertainty for the time being, which is more cautious. Particularly, our modeling method here keeps the large uncertainty of discontinuity criterion when the noise intensity is large. This will be helpful for the final fusion-based detection to decrease the miss-detections and false alarms.

3) *Fusion based detection* The generated BBAs  $m_1$  and  $m_2$  can be combined, e.g., using Eq. (5) to obtain  $m(\cdot) = [m_1 \oplus m_2](\cdot)$ , which is a combined BBA for the noise detection representing the simultaneous use of the extreme property and discontinuity property. Once  $m$  is obtained, we use the pignistic probability transformation in Eq. (6) to transform  $m$  into a probability measure BetP. If  $\text{BetP}(N) \geq 0.5$ , the center pixel should be detected as the noise.

Here we use an example to illustrate our detection procedure. A detection window is shown in Fig. 9.

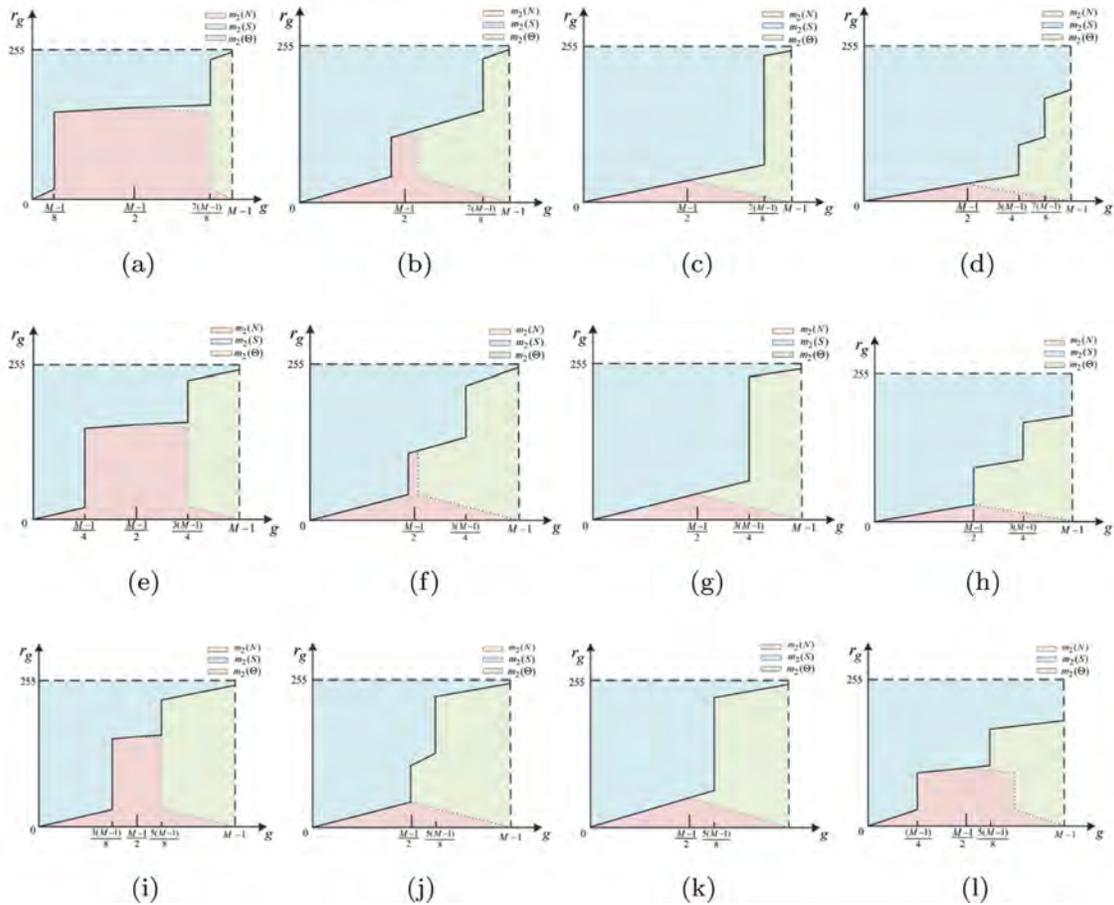


Figure 8. Illustration of  $m_2$  in cases with different noise densities. (From top to bottom, from left to right) The first row to third row show the cases when the noise densities are 25%, 50% and 75% respectively. The first column to fourth column show the cases when the center pixels are the impulse noise, edge pixel, signal pixel in the bright or dark area and common signal pixel without extreme property respectively.

202	203	203	201	206
201	202	8	204	204
203	200	8	0	202
200	9	201	204	204
200	201	200	204	202

Figure 9. The illustration of a detection window.

In this window, intensities of signal pixels range from 200 to 206 and several pixels are corrupted by the impulse noise with intensities of 0, 8 or 9. Assume the intensity range of the whole image is  $[0, 255]$ . The value of  $\varepsilon$  in (10) is 0.1.

According to the modeling method I, the generated BBAs

are:

$$\begin{cases} m_1(N) = 0.8416, \\ m_1(S) = 0.0933, \\ m_1(\Theta) = 0.0651, \end{cases} \quad \text{and} \quad \begin{cases} m_2(N) = 0.5696, \\ m_2(S) = 0.3320, \\ m_2(\Theta) = 0.0984. \end{cases}$$

The combined BBA is:

$$\begin{cases} m(N) = 0.8978, \\ m(S) = 0.0926, \\ m(\Theta) = 0.0096. \end{cases}$$

Then, we obtain the pignistic probability  $\text{BetP}(N) = 0.9026$  and the center pixel is finally detected as the impulse noise since  $\text{BetP}(N) > 0.5$ .

*B-2 Evidential modeling method II and fusion based detection*

The impulse noise detection with two criteria including the extreme property and discontinuity property can be viewed as a multi-criteria (to be more accurately bi-criteria) decision making problem. Therefore, we can use the cautious ordered weighted averaging with evidential reasoning (COWA-ER) method [34] to generate BBAs and to implement the fusion-based noise detection.

## 1) COWA-ER method

In the noise detection problem, for a given pixel, the finite set of alternatives  $\Theta = \{\theta_1, \theta_2\} = \{N, S\}$ . According to the analyses in method I, the pessimistic and optimistic valuations of the expected payoffs to these alternatives obtained from the two detection criteria (extreme criterion and discontinuity criterion) are:

$$\begin{cases} e_{\min}(N) = \min\left\{\frac{d_c - d_0}{d_{\text{ext}} - d_0 + \varepsilon}, \frac{\text{ROAD}_{\frac{M-1}{2}}(x_{i,j})}{\frac{M-1}{2} \times (I_{\max} - I_{\min})}\right\}, \\ e_{\max}(N) = \max\left\{\frac{d_c - d_0}{d_{\text{ext}} - d_0 + \varepsilon}, \frac{\text{ROAD}_{\frac{M-1}{2}}(x_{i,j})}{\frac{M-1}{2} \times (I_{\max} - I_{\min})}\right\}, \\ e_{\min}(S) = \min\left\{1 - \frac{d_c - d_{\text{med}}}{d_{\text{ext}} - d_{\text{med}}}, 1 - \frac{\text{ROAD}_{\frac{M-1}{2}}(x_{i,j})}{(M-1) \times (I_{\max} - I_{\min})}\right\}, \\ e_{\max}(S) = \max\left\{1 - \frac{d_c - d_{\text{med}}}{d_{\text{ext}} - d_{\text{med}}}, 1 - \frac{\text{ROAD}_{\frac{M-1}{2}}(x_{i,j})}{(M-1) \times (I_{\max} - I_{\min})}\right\}. \end{cases} \quad (13)$$

Then, the expected payoff matrix is constructed as:

$$E = \begin{bmatrix} E[N] \\ E[S] \end{bmatrix} = \begin{bmatrix} [e_{\min}(N), e_{\max}(N)] \\ [e_{\min}(S), e_{\max}(S)] \end{bmatrix}. \quad (14)$$

Here, the expected payoffs  $E[N]$  and  $E[S]$  are imprecise since they belong to the interval  $[e_{\min}(\cdot), e_{\max}(\cdot)]$  where the lower and upper bounds represent the pessimistic and optimistic attitudes, respectively.

Then, divide each bound of intervals by the max of the bounds, i.e.,  $e_{\text{MAX}} = \max\{e_{\max}(N), e_{\max}(S)\}$ , to get a new normalized imprecise expected payoff vector  $E^{\text{Imp}}$ :

$$E^{\text{Imp}} = \begin{bmatrix} [e_{\min}(N)/e_{\text{MAX}}, e_{\max}(N)/e_{\text{MAX}}] \\ [e_{\min}(S)/e_{\text{MAX}}, e_{\max}(S)/e_{\text{MAX}}] \end{bmatrix} = \begin{bmatrix} [a_1, b_1] \\ [a_2, b_2] \end{bmatrix}. \quad (15)$$

In the final, convert the normalized imprecise expected payoff vector  $E^{\text{Imp}}$  into BBAs according to a very natural and simple transformation [34], [36]. The generation of a BBA associated to the hypothesis  $\theta_i$ , ( $\theta_1 = N$ ,  $\theta_2 = S$ ) from any imprecise value  $[a_i, b_i] \subseteq [0, 1]$  is generated as:

$$\begin{cases} m_i(\theta_i) = a_i, \\ m_i(\bar{\theta}_i) = 1 - b_i, \\ m_i(\theta_i \cup \bar{\theta}_i) = b_i - a_i. \end{cases} \quad (16)$$

$\bar{\theta}_i$  is the complement of  $\theta_i$  in  $\Theta$ . With such a conversion, one sees that  $Bel(\theta_i) = a_i$ ,  $Pl(\theta_i) = b_i$  and the uncertainty is represented by the length of the interval  $[a_i, b_i]$ .

## 2) Fusion based detection

By using the COWA-ER method, we can obtain two BBAs:  $m_1$  and  $m_2$ . The generated BBAs can be combined using Eq. (5), that is  $m(\cdot) = [m_1 \oplus m_2](\cdot)$ . Once  $m$  is computed, we use the pignistic probability transformation in (6) to transform  $m$  into a probability measure BetP. If  $\text{BetP}(N) \geq 0.5$ , the center pixel should be detected as the impulse noise.

Here we consider the same example showed in Fig. 9. The value of  $\varepsilon$  in Eq. (13) is 0.1. The BBAs generated from modeling method II are:

$$\begin{cases} m_1(N) = 0.6768, \\ m_1(S) = 0, \\ m_1(\Theta) = 0.3232, \end{cases} \quad \text{and} \quad \begin{cases} m_2(N) = 0.6055, \\ m_2(S) = 0.1109, \\ m_2(\Theta) = 0.2836. \end{cases}$$

The combined BBA is:

$$\begin{cases} m(N) = 0.8622, \\ m(S) = 0.0388, \\ m(\Theta) = 0.0990. \end{cases}$$

Then, we get the pignistic probability  $\text{BetP}(N) = 0.9117$  and the center pixel is finally detected as the impulse noise because  $\text{BetP}(N) > 0.5$ .

When modeling the uncertainty of noise detection, the two proposed methods use the same information (extreme criterion and discontinuity criterion) but generate belief functions in different ways. Either of these two methods can be an alternative to the other in many cases but they might generate different detection results in some special situations.

## B-3 Different detection results with contradictory evidences

In many cases, the two proposed methods generate same detection results since they describe the extreme criterion and discontinuity criterion in very similar ways and both of their combined evidences will assign a larger belief to the same candidate (noise or signal). However, when these two evidence sources are highly contradictory (extreme criterion and discontinuity criterion give very different supports to the target pixel), the two proposed methods might generate different detection results as illustrated in the following two different examples that the evidence sources are highly contradictory.

## 1) Detection results for situation 1

Situation 1 describes the situation when the target pixel is a signal in a dark area close to an edge where the pixels at the other side of the edge have higher intensities as shown in Fig. 10.

1	1	2	2	1
0	0	0	1	2
9	3	2	3	4
50	41	28	18	15
26	34	44	45	41

Figure 10. Highly contradictory situation 1.

According to the extreme criterion and discontinuity criterion, the two BBAs generated by evidential modeling method I using Eq. (10) and Eq. (12) are:

$$\begin{cases} m_1(N) = 0.9548, \\ m_1(S) = 0.0235, \\ m_1(\Theta) = 0.0217, \end{cases} \quad \text{and} \quad \begin{cases} m_2(N) = 0.0039, \\ m_2(S) = 0.9440, \\ m_2(\Theta) = 0.0521. \end{cases}$$

In this situation, the proposition that the target pixel should be detected as noise obtained very different supports from the extreme criterion ( $m_1(N)$  is large) and discontinuity criterion

( $m_2(N)$  is small) since the target pixel's intensity is very close to the extreme value 0, but at the same time, the target pixel has many neighborhoods have the similar intensities. After the evidence combination and probability transformation, we finally get  $BetP(N) = 0.5491$  and the target pixel is false alarmed as noise since  $BetP(N) > 0.5$ .

The evidential modeling method II deals with these two highly contradictory evidences in different ways. According to Eq. (13) and Eq. (14), the expected payoff matrix is generated as:

$$E = \begin{bmatrix} E[N] \\ E[S] \end{bmatrix} = \begin{bmatrix} [0.0039, 0.9548] \\ [0.0235, 0.9440] \end{bmatrix}.$$

Then, we get the normalized expected payoff vector:

$$E^{Imp} = \begin{bmatrix} E[N] \\ E[S] \end{bmatrix} = \begin{bmatrix} [0.0041, 1.0000] \\ [0.0246, 0.9887] \end{bmatrix}.$$

The generated BBAs are:

$$\begin{cases} m_1(N) = 0.0041, \\ m_1(S) = 0, \\ m_1(\Theta) = 0.9959, \end{cases} \quad \text{and} \quad \begin{cases} m_2(N) = 0.0113, \\ m_2(S) = 0.0246, \\ m_2(\Theta) = 0.9641. \end{cases}$$

After the evidence combination and probability transformation, we finally get  $BetP(N) = 0.4954$  and the target pixel is successfully detected as signal since  $BetP(N) < 0.5$ . For this example, the detection result generated by the proposed method II is more reasonable.

## 2) Detection results for situation 2

Situation 2 describes the highly corrupted situation when the target pixel is noise and the neighborhood signal pixels (colored with green) have similar intensities with the target pixel as shown in Fig. 11.

249	223	252	7	7
220	219	248	249	253
6	252	246	0	9
254	246	2	209	245
219	219	251	245	247

Figure 11. Highly contradictory situation 2.

According to Eq. (10) and Eq. (12), the two BBAs generated by evidential modeling method I are:

$$\begin{cases} m_1(N) = 0.7914, \\ m_1(S) = 0.1049, \\ m_1(\Theta) = 0.1037, \end{cases} \quad \text{and} \quad \begin{cases} m_2(N) = 0.0141, \\ m_2(S) = 0.7296, \\ m_2(\Theta) = 0.2564. \end{cases}$$

In this situation, the proposition that the target pixel should be detected as noise obtained very different supports from the extreme criterion ( $m_1(N)$  is large) and discontinuity criterion

( $m_2(N)$  is small) since the target pixel's intensity is close to the extreme value 255, but at the same time, there are many neighborhoods have the similar intensities with the target pixel. After the evidence combination and probability transformation, we finally get  $BetP(N) = 0.5432$  and the target pixel is successfully detected as noise since  $BetP(N) > 0.5$ .

The evidential modeling method II deals with these two highly contradictory evidences in different ways. According to Eq. 13 and Eq. 14, the expected payoff matrix is generated as:

$$E = \begin{bmatrix} E[N] \\ E[S] \end{bmatrix} = \begin{bmatrix} [0.0141, 0.7914] \\ [0.1049, 0.7296] \end{bmatrix}.$$

Then, we get the normalized expected payoff vector:

$$E^{Imp} = \begin{bmatrix} E[N] \\ E[S] \end{bmatrix} = \begin{bmatrix} [0.0178, 1.0000] \\ [0.1325, 0.9219] \end{bmatrix}.$$

The generated BBAs are:

$$\begin{cases} m_1(N) = 0.0178, \\ m_1(S) = 0, \\ m_1(\Theta) = 0.9822, \end{cases} \quad \text{and} \quad \begin{cases} m_2(N) = 0.0781, \\ m_2(S) = 0.1325, \\ m_2(\Theta) = 0.7894. \end{cases}$$

After the evidence combination and probability transformation, we finally get  $BetP(N) = 0.4809$  and the target pixel is miss-detected as signal since  $BetP(N) < 0.5$ . For this example, the detection result generated by the proposed method I is more reasonable.

From the above, the two proposed methods are likely to generate different detection results in highly contradictory situations.

## IV. ADAPTIVE MEDIAN FILTERING

After the noise detection, we focus on the filter implementation. It should be better that only the corrupted pixels will undergo the filtering. The size of the filtering window influences the filtering performance a lot, and the optimal window size is usually determined by the detection result. Therefore, in this paper we further propose an adaptive switch median filtering method, which adaptively determines the size of filtering window according to the detection result.

For a given pixel at  $(i, j)$ , the filtering window with a size of  $w_F \times w_F$  centered at it is:

$$W_F(i, j) = \left\{ x_{i-s, j-t} \mid -\frac{(w_F - 1)}{2} \leq s, t \leq \frac{(w_F - 1)}{2} \right\}, \quad (17)$$

where  $x_{i-s, j-t}$  is the intensity of the pixel at  $(i - s, j - t)$ .

Generally, in order to preserve details better, the size of filtering window should be as small as possible if there are enough signal pixels in the filtering window to help determine the filtered value. In the current filtering window  $W_F(i, j)$ , the proportion of the detected signal pixels is:

$$S_{pro}^{W_F} = \frac{S_{num}^{W_F}}{w_F \times w_F}, \quad (18)$$

where  $S_{num}^{W_F}$  is the number of the detected signal pixels in  $W_F(i, j)$ . If the proportion of signal pixels in the current filtering window, i.e.,  $S_{pro}^{W_F}$  is small, the size of the filtering

Table I  
ESTIMATION RESULTS OF NOISE DENSITY (%).

Actual	10	20	30	40	50	60	70	80	90
Method I	10.01	19.98	29.99	40.00	50.01	60.01	70.00	80.01	90.01
Method II	10.01	19.98	29.98	39.99	49.99	59.99	70.00	80.00	90.00

window should be expanded to see if the proportion is large enough in a larger filtering window.

When the noise density is large,  $S_{\text{pro}}^{W_F}$  is likely to be small. Thus, the minimum  $S_{\text{pro}}^{W_F}$  required for not extending the filtering window size should be reduced with the increase of the noise density to avoid over smoothing. Therefore, the noise density should be estimated first.

#### A. Noise density estimation

The noise density is estimated according to the noise detection result:

$$\hat{d}_N = \frac{N_{\text{num}}}{P_{\text{num}}}. \quad (19)$$

Here,  $N_{\text{num}}$  is the total number of the detected noise pixels and  $P_{\text{num}}$  is the total number of the pixels in the image.

The performance of noise density estimation for corrupted Lena images are presented in Table I, where method I and method II represent the two proposed evidential modeling methods respectively. In this experiment, noise pixels take values in the sets of  $S_1 = \{0, 1, \dots, 10\}$  and  $S_2 = \{245, 246, \dots, 255\}$ , i.e.,  $\alpha = 10$ . The values of  $\varepsilon$  in Eq. (10) and Eq. (13) are 0.1. The size of  $W_D$  is  $11 \times 11$  based on a great deal of tests. According to Table I, the estimation results are very close to the actual noise densities indicating that our detection methods are effective, and they can be used to determine the size of filtering window.

#### B. Filtering method

According to the estimated noise density  $\hat{d}_N$  and the proportion of the detected signal pixels  $S_{\text{pro}}^{W_F}$ , the condition of judging whether the current filtering window should be expanded or not, is set as:

$$S_{\text{pro}}^{W_F} > (1 - \hat{d}_N) \times \beta. \quad (20)$$

Here,  $\beta$  is a scale factor taking value in the range of (0, 1). We set it as 1/4 based on a great deal of tests on various images. When the noise density is small, the minimum required  $S_{\text{pro}}^{W_F}$  for not expanding the current filtering window is large; when the noise density is large, the minimum required  $S_{\text{pro}}^{W_F}$  is small. Our filtering method can be outlined below:

- Step 1: Set the initial size of filtering window  $w_F \times w_F$  to  $3 \times 3$  and set the maximum window size to  $w_F^{\text{max}} \times w_F^{\text{max}}$ .
- Step 2: Set a filtering window  $W_F(i, j)$  centered at the target pixel at  $(i, j)$  with current size of  $w_F \times w_F$ .
- Step 3: If the proportion of the detected signal pixels in the filtering window, i.e.,  $S_{\text{pro}}^{W_F}$  satisfies the criterion in Eq. (20), go to Step 5).
- Step 4: Extend the filtering window size to  $(w_F + 1) \times (w_F + 1)$  and repeat Steps 2) - 3) until the current filtering window size reaches  $w_F^{\text{max}} \times w_F^{\text{max}}$ .

- Step 5: Apply a median filtering to the current filtering window. The output intensity  $Y_{i,j} = \text{median}\{x_{i-s,j-t} | x_{i-s,j-t} \in W_F^D(S)\}$ , where  $W_F^D(S)$  is the set of all detected signal pixels in the current filtering window.

The maximum window size is empirically given in Table II based on a large quantities of tests on various images. In Table II, different window sizes are suggested for different noise density levels.

Table II  
RECOMMENDED MAXIMUM SIZE OF FILTERING WINDOW.

Estimated noise density	$w_F^{\text{max}} \times w_F^{\text{max}}$
$0\% < \hat{d}_N < 30\%$	$3 \times 3$
$30\% < \hat{d}_N \leq 50\%$	$5 \times 5$
$50\% < \hat{d}_N \leq 70\%$	$7 \times 7$
$\hat{d}_N > 70\%$	$9 \times 9$

## V. EXPERIMENTS

The adaptive switching median filtering method we proposed includes two components: the impulse noise detection and the adaptive filtering process. Since the noise detection plays a key role in the final denoising performance, we first evaluate the performance of the noise detection. Then, we evaluate the filtering performance of our proposed adaptively median filtering and the whole denoising performance of our proposed ASMF-DBER method, respectively. Furthermore, the computational cost and sensitivity of the parameters' setting of ASMF-DBER will be discussed. We will also check the adaptability of our ASMF-DBER for the value of  $\alpha$  in the impulse noise model, which in fact controls the intensity range that the noise pixels take values in.

Experiments are carried out using several monochrome images (Fig. 12). The experiment results of several existing methods, i.e., BDND [18], IBDND [19], ACWM [20], ASWM [23], ROR-NLM [25] and WCSR [24] are also provided for comparison.

#### A. Performance evaluation of noise detection

For the two proposed noise detection methods (method I and method II) based on two different evidential modeling methods respectively, we evaluate their performances using corrupted Lena image and the results are shown in TABLE III. The performances of ACWM, BDND, ASWM and ROR-NLM methods are also provided for comparison. The performance evaluation indices used here include the false alarm rate (FAR), miss-detection rate (MDR) and accuracy rate (AR):

$$\text{FAR} = \frac{FA_{\text{num}}}{SA_{\text{num}}}, \quad (21)$$

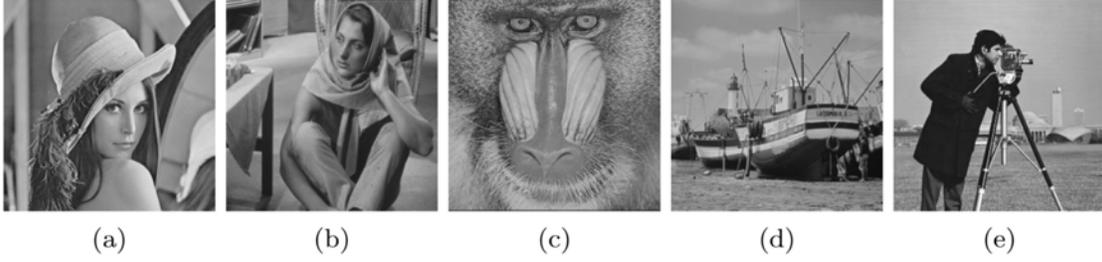


Figure 12. Monochrome images for experiments. (a) Lena. (b) Barbara. (c) Baboon. (d) Boat. (e) Cameraman.

 Table III  
 COMPARISON OF THE NOISE DETECTION PERFORMANCES FOR CORRUPTED LENA IMAGES (%).

Noise density	Performance	ACWM	BDND	ASWM	ROR-NLM	Method I	Method II
10%	FAR	0.291	0.003	0.654	1.020	0.004	0.003
	MDR	0.302	0.023	0.317	0.019	0.237	0.191
	AR	99.708	<b>99.995</b>	99.380	99.080	99.973	99.978
20%	FAR	0.608	0.002	0.927	1.176	0.004	0.002
	MDR	0.967	0.264	0.485	0.086	0.116	0.122
	AR	99.320	99.946	99.161	99.042	<b>99.974</b>	<b>99.974</b>
30%	FAR	1.263	0.004	1.372	1.341	0.003	0.002
	MDR	2.995	1.123	0.855	0.157	0.046	0.107
	AR	98.217	99.660	98.783	99.014	<b>99.984</b>	99.967
40%	FAR	2.923	0.003	2.370	1.433	0.006	0.002
	MDR	6.975	2.941	1.924	0.414	0.004	0.042
	AR	95.456	98.822	97.808	98.975	<b>99.995</b>	99.982
50%	FAR	6.191	0.010	13.893	5.663	0.014	0.002
	MDR	12.864	6.057	3.492	1.634	0.003	0.027
	AR	90.473	96.967	91.308	96.351	<b>99.992</b>	99.986
60%	FAR	12.063	0.014	13.892	5.665	0.012	0.002
	MDR	20.921	10.387	11.785	6.513	0	0.005
	AR	82.622	93.762	87.372	93.826	99.995	<b>99.996</b>
70%	FAR	21.176	0.047	29.166	18.784	0.037	0.004
	MDR	30.789	15.587	22.981	18.742	0	0.002
	AR	72.095	89.075	75.164	81.245	99.989	<b>99.997</b>
80%	FAR	33.869	0.137	52.535	43.370	0.064	0.006
	MDR	41.782	21.976	36.742	35.514	0	0.001
	AR	59.801	82.392	60.099	62.915	99.987	<b>99.998</b>
90%	FAR	50.239	0.852	76.716	74.130	0.114	0.004
	MDR	52.852	29.431	38.678	48.875	0.001	0.001
	AR	47.409	73.427	57.518	48.600	99.988	<b>99.999</b>

$$\text{MDR} = \frac{MD_{\text{num}}}{N_{\text{num}}^A}, \quad (22)$$

$$\text{AR} = \frac{P_{\text{num}} - FA_{\text{num}} - MD_{\text{num}}}{P_{\text{num}}}. \quad (23)$$

Here,  $FA_{\text{num}}$  is the number of the actual signal pixels being detected as the noise,  $MD_{\text{num}}$  is the number of the actual noise pixels being detected as the signal,  $S_{\text{num}}^A$  is the number of the actual signal pixels,  $N_{\text{num}}^A$  is the number of the actual noise pixels, and  $P_{\text{num}}$  is the number of pixels in the image.

In this experiment,  $\alpha = 10$ , i.e., noise pixels take values in  $S_1 = \{0, 1, \dots, 10\}$  and  $S_2 = \{245, 246, \dots, 255\}$ . Values of  $\varepsilon$  in Eq. (10) and Eq. (13) are 0.1. The size of  $W_D$  is empirically determined as  $11 \times 11$  based on a great deal of tests.

As shown in Table III, when the noise density is no larger than 60%, the accuracy rates of these methods are all larger than 80%. When the noise density is larger than 60%, the accuracy rates of ACWM, BDND, ASWM and ROR-NLM methods drop rapidly. However, our proposed methods still achieve high accuracy rates ( $\geq 90\%$ ).

### B. Performance of filtering

To evaluate the filtering performance, we compare the filtering performance of the proposed adaptive median filtering method with the standard median filtering (SMF) used in ACWM, ASWM and the filters used in BDND, IBDND (adaptive weighted median filter), ROR-NLM and WCSR, respectively. In this experiment,  $\alpha = 10$  and all the filters are used on the detected noise pixels generated by the proposed noise detection method I. The experimental result is shown in Fig. 13, where  $F_{\text{ROR-NLM}}$ ,  $F_{\text{BDND}}$ ,  $F_{\text{IBDND}}$  and  $F_{\text{WCSR}}$  denote the filters used in ROR-NLM, BDND, IBDND and WCSR algorithms respectively.

According to Fig. 13, when the noise density is no larger than 30%, the proposed filter has similar performance with the filters used in IBDND and WCSR. With the increase of the noise density, the proposed filter generates better performance than other filters.

### C. Performance of denoising

To verify the whole denoising performance of our proposed ASMF-DBER, we compare the denoising performance of our

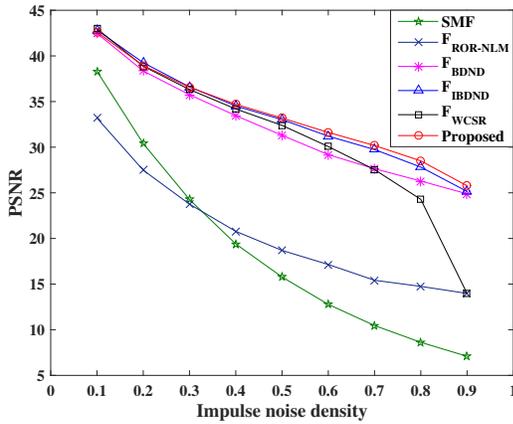


Figure 13. Comparison of filtering performances using PSNR for corrupted Lena images.

proposed ASMF-DBER with ACWM, ASWM, ROR-NLM, BDND, IBDND and WCSR using PSNR and SSIM as shown in Fig. 14 and Fig. 15, respectively. In this experiment,  $\alpha = 10$  and the size of detection window  $W_D$  is  $11 \times 11$ . ASMF-DBER I and II represent the denoising results based on the two proposed detection methods, respectively.

When the noise density is low ( $\leq 20\%$ ), BDND, IBDND, and the proposed methods have similar denoising performances since they all can obtain high detection accuracy rates in low corrupted situations (as illustrated in Table III for Lena image) and have similar denoising performances when the noise detection results are accurate enough (as shown in Fig. 13 when only the actual noise pixels are filtered).

BDND and IBDND have better performance for Lena, Baboon and Boat images when the noise density is 10%. These images have no intensities close to extreme values (0 and 255). BDND and IBDND method can obtain better performance easily since they only uses extreme criterion when detecting impulse noise.

WCSR method has very good performance on Barbara image when the noise density is 10%. The reason is that Barbara image has big areas with regular texture. When the noise density is low, since the noise detection result is accurate enough, WCSR can reconstruct the texture very well using the trained dictionaries.

With the increase in noise density, the PSNR of BDND, IBDND and WCSR are much lower than that they achieved in Fig. 13 when only the actual noise pixels, but not the detected noise pixels, are filtered. That means, when carrying out the filtering on the detected impulse noise pixels, the detection result affects the whole denoising performance significantly. BDND, IBDND and WCSR fail to achieve satisfied filtering performances because of their poor detection results.

The subjective quality comparisons of filtered images are illustrated from Fig. 16 to Fig. 19. The false alarmed pixels and miss-detected pixels of the two proposed methods are colored with red and green, respectively. Except for Cameraman image, other test images do not have many false alarms and miss-

detections. In order to highlight the colored pixels in these images, we circled the colored pixels using the corresponding colors (red for false alarms and green for miss-detections).

From the comparisons of quantitative results and visually subjective qualities, we can see that the proposed ASMF-DBER algorithms obtain superior denoising results compared with other switch median filters and the sparse representation based method. Particularly, in the high noise density cases, ASMF-DBER has obvious advantages over others.

For Cameraman image, the pixels around the edge of the “cameraman” would obtain highly contradictory evidence supports from the extreme criterion and discontinuity criterion, as the highly contradictory situation 1 (Fig. 10), and the proposed two detection methods are likely to obtain different detection results. In Fig. 17, ASMF-DBER II has more false alarms than ASMF-DBER I at these pixels, so that the denoising performance of ASMF-DBER I for Cameraman image is not so good as ASMF-DBER II, as shown in Fig. 14(e) and Fig. ??(e).

Among these algorithms, WCSR has the most parameters (8 parameters) to be determined and some of them are sensitive with the noise density, what is a challenge for WCSR to obtain a satisfied denoising result.

From the above colored incorrect detections of the proposed two methods and Table III, we can find that ASMF-DBER I generates more false alarms than ASMF-DBER II and ASMF-DBER II generates more miss-detections than ASMF-DBER I. Therefore, in practical applications, if the user relatively more emphasizes low miss-detection rate, we suggest ASMF-DBER I; if the user relatively more emphasizes low false-alarm rate, we suggest ASMF-DBER II.

#### D. Sensitivity of parameters' setting

There are two parameters to determine in our method. One is the detection window size and the other one is  $\beta$  in Eq. 20 used for deciding whether the current filtering window should be expanded or not. To discuss the sensitivity of the setting of these two parameters, we compare the denoising performances of all the combinations of the two parameters. The comparison results for the two proposed denoising methods are shown in Table IV and Table V, respectively. In this experiment,  $\beta$  changes with an incremental step  $1/8$  from  $1/8$  to  $7/8$ . The detection window size is set as  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$  or  $13 \times 13$ .

From Table IV and Table V, the filtering performance is not very sensitive to the setting of  $\beta$ . When the noise density is 10% or 20%, all the  $\beta$  generate the same performance since the limited maximum filtering window is  $3 \times 3$  when the estimated noise density is no larger than 30% according to Table II. With the increase of the noise density, the denoising performance becomes poorer when selecting small size of detection window. When the noise density is higher than 70%, large detection windows (larger than  $7 \times 7$ ) achieve obvious better denoising performance than small detection windows (no larger than  $7 \times 7$ ). When the size of detection window is

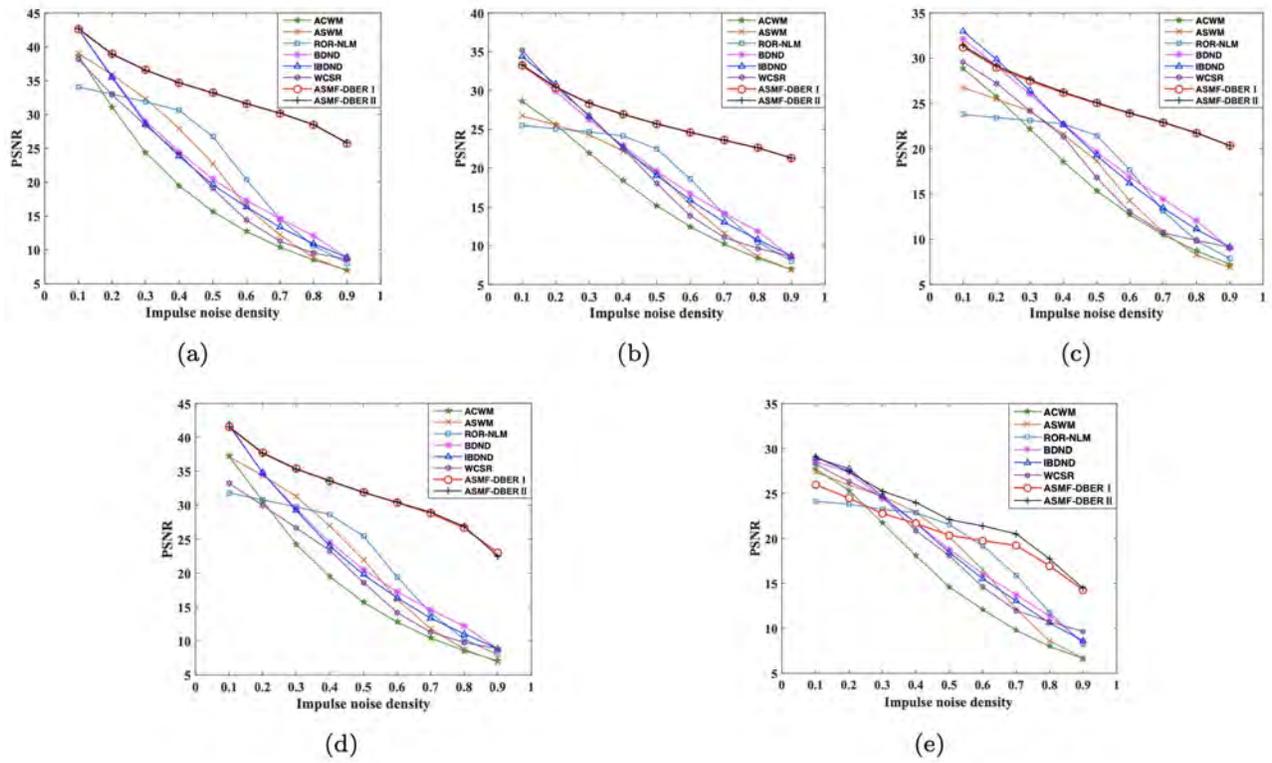


Figure 14. Comparisons of denoising performances using PSNR. (a) Lena. (b) Barbara. (c) Baboon. (d) Boat. (e) Cameraman.

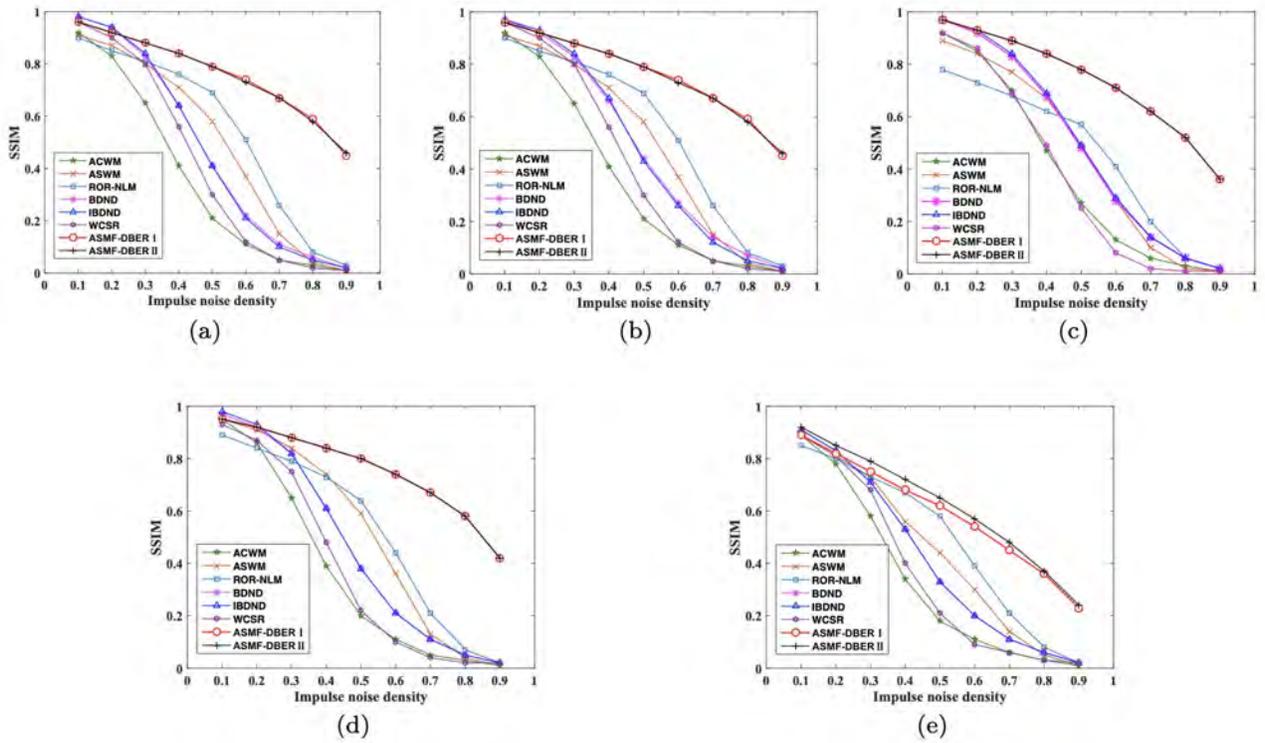


Figure 15. Comparisons of denoising performances using SSIM. (a) Lena. (b) Barbara. (c) Baboon. (d) Boat. (e) Cameraman.

Table IV  
DENOISING PERFORMANCES OF ASMF-DBER I FOR DIFFERENT DETECTION WINDOW SIZE AND  $\beta$ .

Noise	density	10%	20%	30%	40%	50%	60%	70%	80%	90%
5 × 5	$\beta = 1/8$	41.50	38.27	35.96	34.22	32.56	30.96	29.07	23.04	13.36
	$\beta = 1/4$	41.50	38.27	35.96	34.26	32.59	30.96	29.09	23.07	13.36
	$\beta = 3/8$	41.50	38.27	35.96	34.24	32.59	31.08	29.40	23.08	13.39
	$\beta = 1/2$	41.50	38.27	35.96	34.24	32.43	31.08	29.42	23.13	13.39
	$\beta = 5/8$	41.50	38.27	35.96	34.16	32.43	30.68	29.37	25.30	13.46
	$\beta = 3/4$	41.50	38.27	35.96	33.73	32.30	30.63	28.84	25.30	13.46
	$\beta = 7/8$	41.50	38.27	35.96	33.73	32.30	30.31	28.73	25.36	13.45
7 × 7	$\beta = 1/8$	41.77	38.56	36.12	34.35	32.63	31.04	29.59	27.61	20.34
	$\beta = 1/4$	41.77	38.56	36.12	34.40	32.66	31.04	29.60	27.67	20.35
	$\beta = 3/8$	41.77	38.56	36.12	34.37	32.66	31.15	29.60	27.67	20.35
	$\beta = 1/2$	41.77	38.56	36.12	34.37	32.50	31.15	29.59	27.70	20.73
	$\beta = 5/8$	41.77	38.56	36.12	34.30	32.50	30.73	29.54	27.51	20.80
	$\beta = 3/4$	41.77	38.56	36.12	33.82	32.36	30.68	28.88	27.49	20.80
	$\beta = 7/8$	41.77	38.56	36.12	33.82	32.36	30.35	28.70	27.38	21.08
9 × 9	$\beta = 1/8$	41.96	38.70	36.18	34.37	32.63	31.05	29.58	27.70	24.28
	$\beta = 1/4$	41.96	38.70	36.18	34.42	<b>32.67</b>	31.05	<b>29.61</b>	27.73	24.33
	$\beta = 3/8$	41.96	38.70	36.18	34.40	<b>32.67</b>	<b>31.16</b>	29.59	27.74	24.33
	$\beta = 1/2$	41.96	38.70	36.18	34.40	32.48	<b>31.16</b>	29.59	27.73	24.66
	$\beta = 5/8$	41.96	38.70	36.18	34.30	32.48	30.74	29.53	27.43	24.70
	$\beta = 3/4$	41.96	38.70	36.18	33.82	32.21	30.69	28.88	27.40	24.70
	$\beta = 7/8$	41.96	38.70	36.18	33.82	32.12	30.36	28.70	27.28	24.66
11 × 11	$\beta = 1/8$	41.92	<b>38.76</b>	<b>36.23</b>	34.38	32.62	31.05	29.58	27.71	25.07
	$\beta = 1/4$	41.92	<b>38.76</b>	<b>36.23</b>	<b>34.43</b>	32.65	31.05	29.61	<b>27.75</b>	25.13
	$\beta = 3/8$	41.92	<b>38.76</b>	<b>36.23</b>	34.41	32.65	<b>31.16</b>	29.59	<b>27.75</b>	25.13
	$\beta = 1/2$	41.92	<b>38.76</b>	<b>36.23</b>	34.41	32.47	<b>31.16</b>	29.59	27.74	25.25
	$\beta = 5/8$	41.92	<b>38.76</b>	<b>36.23</b>	34.31	32.47	30.74	29.53	27.45	<b>25.29</b>
	$\beta = 3/4$	41.92	<b>38.76</b>	<b>36.23</b>	33.83	32.18	30.69	28.88	27.44	<b>25.29</b>
	$\beta = 7/8$	41.92	<b>38.76</b>	<b>36.23</b>	33.83	32.10	30.36	28.70	27.32	25.18
13 × 13	$\beta = 1/8$	<b>41.99</b>	38.74	36.21	34.37	32.62	31.04	29.56	27.69	25.09
	$\beta = 1/4$	<b>41.99</b>	38.74	36.22	34.41	32.64	31.04	29.58	27.73	25.15
	$\beta = 3/8$	<b>41.99</b>	38.74	36.22	34.35	32.64	31.15	29.55	27.74	25.15
	$\beta = 1/2$	<b>41.99</b>	38.74	36.20	34.35	32.47	31.15	29.55	27.72	25.25
	$\beta = 5/8$	<b>41.99</b>	38.74	36.20	34.23	32.46	30.72	29.48	27.41	<b>25.29</b>
	$\beta = 3/4$	<b>41.99</b>	38.74	35.90	33.76	32.18	30.67	28.80	27.39	<b>25.29</b>
	$\beta = 7/8$	<b>41.99</b>	38.74	35.52	33.76	32.10	30.33	28.62	27.29	25.18

Table V  
DENOISING PERFORMANCES OF ASMF-DBER II FOR DIFFERENT DETECTION WINDOW SIZE AND  $\beta$ .

Noise	density	10%	20%	30%	40%	50%	60%	70%	80%	90%
5 × 5	$\beta = 1/8$	41.77	38.29	35.87	33.87	32.31	30.63	28.38	22.67	13.23
	$\beta = 1/4$	41.77	38.29	35.87	33.89	32.34	30.63	28.39	22.70	13.23
	$\beta = 3/8$	41.77	38.29	35.87	33.78	32.34	30.70	28.90	22.70	13.26
	$\beta = 1/2$	41.77	38.29	35.87	33.78	32.17	30.70	28.91	22.75	13.26
	$\beta = 5/8$	41.77	38.29	35.87	33.66	32.17	30.40	28.90	24.82	13.32
	$\beta = 3/4$	41.77	38.29	35.87	33.26	32.08	30.35	28.49	24.82	13.32
	$\beta = 7/8$	41.77	38.29	35.87	33.26	32.08	30.18	28.40	24.95	13.65
7 × 7	$\beta = 1/8$	42.04	38.56	36.11	34.09	32.53	30.98	29.44	27.45	19.78
	$\beta = 1/4$	42.04	38.56	36.11	34.11	32.56	30.99	29.44	27.51	19.78
	$\beta = 3/8$	42.04	38.56	36.11	34.00	32.56	31.00	29.35	27.51	19.78
	$\beta = 1/2$	42.04	38.56	36.11	34.00	32.38	31.00	29.35	27.51	20.18
	$\beta = 5/8$	42.04	38.56	36.11	33.86	32.38	30.62	29.32	27.41	20.24
	$\beta = 3/4$	42.04	38.56	36.11	33.44	32.27	30.57	28.76	27.40	20.24
	$\beta = 7/8$	42.04	38.56	36.11	33.44	32.27	30.19	28.64	27.31	21.50
9 × 9	$\beta = 1/8$	42.13	38.73	36.24	34.16	32.58	31.04	29.49	27.60	24.56
	$\beta = 1/4$	42.13	38.73	36.24	34.18	32.62	31.04	29.49	27.64	24.60
	$\beta = 3/8$	42.13	38.73	36.24	34.06	32.62	31.06	29.40	27.65	24.60
	$\beta = 1/2$	42.13	38.73	36.24	34.06	32.43	31.05	29.39	27.64	24.86
	$\beta = 5/8$	42.13	38.73	36.24	33.93	32.43	30.66	29.37	27.46	24.93
	$\beta = 3/4$	42.13	38.73	36.24	33.50	32.32	30.62	28.79	27.44	24.93
	$\beta = 7/8$	42.13	38.73	36.24	33.50	32.32	30.22	28.64	27.35	24.88
11 × 11	$\beta = 1/8$	42.15	38.77	<b>36.27</b>	34.18	32.61	31.05	<b>29.50</b>	27.60	24.97
	$\beta = 1/4$	42.15	38.77	<b>36.27</b>	<b>34.20</b>	<b>32.64</b>	31.06	<b>29.50</b>	<b>27.65</b>	25.01
	$\beta = 3/8$	42.15	38.77	<b>36.27</b>	34.09	<b>32.64</b>	<b>31.07</b>	29.41	<b>27.65</b>	25.01
	$\beta = 1/2$	42.15	38.77	<b>36.27</b>	34.09	32.45	31.06	29.40	27.64	25.12
	$\beta = 5/8$	42.15	38.77	<b>36.27</b>	33.95	32.45	30.68	29.38	27.46	<b>25.17</b>
	$\beta = 3/4$	42.15	38.77	<b>36.27</b>	33.52	32.34	30.63	28.79	27.44	<b>25.17</b>
	$\beta = 7/8$	42.15	38.77	<b>36.27</b>	33.52	32.34	30.24	28.67	27.35	25.09
13 × 13	$\beta = 1/8$	<b>42.20</b>	<b>38.79</b>	<b>36.27</b>	34.19	32.60	31.05	<b>29.50</b>	27.60	24.97
	$\beta = 1/4$	<b>42.20</b>	<b>38.79</b>	<b>36.27</b>	<b>34.20</b>	32.63	31.06	<b>29.50</b>	<b>27.65</b>	25.01
	$\beta = 3/8$	<b>42.20</b>	<b>38.79</b>	<b>36.27</b>	34.09	32.63	<b>31.07</b>	29.41	<b>27.65</b>	25.01
	$\beta = 1/2$	<b>42.20</b>	<b>38.79</b>	<b>36.27</b>	34.09	32.44	31.06	29.41	27.64	25.12
	$\beta = 5/8$	<b>42.20</b>	<b>38.79</b>	<b>36.27</b>	33.96	32.44	30.68	29.36	27.46	25.16
	$\beta = 3/4$	<b>42.20</b>	<b>38.79</b>	<b>36.27</b>	33.52	32.33	30.63	28.71	27.45	25.16
	$\beta = 7/8$	<b>42.20</b>	<b>38.79</b>	<b>36.27</b>	33.52	32.33	30.24	28.54	27.36	25.09



Figure 16. Denoising results for Lena image (noise density is 30%). (a) Corrupted image. (b) ASWM. (c) ROR-NLM. (d) BDND. (e) IBDND. (f) WCSR. (g) ASMF-DBER I. (h) ASMF-DBER II. (i) Colored detection results of ASMF-DBER I. (j) Colored detection results of ASMF-DBER II.

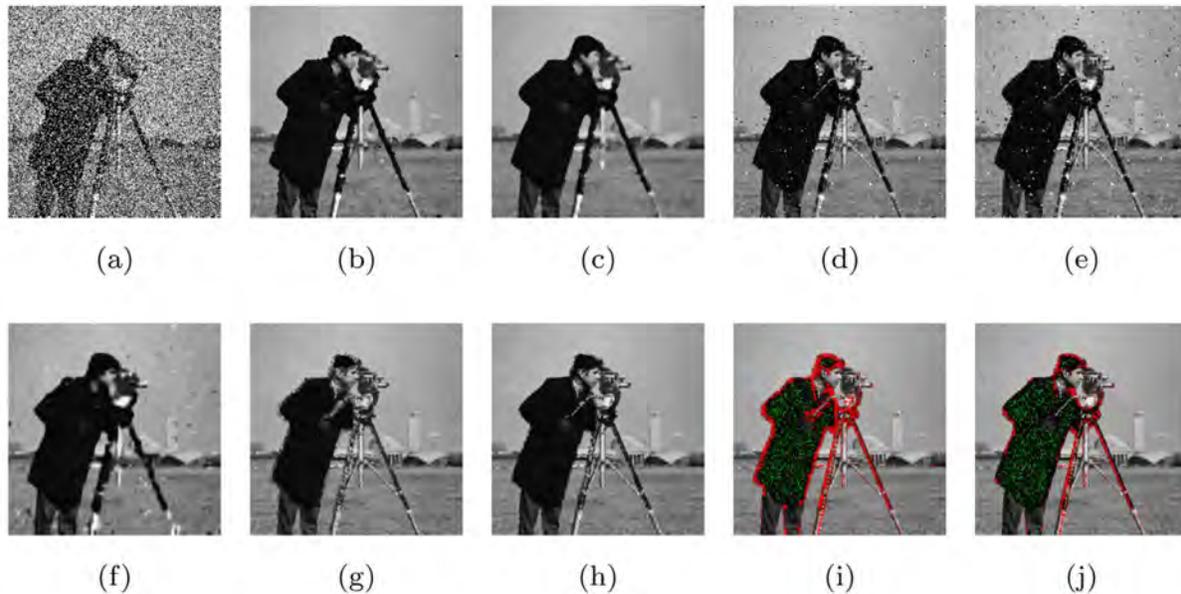


Figure 17. Denoising results for Cameraman image (noise density is 40%). (a) Corrupted image. (b) ASWM. (c) ROR-NLM. (d) BDND. (e) IBDND. (f) WCSR. (g) ASMF-DBER I. (h) ASMF-DBER II. (i) Colored detection results of ASMF-DBER I. (j) Colored detection results of ASMF-DBER II.

set as  $11 \times 11$ , and  $\beta$  is set as  $1/4$ , we can usually obtain the best denoising performance.

### E. Computational cost

The computational cost is an important index to evaluate an algorithm. We timed the computational costs of ACWM, ASWM, ROR-NLM, BDND, IBDND, WCSR and the proposed methods by running the algorithms on a Windows 7 Enterprise system equipped with Intel Core i7-4790 CPU at 3.60 GHz and 8.00 GB DDR-III memory. The comparison

of their average execution time for corrupted Lena images with size of  $512 \times 512$  are shown in Table VI. Each average execution time is calculated from 10 runs of experiments. According to Table VI, the computational cost of the proposed methods varies from 80 to 130 seconds with the increase of noise density. The sparse representation based method WCSR is most time consuming (more than 3000 seconds) and the proposed methods need more computational cost compared with ACWM, ASWM, BDND and IBDND algorithms.

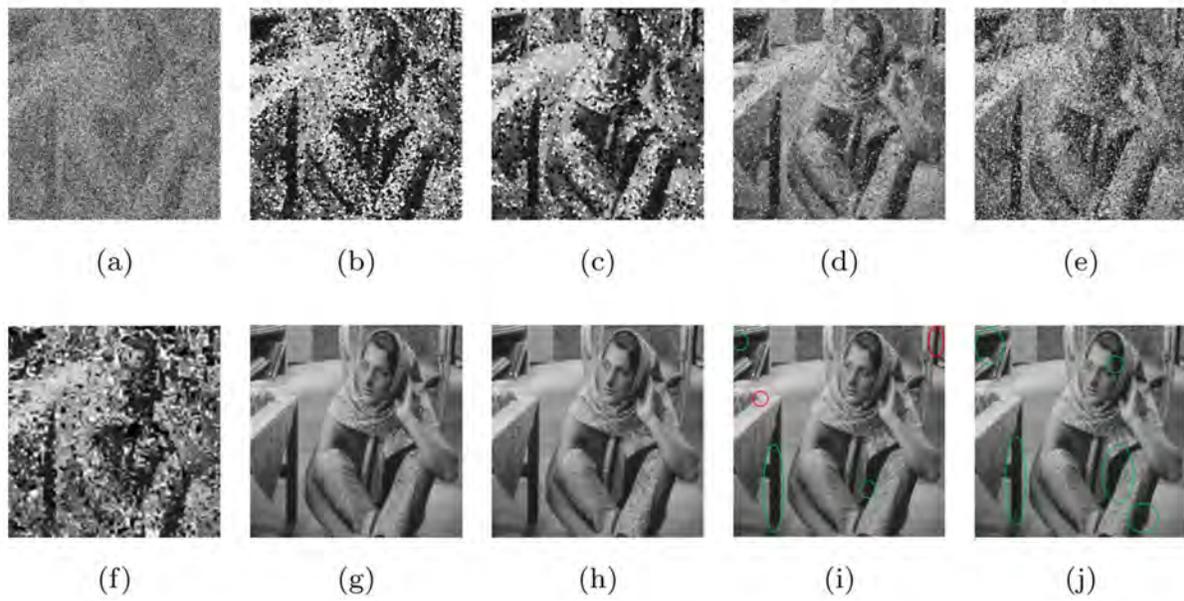


Figure 18. Denoising results for Barbara image (noise density is 70%). (a) Corrupted image. (b) ASWM. (c) ROR-NLM. (d) BDND. (e) IBDND. (f) WCSR. (g) ASMF-DBER I. (h) ASMF-DBER II. (i) Colored detection results of ASMF-DBER I. (j) Colored detection results of ASMF-DBER II.

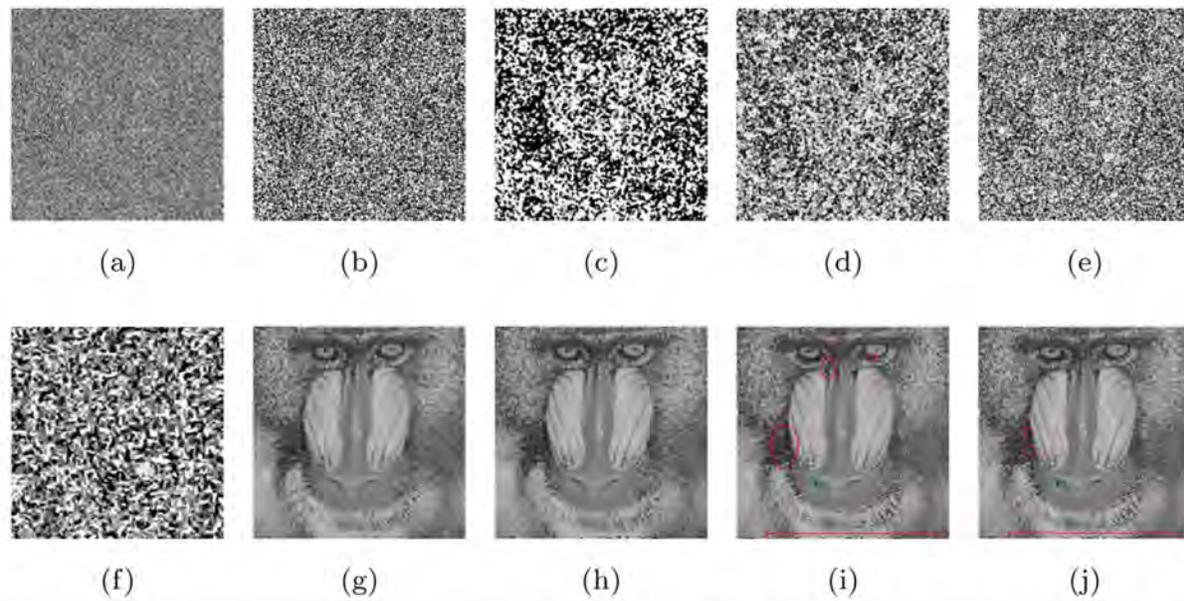


Figure 19. Denoising results for Baboon image (noise density is 90%). (a) Corrupted image. (b) ASWM. (c) ROR-NLM. (d) BDND. (e) IBDND. (f) WCSR. (g) ASMF-DBER I. (h) ASMF-DBER II. (i) Colored detection results of ASMF-DBER I. (j) Colored detection results of ASMF-DBER II.

Table VI

AVERAGE EXECUTION TIME OF EIGHT ALGORITHMS FOR CORRUPTED LENA IMAGE WITH DIFFERENT NOISE DENSITIES (UNIT SECOND).

Methods	10%	20%	30%	40%	50%	60%	70%	80%	90%
ACWM	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
ASWM	27.9	28.9	29.3	29.4	36.7	37.5	44.1	50.3	63.2
ROR-NLM	68.5	84.5	100.1	115.7	131.5	144.2	151.0	147.2	137.6
BDND	75.6	74.7	74.9	75.5	75.9	76.6	77.5	77.4	77.7
IBDND	75.8	75.0	74.4	76.0	76.2	76.7	77.2	76.9	77.2
WCSR	3281.3	3288.5	3297.8	3293.3	3328.3	3298.6	3312.7	3326.0	3219.6
ASMF-DBER I	87.8	88.4	95.7	98.5	101.3	111.3	111.5	124.9	126.3
ASMF-DBER II	88.3	88.8	96.3	99.2	102.0	113.1	114.0	125.4	127.1

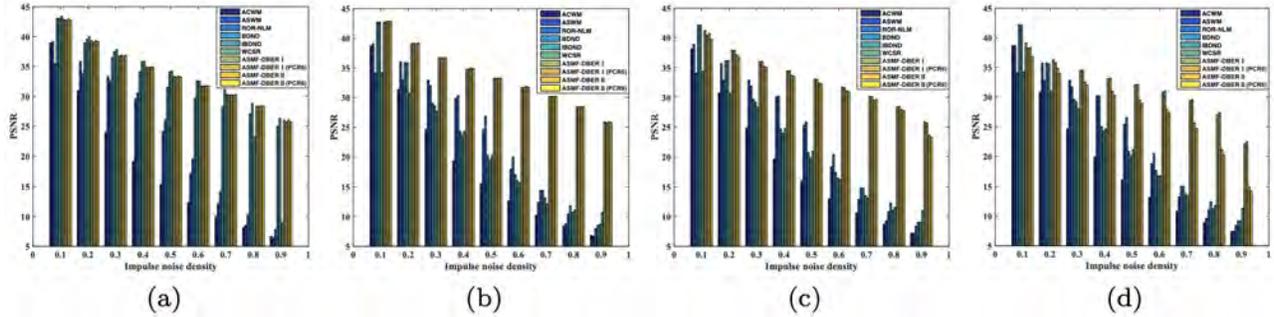


Figure 20. Comparisons of denoising performance using PSNR for Lena images corrupted by the impulse noise with various values of  $\alpha$ . (a)  $\alpha = 0$ . (b)  $\alpha = 5$ . (c)  $\alpha = 15$ . (d)  $\alpha = 20$ .

To some extent, the better denoising performance of the proposed methods is at the cost of more computational cost.

#### F. Adaptability for different impulse noise models

In order to further check the adaptability of ASMF-DBER for the different values of  $\alpha$  in noise model, i.e., the different intensity ranges for impulse noise, we use PSNR for Lena images corrupted by the impulse noise with other values of  $\alpha$ , the quantitative results are shown in Fig. 20. Furthermore, we also compare the results obtained using a recent alternative fusion rule PCR6 [37] (Proportional Conflict Redistribution rule No. 6) when combining the generated BBAs  $m_1$  and  $m_2$  in our ASMF-DBER I and ASMF-DBER II methods. These two results are denoted by ASMF-DBER I (PCR6) and ASMF-DBER II (PCR6) respectively.

As shown in Fig. 20, the PSNR of ASMF-DBER results are relatively high when  $\alpha$  varies between 0 and 15. Although they drop slightly when  $\alpha = 20$ , they are still higher than other methods in general.

## VI. CONCLUSION

To deal with the problem of the impulse noise reduction, first, we propose two impulse noise detection methods based on evidential reasoning before filtering. Second, we design an adaptive switching median filtering method, which adaptively determines the size of filtering window according to detection results. The subjective and objective analyses from our experimental results verify that our new proposed detection approaches and related filtering algorithms have superior performance compared with existing algorithms.

The generation of BBAs is crucial in evidential reasoning, however there is no general theoretical method for BBA

generation. In this paper, we use two types of BBA generation methods in evidential modeling for the uncertainties encountered in the impulse noise detection and have evaluated their performances. In future work, we will focus on other BBA generation methods, which can better depict the uncertainty encountered in the impulse noise detection. Other evidence combination rules will also be used to make comparisons. We will also do more theoretical analyses on the determination of parameters used in our algorithm. Furthermore, we will apply our impulse noise detection method to sparse representation based filtering approach to deal with more complicated noise models, such as the impulse/Gaussian mixed noise.

## ACKNOWLEDGMENT

This work was supported by the Grant for State Key Program for Basic Research of China (973) (No. 2013CB329405), National Natural Science Foundation (Nos. 61573275, 61671370), Science and Technology Project of Shaanxi Province (No. 2013KJXX-46), Postdoctoral Science Foundation of China (No. 2016M592790), and Fundamental Research Funds for the Central Universities (No. xjj2014122, xjj2016066).

## REFERENCES

- [1] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Englewood Cliffs, NJ, USA, Prentice-Hall, 2002.
- [2] J. Dezert, Z. Liu, G. Mercier, *Edge Detection in Color Images Based on DSMT*, in Proc. of Int. Conf. on Information Fusion (Fusion 2011), Chicago, Jul. 2011.
- [3] Z. Zhang, D. Han, Y. Yang, *Image segmentation based on evidential Markov random field model*, in Proc. of IEEE Int. Conf. on Control, Automation and Information Sciences, Changshu, pp. 239-244, Oct. 2015.

- [4] Y. Wang , L. Deng , *Modeling object recognition in visual cortex using multiple firing  $k$ -means and non-negative sparse coding*, Signal Processing, Vol. 124, pp. 198–209, 2016.
- [5] F. Nellro , M.J. Thurley , H. Jonsson , C. Andersson, *Automated measurement of sintering degree in optical microscopy through image analysis of particle joins*, Pattern Recognition, Vol. 48(11), pp. 3451–3465, 2015.
- [6] M.S. Zarchi , R.T. Tan , C.V. Gemeren , A. Monadjemi , R.C. Veltkamp, *Understanding image concepts using ISTOP model*, Pattern Recognition, Vol. 53, pp. 174–183, 2016.
- [7] I. Pitas, A.N. Venetsanopoulos, *Order statistics in digital image processing*, Proc. IEEE, Vol. 80(12), pp. 1893–1921, Dec. 1992.
- [8] D.R.K. Brownrigg, *The weighted median filter*, Commun. ACM, Vol. 27(8), pp. 807–818, Aug. 1984.
- [9] S. J. Ko, Y. H. Lee, *Center weighted median filters and their applications to image enhancement*, IEEE Trans. on Circuits Syst., Vol. 38(9), pp. 984–993, Sep. 1991.
- [10] T. Sun, Y. Neuvo, *Detail-preserving median based filters in image processing*, Pattern Recognit. Lett., Vol. 15(4), pp. 341–347, Apr. 1994.
- [11] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer Publishing Company, Incorporated, 2010.
- [12] M. Aharon, M. Elad, A. Bruckstein, *rmK-SVD: an algorithm for designing overcomplete dictionaries for sparse representation*, IEEE Trans. on Signal Processing, Vol. 54(11), pp. 4311–4322, Nov. 2006.
- [13] M. Elad, M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. on Image Processing, Vol. 15(12), pp. 3736–3745, Dec. 2006.
- [14] S. Wang , Y. Xia , Q. Liu , P. Dong , D.D. Feng , J. Luo, *Fenchel duality based dictionary learning for restoration of noisy images*, IEEE Trans. on Image Processing, Vol. 22(12), pp. 5214–5225, 2013.
- [15] S. Wang, Q. Liu, Y. Xia, et al., *Dictionary learning based impulse noise removal via  $l_1$ - $l_1$  minimization*, Signal Processing, Vol. 93(9), pp. 2696–2708, 2013.
- [16] S. Xu , X. Yang , S. Jiang, *A fast nonlocally centralized sparse representation algorithm for image denoising*, Signal Processing, Vol. 131, pp. 99–112, 2017.
- [17] J. Liu, X.-C. Tai, H. Huang, Z. Huan, *A weighted dictionary learning model for denoising images corrupted by mixed noise*, IEEE Trans. on Image Processing, Vol. 22(3), pp. 1108–1120, Mar. 2013.
- [18] P.-E. Ng, K.-K. Ma, *A switching median filter with boundary discriminative noise detection for extremely corrupted images*, IEEE Trans. on Image Processing, Vol. 15(6), pp. 1506–1516, June 2006.
- [19] I.F. Jafar, R.A. AlNa'mneh, K.A. Darabkh, *Efficient improvements on the BDND filtering algorithm for the removal of high-density impulse noise*, IEEE Trans. on Image Processing, Vol. 22(3), pp. 1223–1232, Mar. 2013.
- [20] T. Chen, H. Wu, *Adaptive impulse detection using center-weighted median filters*, IEEE Signal Processing Letters, Vol. 8(1), pp. 1–3, Jan. 2001.
- [21] Y. Dong, S. Xu, *A new directional weighted median filter for removal of random-valued impulse noise*, IEEE Signal Processing Letters, Vol. 14(3), pp. 193–196, Mar. 2007.
- [22] U. Ghaneker, A. K. Singh, R. Pandey, *A contrast enhancement-based filter for removal of random valued impulse noise*, IEEE Signal Processing Letters, Vol. 17(1), pp. 47–50, Jan. 2010.
- [23] S. Akkoul, R. Ledec, R. Leconge, *A new adaptive switching median filter*, IEEE Signal Processing Letters, Vol. 17(6), pp. 587–590, Apr. 2010.
- [24] C. Chen, L. Liu, L. Chen, Y. Tang, Y. Zhou, *Weighted Couple Sparse Representation With Classified Regularization for Impulse Noise Removal*, IEEE Trans. on Image Processing, Vol. 24(11), pp. 4014–4026, Nov. 2013.
- [25] B. Xiong, Z. Yin, *A universal denoising framework with a new impulse detector and nonlocal means*, IEEE Trans. on Image Processing, Vol. 21(4), pp. 1663–1675, Apr. 2012.
- [26] H. L. Eng, K. K. Ma, *Noise adaptive soft-switching median filter*, IEEE Trans. Image Processing, Vol. 10(2), pp. 1663–1675, Feb. 2001.
- [27] Z. Zhou, *Cognition and removal of impulse noise with uncertainty*, IEEE Trans. on Image Processing, Vol. 21(7), pp. 3157–3167, July 2012.
- [28] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [29] R. Garnett, T. Huegerich, C. Chui, W. He, *A universal noise removal algorithm with an impulse detector*, IEEE Trans. on Image Processing, Vol. 14(11), pp. 1747–1754, Nov. 2005.
- [30] T.L. Tan, A. Pearce, T. Carter, A. Holt, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 2002 .
- [31] Y. Yang, D. Han, *A new distance-based total uncertainty measure in the theory of belief functions*, Knowledge-Based Systems, Vol. 94, pp. 114–123, Feb. 2016.
- [32] P. Smets, *The transferable belief model*, Artificial Intelligence, Vol. 66(2), pp. 191–234, 1994.
- [33] D. Han, J. Dezert, Z. Duan, *Evaluation of probability transformations of belief functions for decision making*, IEEE Trans. Systems Man and Cybernetics: Systems, Vol. 46(1), pp. 93–108, Apr. 2016.
- [34] J.-M. Tacnet, J. Dezert, *Cautious OWA and evidential reasoning for decision making under uncertainty*, in Proc. Int. Conf. on Information Fusion (Fusion 2011), pp. 2074–2081, Chicago, USA, July, 2011.
- [35] A. Irpino, R. Verde, *Dynamic clustering of interval data using a Wasserstein-based distance*, Pattern Recognition Letters, Vol. 29(11), pp. 1648–1658, Aug. 2008.
- [36] J. Dezert, *Autonomous Navigation with Uncertain Reference Points using the PDAF*, Multitarget-Multisensor Tracking: Applications and Advances, Vol 2, pp. 271–324, Artech House, Norwood, 1991.
- [37] F. Smarandache, J. Dezert, *Advances and Applications of DSMT for Information Fusion: Collected Works*, Vols. 1–4, American Research Press, Rehoboth, NM, USA, 2004–2015.



# Drawbacks of Uncertainty Measures Based on the Pignistic Transformation

Joaquín Abellán<sup>a</sup>, Eloi Bossé<sup>b</sup>

<sup>a</sup>Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain.

<sup>b</sup>Image & Information Processing Department (iTi), IMT-Atlantique, Brest, France.

Emails: jabellan@decsai.ugr.es, ebosse861@gmail.com

Originally published as: J. Abellán, E. Bosse, *Drawbacks of Uncertainty Measures Based on the Pignistic Transformation*, IEEE Trans. On Systems, Man, and Cybernetics: Systems, Vol. 48(3), pp. 382-388, March 2018, and reprinted with permission.

**Abstract**—Recently, a measure of total uncertainty (TU) in Dempster-Shafer Theory (DST), based on the pignistic distribution called Ambiguity Measure (AM), have been modified. The resulting new measure has been simply referred as Modified Ambiguity Measure (MAM). In the literature, it has been shown that AM, in addition to showing some undesirable behaviors, has important drawbacks related to two essential properties for such measures: subadditivity and monotonicity. The MAM measure has been developed to solve the AM subadditivity problem, but this paper demonstrates that MAM suffers the same drawback as AM with respect to monotonicity. A measure of uncertainty that cannot meet the monotonicity requirement has a major drawback for its exploitation in operational contexts such as in analytics, information fusion and decision support. This paper aims at identifying and discussing drawbacks of this type of measures (AM, MAM). Our main motivation is to insist upon the important requirement of monotonicity that a TU measure should possess to improve its potential of being used and trusted in applications. This discussion is due time since the monotonicity problem needs first to be solved to avoid building too high expectations for usefulness and potential exploitation of such measures in operational communities.

**Keywords:** Imprecise probabilities, theory of evidence, measures of uncertainty, conflict, non-specificity, pignistic probability.

## I. INTRODUCTION

Dempster-Shafer's theory (DST) extends the classical probability theory (PT). In DST, more types of uncertainty can be represented than in PT. These types of uncertainty found in DST are called *conflict*, *randomness* or *discord*; and *non-specificity* respectively (see Yager [25]). Klir and Wierman [18] present a total uncertainty (TU) measure in DST that has been justified by an axiomatic approach considering TU in probability theory as a reference. They also attach to that TU definition, a set of five desired properties that TU must verify. Abellán and Masegosa [7] extend that set to add the important property of monotonicity as well as other behavioral properties related to TU.

In DST, upper (or maximum) of entropy is the only function that verifies all the basic required properties listed later in Section III: P1-P5. Jousselme et al. [15] presented a new TU measure in DST, called AM, based on the pignistic distribution. The authors, in 2006, proved that AM verifies the needed properties (P1-P5) and that AM sorts out other shortcomings of upper entropy. However, Klir and Lewis [19] found that AM

function does not, in fact, verify requirement P4: subadditivity. Abellán and Masegosa [7] presented an extension of the set of the required properties (P1-P5) that a TU measure in DST must verify. They extended the set (P1-P5) from Klir and Wierman [18] and added some desirable behaviors that TU should have. In 2008, Abellán and Masegosa [7] showed that AM does not verify the important property of monotonicity (P6) in addition to present some undesirable behaviors. Recently, Shahpari and Seyedin [22] have presented a modified function of AM, called MAM. They claim that MAM verifies the required properties (P1-P5) as well as behaves correctly in applications. That claim motivates our discussion here about the drawbacks of such measures.

This paper aims at identifying drawbacks of such measures of uncertainty (AM, MAM) based on the pignistic transformation of a basic probability assignment in DST. Our main motivation is to insist upon the important requirement of monotonicity (P6) that a TU measure should possess to improve its potential to be used and trusted in applications such as in analytics, information fusion and decision support. Defined as they are in Jousselme et al. [15] and Shahpari and Seyedin [22], AM and MAM will produce incorrect results and undesirable behaviors if used in applications.

The paper is organized as follows. Section II reviews briefly the representation of information and uncertainty within the framework of the Dempster-Shafer Theory (DST). Section III discusses the drawbacks of AM and MAM with respect to the required properties of a total uncertainty (TU) in DST. Section IV presents a discussion on desirable behavioral requirements of a TU measure in DST. We conclude in Section V.

## II. INFORMATION REPRESENTATION IN THEORY OF EVIDENCE

### A. Dempster-Shafer theory of Evidence

Let  $X$  be a finite set considered as a set of possible situations,  $|X| = n$ ,  $\wp(X)$  the power set of  $X$  and  $x$  any element in  $X$ . Dempster-Shafer theory of evidence (Dempster [9], Shafer [23]) is based on a function called *basic probability assignment* (b.p.a.), that is a mapping  $m : \wp(X) \rightarrow [0, 1]$ , such that  $m(\emptyset) = 0$  and  $\sum_{A \subseteq X} m(A) = 1$ . A set  $A$  such that  $m(A) > 0$  is called a *focal element* of  $m$ .

Let  $X, Y$  be finite sets. Let  $X \times Y$  be the product space of those sets and  $m$  a b.p.a. on  $X \times Y$ . We use  $m^{\downarrow X}$  to note to the marginal b.p.a. on  $X$ , (and similarly on  $Y$ ,  $m^{\downarrow Y}$ ), and it is defined as follows:

$$m^{\downarrow X}(A) = \sum_{R|A=R^{\downarrow X}} m(R), \quad \forall A \subseteq X \quad (1)$$

where  $R^{\downarrow X}$  is the set projection of  $R$  on  $X$ .

Associated with each basic probability assignment, there exists two functions: a belief function,  $Bel$ , and a plausibility function,  $Pl$ :  $Bel(A) = \sum_{B \subseteq A} m(B)$ ,  $Pl(A) = \sum_{A \cap B \neq \emptyset} m(B)$ . They can be considered as the lower and upper probability of  $A$ , respectively.

We may note that belief and plausibility are interrelated for all  $A \in \wp(X)$ , by  $Pl(A) = 1 - Bel(A^c)$ , where  $A^c$  denotes the complement of  $A$ . Furthermore,  $Bel(A) \leq Pl(A)$ .

On each b.p.a. on a finite set  $X$ , there exists a set of associated probability distributions  $p$  on  $X$ , of the following way:

$$K_m = \{p \mid Bel(A) \leq p(A), \forall A \in \wp(X)\} \quad (2)$$

We remark that  $Bel(A) \leq p(A)$  is, in this case, equivalent to  $Bel(A) \leq p(A) \leq Pl(A)$ .  $K_m$  is a closed and convex set of probability distributions, also called a credal set in the literature.

### B. Measures of uncertainty in DST

The entropy function (Shannon [24]) on probability theory is defined by the following continuous function:

$$S(p) = - \sum_{x \in X} p(x) \log_2(p(x)), \quad (3)$$

where  $p = (p(x))_{x \in X}$  is a probability distribution on  $X$ ,  $p(x)$  is the probability of value  $x$  and  $\log_2$  is used to quantify the value in bits, but in the literature is used  $\log$  and  $\log_2$  indifferently. The value  $S(p)$  (also used as  $S(p(x_1), p(x_2), \dots)$ ) quantifies the only type of uncertainty presented on probability theory. This measure in PT verifies a large set of properties (see Shannon [24], Klir and Wierman [18]).

In DST, Yager [25] found different two types of uncertainty: the first one appears when a b.p.a. has positive masses on sets with empty intersections; and the other one appears when the b.p.a. has positive masses on sets with cardinality greater than one. Those types of uncertainty are normally called *conflict* and *non-specificity*, respectively.

Dubois and Prade [10], introduced in DST a function based on the classical Hartley measure (Hartley [11]) on classical set theory. That measure, noted as  $I$ , represents a measure of non-specificity for a b.p.a. It is expressed in the following way:

$$I(m) = \sum_{A \subseteq X} m(A) \log(|A|). \quad (4)$$

This measure  $I(m)$  has its minimum value (zero) for a b.p.a.  $m$  that is a probability distribution. Its maximum value,

$\log(|X|)$ , is obtained for a b.p.a.,  $m$ , where  $m(X) = 1$  and  $m(A) = 0, \forall A \subset X$ . We can see in the literature that  $I$  verifies a large set of needed properties for such a type of measure.

In the 90's, some measures were introduced with the aim to quantify the conflict degree that a b.p.a. expresses (see Klir and Wierman [18]). Yager [25] introduced the following function:

$$E(m) = - \sum_{A \subseteq X} m(A) \log Pl(A). \quad (5)$$

But this function does not verify in DST all the required properties.

The measure:  $S^*(m)$  equal to the maximum/upper of the entropy (upper entropy) on the set of probability distributions verifying  $Bel(A) \leq \sum_{x \in A} p(x) \leq Pl(A), \forall A \subseteq X$ , was proposed by Harmanec and Klir [12], [13]. This set of probability distributions is the credal set associated with a b.p.a.  $m$ , that we have noted as  $K_m$  in Eq. (2).

$S^*$  is considered as a total uncertainty measure in DST: a measure that quantifies both types of uncertainty: conflict and non-specificity. But in Harmanec and Klir [13] was not separated in those corresponding parts. It can be seen in Abellán, Klir and Moral [6], that this measure can be separating coherently in conflict and non-specificity parts. In DST those parts are similar to the ones for general credal sets. It can be consider

$$S^*(m) = S_*(m) + (S^* - S_*)(m), \quad (6)$$

with  $S^*(m)$  the maximum entropy and  $S_*(m)$  the minimum entropy on the credal set  $K_m$  associated to a b.p.a.  $m$ .  $S_*(m)$  coherently quantifying the conflict part and  $(S^* - S_*)(m)$  its non-specificity part. That measure has been successfully used in applications (see Abellán and Moral [3])

To quantify conflict and non-specificity (ambiguity) in DST, Jousselme et al. [15] presented a measure based on the pignistic distribution on DST. Let  $m$  be a b.p.a. on a finite set  $X$ , then the pignistic probability distribution  $BetP_m$ , is defined on all the subsets  $A$  in  $X$  as follows:

$$BetP_m(A) = \sum_{B \subseteq X} m(B) \frac{|A \cap B|}{|B|}. \quad (7)$$

For a singleton set  $A = \{x\}$ , we have  $BetP_m(\{x\}) = \sum_{x \in B} [m(B)/|B|]$ . Hence, as the authors says [15], the *Ambiguity Measure* (AM) for a b.p.a.  $m$  on a finite set  $X$  can be defined as:

$$AM(m) = \sum_{x \in X} BetP_m(x) \log(BetP_m(x)), \quad (8)$$

i.e., the entropy of the  $BetP_m$  probability distribution.

Recently, Shahpari and Seyedin [22] have presented a modified function of AM, based on the AM drawbacks identified in Klir and Lewis [19]. This function is called MAM and unfortunately presents some mathematical shortcomings in its definition as explained below. Shahpari and Seyedin exposed

that MAM coincides with AM on 1-dimensional space but in the case of 2-dimensional space, they used a different definition for the pignistic distribution without providing the essential justification. Here are the details. Let  $X, Y$  be finite sets, and  $m$  a b.p.a. on  $X \times Y$ . On  $X \times Y$ , MAM coincides with AM function, via the probability distribution  $MBet_{m_{XY}} = Bet_{m_{XY}}$  (MAM is the entropy of that probability distribution). In the case of 2-dimensional space, Shahpari and Seyedin use the following function on each marginal b.p.a.:

$$MBet_{m_X}(x_i) = \sum_{B \in \wp(X), B \ni x_i} \sum_{A \in \wp(X \times Y), B = A \downarrow X} \frac{m_{XY}(A) \#(x_i \in A)}{|A|}, \forall x_i \in X \quad (9)$$

with  $\#(x_i \in A)$  the number of appearances of  $x_i$  in the set  $A$ , and  $|A|$  the cardinal of  $A$ . Similarly they define the values  $MBet_{m_Y}(y_i), \forall y_i \in Y$ .

To simplify, this expression can be reduced to the following one:

$$MBet_{m_X}(x_i) = \sum_{A \in \wp(X \times Y), A \downarrow X \ni x_i} \frac{m_{XY}(A) \#(x_i \in A)}{|A|}, \forall x_i \in X \quad (10)$$

MAM function is then the entropy of those probability distributions on  $X$  and  $Y$  respectively. Authors supposed that  $MBet_{m_X}$  is a probability distribution on  $X$  that belongs to the credal set associated with the marginal of  $m$  on  $X$  ( $K_{m \downarrow X}$ ). That essential assertion has not been verified or proved by the authors. The following property below is our way to prove that assertion.

Proposition 1: With the above notation, the probability distribution expressed by Eq. (9) (or Eq. (10)) belongs to the associated credal set of the marginal b.p.a. on the space  $X$ . As  $MBet_{m_X}$  is a probability distribution on  $X$ , it is only necessary to prove the following expression  $\forall B \subseteq X$ :

$$Bel_{m \downarrow X}(B) \leq MBet_{m_X}(B) = \sum_{x_i \in B} MBet_{m_X}(x_i).$$

Without lost of generality, let  $B$  be a set in  $\wp(X)$  such that  $B = \{x_1, x_2, \dots, x_r\}$ ; and  $A_1, A_2, \dots, A_t$  be all the focal sets of  $m$ ,  $A_i$  in  $\wp(X \times Y)$ , such that  $A_i \downarrow X \subseteq B$ .

We know that

$$MBet_{m_X}(x_1) = \frac{m(A_1) \#(x_1 \in A_1)}{|A_1|} + \dots + \frac{m(A_t) \#(x_1 \in A_t)}{|A_t|} + \alpha_1$$

$$MBet_{m_X}(x_r) = \frac{m(A_1) \#(x_r \in A_1)}{|A_1|} + \dots + \frac{m(A_t) \#(x_r \in A_t)}{|A_t|} + \alpha_r$$

with all  $\alpha_i \geq 0, i = 1, \dots, r$ . We remark the following considerations in the above expressions of  $MBet_{m_X}(x_i), i = 1, \dots, r$ :

- The values  $\#(x_i \in A_j)$  can be 0 if  $x_i \notin A_j$
- The  $\alpha_i$  values come from the sets  $C \in \wp(X \times Y)$  such that  $x_i \in C$  but  $C \downarrow X \not\subseteq B$ .

If we sum all the above expressions, we have that the first terms sums to  $m(A_1)$ , the second ones to  $m(A_2)$ , and so on

with all the rest. Hence,

$$MBet_{m_X}(B) = \sum_{x_i \in B} MBet_{m_X}(x_i) = \sum_j m(A_j) + \alpha,$$

with  $\alpha = \sum_i \alpha_i \geq 0$ .

Finally

$$Bel_{m \downarrow X}(B) = \sum_j m(A_j) \leq MBet_{m_X}(B)$$

With the Property 1 we show that  $MBet_{m_X}$  is compatible with the correct definition of a marginal b.p.a. of  $m$  on  $X$ . But, as we will show in Section III, it would require some modification in the definition of the MAM measure, for being mathematically correct.

### III. DRAWBACKS OF MEASURES BASED ON PIGNISTIC TRANSFORMATION IN DST

In Klir and Wierman [18] are exposed requirements for uncertainty measures in DST that quantify both types of uncertainty, i.e. total uncertainty measures. These requirements, in form of properties are the following ones:

- (P1) Probabilistic consistency: A total uncertainty measure must be equal to the Shannon entropy in the case that the focal elements of a b.p.a.  $m$  are all singleton ones:

$$TU(m) = \sum_{x \in X} m(x) \log m(x). \quad (11)$$

- (P2) Set consistency: When exist a set  $A$  such that  $m(A) = 1$  then  $TU$  must collapse to the Hartley measure:

$$TU(m) = \log |A|. \quad (12)$$

- (P3) Range: The range of  $TU(m)$  must be  $[0, \log |X|]$ .  
 (P4) Subadditivity: Let  $m$  be a b.p.a. on the space  $X \times Y$ ,  $m \downarrow X$  and  $m \downarrow Y$  its marginal b.p.a.s on  $X$  and  $Y$  respectively, then  $TU$  must satisfy the following inequality:

$$TU(m) \leq TU(m \downarrow X) + TU(m \downarrow Y). \quad (13)$$

- (P5) Additivity: Let  $m$  be a b.p.a. on the space  $X \times Y$ ,  $m \downarrow X$  and  $m \downarrow Y$  its marginal b.p.a.s on  $X$  and  $Y$  respectively such that these marginal are not interactive ( $m(A \times B) = m \downarrow X(A) m \downarrow Y(B)$ , with  $A \subseteq X, B \subseteq Y$  and  $m(C) = 0$  if  $C \neq A \times B$ ), then  $TU$  must satisfy the equality:

$$TU(m) = TU(m \downarrow X) + TU(m \downarrow Y). \quad (14)$$

These requirements pretend to extend those of Shannon's entropy in probability theory. In DST there are two different types of uncertainty; one more than in classic PT. The requirement of range is in some way debatable. In literature, one can find arguments in favor of a larger range. In PT never we can find a probability distribution that contains the information of other probability distribution. But, this situation can appear in DST: one b.p.a. can contains the information of another b.p.a., as we can see in the following example.

Example 1: In the analysis about the mortality of a traffic accident can be related to 3 main causes: (1) External causes (EC), i.e. poor visibility, bad road and/or atmospheric conditions; (2) overspeed (OS); (3) driver ill conditions due to consumption of harmful substances (D). Let assume the following two situations in the analysis of a particular accident.

*Situation1*: we have 3 Evidences ( $E_1$ ,  $E_2$  and  $E_3$ ) on the causes of the mortality of an accident. An expert expresses his knowledge in form of the following b.p.a. on the universal  $X = \{EC, OS, D\}$ :

$$E_1 \longrightarrow m_1(\{EC, OS\}) = \beta_1,$$

$$E_2 \longrightarrow m_1(\{EC, D\}) = \beta_2,$$

$$E_3 \longrightarrow m_1(\{OS, D\}) = \beta_3,$$

where  $\beta_i > 0$ ,  $i = 1, 2, 3$ , and  $\sum_i \beta_i = 1$ .

*Situation2*: thinking better on the Evidences, now the expert finds that there exist some reasons to not discard  $D$  in the Evidence  $E_1$ , i.e. a lost in information is produced. Hence, he thinks that he must to change his b.p.a to the following one:

$$E_1 \longrightarrow m_2(\{EC, OS, D\}) = \beta_1,$$

$$E_2 \longrightarrow m_2(\{EC, D\}) = \beta_2,$$

$$E_3 \longrightarrow m_2(\{OS, D\}) = \beta_3.$$

It is clear that *Situation2* represents a situation with a greater level of uncertainty than *Situation1*. Here, we have  $Bel_2(A) \leq Bel_1(A)$  and  $Pl_1(A) \leq Pl_2(A)$ ,  $\forall A \subseteq X$ , implying a bigger level of uncertainty for  $m_2$ .

*Example 1* expresses a situation that must be taken into account for a total uncertainty measure in DST. Hence, for a TU in DST is necessary to verify the following property:

(P6) Monotonicity: A total uncertainty measure in DST must take into account the decreasing or increasing in information.

Formally, let 2 b.p.a.s be on a finite set  $X$ ,  $m_1$  and  $m_2$ , verifying that  $Bel_2(A) \leq Bel_1(A)$ ,  $\forall A \subseteq X$ , then it must be verified that:

$$TU(m_1) \leq TU(m_2). \quad (15)$$

The above definition of monotonicity is called *weak inclusion*.

Using the results of Klir and Wierman [18], Jousselme et al. [15], Klir and Lewis [19], Abellán and Masegosa [7] and Shahpari and Seyedin [22], we have the initial following properties for  $S^*$ ,  $AM$  and  $MAM$  functions:

$S^*$ : P1, P2, P3, P4, P5 and P6.

$AM$ : P1, P2, P3 and P5.

$MAM$ : P1, P2, P3, P4<sup>-</sup> and P5.

$S^*$  is the only measure that satisfies all the proposed properties. We mark P4 for MAM as P4<sup>-</sup> because that property is satisfies under a controversial way as we explain below.

#### A. Drawbacks of the AM measure

In Jousselme et al. it is showed that  $AM$  function satisfies P4 requirement. In Klir and Lewis [19] it is showed that there

exists an error in their proof about the P4 requirement, and the  $AM$  function does not verifies that property.

Abellán and Masegosa [7] proved that also  $AM$  function does not verifies P6 property. Extended details can be found in those two references.

#### B. Drawbacks of the MAM measure

MAM measure coincides on 1-dimensional space with AM measure, then the example provided in Abellán and Masegosa [7] about the non-monotonicity for AM is valid for MAM. Consequently, MAM does not verifies P6 property. Moreover, if we consider *Example 1* with values  $\beta_1 = 0.3$ ,  $\beta_2 = 0.5$  and  $\beta_3 = 0.2$ , we have the following values for MAM:

$$1.049 = MAM(m_2) < MAM(m_1) = 1.081$$

These values imply that decreasing the real information, MAM can give us lower values of uncertainty. This drawback must not be allowed for such a type of measure.

MAM measure satisfies the Property P4 in a controversial way, as discussed below.

When looking for a measure that verifies P4 property, the definition of MAM function is not mathematically totally correct as eluded in the previous section. The authors says that “In cases where the projection process is not used, the modified pignistic probability is computed employing (13)”, where “(13)” is the equation of the AM function (Eq. (8)).

It means that on (1-dimensional)  $m^{\perp X}$  one would use the AM function but when the pignistic probability comes from a join b.p.a then one must used Eq. (9). That is questionable.

Let discuss this question by the following example:

Example 2: Let  $m_1$  be a b.p.a. on space  $X = \{x_1, x_2, x_3\}$ , and masses

$$m_1(\{x_1, x_2\}) = 0.6, m_1(\{x_3\}) = 0.1, m_1(X) = 0.3$$

Then MAM quantifies the information of  $m_1$  as the entropy,  $S$ , of the probability distribution (0.4, 0.4, 0.2).

On the other hand, if we consider the space  $Y = \{y_1, y_2\}$  and  $m$  on  $X \times Y$  with masses

$$m(\{z_{11}, z_{12}, z_{21}\}) = 0.6, m(\{z_{31}, z_{32}\}) = 0.1,$$

$$m(X \times Y) = 0.3,$$

where  $z_{ij} = (x_i, y_j)$ . Here the MAM function quantifies the information of  $m$  on  $X$  as  $S(0.5, 0.3, 0.2)$ . But, in this case,  $m^{\perp X} = m_1$ . This incongruence in the definition of MAM can be arranged by a different definition of marginal as described below.

Shahpari and Seyedin [22] do not use the standard definition of marginal of a b.p.a. and its properties, with the aim that MAM function verifies the property P4. They do not use the standard definition of subadditivity for a measure  $M$  where it must be verified that, being  $m$  a b.p.a. on  $X \times Y$   $M(m) \leq M(m^{\perp X}) + M(m^{\perp Y})$ , having correct definitions of marginal b.p.a.s. As said in Shahpari and Seyedin [22]: “some probabilistic information is lost in the classic DST projection process”. One avenue would have been to change

the standard definition of marginal in DST. The standard definition of marginal is an extension of the one used in PT. Shahpari and Seyedin choose instead to define a different 'pignistic' probability distribution on the marginal b.p.a.s on  $X$  and  $Y$ . It is clear that  $Betm(m_X) \in K_{m^{\downarrow X}}$ , but that  $MBetm(m_X) \in K_{m^{\downarrow X}}$  has not been verified in the work of Shahpari and Seyedin. We prove that assertion in Section II.

Suppose that we have a general b.p.a.  $m$  on the space  $X \times Y \times T$ . If we need to work on  $X \times Y$  space, we use the standard marginal on  $X \times Y$ ,  $m^{\downarrow XY}$ . But to use the 'pignistic' values in this case, should we use  $MBetm_{XY}$  or  $Betm_{XY}$ ?, i.e. the one from  $m$  or the direct one from  $m^{\downarrow XY}$ . If we need to use the similar values on  $X$  or  $Y$ , should we use the values from  $m$  or the values from  $m^{\downarrow XY}$ ?. These questions appear because of the controversial definition of MAM.

From Shahpari and Seyedin [22], with  $S$  the entropy function, we know that:

$$S(MBetm) \leq S(MBetm_{XY}) + S(MBetm_T), \quad (16)$$

$$S(MBetm) \leq S(MBetm_X) + S(MBetm_Y) + S(MBetm_T). \quad (17)$$

But how is  $S(MBetm_{XY})$  with respect to  $S(MBetm_X) + S(MBetm_Y)$ ?. It should be checked that similar inequality is verified.

How is the relation between  $MBetm_X$ , obtained from  $m$  and the similar one obtained from  $m^{\downarrow XY}$ ?. They can be different.

All questions raised above will not appear if, for instance:

- A new definition of the marginal of a b.p.a., compatible with PT and the definition of MAM function, is used.
- The definition MAM on each b.p.a. is independent from its origin, i.e it should not be taken into account if it comes from a joint b.p.a. or not.

The problem of the marginal found by Shahpari and Seyedin is an intrinsic one of the DST as extension of PT, because on a joint space  $X \times Y$  has sense that a b.p.a has as focal element a set  $C$ , with  $C \neq A \times B$ , for any  $A \subseteq X$  and  $B \subseteq Y$ .

Hence, under our point of view, the definition of the function MAM has some drawbacks but the critical one is that it does not verify the monotonicity property P6 and consequently, limits considerably its usage in applications.

#### IV. BEHAVIORAL REQUIREMENTS FOR TOTAL UNCERTAINTY MEASURES IN DST

Jousselmé et al. [15] discuss some shortcomings of  $S^*$  function (upper entropy) in DST and provide a comparison with the behavior of the  $AM$  function. These shortcomings have been presented by Klir and collaborators in the literature and can be expressed of the following way:

- (1) The measure must be computable directly or via algorithms.

- (2) The measure must be separated coherently in the two types of uncertainty coexisting in DST: conflict and non-specificity.
- (3) The measure must be sensitivity to changes of evidence.

These considerations are analyzed in Abellán and Masegosa [7] adding for (3) that a TU should be sensitive to changes in evidence directly or via its parts. They found that it is possible that an increment of conflict causes a decrease in non-specificity and vice versa. It is shown that we can have similar total uncertainty value with different conflict and non-specificity parts. Hence, Abellán and Masegosa [7] showed that a set of Behavioral Requirements (BR) for a TU in DST could be exposed in the following way:

- (BR1) A TU should have a direct calculus or via an algorithm.
- (BR2) A TU must not conceal the two types of uncertainty coexisting in the evidence theory: conflict and non-specificity.
- (BR3) A TU must be sensitive in changes of evidence, directly or via its parts of conflict and non-specificity.

Based on the "Generalized Information theory", of Klir [16], we know that there exists situations where the information can be expressed by more general models. Hence, the *Principle of uncertainty invariance* expresses that "the amount of uncertainty (and information) must be preserved when a representation of uncertainty in one mathematical theory is transformed into its counterpart in another theory". Via this principle, Abellán and Masegosa [7], considered other requirement for a TU in DST, that one can called *Extensibility*:

- (BR4) The extension of a TU in DST on more general theories must be possible.

The above requirements (BR1-BR4) have been analyzed in Abellán and Masegosa [7] for  $S^*$  and  $AM$  functions. Here, we verify for MAM, comparing with  $AM$  and  $S^*$ . As these requirements do not need the use on a joint space, the case of MAM coincides with  $AM$ :

**BR1:**  $AM$  and  $MAM$  functions have a simpler calculation than the  $S^*$ , it is only necessary to obtain the pignistic probability distribution of a b.p.a. The calculation of  $S^*$  in DST has a high computational complexity. The algorithm of Meyerowitz et al [21] was the first to obtain this value. Recently, the computation of this algorithm has been reduced by Liu et al. [20]. Hence, we could conclude that the calculation of all TU in DST can be implemented in a simple way, though they have different complexity.

**BR2:** Abellán, Klir and Moral [6] separate  $S^*$  in two parts:  $S_*$  (minimum of entropy) as a conflict measure and  $S^* - S_*$  as a non-specificity measure. In Abellán and Moral [4], it is shown a branch and bound algorithm to get  $S_*$  on DST and more general theories.  $AM$  and  $MAM$  functions do not present a clear separation between conflict and non-specificity. In Jousselmé et al. [15],  $AM$  is presented as a special case of the function  $\delta S^* + (1 - \delta)I$ , for an unknown  $\delta \in (0, 1)$ . So,

using  $AM(m)$  value, one can not know which quantity corresponds to conflict and which one to non-specificity. **BR3**: In Abellán and Masegosa [7], we see that the problem on the sensitivity to changes of evidence of  $S^*$ , raised in Klir and Smith [17], is not totally justified if we consider its parts. In the examples shown in Klir and Smith [17], an increasing in the conflict part produces a decreasing in the non-specificity part and viceversa. In Jousselme et al. [15], we see that this problem does not appears for  $AM$  (and then for  $MAM$ ), but we do not know what happen with the variations of conflict or non-specificity in those measures.

**BR4**: Using the results of Abellán and Moral [1], [2] and in Abellán, Klir and Moral [6],  $S^*$  can be extended to more general theories than DST, verifying similar sets of properties. In addition, On these theories exist algorithms to obtain  $S^*$  and  $S_*$ , as we can see in Abellán and Moral [4], [5]. The extensions of  $AM$  and  $MAM$  functions on more general theories is still an open question. One possibility for this extension could be to use the Möbius transform as for  $I$  function (see Abellán and Moral [1]), though its calculation would be more complex. As we can see in Abellán, Klir and Moral [6] that the generalization of some uncertainty measures defined in DST could have many problems when they are extended on more general theories. That is the case of  $E$  measure in DST. The question here might be: is it worthy if property P6 is not met?

In addition, Abellán [8] showed, via an example, that in situations where there is no conflict,  $S^*$  could present some questionable behavior that could be considered as not totally correct. In these situations,  $S_*$ , that quantifies the part of conflict, is equal to 0. One could think that intuitively  $S^*$  does not reflect a proper behavior. In that example, 2 b.p.a.s with a clear difference of information are used:  $m_1$  and  $m_2$ , with  $K_{m_1} \subset K_{m_2}$  and  $S^*(m_1) = S^*(m_2)$ .

If we apply  $AM$  or  $MAM$  to similar situations, we get an ill behavior for those measures. The monotonicity property P6 is broken as illustrated by the following example.

Example 3: Let  $m_1$  and  $m_2$  the following b.p.a.s on a finite set  $X = \{x_1, x_2, x_3\}$ , with values:

$$m_1(\{x_1, x_3\}) = 0.5, m_1(\{x_2, x_3\}) = 0.5$$

$$m_2(\{x_2, x_3\}) = 0.5, m_2(\{x_1, x_2, x_3\}) = 0.5$$

It can be checked that  $K_{m_1} \subset K_{m_2}$  ( $Bel_2(A) \leq Bel_1(A), \forall A \subseteq X$ ) and a clear decreasing in information appears when we pass from  $m_1$  to  $m_2$ .

Example 3 presents a case where there is no conflict among their focal sets in both b.p.a.s. In this example  $S^*(m_1) = S^*(m_2)$ ; but

$$1.040 = AM(m_1) > AM(m_2) = 1.028$$

and similar situation appears for  $MAM$ . It clearly shows again an incorrect behavior for the measures based on the pignistic probability distribution.

## V. CONCLUSIONS

This paper discussed drawbacks of total uncertainty (TU) measures drafted in DempsterShafer Theory (DST) and based upon the pignistic transformation. Those measures have been labelled as (AM, MAM) in this paper. The first step was to recall the basic requirements that those measures should meet. Previous work presented analysis about shortcomings of upper entropy which is the only measure that meet those basic requirements within DST. Previous work also added some desirable behavioral requirements as well as an important property to the basic set: monotonicity. Both AM and MAM fail that critical requirement. In addition, we have demonstrated that AM and MAM defined as they are in Jousselme et al. [15] and Shahpari and Seyedin [22], will produce incorrect results and undesirable behaviors if used in applications. Some drawbacks have been corrected but several questions raised, for instance, in Section III would require more extensive work not included here. We provided some hints to solve some issues but the main problem is that both AM and MAM do not meet the important requirement of monotonicity that limit considerably the interest towards their operational applications.

## ACKNOWLEDGEMENTS

We acknowledge Prof. George Klir, who, by his interactions and previous remarks on the measures of uncertainty, inspires and encourages us to pursue further, the analysis of the drawbacks of such measures.

## REFERENCES

- [1] J. Abellán, S. Moral, *A Non-specificity Measure for Convex Sets of Probability Distributions*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 8, pp. 357–367, 2000.
- [2] J. Abellán, S. Moral, *Maximum of entropy for credal sets*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 18, pp. 1215–1225, 2003.
- [3] J. Abellán, S. Moral, *Upper Entropy of Credal Sets. Applications to Credal Classification*, Int. J. of Approximate Reasoning, Vol. 39, pp. 235–255, 2005.
- [4] J. Abellán, S. Moral, *Difference of entropies as a non-specificity measure for credal sets*, Int. J. of General Systems, Vol. 34(3), pp. 201–214, 2005.
- [5] J. Abellán, S. Moral, *An algorithm that computes the upper entropy for order-2 capacities*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 13(4), pp. 467–467, 2005.
- [6] J. Abellán, G.J. Klir, S. Moral, *Disaggregated total uncertainty measure for credal sets*, Int. J. of General Systems, Vol. 35(1), pp. 29–44, 2006.
- [7] J. Abellán, A. Masegosa, *Requirements for total uncertainty measures in Dempster-Shafer theory of evidence*, Int. J. of General Systems, Vol. 37(6), pp. 733–747, 2008.
- [8] J. Abellán, *Combining nonspecificity measures in Dempster-Shafer theory of evidence*, Int. J. of General Systems, Vol. 40(6), pp. 611–622, 2011.
- [9] A.P. Dempster, *Upper and Lower Probabilities Induced by a Multivaluated Mapping*, Ann. Math. Statistic, Vol. 38, pp. 325–339, 1967.
- [10] D. Dubois, H. Prade, *A Note on Measure of Specificity for Fuzzy Sets*, BUSETAL, Vol. 19, pp. 83–89, 1984.
- [11] R.V.L. Hartley, *Transmission of information*, The Bell Systems Technical Journal, Vol. 7, pp. 535–563, 1928.
- [12] D. Harmanec, G.J. Klir, *Measuring total uncertainty in Dempster-Shafer Theory: a novel approach*, Int. J. General System, Vol. 22, pp. 405–419, 1994.

- [13] D. Harmanec, G.J. Klir, *Principle of uncertainty revisited*, in Proc. of 4th Intern. Fuzzy Systems and Intelligent Control Conf., Maui, Hawaii, pp. 331–339, 1996.
- [14] M. Higashi, G.J. Klir, *Measures of uncertainty and information based on possibility distributions*, Int. J. General System, Vol. 9, pp. 43–58, 1983.
- [15] A.-L. Jousselme, C. Liu, D. Grenier, E. Bossé, *Measuring Ambiguity in the Evidence Theory*, IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 36(5), pp. 890–903, 2006.
- [16] G.J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley, Hoboken, NJ, 2006.
- [17] G.J. Klir, R.M. Smith, *Recent developments in generalized information theory*, Int. J. Fuzzy Systems of Mathematics and Artificial Intelligence, Vol. 1(1), pp. 1–13, 1999.
- [18] G.J. Klir, M.J. Wierman, *Uncertainty-Based Information*, Phisica-Verlag, 1998.
- [19] G.J. Klir, H.W. Lewis, *Remarks on “Measuring Ambiguity in the Evidence Theory”*, IEEE Trans. on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol. 38(4), pp. 995–999, 2008.
- [20] C. Liu, D. Grenier, A.-L. Jousselme, E. Bossé, *Reducing Algorithm Complexity for Computing an Aggregate Uncertainty Measure*, IEEE Trans. on Systems, Man and Cybernetics, Part A, Vol. 37(5), pp. 669–679, 2007.
- [21] A. Meyerowitz, F. Richman, E.A. Walker, *Calculating maximum-entropy probabilities densities for belief functions*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 2, pp. 377–389, 1994.
- [22] A. Shahpari, S.A.Seyedin, *Using mutual aggregate uncertainty measures in a threat assessment problem constructed by Dempster-Shafer network*, IEEE Trans. on Systems, Man and Cybernetics: Systems, Vol. 45(6), pp. 877–886, 2015.
- [23] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
- [24] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.
- [25] R.R. Yager, *Entropy and Specificity in a Mathematical Theory of Evidence*, Int. J. of General Systems, Vol. 9, pp. 249–260, 1983.



# On Decombination of Belief Function

Deqiang Han<sup>a</sup>, Yi Yang<sup>b</sup>, Jean Dezert<sup>c</sup>

<sup>a</sup>Center for Information Engineering Science Research, Xi'an Jiaotong University, Xi'an, China.

<sup>b</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an, China.

<sup>c</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

Emails: deqhan@xjtu.edu.cn, jiafeiyi@xjtu.edu.cn, jean.dezert@onera.fr

Originally published as: D. Han, Y. Yang, J. Dezert, *On Decombination of Belief Function*, in Proc. of Int. Conf. on Information Fusion (Fusion 2019), Ottawa, Canada, July 2–5, 2019, and reprinted with permission.

**Abstract**—The evidence combination is a kind of decision-level information fusion in the theory of belief functions. Given two basic belief assignments (BBAs) originated from different sources, one can combine them using some combination rule, e.g., Dempster's rule to expect a better decision result. If one only has a combined BBA, how to determine the original two BBAs to combine? This can be considered as a defusion of information. This is useful, e.g., one can analyze the difference or dissimilarity between two different information sources based on the BBAs obtained using evidence decombination. Therefore, in this paper, we research on such a defusion in the theory of belief functions. We find that it is a well-posed problem if one original BBA and the combined BBA are both available, and it is an under-determined problem if both BBAs to combine are unknown. We propose an optimization-based approach for the evidence decombination according to the criteria of divergence maximization. Numerical examples are provided illustrate and verify our proposed decombination approach, which is expected to be used in applications such the difference analysis between information sources in information fusion systems when the original BBAs are discarded, and performance evaluation of combination rules.

**Index Terms**—information fusion, decombination, belief functions, combination, divergence maximization

## I. INTRODUCTION

The theory of belief functions, which is also known as the Dempster-Shafer evidence theory [1], has been widely used in many information fusion based applications including the pattern classification [2], [3], multi criteria decision making (MCDM) [4], fault diagnosis [5] and image processing [6].

The information fusion in the theory of belief functions is implemented by evidence combination based on some combination rule, e.g., the well-known Dempster's rule. There have also emerged various alternative combination rules including Yager's rule [7], Dubois & Prade's rule [8], Smets' rule [9], Murphy's rule [10], Florea's rule [11], proportional conflict redistribution 5 (PCR5), and PCR6 [12], [13], etc.

The inverse process of the information fusion, which can also be called as information “defusion” or “decombination”, is also meaningful in information processing and analysis. Like the blind source separation (BSS) [14] and independent

component analysis [15], which aim to recover independent sources given only the observations that are unknown linear mixtures of the unobserved independent source signals, can be considered as a process of information decombination. One can analyze the original information sources and judge their relationship based on the results obtained using decombination. The community of belief functions theory seldom research on the information decombination problem, which means that given a combined BBA, how to determine the original BBAs for the combination. In Smets' work [16], the concept of decomposition of evidence was proposed, which focuses on decomposing any BBA (not always assumed to a combined BBA) into many simple support function of BBAs. He also proposed the inverse operation of evidence combination, which only focus on the following case: given a combined BBA and one BBA participating the combination, how to restore another BBA participating the combination. In this paper, we focus on the information decombination (separation) or evidence decombination in the theory of belief functions. For simplicity, here we only concern the evidence decombination for two information sources. We find that given the combined BBA together with one original BBA, it is well-posed, that is, the other BBA can be uniquely determined. However, it turns out to be an under-determined problem (with multiple solutions) if both BBAs participating the combination are unknown and the combined BBA is given. The optimization (maximization) based decombination method is proposed accordingly, where the objective function is the distance between the two original BBAs (unknown variables to determine). Examples and experiments are provided to illustrate and verify our proposed information decombination method for the belief function.

## II. BASICS OF BELIEF FUNCTIONS THEORY

The basic concept in the theory of belief functions [1] is the frame of discernment (FOD), which is determined by what we want to know and what we know. Elements in an FOD are mutually exclusive and exhaustive.  $m : 2^\Theta \rightarrow [0, 1]$  is defined as a basic belief assignment (BBA, also called a mass function) defined on the FOD  $\Theta$  satisfying

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad m(\emptyset) = 0 \quad (1)$$

This work was supported by the National Natural Science Foundation (No. 61573275, No. 61671370), Postdoctoral Science Foundation of China (No. 2016M592790), Postdoctoral Science Research Foundation of Shaanxi Province (No. 2016BSHEDZZ46), and Fundamental Research Funds for the Central Universities (No. xjj2016066)

where  $2^\Theta$  denotes the powerset of  $\Theta$ . if  $\forall m(A) > 0$ , then  $A$  is called a focal element of  $m(\cdot)$ . If a BBA only has singleton focal elements, then it is called a Bayesian BBA.

Given a BBA  $m(\cdot)$ , its corresponding belief function ( $Bel$ ) and plausibility function ( $Pl$ ) are respectively defined as

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (3)$$

The belief  $Bel(A)$  represents the justified specific support for the focal element (or proposition)  $A$ , while the plausibility  $Pl(A)$  represents the potential specific support for  $A$ . The length of the belief interval  $[Bel(A), Pl(A)]$  represents the imprecision degree of  $A$ .

The evidence combination is the fusion of the BBAs originated from different sources. Two independent BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  can be combined using Dempster's rule of combination [1] defined by

$$m(A) = \begin{cases} 0, & A = \emptyset \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i)m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)}, & A \neq \emptyset \end{cases} \quad (4)$$

Dempster's rule in general can be considered as a multiplicative and conjunctive fusion rule. Dempster's rule of combination has been criticized for its counter-intuitive behaviors [17], [18], especially in high conflict cases. Many alternative combination rules have been proposed accordingly. See [12], [19], [20] for details. Other researchers like Haenni [21] think that the conflict results from a fault in the framing of problem.

Distance of evidence is for measuring the dissimilarity between BBAs. The most commonly used and strict distance of evidence is Jousselme's distance [22] defined as follows.

$$d_J(m_1, m_2) \triangleq \sqrt{0.5 \cdot (m_1 - m_2)^T \mathbf{Jac} (m_1 - m_2)} \quad (5)$$

where the elements  $Jac(A, B)$  of Jaccard's weighting matrix  $\mathbf{Jac}$  are defined as

$$\mathbf{Jac}(A, B) = |A \cap B| / |A \cup B| \quad (6)$$

Here  $A, B$  are focal elements of  $m_1$  and  $m_2$ , respectively. Jaccard's matrix has been proved to be positive-definite [23], therefore, Jousselme's distance is a strict metric satisfying four requirements of the distance metric including the non-negativity, non-degeneracy, symmetry, and triangular inequality.

### III. EVIDENCE DECOMBINATION IN BELIEF FUNCTIONS THEORY

The evidence combination can be considered as a procedure of information fusion<sup>1</sup> as shown in Fig. 1.



Fig. 1. Evidence combination - Information Fusion.

Given a BBA obtained after the combination, if one wants to know the possible original BBAs, then the evidence decombination is needed, which can be considered as a procedure of information decombination as shown in Fig. 2.



Fig. 2. Evidence decombination - Information Decombination or "Defusion".

In this paper, we focus on determining the original BBAs given a combined BBA. First, we analyze the relationship between the combined BBA and the original ones. For simplicity, we only suppose that there are two original BBAs in this paper.

#### A. Relation between Combined BBA and Original Ones according to Dempster's Rule

According to the Dempster's rule in Eq.(4), one can obtain the following equations. Suppose that  $m_1(\cdot)$  and  $m_2(\cdot)$  are two BBAs defined on the FOD  $\Theta = \{\theta_1, \dots, \theta_n\}$ . For each BBA, there are at most  $2^n - 1$  focal elements as shown below.

$$\begin{matrix} m_1 & m_2 \\ \begin{bmatrix} \{\theta_1\} \\ \{\theta_2\} \\ \{\theta_1, \theta_2\} \\ \vdots \\ \Theta \end{bmatrix} & \begin{bmatrix} \{\theta_1\} \\ \{\theta_2\} \\ \{\theta_1, \theta_2\} \\ \vdots \\ \Theta \end{bmatrix} \end{matrix}$$

Define a matrix  $R^{(k)}$  for each  $k = 1, \dots, 2^n - 1$  where

$$R^{(k)}(i, j) = \begin{cases} 1, & \text{if } C_k = A_i \cap B_j \\ 0, & \text{if } C_k \neq A_i \cap B_j \end{cases} \quad (7)$$

where  $A_i$  is the focal element of  $m_1(\cdot)$ , and where  $B_j$  is the focal element of  $m_2(\cdot)$ . The combined BBA is  $m(\cdot) = m_1(\cdot) \oplus m_2(\cdot)$ , and  $C_k$  is the focal element of  $m(\cdot)$ . Note that  $i, j, k = 1, \dots, 2^n - 1$ . According to Dempster's rule, the mass assignment of focal element  $C_k$  in the combined BBA is

$$m(C_k) = \frac{\begin{bmatrix} m_1(\{\theta_1\}) \\ m_1(\{\theta_2\}) \\ \vdots \\ m_1(\Theta) \end{bmatrix}^T R^{(k)} \begin{bmatrix} m_2(\{\theta_1\}) \\ m_2(\{\theta_2\}) \\ \vdots \\ m_2(\Theta) \end{bmatrix}}{1 - K} \quad (8)$$

<sup>1</sup>or information compression because from two BBAs we get one.

where  $k = 1, \dots, 2^n - 1$  and  $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)$  denotes the conflict coefficient. For simplicity in the sequel, we denote the mass value vector as

$$\mathbf{m}_1 = \begin{bmatrix} m_1(\{\theta_1\}) \\ m_1(\{\theta_2\}) \\ m_1(\{\theta_1, \theta_2\}) \\ m_1(\{\theta_3\}) \\ m_1(\{\theta_1, \theta_3\}) \\ m_1(\{\theta_2, \theta_3\}) \\ m_1(\Theta) \end{bmatrix}^T, \mathbf{m}_2 = \begin{bmatrix} m_2(\{\theta_1\}) \\ m_2(\{\theta_2\}) \\ m_2(\{\theta_1, \theta_2\}) \\ m_2(\{\theta_3\}) \\ m_2(\{\theta_1, \theta_3\}) \\ m_2(\{\theta_2, \theta_3\}) \\ m_2(\Theta) \end{bmatrix}^T$$

Then, Eq. (8) can be rewritten as

$$m(C_k) = \frac{\mathbf{m}_1^T R^{(k)} \mathbf{m}_2}{1 - K}$$

1) *Case I:* In this case, the combined BBA  $m(\cdot)$  is available, and both original BBAs are unknown. That is,  $m_1(A_i)$  ( $i = 1, \dots, 2^n - 1$ ) and  $m_2(B_j)$  ( $j = 1, \dots, 2^n - 1$ ) are unknown variables to determine, then the quantity of the unknown variable is  $2^n - 1 \times 2 = 2^{n+1} - 2$ . For the BBA, there exists

$$\sum_{i=1}^{2^n-1} m_1(A_i) = 1 \quad (9)$$

$$\sum_{j=1}^{2^n-1} m_2(B_j) = 1 \quad (10)$$

Considering Eqs. (8)-(10), we have  $2^n - 1 + 2 = 2^n + 1$  simultaneous equations. As aforementioned, to determine all the mass values of  $m_1(\cdot)$  and  $m_2(\cdot)$ , we have  $2^{n+1} - 2$  unknown variables. That is, the quantity of the unknown variables is larger than that of the equations. Therefore, this is an under-determined problem with multiple solutions in general.

2) *Case II:* In this case, the combined BBA  $m(\cdot)$  and one original BBA (e.g.,  $m_1(\cdot)$ ) are available, while another original BBA (e.g.,  $m_2(\cdot)$ ) is unknown. To determine  $m_2(\cdot)$ , we have  $2^n - 1$  unknown variables. By considering Eqs. (8) and (10), we have  $2^n$  simultaneous equations. That is, the quantity of the unknown variables is less than that of the equations. Therefore, this is an over-determined problem, and then  $m_2(\cdot)$  can be determined uniquely.

### B. Optimization Based Evidence Decombination

As aforementioned, given a combined BBA, to determine the two original BBAs is an under-determined problem, for which, the optimization-based approach is feasible. Then, the key issue is to select an appropriate criterion to establish the objective function for the optimization.

In fact, the evidence decombination is like the blind source separation (BSS), where the divergence between different sources are used for the optimization based source separation, e.g, minimization of the mutual information (MMI) [24], which represents the largest divergence. Therefore, in this paper, we use for reference the criterion in BSS to design the objective function in optimization based evidence decombination. Here we use the distance of evidence to

describe the divergence between BBAs. Furthermore, we use the simultaneous equations including the Eqs (8)-(10) together with inequalities (to assure a legal BBA<sup>2</sup> with the mass value lies in  $[0,1]$ ) as the constraints for the distance maximization to implement the evidence decombination as illustrated in Eq. (11).

$$\max_{\mathbf{m}_1, \mathbf{m}_2} d_J(\mathbf{m}_1, \mathbf{m}_2) = \sqrt{0.5 \cdot (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{J} \mathbf{a} c (\mathbf{m}_1 - \mathbf{m}_2)}$$

$$s.t. \begin{cases} m(C_k) = \frac{\mathbf{m}_1^T R^{(k)} \mathbf{m}_2}{1 - K} \\ \sum_{i=1}^{2^n-1} m_1(A_i) = 1 \\ \sum_{j=1}^{2^n-1} m_2(B_j) = 1 \\ 0 \leq m_1(A_i) \leq 1, \forall i = 1, \dots, 2^n - 1 \\ 0 \leq m_2(B_j) \leq 1, \forall j = 1, \dots, 2^n - 1 \end{cases} \quad (11)$$

By solving<sup>3</sup> the constrained maximization problem in Eq. (11), one can obtain a pair of BBAs that are farthest to each other, and that provide the combined BBA when fused with Dempster's rule.

## IV. NUMERICAL EXAMPLES OF EVIDENCE DECOMBINATION BASED ON OPTIMIZATION

In this section we give different examples illustrating how BBAs decombination can be obtained based on optimization of evidence decombination.

### A. Example 1

Suppose that the FOD is  $\{\theta_1, \theta_2, \theta_3\}$ . A BBA obtained after the combination of two unknown BBAs is

$$\begin{aligned} m(\{\theta_1\}) &= 0.1, m(\{\theta_2\}) = 0.2, m(\{\theta_1, \theta_2\}) = 0.1, \\ m(\{\theta_3\}) &= 0.1, m(\{\theta_1, \theta_3\}) = 0.1, \\ m(\{\theta_2, \theta_3\}) &= 0.3, m(\Theta) = 0.1. \end{aligned}$$

The equality constraints for the maximization problem include

$$m(\{\theta_1\}) = 0.1 = \frac{\begin{bmatrix} m_1(\{\theta_1\}) \\ m_1(\{\theta_2\}) \\ m_1(\{\theta_1, \theta_2\}) \\ m_1(\{\theta_3\}) \\ m_1(\{\theta_1, \theta_3\}) \\ m_1(\{\theta_2, \theta_3\}) \\ m_1(\Theta) \end{bmatrix}^T R^{(1)} \begin{bmatrix} m_2(\{\theta_1\}) \\ m_2(\{\theta_2\}) \\ m_2(\{\theta_1, \theta_2\}) \\ m_2(\{\theta_3\}) \\ m_2(\{\theta_1, \theta_3\}) \\ m_2(\{\theta_2, \theta_3\}) \\ m_2(\Theta) \end{bmatrix}}{1 - K}$$

where

$$R^{(1)} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

It can be rewritten to a simpler form as

$$m(\{\theta_1\}) = 0.1 = \frac{\mathbf{m}_1^T R^{(1)} \mathbf{m}_2}{1 - K}$$

For other focal elements,

$$m(\{\theta_2\}) = 0.2 = \frac{\mathbf{m}_1^T R^{(2)} \mathbf{m}_2}{1 - K}$$

<sup>2</sup>to obtain admissible BBAs with values in  $[0,1]$  and their sum equals to one.

<sup>3</sup>Here we use the global optimization toolbox in Matlab<sup>TM</sup>.

where

$$R^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$m(\{\theta_1, \theta_2\}) = 0.1 = \frac{\mathbf{m}_1^T R^{(3)} \mathbf{m}_2}{1 - K}$$

where

$$R^{(3)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$m(\{\theta_3\}) = 0.1 = \frac{\mathbf{m}_1^T R^{(4)} \mathbf{m}_2}{1 - K}$$

where

$$R^{(4)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$m(\{\theta_1, \theta_3\}) = 0.1 = \frac{\mathbf{m}_1^T R^{(5)} \mathbf{m}_2}{1 - K}$$

where

$$R^{(5)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$m(\{\theta_2, \theta_3\}) = 0.3 = \frac{\mathbf{m}_1^T R^{(6)} \mathbf{m}_2}{1 - K}$$

where

$$R^{(6)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$m(\Theta) = 0.1 = \frac{\mathbf{m}_1^T R^{(7)} \mathbf{m}_2}{1 - K}$$

where

$$R^{(7)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and the two equations in Eqs.(9) and Eqs.(10). The inequality constraints are

$$\begin{aligned} 0 &\leq m_1(\{\theta_1\}) \leq 1 \\ 0 &\leq m_1(\{\theta_2\}) \leq 1 \\ 0 &\leq m_1(\{\theta_1, \theta_2\}) \leq 1 \\ 0 &\leq m_1(\{\theta_3\}) \leq 1 \\ 0 &\leq m_1(\{\theta_1, \theta_3\}) \leq 1 \\ 0 &\leq m_1(\{\theta_2, \theta_3\}) \leq 1 \\ 0 &\leq m_1(\Theta) \leq 1 \end{aligned}$$

and

$$\begin{aligned} 0 &\leq m_2(\{\theta_1\}) \leq 1 \\ 0 &\leq m_2(\{\theta_2\}) \leq 1 \\ 0 &\leq m_2(\{\theta_1, \theta_2\}) \leq 1 \\ 0 &\leq m_2(\{\theta_3\}) \leq 1 \\ 0 &\leq m_2(\{\theta_1, \theta_3\}) \leq 1 \\ 0 &\leq m_2(\{\theta_2, \theta_3\}) \leq 1 \\ 0 &\leq m_2(\Theta) \leq 1 \end{aligned}$$

According to the constrained maximization in Eq. (11), one can obtain two BBAs as follows:

$$\begin{aligned} m_a(\{\theta_1\}) &= 0, m_a(\{\theta_2\}) = 0, m_a(\{\theta_1, \theta_2\}) = 0.0323, \\ m_a(\{\theta_3\}) &= 0.1612, m_a(\{\theta_1, \theta_3\}) = 0.1612, \\ m_a(\{\theta_2, \theta_3\}) &= 0.4840, m_a(\Theta) = 0.1613. \end{aligned}$$

and

$$\begin{aligned} m_b(\{\theta_1\}) &= 0.0834, m_b(\{\theta_2\}) = 0, m_b(\{\theta_1, \theta_2\}) = 0.3666, \\ m_b(\{\theta_3\}) &= 0, m_b(\{\theta_1, \theta_3\}) = 0.0001, \\ m_b(\{\theta_2, \theta_3\}) &= 0, m_b(\Theta) = 0.5499. \end{aligned}$$

It is easy to verify that the combination result  $m_a(\cdot) \oplus m_b(\cdot)$  is the same as the given BBA  $m(\cdot)$ .

### B. Example 2

Suppose that there are two BBAs defined on the FOD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ :

$$\begin{aligned} m_1(\{\theta_1\}) &= 0.6, m_1(\{\theta_2\}) = 0.2, \\ m_1(\{\theta_2, \theta_3\}) &= 0.1, m_1(\Theta) = 0.1. \end{aligned}$$

and

$$\begin{aligned} m_2(\{\theta_1\}) &= 0.2, m_2(\{\theta_2\}) = 0.6, \\ m_2(\{\theta_2, \theta_3\}) &= 0.1, m_2(\Theta) = 0.1. \end{aligned}$$

By calculating the Jousselme's distance in Eq. (5), one obtains that

$$d_J(m_1, m_2) = 0.4.$$

With Dempster's rule of combination, one obtains that  $m(\cdot) = m_1(\cdot) \oplus m_2(\cdot)$  with

$$\begin{aligned} m(\{\theta_1\}) &= 0.3846, m(\{\theta_2\}) = 0.5385, \\ m(\{\theta_2, \theta_3\}) &= 0.0577, m(\Theta) = 0.0192. \end{aligned}$$

According to the evidence decombination approach in Eq (11), one obtains that

$$\begin{aligned} m_a(\{\theta_1\}) &= 0, m_a(\{\theta_2\}) = 0.8750, \\ m_a(\{\theta_2, \theta_3\}) &= 0.0851, m_a(\Theta) = 0.0400. \end{aligned}$$

and

$$\begin{aligned} m_b(\{\theta_1\}) &= 0.9399, m_b(\{\theta_2\}) = 0, \\ m_b(\{\theta_2, \theta_3\}) &= 0.0131, m_b(\Theta) = 0.0470. \end{aligned}$$

It is easy to verify that the combination result  $m_a(\cdot) \oplus m_b(\cdot) = m(\cdot)$ , which is the same as  $m_1(\cdot) \oplus m_2(\cdot)$ .

By calculating the Jousselme's distance given by Eq. (5), one can verify that

$$d_J(m_a, m_b) = 0.9265 > d_J(m_1, m_2) = 0.4.$$

### C. Example 3

A given combined BBA is the same as that in Example 2.

$$\begin{aligned} m(\{\theta_1\}) &= 0.3846, m(\{\theta_2\}) = 0.5385, \\ m(\{\theta_2, \theta_3\}) &= 0.0577, m(\Theta) = 0.0192. \end{aligned}$$

Moreover, suppose that we have additional information and we also know  $m_1(\cdot)$ :

$$\begin{aligned} m_1(\{\theta_1\}) &= 0.6, m_1(\{\theta_2\}) = 0.2, \\ m_1(\{\theta_2, \theta_3\}) &= 0.1, m_1(\Theta) = 0.1. \end{aligned}$$

Then, we try to use the BBA decombination to calculate the  $\hat{m}_2(\cdot)$  and to check whether it is the same as  $m_2(\cdot)$  in

Example 2. Here is just the case II as aforementioned in Sect III.A. Therefore,  $\hat{m}_2(\cdot)$  should be unique. So, there should exist  $m_2(\cdot) = \hat{m}_2(\cdot)$ . It is an over-determined problem, and we can still use the optimization to solve  $\hat{m}_2(\cdot)$  by modifying the optimization problem to

$$\begin{aligned} \max_{\hat{m}_2} d_J(m_1, \hat{m}_2) &= \sqrt{0.5 \cdot (m_1 - \hat{m}_2)^T \mathbf{Jac}(m_1 - \hat{m}_2)} \\ \text{s.t.} \quad &\begin{cases} m(C_k) = \frac{\mathbf{m}_1^T R^{(k)} \hat{\mathbf{m}}_2}{2^n - 1} \\ \sum_{j=1}^{2^n - 1} \hat{m}_2(B_j) = 1 \\ 0 \leq \hat{m}_2(B_j) \leq 1, \forall j = 1, \dots, 2^n - 1 \end{cases} \end{aligned} \quad (12)$$

where

$$\hat{\mathbf{m}}_2 = \begin{bmatrix} \hat{m}_2(\{\theta_1\}) \\ \hat{m}_2(\{\theta_2\}) \\ \vdots \\ \hat{m}_2(\Theta) \end{bmatrix}$$

By solving Eq. (12), one obtains

$$\begin{aligned} \hat{m}_2(\{\theta_1\}) &= 0.2, \hat{m}_2(\{\theta_2\}) = 0.6, \\ \hat{m}_2(\{\theta_2, \theta_3\}) &= 0.1, \hat{m}_2(\Theta) = 0.1. \end{aligned}$$

That is, given a combined BBA and one original BBA, another original one can be determined uniquely.

## V. FURTHER ANALYSIS ON EVIDENCE DECOMBINATION

### A. Divergence Minimization or Maximization?

In the evidence decombination shown in Eq. (11), distance maximization is adopted. This is inspired by the minimization of mutual information (i.e., the maximization of divergence) between sources in Blind Source Separation (BSS), which aims to bring out more independent components [24]. One can also try to implement the evidence decombination based on the distance minimization. Based on our analysis, we find that if the distance minimization is used, the minimum distance will be zero and the BBAs of two sources are identical.

Suppose that  $m_1(\cdot) = m_2(\cdot) = m_0(\cdot)$ , one can rewrite the constraints in Eq. (11) as

$$\begin{cases} m(C_k) = \frac{\mathbf{m}_0^T R^{(k)} \mathbf{m}_0}{2^n - 1} \\ \sum_{i=1}^{2^n - 1} m_0(A_i) = 1 \\ 0 \leq m_0(B_j) \leq 1, \forall j = 1, \dots, 2^n - 1 \end{cases} \quad (13)$$

where

$$\mathbf{m}_0 = \begin{bmatrix} m_0(\{\theta_1\}) \\ m_0(\{\theta_2\}) \\ \vdots \\ m_0(\Theta) \end{bmatrix}$$

As we see in Eq. (13), there are  $2^n - 1$  unknown variables (mass values for  $2^n - 1$  focal elements in  $m_0(\cdot)$ ) to determine. There are  $2^n - 1 + 1 = 2^n$  simultaneous equations in total. Therefore, if the solution exists, in general this is an over-determined problem which has the unique solution.

Here we provide an example to verify this, where the combined BBA is still as chosen in Example 2, which is

$$\begin{aligned} m(\{\theta_1\}) &= 0.3846, m(\{\theta_2\}) = 0.5385, \\ m(\{\theta_2, \theta_3\}) &= 0.0577, m(\Theta) = 0.0192. \end{aligned}$$

According to Eq. (11) and change the maximization to minimization, we obtain that  $m_1(\cdot) = m_2(\cdot) = m_0(\cdot)$ , which is

$$\begin{aligned} m_0(\{\theta_1\}) &= 0.3877, m_0(\{\theta_2\}) = 0.3958, \\ m_0(\{\theta_2, \theta_3\}) &= 0.1082, m_0(\Theta) = 0.1082. \end{aligned}$$

It is easy to verify that  $m_0(\cdot) \oplus m_0(\cdot) = m(\cdot)$ .

We prefer the criterion of distance maximization, since it can bring out more distinct (likely to be more independent) evidences.

Note that since we select the maximization, to assure to find the unique global optimal, the objective should be upper-convex. However, the objective function, i.e., the distance of evidence cannot satisfy this. Therefore, in our work in this paper, intelligent optimization algorithms [25] (e.g., the particle swarm algorithm and genetic algorithm) are adopted for the maximization to achieve a better solution.

### B. Possible Applications

Note that given a combined BBA  $m(\cdot)$ ,  $m_a(\cdot)$  and  $m_b(\cdot)$  after the evidence decombination. However, we do not know the specific correspondence between  $\{m_a(\cdot), m_b(\cdot)\}$  and  $\{m_1(\cdot), m_2(\cdot)\}$ . That is,  $m_a(\cdot)$  could correspond to  $m_1(\cdot)$  or  $m_2(\cdot)$ , and  $m_b(\cdot)$  could also correspond to  $m_1(\cdot)$  or  $m_2(\cdot)$ . Therefore, it cannot be used for analyzing or evaluating specific single sensor; however, the evidence decombination is expected to be used in applications like divergence evaluation between sensors, which is helpful for the sensor management. Given a BBA, if one can decombine it into two BBAs, then the maximum difference between corresponding information sources can be evaluated by calculating the distance between the two BBAs.

Another possible application is the evaluation of different combination rules. Here, we only use the Dempster's rule to construct the evidence decombination. In fact, other alternative combination rules can also be used for finding evidence decombination, where the difference between most of existing rules of combinations available in the literature lies in the choice of matrix  $R^{(k)}$  in Eq. (7). Then, given a BBA, one can use different decombination methods corresponding to different combination rule to bring out different pairs of BBAs. One can calculate the distance between two BBAs in each pair to represent the aggregation capability of the corresponding combination rule. That is, an evidence decombination approach can bring out a more divergent BBA pair, then the decombination method's corresponding combination rule can aggregate (combine) a more divergent BBA pair to the same BBA compared with other rules. So we say that it has a better aggregation capability.

## VI. CONCLUSIONS

In this paper, an evidence decombination approach is proposed, where the distance maximization criterion is adopted in the evidence decombination. Some numerical examples and related analysis are provided to illustrate our proposed method and the possible applications are forecasted.

In this paper, the distance of evidence used in the optimization is Jousselme's distance. In our future work, we will try other strict distance metric [26], [27] in the theory of belief functions for comparison. Currently, the objective function is the distance of evidence. In future work, we will try to use the difference between BBAs' uncertainty measure values [28], [29]. Furthermore, we only consider two sources of evidence for the evidence decombination for simplicity. In our future work, we will try to design more sources (larger than two) for the evidence decombination. This paper is only a preliminary work on the evidence decombination, in future research work, we will try to apply the proposed method in various appropriate applications.

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton university press Princeton, 1976, vol. 1.
- [2] Z. Liu, Q. Pan, and J. Dezert, "Evidential classifier for imprecise data based on belief functions," *Knowledge-Based Systems*, vol. 52, pp. 246–257, Nov. 2013.
- [3] Y. Yang, D. Han, and J. Dezert, "An angle-based neighborhood graph classifier with evidential reasoning," *Pattern Recognition Letters*, vol. 71, pp. 78–85, 2016.
- [4] D. Han, J. Dezert, J.-M. Tacnet, and C. Han, "A fuzzy-cautious OWA approach with evidential reasoning," in *Proceedings of the 15th International Conference on Information Fusion (FUSION)*, Singapore, July 2012, pp. 278–285.
- [5] X. Xu, Z. Zhang, D.-L. Xu, and Y.-W. Chen, "Interval-valued evidence updating with reliability and sensitivity analysis for fault diagnosis," *Int. J. Computational Intelligence Systems*, vol. 9, no. 3, pp. 396–415, 2016.
- [6] Z. Zhang, D. Han, J. Dezert, and Y. Yang, "A new adaptive switching median filter for impulse noise reduction with pre-detection based on evidential reasoning," *Signal Processing*, vol. 147, pp. 173–189, 2018.
- [7] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.
- [8] D. Dubois and H. Prade, "Representation and combination of uncertainty with belief functions and possibility measures," *Computational Intelligence*, vol. 4, no. 3, pp. 244–264, 1988.
- [9] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191–234, 1994.
- [10] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decision Support Systems*, vol. 29, no. 1, pp. 1–9, 2000.
- [11] M. C. Florea, A.-L. Jousselme, É. Bossé, and D. Grenier, "Robust combination rules for evidence theory," *Information Fusion*, vol. 10, no. 2, pp. 183–197, 2009.
- [12] F. Smarandache and J. Dezert, *Advances and Applications of DSMT for Information Fusion-Collected Works-Volume 3*. American Research Press, 2009.
- [13] —, *Advances and Applications of DSMT for Information Fusion-Collected Works-Volume 4*. American Research Press, 2015.
- [14] D. G. Fantinato, L. T. Duarte, Y. Deville, R. Attux, C. Jutten, and A. Neves, "A second-order statistics method for blind source separation in post-nonlinear mixtures," *Signal Processing*, vol. 155, pp. 63 – 72, 2019.
- [15] P. Comon and C. Jutten, *Handbook of Blind Source Separation, Independent Component Analysis and Applications*. Oxford, UK: Academic Press, 02 2010.
- [16] P. Smets, "The canonical decomposition of a weighted belief," in *Proceedings of the International Joint Conference on Artificial Intelligence*. Montreal, Quebec, Canada: AAAI, Aug 20th - Aug 25th 1995, pp. 1–6.
- [17] L. A. Zadeh, "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination," *AI Magazine*, vol. 7, no. 2, pp. 85–90, 1986.
- [18] J. Dezert and A. Tchamova, "On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule," *International Journal of Intelligent Systems*, vol. 29, no. 3, pp. 223–252, 2014.
- [19] D. Han, Y. D. Deng, and C. Han, "Sequential weighted combination for unreliable evidence based on evidence variance," *Decision Support Systems*, vol. 56, pp. 387–393, 2013.
- [20] X. Deng, D. Han, J. Dezert, Y. Deng, and Y. Shyr, "Evidence combination from an evolutionary game theory perspective," *IEEE Transactions on Cybernetics*, vol. 46, no. 9, pp. 2070–2082, 2016.
- [21] R. Haenni, "Shedding new light on Zadeh's criticism of Dempster's rule of combination," in *Proceedings of the 7th International Conference on Information Fusion, ISIF*. Philadelphia, Pennsylvania, USA: IEEE, June 27th-July 1st 2006, pp. 1–6.
- [22] A.-L. Jousselme, D. Grenier, and É. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.
- [23] M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the Jaccard index matrix," *International Journal of Approximate Reasoning*, vol. 54, no. 5, pp. 615–626, 2013.
- [24] M. Babaie-Zadeh and C. Jutten, "A general approach for mutual information minimization and its application to blind source separation," *Signal Processing*, vol. 85, no. 5, pp. 975 – 995, 2005, information Theoretic Signal Processing.
- [25] R. Kruse, C. Borgelt, F. Klawonn, C. Moewes, M. Steinbrecher, and P. Held, *Computational Intelligence: A Methodological Introduction*. London, UK: Springer, 2013.
- [26] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, no. 2, pp. 118–145, 2012.
- [27] D. Han, J. Dezert, and Y. Yang, "Belief interval-based distance measures in the theory of belief functions," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 6, pp. 833–850, June 2018.
- [28] A.-L. Jousselme, C. Liu, D. Grenier, and É. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 36, no. 5, pp. 890–903, 2006.
- [29] Y. Yang and D. Han, "A new distance-based total uncertainty measure in the theory of belief functions," *Knowledge-Based Systems*, vol. 94, pp. 114–123, 2016.

# User-specified Optimization Based Transformation of Fuzzy Membership into Basic Belief Assignment

Xiaojing Fan<sup>a</sup>, Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>

<sup>a</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>c</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an, China.

Emails: [jjingxiulian@126.com](mailto:jjingxiulian@126.com), [deqhan@xjtu.edu.cn](mailto:deqhan@xjtu.edu.cn), [jean.dezert@onera.fr](mailto:jean.dezert@onera.fr), [jiafeiyy@mail.xjtu.edu.cn](mailto:jiafeiyy@mail.xjtu.edu.cn)

Originally published as: X. Fan, D. Han, J. Dezert, Y. Yang, *User-Specified Optimization Based Transformation of Fuzzy Membership Into Basic Belief Assignment*, in Proc. of Int. Conf. on Information Fusion (Fusion 2019), Ottawa, Canada, July 2-5, 2019, and reprinted with permission.

**Abstract**—To combine different types of uncertain information from different sources under different frameworks, we need transformations between different frameworks. For the transformation of a fuzzy membership function (FMF) into a basic belief assignment (BBA), several approaches have been proposed. Among these approaches, the uncertainty optimization based transformations can provide BBAs without predefining focal elements. However, these two transformations, which respectively use the uncertainty maximization and minimization criteria, emphasize the extreme cases of uncertainty. We expect to obtain a BBA, which is the trade-off between the two BBAs obtained by solving the uncertainty maximization and minimization, to avoid extreme attitudinal bias. In this paper, we propose two transformations of an FMF into a BBA by using a user-specified weighting factor to obtain such a trade-off (or balanced) BBA. Some examples and related analyses are provided to show the rationality and effectiveness of the proposed transformations.

**Index Terms**—evidence theory, basic belief assignment, fuzzy membership function, optimization, transformation

## I. INTRODUCTION

In the information fusion, we need to deal with a large amount of uncertain information. Various types of uncertainty theories have been proposed to deal with different types of uncertainty, e.g., the probability theory, fuzzy set theory [1], possibility theory [2], rough set theory [3] and Dempster-Shafer evidence theory (DST) [4] etc. When we fuse the information from different sources under different theoretical frameworks, we need the transformation between different frameworks.

For the information represented by the FMF and BBA, we can transform an FMF into a BBA. Then, we can combine the BBAs to implement the information fusion. There have been proposed many transformations of an FMF into a BBA [5]–[9]. In [5], Bi et al. proposed a transformation that normalizes a given FMF to generate a BBA with singleton focal elements only. By using the  $\alpha$ -cut approach, Florea et al. [6] transformed an FMF into a BBA with focal elements nested in order. However, these two approaches above have to predefine the focal elements, which lack of intuitiveness and

objectiveness. Han et al. [7] proposed two approaches without predefining focal elements. These two approaches can provide BBAs by solving constrained uncertainty maximization and minimization.

For the two transformations of Han et al. [7], both two objective functions are the ambiguity measure ( $AM$ ) and their constraints are mainly constructed based on the given FMF. Their rationality and effectiveness are both justified in [7]. During the process of solving optimization problems, these two transformations emphasize on the minimum and maximal uncertainties of the BBA, respectively. We think that the BBA being the trade-off (or balanced) between the two BBAs obtained by solving the uncertainty maximization and minimization is more preferred, which might avoid being “one-sided” on the uncertainty degree. In this paper, we propose two approaches by using a user-specified weighting factor to determine BBAs. One transformation is the weighted average by using the user-specified weighting factor with the two BBAs obtained by optimization based transformations [7]. The other transformation brings out a trade-off BBA by solving a constrained minimization problem. The objective function is based on the user-specified weighting factor, the distance of evidence and the two BBAs obtained by uncertainty optimization. The constraints are mainly based on the given FMF. That is, each of our proposed transformations can transform an FMF into a BBA, which can be considered as the trade-off between the two BBAs obtained with uncertainty optimization. Some examples and related analyses are provided to justify the proposed transformations.

## II. PRELIMINARY

### A. Basics of the Theory of Belief Functions

The theory of belief functions [4], introduced historically by Shafer in DST, is a powerful framework for uncertainty modeling and reasoning. Let  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  be the frame of discernment (FOD). Under the closed world assumption, the FOD is mutually exclusive and exhaustive. The BBA (also called a mass function) is defined on the power set of  $\Theta$ , which can be denoted by a function  $m : 2^\Theta \rightarrow [0, 1]$  satisfying

$$\sum_{A \subseteq \Theta} m(A) = 1, m(\emptyset) = 0 \quad (1)$$

This work was supported by the National Natural Science Foundation (Nos. 61573275, 61671370), Postdoctoral Science Foundation of China (No. 2016M592790), Postdoctoral Science Research Foundation of Shaanxi Province (No. 2016BSHEDZZ46) and Fundamental Research Funds for the Central Universities (No. xjj2016066).

where  $\emptyset$  denotes the empty set.  $\forall A$ , if  $m(A) > 0$ , then  $A$  is called a focal element.  $m(A)$  denotes the evidence support to the proposition  $A$ .

The belief function  $Bel$  for all  $A \subseteq \Theta$ , as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (2)$$

The plausibility function  $Pl$  for all  $A \subseteq \Theta$ , as:

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (3)$$

Suppose there are two independent BBAs  $m_1$  and  $m_2$  on the same FOD. Historically Shafer proposed Dempster's rule to combine two (or more) BBAs. Dempster's rule of combination is

$$m(A) = \begin{cases} 0, & A = \emptyset \\ \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - K}, & A \neq \emptyset \end{cases} \quad (4)$$

where  $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$  represents the conflict coefficient between two BBAs. There exist other alternative combination rules [10], [11].

### B. Uncertainty Measure of a BBA

The uncertainty of a BBA includes two types: the discord and the non-specificity. Different measures of uncertainty [12]–[16] have been proposed, e.g., the non-specificity measure [14], the ambiguity measure ( $AM$ ) [15] and the aggregated uncertainty ( $AU$ ) [16]. The definition of  $AM$  is as follows:

$$AM(m) = - \sum_{\theta \in \Theta} BetP_m(\theta) \log_2(BetP_m(\theta)) \quad (5)$$

where  $BetP_m(\theta) = \sum_{\theta \in A \subseteq \Theta} m(A)/|A|$  is the pignistic probability [17].  $|A|$  denotes the cardinality of the set  $A$ .

## III. TRANSFORMATION OF FMF INTO BBA

### A. Concept of Fuzzy Set

Fuzzy sets [1] were proposed by Zadeh to describe the concepts without precise definitions. Let  $\Theta$  be the universe of discourse (equivalent to FOD in the belief functions). A fuzzy membership function is denoted by  $\mathbf{u} = \mu(\theta)$ ,  $\theta \in \Theta$ . For  $\mu : \Theta \rightarrow [0, 1]$ ,  $\mu(\theta) \in [0, 1]$  is called the degree of membership for  $\theta$ .

### B. Traditional Transformations of FMF into BBA

a) *Transformations with the predefinition of focal elements:* For a given FMF, two available types of transformations below can provide a BBA, which have to predefine the focal elements. Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  and the given FMF is  $\boldsymbol{\mu} = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . The obtained BBA is represented by  $m$ .

In the work of Bi et al. [5], the BBA is determined as follows:

$$m(\{\theta_i\}) = \mu(\theta_i) / \sum_{j=1}^n \mu(\theta_j) \quad (6)$$

This approach predefines all focal elements as singletons, and it is the result of normalization for the given FMF.

Another transformation with the predefinition of focal elements is the work of Florea et al. [6] by using the  $\alpha$ -cut approach. Suppose that  $\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)$  are sorted into ascending order as  $0 = \alpha_0 < \alpha_1 < \alpha_2 < \dots < \alpha_M \leq 1$ , where  $M \leq |\Theta|$ . The BBA is determined by using the transformation [6] as follows:

$$m(A_j) = \frac{\alpha_j - \alpha_{j-1}}{\alpha_M} \quad (7)$$

where  $A_j = \{\theta_i \in \Theta | \mu(\theta_i) \geq \alpha_j\}$ ,  $i = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, M$ . This transformation predefines the focal elements nested in order for the given FMF.

Both two approaches can transform an FMF into a BBA. However, the transformations with the predefinition of focal elements lack of intuitiveness and objectiveness. For a given FMF, the optimization based transformations can obtain a BBA without predefining the focal elements.

b) *Transformations based on the uncertainty optimization:* In the work of Han et al. [7], the two transformations that have no predefinition of focal elements are obtained by solving the uncertainty maximization and minimization. Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  and the given FMF is  $\boldsymbol{\mu} = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . The obtained BBA is represented by  $m$ .

There exists a relationship [18] between the FMF and the belief function or plausibility function. When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ , the FMF is equivalent to a singleton plausibility function, which is denoted by

$$Pl(\{\theta_i\}) = \sum_{\{\theta_i\} \cap A \neq \emptyset} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \quad (8)$$

It is the necessary and sufficient condition for the FMF to be a singleton plausibility function.

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ , the FMF is equivalent to a singleton belief function, which is denoted by

$$Bel(\{\theta_i\}) = \sum_{A \subseteq \{\theta_i\}} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \quad (9)$$

Similarly, it is the necessary and sufficient condition for the FMF to be a singleton belief function.

The detailed proof of the above relationships are given in [18].

There is a BBA transformed from a given FMF, and the FMF and BBA satisfy Eq. (8) or Eq. (9). Then,  $n$  linear equations for the corresponding relations can be obtained. In addition, one has  $\sum_{A \subseteq \Theta} m(A) = 1$ . There exist  $n + 1$  linear equations. However, except for  $m(\emptyset) = 0$ , in the worst case there are  $2^n - 1$  focal elements to assign the belief. The  $n + 1$  linear equations with respect to  $2^n - 1$  undetermined variables, which is an under-determined problem, i.e., it usually has multiple solutions.

Therefore, to obtain a unique BBA, Han et al. [7] established two uncertainty optimization based transformations for the given FMF. Both two objective functions are  $AM$  and the constraints are mainly based on the given FMF.

The objective function of the uncertainty maximization problem and the corresponding constraints are as follows:

When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ ,

$$\begin{aligned} \max_m & \left\{ - \sum_{i=1}^n [BetP_m(\theta_i) \log_2(BetP_m(\theta_i))] \right\} \\ \text{s.t.} & \begin{cases} \sum_{\{\theta_i\} \cap A \neq \emptyset} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1 \\ 0 \leq m(A) \leq 1 \end{cases} \end{aligned} \quad (10)$$

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ ,

$$\begin{aligned} \max_m & \left\{ - \sum_{i=1}^n [BetP_m(\theta_i) \log_2(BetP_m(\theta_i))] \right\} \\ \text{s.t.} & \begin{cases} \sum_{A \subseteq \{\theta_i\}} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1 \\ 0 \leq m(A) \leq 1 \end{cases} \end{aligned} \quad (11)$$

In the sequel, this transformation is represented by “ $T_{\max}$ ” for convenience.

The objective function of the uncertainty minimization problem and the corresponding constraints are as follows:

When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ ,

$$\begin{aligned} \min_m & \left\{ - \sum_{i=1}^n [BetP_m(\theta_i) \log_2(BetP_m(\theta_i))] \right\} \\ \text{s.t.} & \begin{cases} \sum_{\{\theta_i\} \cap A \neq \emptyset} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1 \\ 0 \leq m(A) \leq 1 \end{cases} \end{aligned} \quad (12)$$

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ ,

$$\begin{aligned} \min_m & \left\{ - \sum_{i=1}^n [BetP_m(\theta_i) \log_2(BetP_m(\theta_i))] \right\} \\ \text{s.t.} & \begin{cases} \sum_{A \subseteq \{\theta_i\}} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1 \\ 0 \leq m(A) \leq 1 \end{cases} \end{aligned} \quad (13)$$

In the sequel, this transformation is represented by “ $T_{\min}$ ” for convenience.

The unique BBA can be determined without predefining focal elements by using “ $T_{\max}$ ” or “ $T_{\min}$ ”. The obtained BBA is the optimal solution of the uncertainty maximization or minimization. During the process of transforming an FMF into a BBA, “ $T_{\max}$ ” and “ $T_{\min}$ ” emphasize on the maximal and minimum uncertainty cases of the obtained BBA, respectively. We think that the BBA being the trade-off between the two BBAs obtained by using “ $T_{\max}$ ” and “ $T_{\min}$ ” is more preferred, which might avoid bias in terms of uncertainty degree.

#### IV. TRANSFORMATIONS WITH USER-SPECIFIED WEIGHTING FACTOR

As aforementioned, we can obtain two BBAs by using “ $T_{\max}$ ” and “ $T_{\min}$ ”, respectively. Based on these two BBAs,

we aim to construct a transformation to determine a trade-off BBA. The trade-off BBA which satisfies the relationship between the FMF and the singleton plausibility or singleton belief. We use a user-specified weighting factor to influence how close the trade-off BBA is to each of the two BBAs above. Suppose that the user-specified weighting factor is represented by  $\alpha$  and  $0 \leq \alpha \leq 1$ . When  $\alpha \rightarrow 0$ , the trade-off BBA is close to the BBA obtained by using “ $T_{\min}$ ”. When  $\alpha \rightarrow 1$ , the trade-off BBA is close to the BBA obtained by using “ $T_{\max}$ ”. To meet the requirements above, we propose two different approaches to determine the trade-off BBAs.

##### A. Weighted Average based Transformation

Let  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  be the FOD. The given FMF is represented by  $\boldsymbol{\mu} = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . Suppose that the BBA obtained by using “ $T_{\min}$ ” is denoted by  $m_{\min}$ , and the BBA obtained by using “ $T_{\max}$ ” is denoted by  $m_{\max}$ . The user-specified weighting factor is denoted by  $\alpha$  ( $0 \leq \alpha \leq 1$ ). The trade-off BBA is denoted by  $m$ . The Weighted Average of  $m_{\min}$  and  $m_{\max}$  can bring out a trade-off BBA as follows:

$$m(A) = (1 - \alpha) \cdot m_{\min}(A) + \alpha \cdot m_{\max}(A) \quad (14)$$

where  $A \subseteq \Theta$ . In the sequel, the transformation based on the weighted average (WA) is denoted by “ $T_{WA}$ ” for convenience.

The BBA obtained in (14) is an admissible BBA and it satisfies the constraints established based on FMF.

According to Eq. (14), the following conditions can be satisfied:

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad 0 \leq m(A) \leq 1 \quad (15)$$

For the transformation of an FMF into a BBA, it is necessary that the obtained BBA satisfies the relationship between the FMF and the singleton plausibility or singleton belief. Although “ $T_{WA}$ ” is a simple and direct transformation of an FMF into a trade-off BBA, it also satisfies the relationship. The proof is provided below.

When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ ,  $m_{\min}$  and  $m_{\max}$  satisfy Eq. (6), respectively, i.e.,  $Pl_{\min}(\{\theta_i\}) = Pl_{\max}(\{\theta_i\}) = \mu(\theta_i)$ ,  $i = 1, 2, \dots, n$ . According to Eq. (14),

$$\begin{aligned} Pl(\{\theta_i\}) &= \sum_{\{\theta_i\} \cap A \neq \emptyset} m(A) \\ &= \sum_{\{\theta_i\} \cap A \neq \emptyset} [(1 - \alpha) \cdot m_{\min}(A) + \alpha \cdot m_{\max}(A)] \\ &= (1 - \alpha) \cdot \sum_{\{\theta_i\} \cap A \neq \emptyset} m_{\min}(A) + \alpha \cdot \sum_{\{\theta_i\} \cap A \neq \emptyset} m_{\max}(A) \quad (16) \\ &= (1 - \alpha) \cdot Pl_{\min}(\{\theta_i\}) + \alpha \cdot Pl_{\max}(\{\theta_i\}) \\ &= (1 - \alpha) \cdot \mu(\theta_i) + \alpha \cdot \mu(\theta_i) \\ &= \mu(\theta_i) \end{aligned}$$

Similarly, when  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ ,  $m_{\min}$  and  $m_{\max}$  satisfy Eq. (7), respectively, i.e.,  $Bel_{\min}(\{\theta_i\}) = Bel_{\max}(\{\theta_i\}) =$

$\mu(\theta_i)$ ,  $i = 1, 2, \dots, n$ . According to Eq. (14),

$$\begin{aligned}
 Bel(\{\theta_i\}) &= \sum_{A \subseteq \{\theta_i\}} m(A) \\
 &= \sum_{A \subseteq \{\theta_i\}} [(1-\alpha) \cdot m_{\min}(A) + \alpha \cdot m_{\max}(A)] \\
 &= (1-\alpha) \cdot \sum_{A \subseteq \{\theta_i\}} m_{\min}(A) + \alpha \cdot \sum_{A \subseteq \{\theta_i\}} m_{\max}(A) \quad (17) \\
 &= (1-\alpha) \cdot Bel_{\min}(\{\theta_i\}) + \alpha \cdot Bel_{\max}(\{\theta_i\}) \\
 &= (1-\alpha) \cdot \mu(\theta_i) + \alpha \cdot \mu(\theta_i) \\
 &= \mu(\theta_i)
 \end{aligned}$$

For the trade-off BBA, when  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ , the given FMF is equivalent to the corresponding singleton plausibility. When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ , the given FMF is equivalent to the corresponding singleton belief. That is, “T<sub>WA</sub>” can transform the given FMF into the trade-off BBA, which satisfies the relationship between the FMF and the BBA.

### B. User-specified Optimization based Transformation

Let the FOD be  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The given FMF is denoted by  $\mu = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . Suppose that  $m_{\min}$  and  $m_{\max}$  denote the BBAs obtained by “T<sub>min</sub>” and “T<sub>max</sub>”, respectively. The user-specified weighting factor is denoted by  $\alpha$  ( $0 \leq \alpha \leq 1$ ). The trade-off BBA is represented by  $m$ .

The user-specified weighting factor is used to influence the similarity between the trade-off BBA and  $m_{\min}$  (or  $m_{\max}$ ). The degree of similarity between two BBAs is represented by the distance of evidence. We can use the Jousselme’s distance [19], which is a strict metric defined as

$$d_J(m_a, m_b) = \sqrt{\frac{1}{2}(m_a - m_b) \mathbf{D}(m_a - m_b)} \quad (18)$$

where  $\mathbf{D}(A, B) = |A \cap B| / |A \cup B|$ ,  $A \subseteq \Theta$ ,  $B \subseteq \Theta$ . According to Eq. (18), the obtained BBA is more similar to  $m_{\min}$ , if  $d_J(m, m_{\min})$  is smaller. If  $d_J(m, m_{\max})$  is smaller, the obtained BBA is more similar to  $m_{\max}$ .

To obtain the trade-off BBA between  $m_{\min}$  and  $m_{\max}$ , a relationship between the user-specified weighting factor and the distance of evidence can be constructed. When  $\alpha$  is given from 0 to 1, with the decreasing of  $d_J(m, m_{\min})$ , the value of  $d_J(m, m_{\max})$  is increasing. Then we can establish the following equation:

$$\frac{d_J(m, m_{\min})}{d_J(m, m_{\max})} = \frac{\alpha}{1-\alpha} \quad (19)$$

The BBA satisfies Eq. (19) may not always exist. If the following function (equivalent to Eq. (19)) achieves the minimum value, then the trade-off BBA is obtained.

$$obj(m) = [(1-\alpha) \cdot d_J(m, m_{\min}) - \alpha \cdot d_J(m, m_{\max})]^2 \quad (20)$$

When  $\alpha$  is given, we can establish a constrained minimization problem to transform an FMF into a BBA. The objective function is Eq. (20) and the constraints are mainly based on

Eq. (6) or Eq. (7). The transformation of an FMF into a trade-off BBA is obtained by solving the user-specified optimization problem as follows:

When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ ,

$$\begin{aligned}
 \min_m & \left\{ [(1-\alpha) \cdot d_J(m, m_{\min}) - \alpha \cdot d_J(m, m_{\max})]^2 \right\} \\
 \text{s.t.} & \begin{cases} \sum_{\substack{A \subseteq \{\theta_i\} \\ \{\theta_i\} \cap A \neq \emptyset}} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1 \\ 0 \leq m(A) \leq 1 \end{cases} \quad (21)
 \end{aligned}$$

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ ,

$$\begin{aligned}
 \min_m & \left\{ [(1-\alpha) \cdot d_J(m, m_{\min}) - \alpha \cdot d_J(m, m_{\max})]^2 \right\} \\
 \text{s.t.} & \begin{cases} \sum_{A \subseteq \{\theta_i\}} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1 \\ 0 \leq m(A) \leq 1 \end{cases} \quad (22)
 \end{aligned}$$

In the sequel, the transformation based on the user-specified optimization (USO) is denoted by “T<sub>USO</sub>” for convenience.

For a given FMF, the trade-off BBA can be obtained by using “T<sub>USO</sub>”, which is a user-specified optimization based transformation.

## V. EXPERIMENTS

In this section, we provide some examples to illustrate how to transform an FMF into a trade-off BBA using our approaches. Here, we use the optimization toolbox in the Matlab™ to solve the optimization problems under constraints.

### A. Example 1

Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The given FMF is  $\mu(\theta_1) = 0.9$ ,  $\mu(\theta_2) = 0.7$ ,  $\mu(\theta_3) = 0.3$ . Suppose that  $m_{\min}$  and  $m_{\max}$  are the BBAs obtained by using “T<sub>min</sub>” and “T<sub>max</sub>”, respectively. We just list the corresponding BBAs for  $\alpha = 0, 0.3, 0.7$  and 1.

This FMF satisfies  $\sum_{i=1}^3 \mu(\theta_i) = 1.9 > 1$ . Therefore, the given FMF is equivalent to the singleton plausibility. The BBAs obtained by using “T<sub>WA</sub>” and “T<sub>USO</sub>” are listed in the Table I and Table II, respectively.

By using “T<sub>WA</sub>” and “T<sub>USO</sub>”, when  $\alpha = 0$ , the obtained BBAs are identical to  $m_{\min}$ , and the values of  $AM$  are the minimum uncertainty. When  $\alpha \rightarrow 0$ , the obtained BBA is similar to  $m_{\min}$ , and its uncertainty is close to the minimum uncertainty.

Similarly, when  $\alpha = 1$ , the obtained BBAs are identical to  $m_{\max}$ , and the values of  $AM$  are the maximal uncertainty. When  $\alpha \rightarrow 1$ , the obtained BBA is similar to  $m_{\max}$ , and its uncertainty is close to the maximal uncertainty.

### B. Example 2

Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ . The given FMF is  $\mu(\theta_1) = 1$ ,  $\mu(\theta_2) = 0.2$ ,  $\mu(\theta_3) = 0.3$ ,  $\mu(\theta_4) = 0.3$ . Suppose that  $m_{\min}$  and  $m_{\max}$  are the BBAs obtained by using “T<sub>min</sub>” and “T<sub>max</sub>”, respectively. We just list the corresponding BBAs for  $\alpha = 0, 0.2, 0.8$  and 1.

TABLE I  
USING “T<sub>WA</sub>” TO OBTAIN BBAS IN EXAMPLE 1.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.3, m(\{\theta_1, \theta_2\}) = 0.3$ $m(\{\theta_2\}) = 0.1, m(\Theta) = 0.3$	1.3367
$\alpha = 0.3$	$m(\{\theta_1\}) = 0.21, m(\{\theta_1, \theta_2\}) = 0.42$ $m(\{\theta_2\}) = 0.07, m(\{\theta_1, \theta_3\}) = 0.06$ $m(\{\theta_3\}) = 0.03, m(\Theta) = 0.21$	1.4003
$\alpha = 0.7$	$m(\{\theta_1\}) = 0.09, m(\{\theta_1, \theta_2\}) = 0.58$ $m(\{\theta_2\}) = 0.03, m(\{\theta_1, \theta_3\}) = 0.14$ $m(\{\theta_3\}) = 0.07, m(\Theta) = 0.09$	1.4730
$\alpha = 1$	$m(\{\theta_1, \theta_2\}) = 0.7, m(\{\theta_3\}) = 0.1$ $m(\{\theta_1, \theta_3\}) = 0.2$	1.5129

TABLE II  
USING “T<sub>USO</sub>” TO OBTAIN BBAS IN EXAMPLE 1.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.3, m(\{\theta_1, \theta_2\}) = 0.3$ $m(\{\theta_2\}) = 0.1, m(\Theta) = 0.3$	1.3367
$\alpha = 0.3$	$m(\{\theta_1\}) = 0.2027, m(\{\theta_1, \theta_2\}) = 0.4145$ $m(\{\theta_2\}) = 0.0828, m(\{\theta_1, \theta_3\}) = 0.0802$ $m(\{\theta_3\}) = 0.0172, m(\Theta) = 0.2026$	1.3969
$\alpha = 0.7$	$m(\{\theta_1\}) = 0.1037, m(\{\theta_1, \theta_2\}) = 0.5962$ $m(\{\theta_2\}) = 0.0001, m(\{\theta_1, \theta_3\}) = 0.0965$ $m(\{\theta_3\}) = 0.0999, m(\Theta) = 0.1036$	1.4828
$\alpha = 1$	$m(\{\theta_1, \theta_2\}) = 0.7, m(\{\theta_3\}) = 0.1$ $m(\{\theta_1, \theta_3\}) = 0.2$	1.5129

According to  $\sum_{i=1}^4 \mu(\theta_i) = 1.8 > 1$ , the FMF is equivalent to the singleton plausibility. In the Table III and Table IV, the BBAs obtained by using “T<sub>WA</sub>” and “T<sub>USO</sub>” are listed, respectively.

When  $\alpha = 0$ , the obtained BBAs are identical to  $m_{min}$ , and the values of AM are the minimum uncertainty. When  $\alpha = 1$ , the obtained BBAs are identical to  $m_{max}$ , and the values of AM are the maximal uncertainty.

In the Table III and Table IV, when  $\alpha \rightarrow 0$ , the obtained BBA is similar to  $m_{min}$ , and its uncertainty is close to the minimum uncertainty. When  $\alpha \rightarrow 1$ , the obtained BBA is similar to  $m_{max}$ , and its uncertainty is close to the maximal uncertainty.

TABLE III  
USING “T<sub>WA</sub>” TO OBTAIN BBAS IN EXAMPLE 2.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.7, m(\Theta) = 0.2$ $m(\{\theta_1, \theta_3, \theta_4\}) = 0.1$	1.0896
$\alpha = 0.2$	$m(\{\theta_1\}) = 0.60, m(\{\theta_1, \theta_2\}) = 0.04$ $m(\{\theta_1, \theta_3\}) = 0.06, m(\{\theta_1, \theta_4\}) = 0.06$ $m(\{\theta_1, \theta_3, \theta_4\}) = 0.08, m(\Theta) = 0.16$	1.2099
$\alpha = 0.8$	$m(\{\theta_1\}) = 0.3, m(\{\theta_1, \theta_2\}) = 0.16$ $m(\{\theta_1, \theta_3\}) = 0.24, m(\{\theta_1, \theta_4\}) = 0.24$ $m(\{\theta_1, \theta_3, \theta_4\}) = 0.02, m(\Theta) = 0.04$	1.5122
$\alpha = 1$	$m(\{\theta_1\}) = 0.2, m(\{\theta_1, \theta_2\}) = 0.2$ $m(\{\theta_1, \theta_3\}) = 0.3, m(\{\theta_1, \theta_4\}) = 0.3$	1.5955

TABLE IV  
USING “T<sub>USO</sub>” TO OBTAIN BBAS IN EXAMPLE 2.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.7, m(\Theta) = 0.2$ $m(\{\theta_1, \theta_3, \theta_4\}) = 0.1$	1.0896
$\alpha = 0.2$	$m(\{\theta_1\}) = 0.6008, m(\{\theta_1, \theta_2\}) = 0.0471$ $m(\{\theta_1, \theta_3\}) = 0.0484$ $m(\{\theta_1, \theta_2, \theta_3\}) = 0.0037$ $m(\{\theta_1, \theta_4\}) = 0.0515$ $m(\{\theta_1, \theta_2, \theta_4\}) = 0.0006$ $m(\{\theta_1, \theta_3, \theta_4\}) = 0.0993, m(\Theta) = 0.1486$	1.2133
$\alpha = 0.8$	$m(\{\theta_1\}) = 0.2990, m(\{\theta_1, \theta_2\}) = 0.1665$ $m(\{\theta_1, \theta_3\}) = 0.2345, m(\{\theta_1, \theta_4\}) = 0.2010$ $m(\{\theta_1, \theta_2, \theta_4\}) = 0.0335$ $m(\{\theta_1, \theta_3, \theta_4\}) = 0.0655$	1.5227
$\alpha = 1$	$m(\{\theta_1\}) = 0.2, m(\{\theta_1, \theta_2\}) = 0.2$ $m(\{\theta_1, \theta_3\}) = 0.3, m(\{\theta_1, \theta_4\}) = 0.3$	1.5955

C. Example 3

Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ . The given FMF is  $\mu(\theta_1) = 0.6, \mu(\theta_2) = 0.1, \mu(\theta_3) = 0.2, \mu(\theta_4) = 0.1$ . We just list the corresponding BBAs for  $\alpha = 0, 0.4, 0.9$  and 1.

This FMF satisfies  $\sum_{i=1}^4 \mu(\theta_i) = 1$ . The FMF is equivalent to the singleton plausibility or singleton belief. In the Table V and Table VI, the BBAs obtained by using “T<sub>WA</sub>” and “T<sub>USO</sub>” are listed, respectively.

In the Table V and Table VI, the BBAs obtained when  $\alpha = 0$  are identical to the BBAs obtained when  $\alpha = 1$ , i.e., the two BBAs obtained by using “T<sub>min</sub>” and “T<sub>max</sub>” are the same. Therefore,  $\forall \alpha \in [0, 1]$ , the obtained BBAs are without the influence of  $\alpha$ . When  $\alpha$  is given from 0 to 1, all the obtained BBAs are Bayesian belief functions and are identical.

TABLE V  
USING “T<sub>WA</sub>” TO OBTAIN BBAS IN EXAMPLE 3.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710
$\alpha = 0.4$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710
$\alpha = 0.9$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710
$\alpha = 1$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710

D. Example 4

Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The given FMF is  $\mu(\theta_1) = 0.6, \mu(\theta_2) = 0.2, \mu(\theta_3) = 0.1$ . Suppose that  $m_{min}$  and  $m_{max}$  are the BBAs obtained by using “T<sub>min</sub>” and “T<sub>max</sub>”, respectively. We just list the corresponding BBAs for  $\alpha = 0, 0.3, 0.8$  and 1.

This FMF satisfies  $\sum_{i=1}^3 \mu(\theta_i) = 0.9 < 1$ . The FMF is equivalent to the singleton belief. The BBAs obtained by using “T<sub>WA</sub>” and “T<sub>USO</sub>” are listed in the Table VII and Table VIII, respectively.

TABLE VI  
 USING "T<sub>USO</sub>" TO OBTAIN BBAs IN EXAMPLE 3.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710
$\alpha = 0.4$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710
$\alpha = 0.9$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710
$\alpha = 1$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.1$ $m(\{\theta_3\}) = 0.2, m(\{\theta_4\}) = 0.1$	1.5710

In the Table VII and Table VIII, when  $\alpha = 0$ , the obtained BBAs are identical to  $m_{min}$ , and the values of AM are the minimum uncertainty. When  $\alpha = 1$ , the obtained BBAs are identical to  $m_{max}$ , and the values of AM are the maximal uncertainty. When  $\alpha \rightarrow 0$ , the obtained BBA is similar to  $m_{min}$ , and its uncertainty is close to the minimum uncertainty. When  $\alpha \rightarrow 1$ , the obtained BBA is similar to  $m_{max}$ , and its uncertainty is close to the maximal uncertainty.

 TABLE VII  
 USING "T<sub>WA</sub>" TO OBTAIN BBAs IN EXAMPLE 4.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.6, m(\{\theta_3\}) = 0.1$ $m(\{\theta_2\}) = 0.2, m(\{\theta_1, \theta_2\}) = 0.1$	1.2362
$\alpha = 0.3$	$m(\{\theta_1\}) = 0.6, m(\{\theta_1, \theta_2\}) = 0.07$ $m(\{\theta_2\}) = 0.2, m(\{\theta_2, \theta_3\}) = 0.03$ $m(\{\theta_3\}) = 0.1$	1.2749
$\alpha = 0.8$	$m(\{\theta_1\}) = 0.6, m(\{\theta_1, \theta_2\}) = 0.02$ $m(\{\theta_2\}) = 0.2, m(\{\theta_2, \theta_3\}) = 0.08$ $m(\{\theta_3\}) = 0.1$	1.3321
$\alpha = 1$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.2$ $m(\{\theta_3\}) = 0.1, m(\{\theta_2, \theta_3\}) = 0.1$	1.3527

 TABLE VIII  
 USING "T<sub>USO</sub>" TO OBTAIN BBAs IN EXAMPLE 4.

$\alpha$	BBA	AM
$\alpha = 0$	$m(\{\theta_1\}) = 0.6, m(\{\theta_3\}) = 0.1$ $m(\{\theta_2\}) = 0.2, m(\{\theta_1, \theta_2\}) = 0.1$	1.2362
$\alpha = 0.3$	$m(\{\theta_1\}) = 0.6, m(\{\theta_1, \theta_2\}) = 0.0697$ $m(\{\theta_2\}) = 0.2, m(\{\theta_1, \theta_3\}) = 0.0006$ $m(\{\theta_3\}) = 0.1, m(\{\theta_2, \theta_3\}) = 0.0297$	1.2748
$\alpha = 0.8$	$m(\{\theta_1\}) = 0.6, m(\{\theta_1, \theta_2\}) = 0.02$ $m(\{\theta_2\}) = 0.2, m(\{\theta_2, \theta_3\}) = 0.08$ $m(\{\theta_3\}) = 0.1$	1.3321
$\alpha = 1$	$m(\{\theta_1\}) = 0.6, m(\{\theta_2\}) = 0.2$ $m(\{\theta_3\}) = 0.1, m(\{\theta_2, \theta_3\}) = 0.1$	1.3527

### E. Example 5

Suppose that a system of classification with three sensors, including displacement sensor  $S_1$ , pressure sensor  $S_2$  and image sensor  $S_3$ . Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . Three sensors are used for measuring the size, weight and state of the sample, respectively. The measurements of sensors are used to

obtain two FMFs and a BBA. According to the parameters and the measurements of the sensor, the FMF is defined as

$$\mu(\theta_j) = \begin{cases} \frac{x - \min_i}{ave_i - \min_i}, & x \in [\min_i, ave_i] \\ \frac{x - \max_i}{ave_i - \max_i}, & x \in (ave_i, \max_i] \\ 0, & \text{others} \end{cases} \quad (23)$$

where  $i = 1, 2$ .  $\min_i$  and  $\max_i$  are the minimum and maximal values of the class  $\theta_j$  ( $j = 1, 2, 3$ ), respectively.  $ave_i$  is the average value of the class  $\theta_j$ .

 TABLE IX  
 THE PARAMETERS AND THE MEASUREMENTS OF SENSORS.

Class	$S_1$			$S_2$		
	$\min_1$	$\max_1$	$ave_1$	$\min_2$	$\max_2$	$ave_2$
$\theta_1$	43.3	58.4	50.8	2.9	4.1	3.4
$\theta_2$	50.9	70.1	59.5	2.0	3.4	2.8
$\theta_3$	49.4	79.3	65.7	2.2	3.8	2.9
Sample	56			3.2		

In the Table IX, the parameters of  $S_1$  and  $S_2$  and the measurements of a sample are listed. The class of this sample is  $\theta_2$ . According to (23), two FMFs are as follows:

$$S_1 : \mu(\theta_1) = 0.3158, \mu(\theta_2) = 0.5930, \mu(\theta_3) = 0.4049;$$

$$S_2 : \mu(\theta_1) = 0.6, \mu(\theta_2) = 0.6667, \mu(\theta_3) = 0.3333.$$

According to the image of  $S_3$ , the expert determined the BBA directly as follows:

$$m_{S_3}(\{\theta_3\}) = 0.51, m_{S_3}(\{\theta_2, \theta_3\}) = 0.38, m_{S_3}(\Theta) = 0.11.$$

 TABLE X  
 USING "T<sub>WA</sub>" TO OBTAIN BBAs IN EXAMPLE 5.

$\alpha$	$m_{S_1}$	$m_{S_2}$
$\alpha = 0$	$m_{S_1}(\{\theta_1\}) = 0.0021$ $m_{S_1}(\{\theta_2\}) = 0.5930$ $m_{S_1}(\{\theta_3\}) = 0.0912$ $m_{S_1}(\{\theta_1, \theta_3\}) = 0.3137$	$m_{S_2}(\{\theta_1\}) = 0.2667$ $m_{S_2}(\{\theta_2\}) = 0.4$ $m_{S_2}(\{\theta_1, \theta_3\}) = 0.0666$ $m_{S_2}(\{\Theta\}) = 0.2667$
$\alpha = 0.3$	$m_{S_1}(\{\theta_1\}) = 0.0759$ $m_{S_1}(\{\theta_2\}) = 0.4989$ $m_{S_1}(\{\theta_1, \theta_2\}) = 0.0203$ $m_{S_1}(\{\theta_3\}) = 0.1115$ $m_{S_1}(\{\theta_1, \theta_3\}) = 0.2196$ $m_{S_1}(\{\theta_2, \theta_3\}) = 0.0738$	$m_{S_2}(\{\theta_1\}) = 0.1967$ $m_{S_2}(\{\theta_2\}) = 0.3$ $m_{S_2}(\{\theta_1, \theta_2\}) = 0.17$ $m_{S_2}(\{\theta_3\}) = 0.09$ $m_{S_2}(\{\theta_1, \theta_3\}) = 0.0466$ $m_{S_2}(\{\theta_2, \theta_3\}) = 0.01$ $m_{S_2}(\{\Theta\}) = 0.1867$
$\alpha = 0.8$	$m_{S_1}(\{\theta_1\}) = 0.1989$ $m_{S_1}(\{\theta_2\}) = 0.3420$ $m_{S_1}(\{\theta_1, \theta_2\}) = 0.0542$ $m_{S_1}(\{\theta_3\}) = 0.1454$ $m_{S_1}(\{\theta_1, \theta_3\}) = 0.0627$ $m_{S_1}(\{\theta_2, \theta_3\}) = 0.1968$	$m_{S_2}(\{\theta_1\}) = 0.08$ $m_{S_2}(\{\theta_2\}) = 0.1334$ $m_{S_2}(\{\theta_1, \theta_2\}) = 0.4534$ $m_{S_2}(\{\theta_3\}) = 0.24$ $m_{S_2}(\{\theta_1, \theta_3\}) = 0.0133$ $m_{S_2}(\{\theta_2, \theta_3\}) = 0.0266$ $m_{S_2}(\{\Theta\}) = 0.0533$
$\alpha = 1$	$m_{S_1}(\{\theta_1\}) = 0.2480$ $m_{S_1}(\{\theta_2\}) = 0.2793$ $m_{S_1}(\{\theta_1, \theta_2\}) = 0.0687$ $m_{S_1}(\{\theta_3\}) = 0.1590$ $m_{S_1}(\{\theta_2, \theta_3\}) = 0.2459$	$m_{S_2}(\{\theta_1\}) = 0.0333$ $m_{S_2}(\{\theta_2\}) = 0.0667$ $m_{S_2}(\{\theta_1, \theta_2\}) = 0.5667$ $m_{S_2}(\{\theta_3\}) = 0.3$ $m_{S_2}(\{\theta_2, \theta_3\}) = 0.0333$

Suppose that  $m_{S_1}$  and  $m_{S_2}$  denote the obtained BBAs transformed from the two FMFs of  $S_1$  and  $S_2$ , respectively.

TABLE XIII  
THE PIGNISTIC PROBABILITIES IN EXAMPLE 5.

$\alpha$	“T <sub>WA</sub> ”	“T <sub>USO</sub> ”
$\alpha = 0$	$BetP(\{\theta_1\}) = 0.0446$ $BetP(\{\theta_2\}) = 0.5731$ $BetP(\{\theta_3\}) = 0.3822$	$BetP(\{\theta_1\}) = 0.0446$ $BetP(\{\theta_2\}) = 0.5731$ $BetP(\{\theta_3\}) = 0.3822$
$\alpha = 0.3$	$BetP(\{\theta_1\}) = 0.0525$ $BetP(\{\theta_2\}) = 0.5648$ $BetP(\{\theta_3\}) = 0.3827$	$BetP(\{\theta_1\}) = 0.0511$ $BetP(\{\theta_2\}) = 0.5627$ $BetP(\{\theta_3\}) = 0.3863$
$\alpha = 0.8$	$BetP(\{\theta_1\}) = 0.0565$ $BetP(\{\theta_2\}) = 0.5567$ $BetP(\{\theta_3\}) = 0.3868$	$BetP(\{\theta_1\}) = 0.0520$ $BetP(\{\theta_2\}) = 0.5570$ $BetP(\{\theta_3\}) = 0.3910$
$\alpha = 1$	$BetP(\{\theta_1\}) = 0.0549$ $BetP(\{\theta_2\}) = 0.5556$ $BetP(\{\theta_3\}) = 0.3895$	$BetP(\{\theta_1\}) = 0.0549$ $BetP(\{\theta_2\}) = 0.5556$ $BetP(\{\theta_3\}) = 0.3895$

TABLE XI  
USING “T<sub>USO</sub>” TO OBTAIN BBAS IN EXAMPLE 5.

$\alpha$	$m_{S_1}$	$m_{S_2}$
$\alpha = 0$	$m_{S_1}(\{\theta_1\}) = 0.0021$ $m_{S_1}(\{\theta_2\}) = 0.5930$ $m_{S_1}(\{\theta_3\}) = 0.0912$ $m_{S_1}(\{\theta_1, \theta_3\}) = 0.3137$	$m_{S_2}(\{\theta_1\}) = 0.2667$ $m_{S_2}(\{\theta_2\}) = 0.4$ $m_{S_2}(\{\theta_1, \theta_3\}) = 0.0666$ $m_{S_2}(\{\Theta\}) = 0.2667$
$\alpha = 0.3$	$m_{S_1}(\{\theta_1\}) = 0.0354$ $m_{S_1}(\{\theta_2\}) = 0.4927$ $m_{S_1}(\{\theta_1, \theta_2\}) = 0.0670$ $m_{S_1}(\{\theta_3\}) = 0.1654$ $m_{S_1}(\{\theta_1, \theta_3\}) = 0.2062$ $m_{S_1}(\{\theta_2, \theta_3\}) = 0.0260$ $m_{S_1}(\{\Theta\}) = 0.0073$	$m_{S_2}(\{\theta_1\}) = 0.1902$ $m_{S_2}(\{\theta_2\}) = 0.2641$ $m_{S_2}(\{\theta_1, \theta_2\}) = 0.2124$ $m_{S_2}(\{\theta_3\}) = 0.0291$ $m_{S_2}(\{\theta_1, \theta_3\}) = 0.1140$ $m_{S_2}(\{\theta_2, \theta_3\}) = 0.1068$ $m_{S_2}(\{\Theta\}) = 0.0834$
$\alpha = 0.8$	$m_{S_1}(\{\theta_1\}) = 0.1918$ $m_{S_1}(\{\theta_2\}) = 0.3190$ $m_{S_1}(\{\theta_1, \theta_2\}) = 0.0844$ $m_{S_1}(\{\theta_3\}) = 0.2152$ $m_{S_1}(\{\theta_2, \theta_3\}) = 0.15$ $m_{S_1}(\{\Theta\}) = 0.0396$	$m_{S_2}(\{\theta_1\}) = 0.0942$ $m_{S_2}(\{\theta_2\}) = 0.1193$ $m_{S_2}(\{\theta_1, \theta_2\}) = 0.4532$ $m_{S_2}(\{\theta_3\}) = 0.2379$ $m_{S_2}(\{\theta_1, \theta_3\}) = 0.0012$ $m_{S_2}(\{\theta_2, \theta_3\}) = 0.0428$ $m_{S_2}(\{\Theta\}) = 0.0514$
$\alpha = 1$	$m_{S_1}(\{\theta_1\}) = 0.2480$ $m_{S_1}(\{\theta_2\}) = 0.2793$ $m_{S_1}(\{\theta_1, \theta_2\}) = 0.0687$ $m_{S_1}(\{\theta_3\}) = 0.1590$ $m_{S_1}(\{\theta_2, \theta_3\}) = 0.2459$	$m_{S_2}(\{\theta_1\}) = 0.0333$ $m_{S_2}(\{\theta_2\}) = 0.0667$ $m_{S_2}(\{\theta_1, \theta_2\}) = 0.5667$ $m_{S_2}(\{\theta_3\}) = 0.3$ $m_{S_2}(\{\theta_2, \theta_3\}) = 0.0333$

The BBAs obtained by using “T<sub>WA</sub>” and “T<sub>USO</sub>” are listed in the Table X and Table XI, respectively. Then, we combine these three BBAs (i.e.,  $m_{S_1}$ ,  $m_{S_2}$  and  $m_{S_3}$ ). The combined BBA is represented by  $m$ . The combined BBAs and the pignistic probabilities are listed in the Table XII and Table XIII, respectively. We just list the corresponding BBAs for  $\alpha = 0, 0.3, 0.8$  and  $1$ .

In the Table XIII, all the classification results are  $\theta_2$  and are correct. When  $\alpha$  is given from 0 to 1,  $m_{S_1}(\{\theta_2\})$  and  $m_{S_2}(\{\theta_2\})$  are decreasing in the Table X and Table XI. With the increasing value of  $\alpha$ ,  $m_{S_1}$  and  $m_{S_2}$  are more close to  $m_{max}$  (i.e., the BBA obtained by using “T<sub>max</sub>” or the BBA obtained when  $\alpha = 1$ ), which is the reason of the decreasing value of  $BetP(\{\theta_2\})$  in the Table XIII.

TABLE XII  
THE COMBINED BBAS IN EXAMPLE 5.

$\alpha$	“T <sub>WA</sub> ”	“T <sub>USO</sub> ”
$\alpha = 0$	$m(\{\theta_1\}) = 0.0276$ $m(\{\theta_2\}) = 0.5731$ $m(\{\theta_3\}) = 0.3652$ $m(\{\theta_1, \theta_3\}) = 0.0340$	$m(\{\theta_1\}) = 0.0276$ $m(\{\theta_2\}) = 0.5731$ $m(\{\theta_3\}) = 0.3652$ $m(\{\theta_1, \theta_3\}) = 0.0340$
$\alpha = 0.3$	$m(\{\theta_1\}) = 0.0429$ $m(\{\theta_2\}) = 0.5531$ $m(\{\theta_1, \theta_2\}) = 0.0024$ $m(\{\theta_3\}) = 0.3637$ $m(\{\theta_1, \theta_3\}) = 0.0168$ $m(\{\theta_2, \theta_3\}) = 0.0212$	$m(\{\theta_1\}) = 0.0408$ $m(\{\theta_2\}) = 0.5547$ $m(\{\theta_1, \theta_2\}) = 0.0069$ $m(\{\theta_3\}) = 0.375$ $m(\{\theta_1, \theta_3\}) = 0.0135$ $m(\{\theta_2, \theta_3\}) = 0.0089$ $m(\{\Theta\}) = 0.0002$
$\alpha = 0.8$	$m(\{\theta_1\}) = 0.0513$ $m(\{\theta_2\}) = 0.5409$ $m(\{\theta_1, \theta_2\}) = 0.0089$ $m(\{\theta_3\}) = 0.3747$ $m(\{\theta_1, \theta_3\}) = 0.0014$ $m(\{\theta_2, \theta_3\}) = 0.0228$	$m(\{\theta_1\}) = 0.0418$ $m(\{\theta_2\}) = 0.5341$ $m(\{\theta_1, \theta_2\}) = 0.0199$ $m(\{\theta_3\}) = 0.3780$ $m(\{\theta_2, \theta_3\}) = 0.0255$ $m(\{\Theta\}) = 0.0007$
$\alpha = 1$	$m(\{\theta_1\}) = 0.0487$ $m(\{\theta_2\}) = 0.5435$ $m(\{\theta_1, \theta_2\}) = 0.0124$ $m(\{\theta_3\}) = 0.3837$ $m(\{\theta_2, \theta_3\}) = 0.0118$	$m(\{\theta_1\}) = 0.0487$ $m(\{\theta_2\}) = 0.5435$ $m(\{\theta_1, \theta_2\}) = 0.0124$ $m(\{\theta_3\}) = 0.3837$ $m(\{\theta_2, \theta_3\}) = 0.0118$

VI. CONCLUSIONS

In this paper, we have proposed two approaches with a user-specified weighting factor to transform a given FMF into a trade-off BBA. These two approaches are both effective approaches for obtaining a trade-off BBA. The users can transform a given FMF into a BBA by their preferred approach. With the cardinality of FOD increasing, the computational complexity of the optimization will become exponential growth. The reason for this is the structure of the belief functions. By using the user-specified weighting factor to influence how close the trade-off BBA is to each of the two BBAs obtained by solving the uncertainty maximization and minimization. The example of using our transformations in the practical application is provided. The numerical examples indicate that the uncertainty of the obtained BBA is between the minimum and maximal uncertainties. In a future work, we will try to use and compare different types of the distance of evidence as objective function to expect a better trade-off BBA.

REFERENCES

- [1] L. A. Zadeh, “Fuzzy Sets,” *Information and Control*, vol. 8, pp. 338-353, June 1965.
- [2] L. A. Zadeh, “Fuzzy sets as a basis for a theory of possibility,” *Fuzzy Sets and Systems*, vol. 1, pp. 3-28, January 1978.
- [3] Z. Pawlak, “Rough sets,” *International Journal of Computer & Information Sciences*, vol. 11, pp. 341-356, October 1982.

- [4] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press Princeton, 1976.
- [5] Y. Bi, D. Bell, H. Wang, G. Guo and K. Greer, "Combining Multiple Classifiers Using Dempsters Rule of Combination for Text Categorization," *Modeling Decisions for Artificial Intelligence*, vol. 3131, pp. 127-138, 2004.
- [6] M. C. Florea, A. L. Jousselme, D. Grenier and É. Bossé, "Approximation techniques for the transformation of fuzzy sets into random sets," *Fuzzy Sets and Systems*, vol. 159, pp. 270-288, February 2008.
- [7] D. Han, C. Han and Y. Deng, "Novel Approaches for the Transformation of Fuzzy Membership Function into Basic Probability Assignment based on Uncertainty Optimization," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 21, pp. 289-322, 2013.
- [8] Y. Yang, X. R. Li and D. Han, "An improved  $\alpha$ -cut approach to transforming fuzzy membership function into basic belief assignment," *Chinese Journal of Aeronautics*, vol. 29, pp.1042-1051, August 2016.
- [9] T. Ma, F. Xiao, "An Improved Method to Transform Triangular Fuzzy Number Into Basic Belief Assignment in Evidence Theory," *IEEE Access*, vol 7, pp. 25308-25322, February 2019.
- [10] R. Ilin, E. Blasch, "Information Fusion with Belief Functions: a Comparison of Proportional Conflict Redistribution PCR5 and PCR6 Rules for Networked Sensors," *International Conference on Information Fusion IEEE: Washington, DC, USA 2015: 2084-2091*.
- [11] F. Smarandache, J. Dezert, "On the consistency of PCR6 with the averaging rule and its application to probability estimation," *Proceedings of the 16th International Conference on Information Fusion IEEE: Istanbul, Turkey, 2013: 1119-1126*.
- [12] U. Höhle, "Entropy with respect to plausibility measures," *Proceedings of the 12th IEEE International Symposium on Multiple-Valued Logic*, pp. 167-169, 1982.
- [13] R. R. Yager, "Entropy and specificity in a mathematical theory of evidence," *International Journal of General Systems*, vol. 9, pp. 249-260, 1983.
- [14] D. Dubois, and H. Prade, "A note on measures of specificity for fuzzy sets," *International Journal of General Systems*, vol. 10, pp. 279-283, 1985.
- [15] A. L. Jousselme, C. Liu, D. Grenier, and É. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 36, pp. 890-903, August 2006.
- [16] D. Harmanec, G. J. Klir, "Measuring total uncertainty in Dempster-Shafer theory: a novel approach," *International Journal of General Systems*, vol. 22, pp. 405-419, 1994.
- [17] P. Smets, R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, pp. 191-234, April 1994.
- [18] J. Kampe de Fariet, "Interpretation of fuzzy membership functions of fuzzy sets in terms of plausibility and belief," *Fuzzy Information and Decision Processes*, pp. 93-98, January 1982.
- [19] A. L. Jousselme, D. Grenier and É. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, pp. 91-101, June 2001.

# Basic Belief Assignment Approximations Using Degree of Non-redundancy for Focal Element

Yi Yang<sup>a</sup>, Deqiang Han<sup>b</sup>, Jean Dezert<sup>c</sup>

<sup>a</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China.

<sup>b</sup>Center for Information Engineering Science Research, Xi'an Jiaotong University, Xi'an 710049, China.

<sup>c</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

Emails: jiafeiyy@mail.xjtu.edu.cn, deqhan@xjtu.edu.cn, jean.dezert@onera.fr

Originally published as: Y. Yang, D. Han, J. Dezert, *Basic Belief Assignment Approximations Using Degree of Non-Redundancy for Focal Element*, Chinese Journal of Aeronautics, Vol. 32(11), pp. 2503–2515, 2019, and reprinted with permission.

**Abstract**—Dempster-Shafer evidence theory, also called the theory of belief function, is widely used for uncertainty modeling and reasoning. However, when the size and number of focal elements are large, the evidence combination will bring a high computational complexity. To address this issue, various methods have been proposed including the implementation of more efficient combination rules and the simplifications or approximations of Basic Belief Assignments (BBAs). In this paper, a novel principle for approximating a BBA into a simpler one is proposed, which is based on the degree of non-redundancy for focal elements. More non-redundant focal elements are kept in the approximation while more redundant focal elements in the original BBA are removed first. Three types of degree of non-redundancy are defined based on three different definitions of focal element distance, respectively. Two different implementations of this principle for BBA approximations are proposed including a batch and an iterative type. Examples, experiments, comparisons and related analyses are provided to validate proposed approximation approaches.

**Keywords:** belief function, combination rule, BBA approximation, focal element redundancy.

## I. INTRODUCTION

Dempster-Shafer Theory (DST) [1] which is also called the theory of belief function, has been widely used in many uncertainty modeling and reasoning related application fields including information fusion [2], pattern classification [3] and Multiple Attributes Decision Making (MADM) [4]. However, DST was criticized because of its limitations [5]. One limitation is its computational complexity [6] in evidence combination, which is influenced by the cardinality of the frame of discernment and the number of focal elements in BBAs to combine. The high computational cost brings a big challenge to the practical use of belief functions.

To reduce the computational cost encountered in evidence combination, many approaches were proposed, which can be in general categorized into the following types. The first type is to design efficient combination algorithms. The representatives of this type include Kennes' method [7], Barnett's approach [8] and Shafer and Logan's implementation for hierarchical evidence [9]. The second type is to simplify the original Basic Belief Assignment (BBA), i.e., to obtain a corresponding approximated BBA. Two major types can be found in the prevailing BBA approximations: (A) To use a BBA with a

simpler and special structure to approximate the original one. For example, one can use the Bayesian BBA [10] and the consonant approximation of a BBA [11]; (B) To limit the quantity or size of focal elements by removing some focal elements by following some criteria (focal elements' size or mass value, or both). Tessem's  $k-l-x$  method [12], Lowrance et al's summarization approach [13], Bauer's D1 approximation [14], Dencœur's inner and outer approximations [15], Monte-Carlo approximation [16], etc. are representatives. They remove focal elements and redistribute the corresponding mass assignment values. In our previous works in recent years, a hierarchical proportional redistribution approach, 17 rank-level based BBA approximation [18], and optimization based approximations [19] were proposed. Shou et al. proposed a BBA approximation based on the correlation coefficient [20].

The work in current paper focuses on reducing the computational cost of evidence combination with BBA approximations. As aforementioned, one can limit the number of focal elements according to some criteria. Intuitively, the rational criterion should relate to the importance or non-redundancy of the focal elements. A focal element with more "common" or "shared information" with other focal elements is more redundant and should be removed first if possible. However, the available criterion is either the focal element's size (i.e., cardinality) or its mass assignment, which has no direct and logical relation with the focal elements' importance or the non-redundancy. Therefore, criteria related to the focal elements' non-redundancy are required for proposing more reliable and efficient BBA approximation approaches. This is the motivation of our work in this paper.

We use the average distance between a given focal element and all other focal elements to define the non-redundancy. Smaller average distance means that the given focal element carries more similar information compared with the others, i.e., it is less non-redundant and should be removed earlier. Different definitions of the distance between focal elements are used in this paper to define different non-redundancies of focal elements. Two strategies of removal (including a batch and an iterative mode) are proposed in the sequel, followed by the re-normalization or redistribution. Numerical examples, simulations and related analyses are provided to show the rationality

and interest of the novel BBA approximation approaches.

This paper extends our previous ideas briefly introduced in [21], where the non-redundancy for focal elements was preliminarily proposed. Comparatively, more definitions for distance of focal element are used to define the non-redundancy of focal elements and different distance definitions are analyzed and compared in this extended version. We also provide more experiments and analyses to provide a precise evaluation of these new approximation methods. These are all added values. The rest of the paper is organized as follows. Section II provides the essentials of DST. Some limitations, especially the computational cost, are pointed out. A brief review of the available works on BBA approximations is provided in Section III. Section III then proposes the non-redundancy of focal elements based on three different types of distance of focal elements. Numerical examples are provided to illustrate and compare different definitions of non-redundancy. Simulations and related analyses are provided in Section IV to verify and evaluate our proposed non-redundancy of focal elements and their performance in BBA approximations. Comparisons between the new proposed approaches and some typical existing ones are also provided. Section V concludes this paper.

## II. PRELIMINARIES OF BBA APPROXIMATION

### A. Basics of Dempster-Shafer evidence theory

In Dempster Shafer evidence theory [1], those elements in the Frame of Discernment (FOD)  $\Theta$  are mutually exclusive and exhaustive. A basic belief assignment (BBA, also called mass function) on a FOD is defined by a mapping  $m : 2^\Theta \mapsto [0, 1]$  satisfying  $m(\emptyset) = 0$  and

$$\sum_{A \in 2^\Theta} m(A) = 1 \quad (1)$$

If  $m(A) > 0$ ,  $A$  is a focal element. Two Bodies of Evidence (BOEs) can be combined using Dempster's rule as

$$m(A) = \begin{cases} 0, & \text{for } A = \emptyset, \\ \frac{1}{1-K} \sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j), & \text{for } A \neq \emptyset. \end{cases} \quad (2)$$

where  $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)$  is the conflict coefficient representing the total conflicting mass assignments between BOEs to combine. Note that Dempster's rule is both commutative and associative. Dempster's rule has also received serious arguments due to its counter-intuitive behaviors [22]. Various alternative combination rules have been proposed. See [23] for more details. These alternatives focus on suppressing the counter-intuitive behaviors of Dempster's rule. However, they also have to face the high computational cost problem with the increase of the FOD's cardinality and that of the focal elements number.

To reduce the high computational cost caused by the evidence combination, one can try to design simpler combination rules, attempt to develop efficient implementations for prevailing rules, or try to simplify (approximate) the original BBA by a simpler one with less focal elements. In this paper, we focus on the BBA approximation, which is deemed more intuitive for human beings to catch the meaning [24].

### B. Brief review of available BBA approximation approaches

An approximation  $f(\cdot)$  of BBA aims to find a simpler BBA  $m_S$  to represent the original BBA  $m$ , i.e.,  $m_S = f(m)$ . The available approaches can be categorized into the following two types: using the BBA with a special structure and reducing the number of focal elements.

#### 1) Using BBA with special structure:

##### (1) Bayesian BBA approximation

A Bayesian BBA approximation outputs a Bayesian BBA with a special structure where all focal elements are singletons. The most representative Bayesian approximation of a BBA is the pignistic probability transformation proposed by Smets [6] and Kennes [7]. Voorbraak [10] uses the normalization of the plausibility for singletons to approximate the original BBA. Sudano [25]–[27] proposed a series of Bayesian approximations based on the proportion between plausibilities or beliefs including the batch mode and the iterative mode. Cuzzolin [28] proposed an intersection approximation for BBA using the proportional repartition of the total non-specific mass assignment for each contribution of the non-specific mass assignments involved. Smarandache and Dezert [23] proposed a Bayesian BBA approximation in the framework of Dezert-Smarandache Theory (DSMT), i.e., the Dezert-Smarandache Probability transformation (DSMP), which can also be applied in DST model. In our previous work [29], a hierarchical DSMP was proposed. More analyses, comparisons and evaluations on these Bayesian approximations can be found in [30].

Note that the Bayesian approximation is usually used for the probabilistic decision but not reducing the computational cost in evidence combination, since any Bayesian BBA approximation makes too lossy approximations.

##### (2) Consonant approximations

Here, the special structure for an approximated BBA is assumed to be consonant support, i.e., the available focal elements are nested in order. The representative works of the consonant approximation include [11], [31].

#### 2) Removing focal elements according to some criteria:

##### (1) Limiting the maximum allowed cardinality of remaining focal elements

In  $k$ -additive approximations [32], the maximum cardinality of available focal elements is no greater than a predefined size  $k$ . In [32], the mass assignments of focal elements with cardinality larger than  $k$  are redistributed to those with cardinality no larger than  $k$ . Such a redistribution mass assignments is done according to the proportions designed based on the average cardinality. In our previous work [19], such a redistribution of mass assignments is implemented via an optimization approach. In our another previous work [17], a BBA approximation with the hierarchical redistribution was proposed. These methods aim to remove the focal elements with larger cardinalities since they bring more computational cost in the combination in general.

##### (2) Limiting the maximum allowed number of remaining focal elements

In this type of approaches, the number of focal elements is reduced by removing some focal elements according to some criteria until the predefined quantity of remaining focal elements is reached

(A)  $k-l-x$  method [12]:

A simplified BBA is obtained according to rules: one should keep no less than  $k$  focal elements; one should keep no more than  $l$  focal elements; one should delete the masses being no greater than  $x$ .

In the  $k-l-x$  method, all focal elements in the original BBA are sorted in a descending order based on their mass assignment values. Then, choose the first  $p$  focal elements such that  $k \leq p \leq l$  and the summation of mass values of those first  $p$  focal elements is no less than  $1-x$ . The removed mass values are redistributed to remaining focal elements (re-normalization).

(B) Summarization method [13]:

Summarization method is similar to the classical  $k-l-x$ , where focal elements with the highest mass values are kept. The removed mass values are accumulated and assigned to the union set of corresponding focal elements. Suppose that  $k$  is the number of focal elements in the desired simplified BBA  $m_S(\cdot)$  of an original BBA  $m(\cdot)$ . Let  $M$  denotes the collection (or set) of  $k-1$  focal elements with the highest mass values. One can obtain the simplified BBA according to

$$m_S(A) = \begin{cases} m(A), & \text{if } A \in M, \\ \sum_{A' \subseteq A, A' \notin M} m(A'), & \text{if } A = A_0, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

where

$$A_0 \triangleq \bigcup_{A' \notin M, m(A') > 0} A' \quad (4)$$

(C) D1 method [14]:

Let  $m(\cdot)$  be the original BBA and  $m_S(\cdot)$  denote the simplified BBA. The desired number of remaining focal elements is  $k$ . Let  $M$  denote the set including  $k-1$  focal elements with the highest mass assignment values in  $m(\cdot)$ , and  $M^-$  be the set including all the other focal elements of  $m(\cdot)$ . D1 method aims to keep all the members of  $M$  and to assign the mass values of those focal elements in set  $M^-$  among the focal elements in  $M$ . The set re-assignment is implemented as follows.

For  $A \in M^-$ , find all the supersets of  $A$  in  $M$  to form the set  $M_A$ . If  $M_A \neq \emptyset$ ,  $m(A)$  will be uniformly re-assigned among those focal elements with smallest size in  $M_A$ . When  $M_A = \emptyset$ , then construct the set  $M'_A$ :

$$M'_A = \{B \in M \mid |B| \geq |A|, B \cap A \neq \emptyset\} \quad (5)$$

If  $M'_A \neq \emptyset$ ,  $m(A)$  is assigned among the focal elements with smallest size in  $M'_A$ . The value assigned to a focal element  $B$  depends on  $|B \cap A|$ . The above procedure will be executed iteratively until all  $m(A)$ ,  $A \in M^-$  have been re-assigned to those focal elements in the set

$M$ . If  $M'_A = \emptyset$  there might be two cases: if  $\Theta \in M$ , the summation of mass assignment values of the focal elements in  $M^-$  will be added to  $m(\Theta)$ ; if  $\Theta \notin M$ , one should set  $\Theta$  as a focal element of  $m_S(\cdot)$  and assign the sum of mass assignment values of the focal elements in the set  $M^-$  to the simplified BBA  $m_S(\Theta)$ .

More details on D1 method with examples can be found in [14].

(D) Joint use of cardinality and mass assignment with rank-level fusion:

In our previous work [18], we jointly use the cardinality and the mass values of focal element to design a rank-level fusion based BBA approximation approach, which is briefly recalled below.

**Step 1.** Sort all the focal elements of an original BBA (with  $L$  focal elements) in an ascending order according to the mass assignment values (an underlying assumption: the focal element with small mass should be deleted first). The rank vector can be obtained as

$$\mathbf{r}_m = [r_m(1), r_m(2), \dots, r_m(L)] \quad (6)$$

Here  $r_m(i)$  is the rank position of the  $i$ -th focal element ( $i = 1, 2, \dots, L$ ) in the original BBA based on mass values.

**Step 2.** Sort all focal elements of the original BBA in a descending order according to the cardinalities (an underlying assumption: the focal element with big cardinality should be deleted first). The rank vector can be obtained as

$$\mathbf{r}_c = [r_c(1), r_c(2), \dots, r_c(L)] \quad (7)$$

Here  $r_c(i)$  denotes the rank position of the  $i$ -th focal elements in the original BBA based on the focal element size.

**Step 3.** By using the rank-level fusion (weighted average), one can obtain a fused rank vector as

$$\mathbf{r}_f = [r_f(1), r_f(2), \dots, r_f(L)] \quad (8)$$

where  $r_f(i) = \alpha r_m(i) + (1-\alpha)r_c(i)$  and  $\alpha \in [0, 1]$  denotes the preference of two different criteria. Such a fused rank can be considered as a relatively comprehensive criterion reflecting both the information of mass values and cardinality.

**Step 4.** Sort  $\mathbf{r}_f$  in an ascending order and find out the focal element with the smallest  $r_f$  value, i.e.,  $r_f(j) = \min r_f(i)$ . Then remove the  $j$ -th focal element in the original BBA.

**Step 5.** Repeat Steps 1–4 until  $k$  focal elements are left. Renormalize the remaining masses of the  $k$  focal elements, and output the approximated BBA in the final.

**Step 6.** Correlation coefficient based BBA approximation (CR-based approximation)

The correlation coefficient is defined as

$$\text{CR}_{\text{BBA}}(m_1, m_2) = \frac{c(m_1, m_2)}{\sqrt{c(m_1, m_1)c(m_2, m_2)}} \quad (9)$$

where

$$c(m_1, m_2) = \sum_{i=1}^{2^n-1} \sum_{j=1}^{2^n-1} m_1(A_i) m_2(A_j) \frac{|A_i \cap B_j|}{|A_i \cup B_j|} \quad (10)$$

is used for BBA approximation. Suppose that original BBA has  $L$  focal elements, and the quantity of desired remaining focal elements is  $k$ .

- i) Remove one focal element  $A_i$  and reassigned its mass value  $m(A_i)$  to the related remaining focal elements according to the redistribution strategy based on singleton relation proposed in Ref. 20 to generate an approximated BBA  $m'_i$ . For each  $A_j$  ( $j = 1, 2, \dots, L$ ) and corresponding  $m'_j$ , calculate  $c(m_i, m'_j)$  using Eq. (10)).
- ii) Sort the  $A_j$  ( $j = 1, 2, \dots, L$ ) according to  $c(m_i, m'_j)$  in an ascending order. Remove  $L - k$  focal elements with top  $L - k$  values of correlation coefficient  $c$ .
- iii) Reassign the mass values of removed focal elements to the remaining focal elements according to the redistribution strategy based on singleton relation proposed in [20]. Then, one obtain the approximated BBA.

Besides the above BBA approximations with a preset quantity of remaining focal elements, Denœux's BBA approximations by the outer and inner approximations [15] using distance between focal elements also preset such a quantity in the approximations. See [15] for details.

Note that the Monte-Carlo based BBA approximation can also be classified into the approximation approaches using the strategy of removing focal elements. See [16] for details.

In this paper, we focus on the BBA approximations through presetting the quantity of remaining focal elements. As aforementioned, existing BBA approximations of this type proposed to remove some focal elements that have smaller mass assignment values, larger cardinalities, or both. Although they have some rational justifications, it is quite dangerous (or risky) to remove those focal elements with small mass values or larger sizes. It may also be unconvincing to remove those focal elements with large cardinality justified only by their bringing possible high computational cost to the combination. Therefore, one should be prudent when using a technique of BBA approximation. It is more convincing to remove those "unimportant" focal elements. The very redundant focal elements can reasonably be considered as "unimportant" (carry duplicate information) and the relatively non-redundant focal elements can reasonably be deemed as important; therefore, we propose to define the degree of non-redundancy for a focal element at first. From this degree of non-redundancy, we can then develop new BBA approximation methods by removing focal elements according to the degree of non-redundancy, and intuitively, the loss of information in terms of distance of evidence might be smaller.

### III. BBA APPROXIMATIONS BASED ON NON-REDUNDANCY OF FOCAL ELEMENTS

In this section, we define the degree of non-redundancy for focal elements based on the distance of focal elements first. Then, we design BBA approximations based on the degree of non-redundancy.

#### A. Non-redundancy of focal elements

Suppose that a BBA  $m(\cdot)$  has  $l > 2$  focal elements. If a focal element  $A_i$  has the largest average distance with other focal elements  $A_j \subseteq \Theta$  ( $j \neq i$ ), then  $A_i$  shares the least common information with other focal elements in the BBA  $m(\cdot)$ , i.e.,  $A_i$  is the most non-redundant one. Therefore, one can define the degree of non-redundancy using the average focal distance between a focal element and the others. Suppose that  $d^F(A_i, A_j)$  is the distance between two focal elements  $A_i$  and  $A_j$ . First, we can compute the distance matrix for all focal elements in BBA  $m(\cdot)$  as

$$\mathbf{Mat}_{FE} = \begin{bmatrix} d^F(A_1, A_1) & d^F(A_1, A_2) & \dots & d^F(A_1, A_l) \\ d^F(A_2, A_1) & d^F(A_2, A_2) & \dots & d^F(A_2, A_l) \\ \vdots & \vdots & \ddots & \vdots \\ d^F(A_l, A_1) & d^F(A_l, A_2) & \dots & d^F(A_l, A_l) \end{bmatrix} \quad (11)$$

Since  $d^F$  is a distance, at least there should exist  $d^F(A_i, A_i) = 0$  and  $d^F(A_i, A_j) = d^F(A_j, A_i)$  where  $i = 1, 2, \dots, l$ . That is, the matrix  $\mathbf{Mat}_{FE}$  is symmetric. Therefore, it is not necessary to compute all elements in  $\mathbf{Mat}_{FE}$ .

For focal element  $A_i$ , we can then define its degree of non-redundancy as

$$\text{nRd}(A_i) \triangleq \frac{1}{l-1} \sum_{j=1}^{l-1} d^F(A_i, A_j) \quad (12)$$

When  $\text{nRd}(A_i)$  is larger,  $A_i$  has a larger non-redundancy (less redundancy); when  $\text{nRd}(A_i)$  takes a smaller value,  $A_i$  has a less non-redundancy (larger redundancy). Then, the problem is how to describe the distance between focal elements. To be more strictly, the "distance" used here should be "dissimilarity", since the distance metric should satisfy all the four requirements including non-degeneracy, symmetry, non-negativity, and the triangular inequality. When there is no confusion raised, we still use the distance in the sequel.

#### B. Distance between focal elements

In general, the distance between two focal elements should use the two aspects of information in focal elements including the mass assignment and focal element (set) as

$$d^F(A_i, A_j) \triangleq f(m(A_i), A_i, m(A_j), A_j) \quad (13)$$

The available distances between focal elements are introduced below.

##### (1) Erkmen's distance:

Erkmen and Stephanou [33] proposed a distance (denoted by  $d_E^F$  here) between focal elements as

$$d_E^F(A_i, A_j) = \frac{|A_i \cup A_j|}{|A_i \cap A_j|} [m(A_i) - m(A_j)] \log_2 \frac{m(A_i)}{m(A_j)} \quad (14)$$

This definition is far from robustness and can bring counter-intuitive results as shown in the following cases.

**Case I:** if  $A_i \cap A_j = \emptyset$ , i.e.,  $|A_i \cap A_j| = 0$ , then  $d_E^F(A_i, A_j) = 0$  cannot be calculated (due to a division by zero). One can also say that it tends to infinity; however, this is not reasonable since in this case the value of distance is dominated by the relationship between focal elements (sets).

**Case II:** if  $m(A_i) = m(A_j)$ , then  $d_E^F(A_i, A_j) = 0$ . This is also counter-intuitive, because the distance value is totally dominated by mass assignments. That is to say, two different focal elements with the same mass value is deemed as identical. Therefore, Erkmen's definition is not appropriate for designing the focal element redundancy.

(2) Dencœux's union distance:

Dencœux [15] proposed a union-operation based distance as

$$\delta_U(A_i, A_j) = [m(A_i) + m(A_j)]|A_i \cup A_j| - m(A_i)|A_i| - m(A_j)|A_j| \quad (15)$$

(3) Dencœux's intersection distance:

Dencœux [15] also proposed an intersection-operation based distance as

$$\delta_\cap(A_i, A_j) = m(A_i)|A_i| + m(A_j)|A_j| - [m(A_i) + m(A_j)]|A_i \cap A_j| \quad (16)$$

Actually, both  $\delta_U$  and  $\delta_\cap$  can be considered as a weighted sum of the Hamming distance [15]. It is not difficult to verify that both  $\delta_U$  and  $\delta_\cap$  have no counter-intuitive results for aforementioned Cases I and II. Therefore, we choose  $\delta_U$  and  $\delta_\cap$  to define the degrees of non-redundancy for the focal element. Here we give further analyses on the two distance definitions  $\delta_U$  and  $\delta_\cap$ .

### C. Analyses on $\delta_U$ and $\delta_\cap$

Suppose that  $m(\cdot)$  is a BBA defined on the FOD  $\Theta$  where  $|\Theta| = n$ . To simplify the analysis, we assume that  $m(\cdot)$  only has two focal elements  $A_1$  and  $A_2$  with mass assignments  $m(A_1) = a$  and  $m(A_2) = 1 - a$ . The behaviors of  $\delta_U$  and  $\delta_\cap$  are analyzed under different situations.

1) *Focal elements' relation:*  $A_1 \subset A_2$ .

In such a case, for  $\delta_\cap$  one gets

$$\begin{aligned} \delta_\cap(A_1, A_2) &= m(A_1)|A_1| + m(A_2)|A_2| \\ &\quad - [m(A_1) + m(A_2)]|A_1 \cap A_2| \\ &= m(A_1)|A_1| + m(A_2)|A_2| \\ &\quad - [m(A_1) + m(A_2)]|A_1| \\ &= (1 - m(A_1))(|A_2| - |A_1|) \end{aligned} \quad (17)$$

As shown in Eq. (17), if  $m(A_1)$  is fixed,  $\delta_\cap$  becomes larger with the enlargement of the difference between focal elements' cardinalities  $|A_2| - |A_1|$ . This makes sense. If the difference of cardinalities i.e.,  $|A_2| - |A_1|$  is fixed,  $\delta_\cap$  becomes smaller with the increase of mass assignment of  $A_1$  (which is contained by  $A_2$ ).

For  $\delta_\cap$ , one gets

$$\begin{aligned} \delta_\cap(A_1, A_2) &= [m(A_1) + m(A_2)]|A_1 \cup A_2| \\ &\quad - m(A_1)|A_1| - m(A_2)|A_2| \\ &= [m(A_1) + m(A_2)]|A_2| \\ &\quad - m(A_1)|A_1| - m(A_2)|A_2| \\ &= m(A_1)(|A_2| - |A_1|) \end{aligned} \quad (18)$$

As shown in Eq. (18), if  $m(A)$  is fixed,  $\delta_U$  becomes larger with the enlargement of the difference between focal elements' cardinalities  $|A_2| - |A_1|$ . This makes sense. If the difference of cardinalities i.e.,  $|A_2| - |A_1|$  is fixed,  $\delta_U$  becomes larger with the increase of mass assignment of  $A_1$  (contained by  $A_2$ ). That is, when  $A_1 \subset A_2$  and  $|A_2| - |A_1|$  are fixed,  $\delta_U$  is positively correlated to the mass of focal element with smaller cardinality ( $A_1$ ), while  $\delta_U$  is positively correlated to the mass of focal element with larger cardinality ( $A_2$ ).

The analyses above can be supported by Example 1 below.

**Example 1.** (Focal elements are nested) Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_5\}$ . Four BBAs are defined on  $\Theta$ , and each has two focal elements as listed in Table I.

Table I  
FOUR BBAS IN EXAMPLE 1.

BBA	$A_1$	$A_2$
$m_1$	$\{\theta_1\}$	$\Theta$
$m_2$	$\{\theta_1, \theta_2\}$	$\Theta$
$m_3$	$\{\theta_1, \theta_2, \theta_3\}$	$\Theta$
$m_4$	$\{\theta_1, \theta_2, \theta_3, \theta_4\}$	$\Theta$

For each BBA, the mass value of  $A_1$  changes from 0.01 to 0.95 with an increase of 0.01 at each step. The values of  $\delta_\cap$  and  $\delta_U$  are shown in Fig. 1.

As shown in Fig. 1,  $\delta_U$  is positively correlated to the mass of focal element with smaller cardinality ( $A_1$ ) while  $\delta_\cap$  is positively correlated to the mass of focal element with larger cardinality ( $A_2$ ). Given a fixed  $m(A_1)$ , with the increase of cardinality of  $A_1$ , i.e., the decrease of  $|A_2| - |A_1|$ , both  $\delta_\cap$  and  $\delta_U$  become smaller.

2) *Focal elements' relation:*  $A_1 \cap A_2 = \emptyset$ .

When  $A_1 \cap A_2 = \emptyset$ , one gets  $|A_1 \cap A_2| = 0$  and

$$\begin{aligned} \delta_\cap(A_1, A_2) &= m(A_1)|A_1| + m(A_2)|A_2| \\ &\quad - [m(A_1) + m(A_2)]|A_1 \cap A_2| \\ &= m(A_1)|A_1| + m(A_2)|A_2| \\ &= m(A_1)(|A_1| - |A_2|) + |A_2| \end{aligned} \quad (19)$$

For  $\delta_U$ , one gets

$$\begin{aligned} \delta_U(A_1, A_2) &= [m(A_1) + m(A_2)]|A_1 \cup A_2| \\ &\quad - m(A_1)|A_1| - m(A_2)|A_2| \\ &= [m(A_1) + m(A_2)](|A_1| + |A_2|) \\ &\quad - m(A_1)|A_1| - m(A_2)|A_2| \\ &= m(A_1)|A_2| - m(A_2)|A_1| \\ &= m(A_1)(|A_2| - |A_1|) + |A_1| \end{aligned} \quad (20)$$

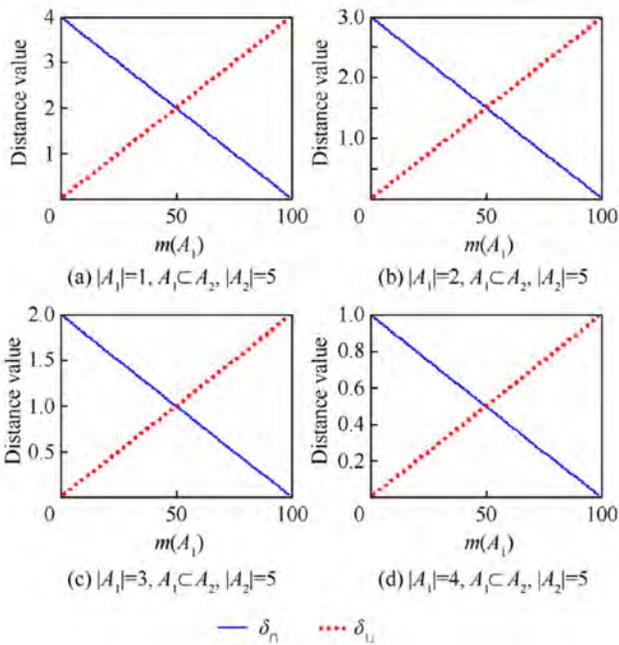


Figure 1. Two distances in Example 1

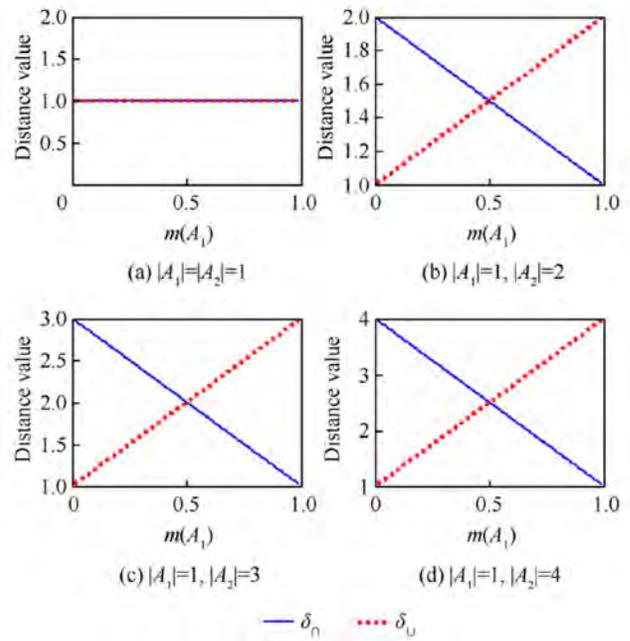


Figure 2. Two distances in Example 2

It can be seen that when  $A_1 \cap A_2 = \emptyset$  if  $|A_1|$  is closer to  $|A_2|$ , both  $\delta_\cap$  and  $\delta_\cup$  are smaller (it means that  $\{\theta_1\}$  is farther from  $\{\theta_2, \theta_3\}$  than from  $\{\theta_3\}$ , which makes some sense). This can be shown in Example 2 below.

**Example 2.** (focal elements have no intersect) Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_5\}$ . Four BBAs are defined on  $\Theta$ , and each has two focal elements as listed in Table II.

Table II  
FOUR BBAs IN EXAMPLE 2.

BBA	$A_1$	$A_2$
$m_1$	$\{\theta_1\}$	$\{\theta_2\}$
$m_2$	$\{\theta_1\}$	$\{\theta_2, \theta_3\}$
$m_3$	$\{\theta_1\}$	$\{\theta_2, \theta_3, \theta_4\}$
$m_4$	$\{\theta_1\}$	$\{\theta_2, \theta_3, \theta_4, \theta_5\}$

In each BBA, the two focal elements have an empty intersection. For each BBA, the mass assignment of  $A_1$  changes from 0.01 to 0.95 with an increase of 0.01 at each step. The values of  $\delta_\cap$  and  $\delta_\cup$  are shown in Fig. 2.

As shown in Fig. 2, when  $|A_1| = |A_2|$  and  $|A_1|$  is fixed, both  $\delta_\cap$  and  $\delta_\cup$  remain unchanged. Given a fixed  $m(A_1)$ , when the difference  $|A_2| - |A_1|$  becomes larger, both  $\delta_\cap$  and  $\delta_\cup$  become larger. When the difference  $|A_2| - |A_1|$  is fixed,  $\delta_\cup$  is positively correlated to the mass of focal element with smaller cardinality ( $A_1$ ), while  $\delta_\cap$  is positively correlated to the mass of the focal element with larger cardinality ( $A_2$ ), i.e., negatively correlated to the mass of the focal element with smaller cardinality.

3) *Focal elements' relation:*  $A_1 \cap A_2 \neq \emptyset$ .

Here  $A_1 \cap A_2 \neq \emptyset$ . Furthermore,  $A_1$  cannot be contained by  $A_2$ , and  $A_2$  cannot be contained by  $A_1$ . We provide an example to show  $\delta_\cap$  and  $\delta_\cup$  behaviors in this situation.

**Example 3.** (focal elements have no intersect) Suppose that the FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_6\}$ . Four BBAs are defined on  $\Theta$ , and each has two focal elements as listed in Table III.

Table III  
FOUR BBAs IN EXAMPLE 3.

BBA	$A_1$	$A_2$
$m_1$	$\{\theta_1, \theta_2\}$	$\{\theta_2, \theta_3\}$
$m_2$	$\{\theta_1, \theta_2\}$	$\{\theta_2, \theta_3, \theta_4\}$
$m_3$	$\{\theta_1, \theta_2\}$	$\{\theta_2, \theta_3, \theta_4, \theta_5\}$
$m_4$	$\{\theta_1, \theta_2\}$	$\{\theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$

For each BBA, the mass assignment of  $A_1$  changes from 0.01 to 0.95 with an increase of 0.01 at each step. The values of  $\delta_\cap$  and  $\delta_\cup$  are shown in Fig. 3.

As we see in Fig. 3, when  $|A_1| = |A_2|$ ,  $\delta_\cap$  and  $\delta_\cup$  equal 1, and they remain unchanged. This is because when  $|A_1| = |A_2| = 2$ , one has  $\delta_\cup(A_1, A_2) = |A_1 \cup A_2| - |A_2|$  and  $\delta_\cap(A_1, A_2) = |A_2| - |A_1 \cap A_2|$ . So,  $\delta_\cap(A_1, A_2) = 1$  and  $\delta_\cup(A_1, A_2) = 1$ .

Given a fixed  $m(A_1)$ , when the difference  $|A_2| - |A_1|$  becomes larger, both  $\delta_\cap$  and  $\delta_\cup$  become larger as shown in Fig. 3. This makes sense, because the uncommon part of  $A_1$  and  $A_2$  becomes large. When the difference  $|A_2| - |A_1|$  is fixed,  $\delta_\cup$  is positively correlated to the mass of focal element with  $A_1$  having a smaller cardinality, while  $\delta_\cap$  is positively correlated to the mass of the focal element  $A_2$  having a larger

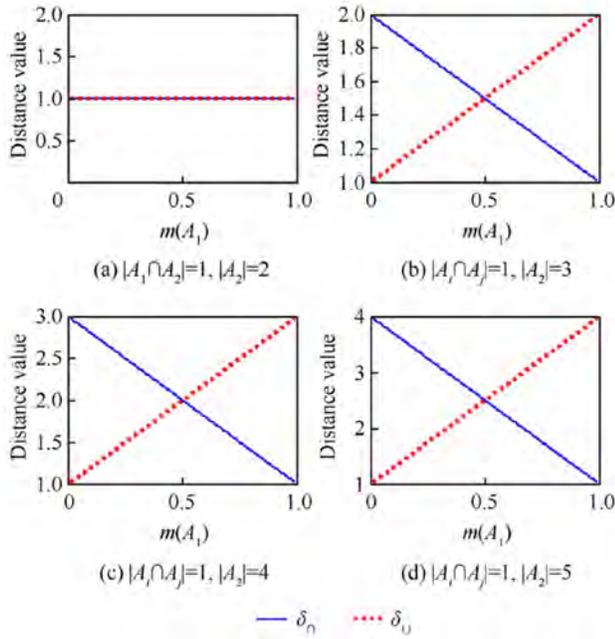


Figure 3. Two distances in Example 3

cardinality, i.e., negatively correlated to the mass of the focal element with smaller cardinality.

#### D. Implementation of BBA approximation using degree of redundancy for focal elements

Based on the degree of non-redundancy in Eq. (12), new BBA approximation methods are proposed in this paper, where the more non-redundant focal elements are kept and the more redundant ones will be removed earlier.

1) *Batch-mode approximation method*: Given an original BBA  $m(\cdot)$  with  $l$  focal elements, in the approximation, we want to keep  $k < l$  focal elements. The batch-mode means that the focal elements quantity is reduced from  $l$  to  $k$  in one run as follows.

**Step 1.** Compute the matrix  $\mathbf{Mat}_{FE}$  first, and then for each  $A_i$  ( $i = 1, 2, \dots, l$ ) compute its non-redundancy value  $nRd(A_i)$ .

**Step 2.** Sort all  $nRd(A_i)$  ( $i = 1, 2, \dots, l$ ) in a descending order.

**Step 3.** Remove the focal elements with ranking positions of bottom  $l - k$ .

**Step 4.** Normalize the mass assignments of the kept  $k$  focal elements and obtain the approximated BBA  $m_S^{BRd}(\cdot)$ .

2) *Iterative-mode approximation method*: Here, we propose to remove iteratively the most redundant focal element (with the least  $nRd$  value) in each step until  $k$  focal elements are kept. This method consists of the following steps:

**Step 1.** Compute the matrix  $\mathbf{Mat}_{FE}$  and the  $nRd$  values for each focal element  $A_i$  ( $i = 1, 2, \dots, l$ ).

**Step 2.** Sort all  $nRd(A_i)$  ( $i = 1, 2, \dots, l$ ) in a descending order.

**Step 3.** Remove the bottom focal element  $A_r$ .

**Step 4.** If the quantity of the kept focal elements is larger than  $k$ , re-compute  $nRd(A_i)$  of the kept focal elements where  $i \neq r$  and go back to Step 3. Otherwise, switch to Step 5.

**Step 5.** Normalize the mass assignments of the kept  $k$  focal elements and obtain the approximated BBA  $m_S^{IRd}(\cdot)$ .

In the iterative-mode, the matrix and degrees of non-redundancy are re-computed in each step after removing a focal element in the precedent step. That is to say, only the non-redundancy values of the current remaining focal elements are involved in each step.

#### E. Illustrative examples

Illustrative examples for presenting the procedure of our proposed non-redundancy degree based BBA approximation approaches are provided here. The specific calculation steps of other major BBA approximation approaches with presetting the number of focal elements are also provided here for comparisons.

**Example 4.** Let us consider a BBA  $m(\cdot)$  defined on  $\Theta = \{\theta_1, \theta_2, \dots, \theta_5\}$  as listed in Table IV.

Table IV  
FOCAL ELEMENTS AND MASS ASSIGNMENTS.

Focal element	Mass value
$A_1 = \{\theta_1, \theta_2\}$	0.50
$A_2 = \{\theta_1, \theta_3, \theta_4\}$	0.30
$A_3 = \{\theta_3\}$	0.10
$A_4 = \{\theta_3, \theta_4\}$	0.05
$A_5 = \{\theta_4, \theta_5\}$	0.05

(1) Using  $k - l - x$  [12]:

Parameters  $k$  and  $l$  are both set to 3, and  $x = 0.1$ . Focal elements  $A_4 = \{\theta_3, \theta_4\}$  and  $A_5 = \{\theta_4, \theta_5\}$  are removed. The kept total mass value is  $1 - 0.05 - 0.05 = 0.9$ ; therefore, the constraint of  $x$  is not violated. All the remaining focal elements' mass assignments are divided by 0.9 for the normalization. The approximated BBA  $m_S^{k-l-x}(\cdot)$  obtained using  $k - l - x$  is shown in Table V. Here  $A'_i$  ( $i = 1, 2, 3$ ) are focal elements<sup>1</sup> in  $m_S^{k-l-x}(\cdot)$ .

Table V  
 $m_S^{k-l-x}(\cdot)$  USING  $k - l - x$  FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.5556
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.3333
$A'_3 = \{\theta_3\}$	0.1111

(2) Using summarization [13]:

Parameter  $k = 3$ . By using the summarization method, one removes focal elements  $A_3 = \{\theta_3\}$ ,  $A_4 = \{\theta_3, \theta_4\}$  and

<sup>1</sup>In the tables VI–XI, table XIII, and tables XV–XVIII we also denote  $A'_i$  the focal elements of approximate BBAs obtained by the different methods.

$A_5 = \{\theta_4, \theta_5\}$ . Their union  $\{\theta_3, \theta_4, \theta_5\}$  is set as a new focal element whose mass assignment is 0.2, since  $m(\{\theta_3\}) + m(\{\theta_3, \theta_4\}) + m(\{\theta_4, \theta_5\}) = 0.2$ . The approximated BBA  $m_S^{\text{Sum}}(\cdot)$  is as shown in Table VI.

Table VI  
 $m_S^{\text{Sum}}(\cdot)$  USING SUMMARIZATION FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.50
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.30
$A'_3 = \{\theta_3, \theta_4, \theta_5\}$	0.20

(3) Using D1 method [14]

The parameter  $k$  is still set to 3 here. When we use D1 method, focal elements  $A_1$  and  $A_2$  belong to  $M$  and  $A_3, A_4$  and  $A_5$  belong to  $M^-$ . The focal element  $A_1 = \{\theta_1, \theta_2\}$  has no intersection with those focal elements in  $M^-$ ; therefore, its value remains unchanged. In  $M$ ,  $A_2$  is the unique superset of  $A_3$  and  $A_4$ , so,  $m(A_3) + m(A_4) = 0.10 + 0.05 = 0.15$  is added to  $A_2$ 's original mass assignment.  $A_2$  covers half of  $A_5$ , so  $m(A_5)/2 = 0.025$  is further added to the mass of  $A_2$ . Finally, the rest mass is assigned to  $\Theta$ . The approximated BBA  $m_S^{\text{D1}}(\cdot)$  is as shown in Table VII.

Table VII  
 $m_S^{\text{D1}}(\cdot)$  USING D1 FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.500
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.475
$A'_3 = \Theta$	0.025

(4) Using Denœux's inner and outer approximations [15]

Since this method uses the focal element distance definition in Eq. (14), here we also use it for comparison. When using the inner approximation [15], the focal elements pair with the smallest distance is removed, and their intersection is considered as a supplemented focal element. Its mass value is the summation of two removed focal elements' mass assignments. Such a procedure is repeated until the preset focal elements quantity is reached. The approximated BBA  $m_S^{\text{Inner}}(\cdot)$  is shown in Table VIII.

Table VIII  
 $m_S^{\text{Inner}}(\cdot)$  USING INNER APPROXIMATION FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.50
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.30
$A'_3 = \emptyset$	0.20

As one sees in Table VIII, the empty set is generated as a focal element, which is not allowed in the classical DST under the closed-world assumption.

The outer approximation is similar to the inner approximation except that the distance used is  $\delta_U$ . The approximated BBA  $m_S^{\text{Outer}}(\cdot)$  is shown in Table IX.

Table IX  
 $m_S^{\text{Outer}}(\cdot)$  USING OUTER APPROXIMATION FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.50
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.45
$A'_3 = \{\theta_4, \theta_5\}$	0.05

(5) Using rank-level fusion based method [18]

The rank of focal elements in  $m(\cdot)$  according to the mass assignments is  $[1, 2, 3, 4, 4]$  (in a descending order). Here  $[1, 2, 3, 4, 4]$  means that  $A_1$  takes the 1st place;  $A_2$  takes the 2nd place;  $A_3$  takes the 3rd place; and  $A_4$  and  $A_5$  both take the 4th place due to their equal mass values.

The rank of focal elements according to their cardinalities in ascending order is  $[2, 3, 1, 2, 2]$ . Here we set  $\alpha = 0.5$ , and approximated BBA  $m_S^{\text{Rank}}(\cdot)$  is shown in Table X.

Table X  
 $m_S^{\text{Rank}}(\cdot)$  USING RANK-LEVEL FUSION BASED APPROXIMATION FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.7692
$A'_2 = \{\theta_3, \theta_4\}$	0.1539
$A'_3 = \{\theta_1, \theta_3, \theta_4\}$	0.0769

(6) Using CR-based approximation

Using the CR-based approximation, the correlation coefficient values are

$$\begin{aligned} c(m, m'_1) &= 0.7096, \\ c(m, m'_2) &= 0.9462, \\ c(m, m'_3) &= 0.9912, \\ c(m, m'_4) &= 0.9462, \\ c(m, m'_5) &= 0.9975. \end{aligned}$$

Then, remove  $A_3$  and  $A_5$ , since they have the top two correlation coefficient values. After the redistribution, the approximated BBA  $m_S^{\text{CR}}(\cdot)$  is shown in Table XI.

Table XI  
 $m_S^{\text{CR}}(\cdot)$  USING RANK-LEVEL FUSION BASED APPROXIMATION FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.5083
$A'_2 = \{\theta_3\}$	0.1208
$A'_3 = \{\theta_1, \theta_3, \theta_4\}$	0.3709

(7) Using the non-redundancy based batch-mode approximation

We want to keep three focal elements, i.e.,  $k = 3$ . Calculate the distance matrix  $\mathbf{Mat}_{FE}$  as

$$\mathbf{Mat}_{FE} = \begin{matrix} & A_1 & A_2 & A_3 & A_4 & A_5 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{matrix} & \begin{bmatrix} 0 & 1.10 & 1.10 & 1.10 & 1.10 \\ 1.10 & 0 & 0.60 & 0.30 & 0.65 \\ 1.10 & 0.60 & 0 & 0.05 & 0.20 \\ 1.10 & 0.30 & 0.05 & 0 & 0.10 \\ 1.10 & 0.65 & 0.20 & 0.10 & 0 \end{bmatrix} \end{matrix}$$

Using this matrix, the degree of non-redundancy for all focal elements of  $m(\cdot)$  are obtained as listed in Table XII.

Table XII  
NON-REDUNDANCY FOR DIFFERENT FOCAL ELEMENTS.

Focal element	Mass value	nRd( $A_i$ )
$A_1 = \{\theta_1, \theta_2\}$	0.50	1.1000
$A_2 = \{\theta_1, \theta_3, \theta_4\}$	0.30	0.6625
$A_3 = \{\theta_3\}$	0.10	0.4875
$A_4 = \{\theta_3, \theta_4\}$	0.05	0.3875
$A_5 = \{\theta_4, \theta_5\}$	0.05	0.5125

Since  $A_3$  and  $A_4$  have the two smallest nRd values, they are two focal elements with the lowest non-redundancy (the highest redundancy). So, they'd better be removed first and their mass assignments are redistributed with the classical normalization step. The approximated BBA  $m_S^{BRd}(\cdot)$  is listed in Table XIII.

Table XIII  
 $m_S^{BRd}(\cdot)$  USING BATCH APPROXIMATION BASED ON REDUNDANCY FOR EXAMPLE 4.

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.5882
$A'_2 = \{\theta_1, \theta_3, \theta_4\}$	0.3530
$A'_3 = \{\theta_4, \theta_5\}$	0.0588

(8) Using the redundancy-based iterative approximation

Here  $k = 3$ , and then two focal elements should be removed. In the iterative mode, we only remove one focal element in each step. Therefore, two steps are required in this example.

In Step 1, we obtain the same degrees of non-redundancy as listed in Table XI. Then,  $A_4$  is removed.

In Step 2, nRd for  $A_i$  ( $i = 1, 2, \dots, 5, i \neq 4$ ) is recalculated according to

$$\text{nRd}(A_i) = \sum_{j=1, j \neq 4, j \neq i}^5 d^F(A_i, A_j)$$

The results are

$$\begin{aligned} \text{nRd}(A_1) &= 1.1000, \\ \text{nRd}(A_2) &= 0.7833, \\ \text{nRd}(A_3) &= 0.6333, \\ \text{nRd}(A_5) &= 0.6500. \end{aligned}$$

Then,  $A_3$  is removed due to its smallest nRd value (i.e., the biggest redundancy among those remaining focal elements). In this example, the BBA  $m_S^{IRd}(\cdot)$  obtained is the same as  $m_S^{BRd}(\cdot)$  listed in Table XII. Note that the batch-mode and the iterative approximations do not always obtain the same results as illustrated in Example 5.

**Example 5.** Assume that the FOD is  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . An original BBA  $m(\cdot)$  is listed in Table XIV, and the quantity of remaining focal elements is set to  $k = 3$ .

Table XIV  
FOCAL ELEMENTS AND MASS VALUES.

Focal element	Mass value
$A_1 = \{\theta_1, \theta_2\}$	0.1780
$A_2 = \{\theta_2, \theta_3\}$	0.2477
$A_3 = \{\theta_2\}$	0.2322
$A_4 = \{\theta_3\}$	0.1759
$A_5 = \Theta$	0.1662

• Using  $\delta_{\cap}$  we can obtain the distance matrix:

$$\mathbf{Mat}_{FE} = \begin{matrix} & A_1 & A_2 & A_3 & A_4 & A_5 \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{matrix} & \begin{bmatrix} 0 & 0.4257 & 0.1780 & 0.5319 & 0.1662 \\ 0.4257 & 0 & 0.2477 & 0.2477 & 0.1662 \\ 0.1780 & 0.2477 & 0 & 0.4081 & 0.3324 \\ 0.5319 & 0.2477 & 0.4081 & 0 & 0.3324 \\ 0.1662 & 0.1662 & 0.3324 & 0.3324 & 0 \end{bmatrix} \end{matrix}$$

All focal elements' degrees of non-redundancy are

$$\begin{aligned} \text{nRd}(A_1) &= 0.3255, \\ \text{nRd}(A_2) &= 0.2718, \\ \text{nRd}(A_3) &= 0.2915, \\ \text{nRd}(A_4) &= 0.3800, \\ \text{nRd}(A_5) &= 0.2493. \end{aligned}$$

Using the batch mode method, focal elements  $A_2$  and  $A_5$  are removed. The approximated BBA is shown in Table XV.

Table XV  
 $m_S^{BRd}(\cdot)$  USING BATCH APPROXIMATION BASED ON REDUNDANCY WITH THE DISTANCE  $\delta_{\cap}$ .

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.5882
$A'_2 = \{\theta_2\}$	0.3530
$A'_3 = \{\theta_3\}$	0.0588

By using the iterative mode method, the degrees of non-redundancy obtained in Step 1 are

$$\begin{aligned} \text{nRd}^I(A_1) &= 0.3255, \\ \text{nRd}^I(A_2) &= 0.2718, \\ \text{nRd}^I(A_3) &= 0.2915, \\ \text{nRd}^I(A_4) &= 0.3800, \\ \text{nRd}^I(A_5) &= 0.2493. \end{aligned}$$

Then, we first remove the focal element  $A_5$  since  $\text{nRd}^I(A_5)$  is the least one. Then recalculate nRd values for remaining focal elements  $A_1, A_2, A_3$  and  $A_4$ :

$$\begin{aligned}\text{nRd}^{II}(A_1) &= 0.3785, \\ \text{nRd}^{II}(A_2) &= 0.3070, \\ \text{nRd}^{II}(A_3) &= 0.2779, \\ \text{nRd}^{II}(A_4) &= 0.3959.\end{aligned}$$

In Step 2,  $\text{nRd}^{II}(A_3)$  is the least one, therefore  $A_3$  is removed. After normalization, we obtain the BBA  $m_S^{\text{IRd}}(\cdot)$  with iterative approximation as shown in Table XVI.

Table XVI  
 $m_S^{\text{IRd}}(\cdot)$  USING ITERATIVE APPROXIMATION BASED ON REDUNDANCY WITH THE DISTANCE  $\delta_{\cap}$ .

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.2959
$A'_2 = \{\theta_2, \theta_3\}$	0.4117
$A'_3 = \{\theta_3\}$	0.2924

- Using  $\delta_{\cup}$ , the distance matrix is

$$\text{Mat}_{\text{FE}} = \begin{matrix} & \begin{matrix} A_1 & A_2 & A_3 & A_4 & A_5 \end{matrix} \\ \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{matrix} & \begin{bmatrix} 0 & 0.4257 & 0.2322 & 0.5298 & 0.1780 \\ 0.4257 & 0 & 0.2322 & 0.1759 & 0.2477 \\ 0.2322 & 0.2322 & 0 & 0.4081 & 0.4644 \\ 0.5298 & 0.1759 & 0.4081 & 0 & 0.3518 \\ 0.1780 & 0.2477 & 0.4644 & 0.3518 & 0 \end{bmatrix} \end{matrix}$$

All focal elements' degrees of non-redundancy are

$$\begin{aligned}\text{nRd}(A_1) &= 0.3414, \\ \text{nRd}(A_2) &= 0.2705, \\ \text{nRd}(A_3) &= 0.3342, \\ \text{nRd}(A_4) &= 0.3664, \\ \text{nRd}(A_5) &= 0.3105.\end{aligned}$$

By using the batch mode method, the focal elements  $A_2$  and  $A_5$  are removed. After applying the normalization, we obtain the approximated BBA as shown in Table XVII.

Table XVII  
 $m_S^{\text{BRd}}(\cdot)$  USING BATCH APPROXIMATION BASED ON REDUNDANCY WITH THE DISTANCE  $\delta_{\cup}$ .

Focal element	Mass value
$A'_1 = \{\theta_1, \theta_2\}$	0.3037
$A'_2 = \{\theta_2\}$	0.3962
$A'_3 = \{\theta_3\}$	0.3001

Using the iterative mode method, degrees of non-redundancy obtained in Step 1 are

$$\begin{aligned}\text{nRd}^I(A_1) &= 0.3414, \\ \text{nRd}^I(A_2) &= 0.2705, \\ \text{nRd}^I(A_3) &= 0.3342, \\ \text{nRd}^I(A_4) &= 0.3664, \\ \text{nRd}^I(A_5) &= 0.3105.\end{aligned}$$

The focal element  $A_2$  is removed first, since it has the smallest nRd value. Then recalculate all nRd values for remaining focal elements  $A_1, A_3, A_4$  and  $A_5$ :

$$\begin{aligned}\text{nRd}^{II}(A_1) &= 0.3133, \\ \text{nRd}^{II}(A_2) &= 0.3682, \\ \text{nRd}^{II}(A_3) &= 0.4299, \\ \text{nRd}^{II}(A_4) &= 0.3314.\end{aligned}$$

In Step 2, the focal element  $A_1$  is removed, since  $\text{nRd}(A_1)$  is the smallest one. After normalization, we can obtain the BBA  $m_S^{\text{IRd}}(\cdot)$  as shown in Table XVIII.

Table XVIII  
 $m_S^{\text{IRd}}(\cdot)$  USING ITERATIVE APPROXIMATION BASED ON REDUNDANCY WITH THE DISTANCE  $\delta_{\cup}$ .

Focal element	Mass value
$A'_1 = \{\theta_2\}$	0.4043
$A'_2 = \{\theta_3\}$	0.3063
$A'_3 = \emptyset$	0.2894

As we can see in Example 5, the results of the batch mode and iterative mode approximations are different. In the next section, we provide experiments and simulations to evaluate our proposed BBA approximation approaches and those available ones.

#### IV. SIMULATIONS FOR EVALUATION

We use the computational cost caused by the evidence combination and the closeness between the approximated BBA and the original one in average to evaluate the performance of approximations. An approximation with less computational cost and larger closeness is desirable. To describe the closeness between BBAs, we use a strict distance of evidence, which is Jousselme's distance ( $d_J$ ) [34]. One can also use other types of strict distance in evidence theory e.g., belief interval based distance of evidence [35].

Suppose that  $m_1, m_2$  are two BBAs defined on  $\Theta$ , with  $|\Theta| = n$ . If  $m_1$  and  $m_2$  are considered as two vectors denoted by  $\mathbf{m}_1$  and  $\mathbf{m}_2$ , respectively, Jousselme's distance of evidence is defined as

$$d_J(\mathbf{m}_1, \mathbf{m}_2) \triangleq \sqrt{0.5(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{J} \mathbf{a} \mathbf{c} (\mathbf{m}_1 - \mathbf{m}_2)}, \quad (21)$$

where  $\mathbf{J} \mathbf{a} \mathbf{c}$  is the so-called Jaccard's weighting matrix whose elements  $J_{ij} = \text{Jac}(A_i, B_j)$  are defined by

$$\text{Jac}(A_i, B_j) = \frac{|A_i \cap B_j|}{|A_i \cup B_j|} \quad (22)$$

It is a most widely used distance of evidence, and it has been proven to be a strict distance metric [36].

Our simulation is based on a Monte Carlo simulation using  $M = 200$  random runs. In  $j$ -th simulation run, the original BBA to approximate  $m^j(\cdot)$  is randomly generated and the different approximation results  $\{m_{S_i}^j(\cdot)\}$  are obtained using the different approximations, where  $i$  denotes the  $i$ -th approximation approach. Here we use  $\oplus$  to denote the evidence combination. We calculate the computational time of the original evidence combination of  $m^j(\cdot) \oplus m^j(\cdot)$  with Dempster's rule, and the computation time of Dempster's combination of each approximated BBA  $m_{S_i}^j(\cdot) \oplus m_{S_i}^j(\cdot)$ . Here we compare our proposed approaches with  $k-l-x$  method ( $S_1$ ), D1 method ( $S_2$ ), Summarization method ( $S_3$ ), Denœux's outer approximation ( $S_4$ ), the rank-level fusion based approximation ( $S_5$ ), and our new degree of non-redundancy based approximations including the batch mode with  $\delta_{\cap}$  ( $S_6$ ), iterative mode with  $\delta_{\cap}$  ( $S_7$ ), batch mode with  $\delta_{\cup}$  ( $S_8$ ), iterative mode with  $\delta_{\cup}$  ( $S_9$ ), and CR-based approximation ( $S_{10}$ ) since all these methods can set the quantity of the remaining focal elements, and they never consider the empty set as a valid focal element (contrarily to inner approximation which will bring troubles for making the comparisons because Jousselme's distance cannot be computed if one allows to put positive mass on empty set because  $|\emptyset| = 0$ ).

In our simulations, the cardinality of the FOD  $\Theta$  is 4. In each random generation, there are  $2^4 - 1 = 15$  focal elements in the original BBA. The number of remaining focal elements for all the approaches used here is set to from 14 down to 2. We randomly generate BBA using Algorithm 1 [37] in Table XIX below.

Table XIX  
ALGORITHM 1: RANDOM GENERATION OF BBA.

Random generation of BBA
<b>Input:</b> $\Theta$ : Frame of discernment; $N_{\max}$ : Maximum number of focal elements
<b>Output:</b> $m$ BBA
Generate $\mathcal{P}(\Theta)$ , which is the power set of $\Theta$ ;
Generate a random permutation of $\mathcal{P}(\Theta) \rightarrow \mathcal{R}(\Theta)$ ;
Generate an integer between 1 and $N_{\max} \rightarrow l$ .
<b>FOR each:</b> First $k$ elements of $\mathcal{R}(\Theta)$ do
Generate a value within $[0, 1] \rightarrow m_i(\cdot)$ , ( $i = 1, 2, \dots, l$ );
<b>END</b>
Normalize the vector $m = [m_1, m_2, \dots, m_l] \rightarrow m'$ ;
$m(A_i) = m'_i$

The average (over 200 runs) combination time and average (over 200 runs) distance values ( $d_j$ ) between the original BBA and the approximated BBA's obtained using different approaches given different remaining focal elements' numbers are shown in Figs. 4 and 5, respectively.

The average (over all runs and all numbers of remaining focal elements) computation time and distance values are shown in Table XX.

Note that the computer for the experiments is with i7-8550CPU, 16 GB LPDDR3 RAM, WINDOWS 10 OS and MATLAB 2013B.

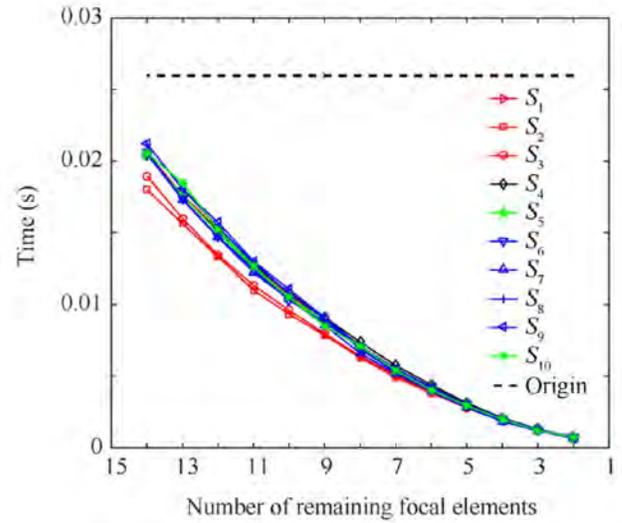


Figure 4. Comparisons between different approximations in terms of computation time.

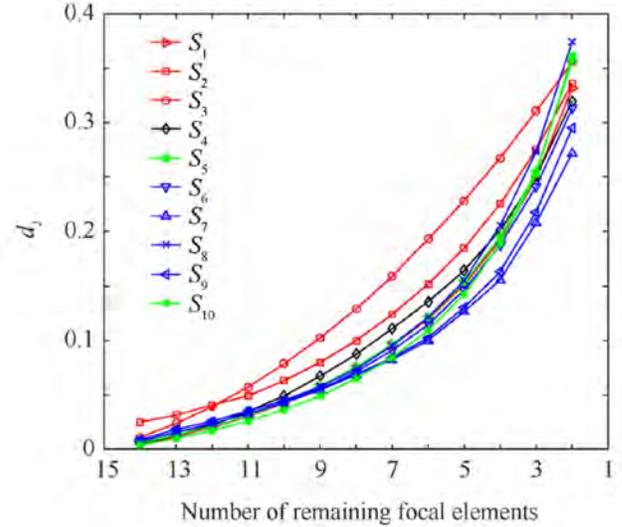


Figure 5. Comparisons between different approximations in terms of  $d_j$ .

Table XX  
COMPARISONS BETWEEN DIFFERENT BBA APPROXIMATIONS IN TERMS OF COMBINATION TIME AND CLOSENESS.

Approach	Time (s)	$d_j$
Original BBA	0.0260	0
$k-l-x$ ( $S_1$ )	0.0085	0.1072
D1 ( $S_2$ )	0.0074	0.1284
Sum ( $S_3$ )	0.0077	0.1512
Outer ( $S_4$ )	0.0086	0.1143
Rank-level ( $S_5$ )	0.0079	0.1104
Batch $\delta_{\cap}$ ( $S_6$ )	0.0083	0.1027
Iterative $\delta_{\cap}$ ( $S_7$ )	0.0082	0.0926
Batch $\delta_{\cup}$ ( $S_8$ )	0.0085	0.1139
Iterative $\delta_{\cup}$ ( $S_9$ )	0.0086	0.0973
CR-based ( $S_{10}$ )	0.0084	0.1031

It can be shown from Table XX, Figs. 4 and 5 that for all the approximations compared here including our proposed four types of approximations based on degree of redundancy for focal elements, the computational time are significantly reduced when compared with original computation time. At the same time, our focal element redundancy based approximations have smaller distance (less loss of information) according to all the distances of evidence used here. In our four new approximations, the iterative mode with  $\delta_{\cap}$  performs the best.

Here we also provide the comparisons of computational cost of different approximation approaches themselves. To obtain the approximation computation time, in each run for different approximation approaches, the average time of approximation with remaining focal elements numbers from 2 to 14 is calculated. Then, each approximation approach’s averaging computation time over 200 runs is listed in Table XXI. The computational complexity of each approach listed is also listed in Table XXI.

Table XXI  
COMPARISON BETWEEN DIFFERENT BBA APPROXIMATIONS IN TERMS OF COMBINATION COMPLEXITY.

Approach	Aver. comp. time (s)	Complexity
$k-l-x$ ( $S_1$ )	0.00075	$O(n \log n)$
D1 ( $S_2$ )	0.00052	$O(n)$
Sum ( $S_3$ )	0.00056	$O(n)$
Outer ( $S_4$ )	0.00150	$O(n^2 + n^2 \log n)$
Rank-level ( $S_5$ )	0.00079	$O(n \log n)$
Batch $\delta_{\cap}$ ( $S_6$ )	0.00110	$O(n^2 + n^2 \log n)$
Iterative $\delta_{\cap}$ ( $S_7$ )	0.00540	$O((n-k)(n^2 + n^2 \log n))$
Batch $\delta_{\cup}$ ( $S_8$ )	0.00100	$O(n^2 + n^2 \log n)$
Iterative $\delta_{\cup}$ ( $S_9$ )	0.00500	$O((n-k)(n^2 + n^2 \log n))$
CR-based ( $S_{10}$ )	1.78150	$O((2n)^{n-k})$

As shown in Fig. 5, the approximated BBA obtained using CR-based method can have smaller distance to the original BBA when the number of remaining focal elements are not so small (from 14 down to 9). However, it is at the price of computational cost. Its computation time is about  $10^2$  times of other approaches compared.

CR-based method use a way like the traversal when selecting the focal elements to remove. Actually, it is not a real traversal, since it removes the  $L-k$  focal elements in a batch, but not one by one. Therefore, when the remaining focal elements number is small, its distance becomes not so small.

Comparatively, according to the experimental results, our proposed approximation approach can achieve smaller distance and at the same time, its time cost is accepted.

Note that with the improvement of the computer’s computing capability, the importance of the mass function approximation will be decreased. However, there still exists some resource-restricted environment or platforms, for example, the embedded system for real time tasks, where the computational resource including the CPU and the RAM are not so adequate and the approximation, which can save computational time, is still important.

On the other hand, the BBA approximation could be considered as a preprocessing of “data”, which can reduce the computational cost. Even if the computational resource is enough, to further reduce the computational cost is still desirable, especially for those real-time applications.

Note that our current performance evaluation on different approximation approaches is based on the experimental results in terms of the statistical averaging combination computational time, and the distance between the approximated BBA and the original one. This makes sense from the engineering or application viewpoints. To comprehensively evaluate different approximation approaches, theoretical analysis and proof are needed, which is also one of the research focuses in our work in the future.

## V. CONCLUSION

Novel methods for BBA approximations are proposed in this paper, where the most redundant focal elements are removed at first. The degree of non-redundancy is defined based on distance between focal elements. Batch and iterative implementations of the BBA approximations are provided. It is experimentally shown that our new BBA approximations can reduce the computational cost of evidence combination with less loss of information, which is described by the distance of evidence. At the same time, the computation time of approximations in our proposed approaches is acceptable.

In our future work, we will focus on designing more comprehensive and rational distance of focal elements, based on which, the degree of focal elements can be calculated. In fact, the non-redundancy represents a type of “importance” for focal elements. We will also try to define some new type of “importance”, based on which the removal of focal elements can be done more rationally executed. As shown in this paper, we evaluate the performance of different BBA approximations using the computation time and the distance of evidence. In future work, we will also explore more comprehensive evaluation criteria and theoretical evaluation in mathematics for the BBA approximation approaches. This is crucial for the design of more effective BBA approximations.

When we use some criterion (e.g., the non-redundancy proposed in this paper) to determine those “unimportant” or “redundant” focal elements, we can combine these focal elements to a new one (with intersection or union operation of these elements) besides removing them. For example, we can combine the most two redundant focal elements to a new focal element by using the operations like intersection, union and other ways to replace the current the removal of redundant focal elements. Furthermore, we can use the method like PCA in the design of BBA approximations for the combination of focal elements to expect a better approximation performance in the future research work.

## ACKNOWLEDGMENT

The authors would like to thank the support on this research by the National Natural Science Foundation of China (Nos. 61671370, 61573275), Postdoctoral Science

Foundation of China (No. 2016M592790), Postdoctoral Science Research Foundation of Shaanxi Province, China (No. 2016BSHEDZZ46) and Fundamental Research Funds for the Central Universities, China (No. xjj201066).

## REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, USA, 1976.
- [2] G. Lin, J. Liang, Y. Qian, *An information fusion approach by combining multigranulation rough sets and evidence theory*, Information Sciences, Vol. 314, pp. 184–199, Sept. 2015.
- [3] G. Dong, G. Kuang, *Target recognition via information aggregation through Dempster-Shafer's evidence theory*, IEEE Geoscience Remote Sensing Letters, Vol. 12(6), pp. 1247–1251, June 2015.
- [4] D. Han, J. Dezert, J.-M. Tacnet, C. Han, *A fuzzy cautious OWA approach with evidential reasoning*, in Proc. of 15th Int. Conf. on Information Fusion (Fusion 2012), Singapore, Sep 7–12, 2012 (Piscataway: IEEE Press 2012, pp. 278–285).
- [5] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, Int. Journal of Intelligent Systems, Vol. 29(3), pp. 223–252, 2014.
- [6] P. Smets, *Practical uses of belief functions*, in Proc. of 15th Int. Conf. on Uncertainty in Artificial Intelligence, Stockholm, Sweden, Jul 30–Aug 1, 1999 (San Francisco: Morgan Kaufmann 1999, pp. 612–21).
- [7] R. Kennes, *Computational aspects of the Möbius transformation of graphs*, IEEE Trans. on Systems, Man and Cybernetics, Vol. 22(2), pp. 201–223, 1992.
- [8] J.A. Barnett, *Computational methods for a mathematical theory of evidence*, in Proc. of 7th Int. Joint Conf. on Artificial Intelligence (IJCAI-81), Vancouver, Canada, Aug. 24–28, 1981 (New York: Springer, 1981, pp. 868–875).
- [9] G. Shafer, R. Logan, *Implementing Dempster's rule for hierarchical evidence*, Artificial Intelligence, Vol. 33(3), pp. 271–298, 1987.
- [10] F. Voorbraak, *A computationally efficient approximation of Dempster-Shafer theory*, Int. J. of Man-Machine Studies, Vol. 30(5), pp. 525–536, 1989.
- [11] F. Cuzzolin, *The geometry of consonant belief functions: Simplicial complexes of necessity measures*, Fuzzy Sets and Systems, Vol. 161(10), pp. 1459–1479, May 2010.
- [12] B. Tessem, *Approximations for efficient computation in the theory of evidence*, Artificial Intelligence, Vol. 61(2), pp. 315–329, 1993.
- [13] J.D. Lowrance, D.G. Thomas, M.S. Thomas, *A framework for evidential-reasoning systems*, in Proc. of the 5th National Conference on Artificial Intelligence, Philadelphia, PA, USA, Aug 11–15, 1986 (Palo Alto: AAAI Press 1986, pp. 896–903).
- [14] M. Bauer, *Approximations for decision making in the Dempster-Shafer theory of evidence*, in Proc. of 12th Int. Conf. on Uncertainty in Artificial Intelligence (UAI 1996), Portland, ME, USA, August 1–4, 1996 (San Francisco: Morgan Kaufmann 1996, pp. 73–80).
- [15] T. Denœux, *Inner and outer approximation of belief structures using a hierarchical clustering approach*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 9(4), pp. 437–460, 2001.
- [16] S. Moral, A. Salmerón, *A Monte Carlo algorithm for combining Dempster-Shafer belief based on approximate pre-computation*, In: A. Hunter, S. Parsons, Editors, in Proc. of 5th European Conf. on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU 1999), London, UK, July 6–9, 1999 (Berlin: Springer 1999, pp. 305–315).
- [17] J. Dezert, D. Han, Z. Liu, J.-M. Tacnet, *Hierarchical proportional redistribution for BBA approximation*, in Proc. of 2nd Int. Conf. on Belief Functions (Belief 2012), Compiègne, France, May 9–11, 2012 (New York: Springer 2012, pp. 275–283).
- [18] Y. Yang, D. Han, C. Han, F. Cao, *A novel approximation of basic probability assignment based on rank-level fusion*, Chinese J. of Aeronautics, Vol. 26(4), pp. 993–999, August 2013.
- [19] D. Han, J. Dezert, C. Han, *New basic belief assignment approximations based on optimization*, in Proc. of 15th Int. Conf. on Information Fusion (Fusion 2012), Singapore, Sept. 7–12, 2012 (Piscataway, NJ: IEEE Press 2012, pp. 286–293).
- [20] S.H. Shou, X.Y. Deng, X. Liu, H.Q. Zheng, J. Wen, *Approximation of basic probability assignment in Dempster-shafer theory based on correlation coefficient*, in Proc. of 20th Int. Conf. on Information Fusion (Fusion 2017), Xi'an, China, July 10–13, 2017 (Piscataway: IEEE Press 2017, pp. 1–7).
- [21] D. Han, Y. Yang, J. Dezert, *Two novel methods for BBA approximation based on focal element redundancy*, in Proc. of 18th Int. Conf. on Information Fusion (Fusion 2015), Washington DC, USA, July 6–9, 2015 (Piscataway: IEEE Press 2015, pp. 428–434).
- [22] F. Smarandache, J. Dezert (Editors), *Applications and advances of DSMT for information fusion*, Vol. 4, Rehoboth: American Research Press, 2015.
- [23] F. Smarandache, J. Dezert (Editors), *Applications and advances of DSMT for information fusion*, Vol. 3, Rehoboth: American Research Press, 2009.
- [24] T. Burger, *Defining new approximations of belief functions by means of Dempster's combination*, in Proc. of 1st Int. Conf. on Belief Functions (Belief 2010), Brest, France, April 1–2, 2010 (New York: Springer 2010, pp. 1–6).
- [25] J. Sudano, *The system probability information content (PIC) relationship to contributing components, combining independent multi-source beliefs, hybrid and pedigree pignistic probabilities*, in Proc. of 5th Int. Conf. on information fusion (Fusion 2002), Annapolis, MD, USA, July 8–11, 2002 (Piscataway: IEEE Press 2002, pp. 1277–1283).
- [26] J. Sudano, *Belief fusion, pignistic probabilities, and information content in fusing tracking attributes*, in Proc. of 2004 IEEE Radar Conf., Philadelphia, PA, USA, April 26–29, 2004 (Piscataway: IEEE Press 2004, pp. 218–224).
- [27] J. Sudano, *Yet another paradigm illustrating evidence fusion (YAPIEF)*, in Proc. of 9th Int. Conf. on Information Fusion (Fusion 2006), Florence, Italy, July 10–13, 2006 (Piscataway: IEEE Press 2006, pp. 1–7).
- [28] F. Cuzzolin, *On the properties of the intersection probability*, in Proc. of the 10th European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2007), Hammamet, Tunisia, Oct. 31–Nov. 2, 2007 (Berlin: Springer 2007, pp. 287–298).
- [29] J. Dezert, D. Han, Z. Liu Z, J.-M. Tacnet, *Hierarchical DSMP transformation for decision-making under uncertainty*, in Proc. of 15th Int. Conf. on Information Fusion (Fusion 2012), Singapore, July 9–12, 2012 (Piscataway: IEEE Press 2012, pp. 294–301).
- [30] D. Han, J. Dezert, Z. Duan, *Evaluation of probability transformations of belief functions for decision making*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 46(1), pp. 93–108, 2016.
- [31] D. Dubois, H. Prade, *Consonant approximations of belief functions*, Int. J. of Approximate Reasoning, Vol. 4(5), pp. 419–449, 1990.
- [32] M. Grabisch, *Upper approximation of non-additive measures by k-additive measures - the case of belief functions*, in Proc. of the 1st Int. Symposium on Imprecise Probabilities and Their Applications (ISIPTA), Gent, Belgium, June 29–July 2, 1999 (Boston: Springer 1999, pp. 158–164).
- [33] A.M. Erkmén, H.E. Stephanou, *Information fractals for evidential pattern classification*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 20(5), pp. 1103–1114, 1990.
- [34] A.-L. Jousselme, D. Grenier, E. Bossé, *A new distance between two bodies of evidence*, Information Fusion, Vol. 2(2), pp. 91–101, 2001.
- [35] D. Han, J. Dezert, Y. Yang, *Belief interval-based distance measures in the theory of belief functions*, IEEE Trans. on Systems, Man, and Cybernetics: Systems, Vol. 48(6), pp. 833–850, 2018.
- [36] M. Bouchard, A.-L. Jousselme, P.-E. Doré, *A proof for the positive definiteness of the Jaccard index matrix*, Int. J. of Approximate Reasoning, Vol. 54(5), pp. 615–626, 2013.
- [37] A.-L. Jousselme, P. Maupin, *Distances in evidence theory: Comprehensive survey and generalizations*, Int. J. of Approximate Reasoning, Vol. 53(2), pp. 118–145, 2012.



# Approximation of Basic Belief Assignment Based on Focal Element Compatibility

Qing Wei<sup>a</sup>, Xinde Li<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Key Laboratory of Measurement and Control of CSE,  
School of Automation, Southeast University, Nanjing, China.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

Emails: qingwei@seu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr

Originally published as: Q. Wei, X. Li, J. Dezert, *Approximation of Basic Belief Assignment Based on Focal Element Compatibility*, in Proc. of Int. Conf. on Information Fusion (Fusion 2019), Ottawa, Canada, July 2-5, 2019, and reprinted with permission.

**Abstract**—The theory of belief functions is an important tool in the field of information fusion. However, the fusion of Basic Belief Assignments (BBAs) requires high computational cost and long computing time when a large number of focal elements are involved in the fusion rules. This problem becomes a bottleneck of application of Belief Functions (BF) in high-dimensional real problems. To overcome this drawback, many approaches were proposed to approximate BBAs to reduce the computational complexity in the fusion process. In this paper, we present a novel method based on the compatibility of focal elements to approximate a BBA by removing some focal elements of the original BBA. Besides, a new mass assignment strategy based on the distance of focal elements is proposed. Several examples, simulations and related analyses are provided to illustrate the interest and efficiency of the proposed method.

**Keywords**—Information fusion, Belief functions, Basic belief assignment, Approximation

## I. INTRODUCTION

The evidence theory was proposed by Dempster in the study of multivalued mapping in 1967 [1] and later promoted by Shafer in 1976 [2] with the introduction of Belief Functions (BF). The theory of belief functions is named also Dempster-Shafer Theory (DST) in the literature. Belief Functions provide an effective method for dealing with the expression and synthesis of uncertain information and they have been widely used in many fields such as image processing [3, 4], target tracking [5], and fault diagnosis [6, 7].

However, the evidence combination will encounter high computational cost when the frame of discernment (FoD) is large. To overcome this drawback, one effective approach to reduce the computational complexity is the BBA approximation. The BBA approximation aims to obtain a simpler BBA by removing some focal elements according to different simplification criteria. In existing works, the simplification criteria can be divided into the following three categories:

- 1) **Simplification based on the mass assignment of a focal element.** The focal elements with smaller mass assignments are deemed unimportant, which should be removed firstly.  $k-l-x$  [8], Summarization [9] and D1 [10] are representatives of this criterion.
- 2) **Simplification based on the cardinality of a focal element.** The focal elements with larger cardinalities

may cause more computational cost.  $k$ -additive approach [11] and hierarchical proportional redistribution approach [12] accomplish the simplification according to this criterion.

- 3) **Hybrid simplification mixing the two previous ones.** Use the previous two criteria jointly to determine which focal elements should be removed at first. Methods like inner and outer approximation [13], rank-level fusion approximation [14], non-redundancy approximation [15], iterative approximation based on distance of evidence [16] and correlation coefficient approximation [17] enter in this hybrid simplification strategy.

In general, the hybrid simplification is the right direction to approximate a BBA due to the one-sidedness of the first and the second simplification criterion.

In this paper, we propose a novel approach using the notion of focal element compatibility. In our method, each focal element has a compatible focal element which can be replaced by it due to the compatibility (based on a similarity measure) between them. To quantify the notion of compatibility, we use the mass value and the cardinality of the set which contains all the focal elements which can replace the given focal element jointly. The focal element with the highest degree of compatibility should be removed at first. Users can preset the number of remaining focal elements. After removing a focal element, the removed mass is redistributed to remaining focal elements to execute the next iteration according to our new mass assignment strategy. Experimental results based on the comparisons with other approximation strategies and related analyses justify that our approach is rational and effective.

This paper is organized as follows. After brief preliminaries on Belief Functions in Section II and classical BBA approximation methods in Section III, we will present the new approximation method based on focal element compatibility in Section IV. Evaluation of it and comparative analysis will be done in Section V with concluding remarks in Section VI.

## II. PRELIMINARIES

### A. Basics of Belief Functions

We consider a frame of discernment (FoD)  $\Theta = \{\theta_1, \dots, \theta_n\}$  whose elements are mutually exclusive and exhaustive. A basic

belief assignment (BBA) over the FoD  $\Theta$  is defined as

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad m(\emptyset) = 0 \quad (1)$$

If  $m(A) > 0$  holds,  $A$  is called a Focal Element (FE). The belief function and plausibility function are defined as follows [2].

$$Bel(A) = \sum_{B \subseteq A} m(B); \quad Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad (2)$$

In DST, two independent bodies of evidence (BOEs) are combined by Dempster's rule as follows.  $\forall A \in 2^\Theta$ :

$$m(A) = \begin{cases} 0, & A = \emptyset \\ \frac{1}{1-K} \sum_{A_i \cap B_j = A} m_1(A_i)m_2(B_j), & A \neq \emptyset \end{cases} \quad (3)$$

where  $K = \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)$  is the conflict coefficient, which represents the total degree of conflict. Other rules of combinations have also been proposed to combine BBAs in the literature [18] but they will be not detailed in this paper since this is out of its scope.

### B. Distance of Focal Elements

We use the definition proposed by Denœux [13] to measure the distance between two focal elements, which is defined as

$$\delta_{\cap}(A_i, A_j) = m(A_i)|A_i| + m(A_j)|A_j| - [m(A_i) + m(A_j)]|A_i \cap A_j| \quad (4)$$

For a given focal element  $A_i$ , if  $\delta_{\cap}(A_i, A_j) = \min_{j' \neq i} \delta_{\cap}(A_i, A_{j'})$ , we will say that  $A_j$  has the highest compatibility degree with  $A_i$ , and  $A_j$  shares the most similar information with  $A_i$ .

## III. BRIEF REVIEW OF BBA APPROXIMATIONS

Some existing BBA approximation approaches are briefly reviewed in this section for the purpose of comparisons with our new method.

1)  **$k-l-x$  approximation [8]**. This method involves three parameters and the approximated BBA is obtained by

- keeping no less than  $k$  focal elements;
- keeping no more than  $l$  focal elements;
- deleting the masses which are no greater than  $x$ .

In  $k-l-x$  algorithm, all original focal elements are sorted according to the mass assignments in a decreasing order. Then, the first  $p$  focal elements are selected such that  $k \leq p \leq l$  and such that the sum of the mass assignments of these  $p$  focal elements is no less than  $1-x$ . The removed mass assignments are redistributed to remaining focal elements by a classical normalization procedure.

2) **Summarization approximation [9]**. This method also keeps focal elements having largest mass values which is similar to the  $k-l-x$  method. The only difference is that the removed mass values are redistributed to their union set. Suppose that  $m(\cdot)$  is the original BBA and  $k$  is the desired number of remaining focal elements in the approximated BBA

$\hat{m}(\cdot)$ . Let  $M$  denote the set of  $k-1$  focal elements with largest mass values in  $m(\cdot)$ . Then  $\hat{m}(\cdot)$  is obtained from  $m(\cdot)$  by

$$\hat{m}(A) = \begin{cases} m(A), & A \in M \\ \sum_{A' \subseteq A, A' \notin M} m(A'), & A = A_0 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where  $A_0$  is

$$A_0 \triangleq \bigcup_{A' \notin M, m(A') > 0} A' \quad (6)$$

3) **D1 approximation [10]**. Suppose that  $m(\cdot)$  is the original BBA and  $k$  is the desired number of remaining focal elements in the approximated BBA  $\hat{m}(\cdot)$ . Let  $M$  denote the set of  $k-1$  focal elements with largest mass values in  $m(\cdot)$  and  $M^-$  be the set including all the other focal elements of  $m(\cdot)$ . D1 method is to keep all the members of  $M$  as the focal elements of  $\hat{m}(\cdot)$  and to assign the mass values of the focal elements in  $M^-$  among the focal elements in  $M$  according to the following procedure.

For a focal element  $A \in M^-$ , in  $M$ , find all the supersets of  $A$  to construct a collection  $M_A$ . If  $M_A$  is not empty, the mass value of  $A$  is uniformly assigned among the focal elements having smallest cardinality in  $M_A$ . When  $M_A$  is empty, then construct  $M'_A$  as

$$M'_A = \{B \in M \mid |B| \geq |A|, B \cap A \neq \emptyset\} \quad (7)$$

Then, if  $M'_A$  is not empty,  $m(A)$  is assigned among the focal elements with smallest cardinality in  $M'_A$ . The value assigned to a focal element  $B$  depends on the value of  $|B \cap A|$ . Such a procedure is iteratively executed until all  $m(A)$  have been assigned to the focal elements in  $M$ .

If  $M'_A$  is empty, there are two possible cases:

- If the total set  $\Theta \in M$ , the sum of mass values of the focal elements in  $M^-$  will be added to  $\Theta$ ;
- If  $\Theta \notin M$ , then let  $\Theta$  be a focal element of  $\hat{m}(\cdot)$  and assign the sum of mass values of the focal elements in  $M^-$  to  $\hat{m}(\Theta)$ .

Note that the number of remaining focal elements is  $k-1$ , if  $\Theta \in M$ .

4) **Rank-level fusion approximation [14]**. This method uses jointly the mass assignments and cardinalities of focal elements to make the simplification. The specific procedure is listed as follows.

- Sort all the focal elements of the original BBA (with  $L$  focal elements) according to the mass assignments (in ascending order which is due to the assumption that the focal element with smallest mass should be removed at first). The rank vector obtained is

$$r_m = [r_m(1), r_m(2), \dots, r_m(L)] \quad (8)$$

- Sort all the focal elements of the original BBA according to the cardinalities (in descending order which is due to the assumption that the focal element with large

cardinality should be removed at first). The rank vector obtained is

$$r_c = [r_c(1), r_c(2), \dots, r_c(L)] \quad (9)$$

- Execute the rank-level fusion and the comprehensive rank vector is

$$r_f = [r_f(1), r_f(2), \dots, r_f(L)] \quad (10)$$

where

$$r_f(i) = \alpha \cdot r_m(i) + (1 - \alpha) \cdot r_c(i) \quad (11)$$

The parameter  $\alpha \in [0, 1]$  is to weight the two different criteria. Finally, we remove the focal element with the smallest  $r_f$  value and do the renormalization of remaining focal elements. Repeat the above steps until only  $k$  focal elements remain and the total mass assignments value to be deleted is no greater than  $x$ .

5) **Correlation coefficient approximation [17]**. The correlation coefficient proposed by Jiang [19] can measure the similarity between two BBAs. In this approximation approach, we remove a focal element  $A_i$  from the original BBA  $m(\cdot)$  and the mass of  $A_i$  is redistributed to remaining focal elements to generate a new BBA  $\hat{m}_i(\cdot)$ . Then, we calculate the correlation coefficient between  $m$  and  $\hat{m}_i$ . We perform the same operation for each focal element and sort all the focal elements in ascending order according to the correlation coefficient. Finally, we remove the largest  $k$  focal elements from the original BBA and do the normalization according to a new assignment strategy.

6) **Iterative approximation based on distance of evidence [16]**. In this algorithm, we remove at first a focal element  $A_i$  from the original BBA  $m(\cdot)$  and we normalize the remaining focal elements to generate a new BBA  $\hat{m}_i(\cdot)$ . Then, we calculate Jousselme's distance between  $m$  and  $\hat{m}_i$ . We perform the same operation for each focal element. Finally, we remove the focal element which generates the new BBA having the closest distance with the original BBA and after a normalization we proceed the next iteration. The above steps are performed iteratively until only  $k$  focal elements remain.

#### IV. NEW BBA APPROXIMATION BASED ON FOCAL ELEMENT COMPATIBILITY

In this section, a novel method for approximating a BBA is proposed. As briefly shown in the previous section, the existing approaches remove some focal elements according to the mass assignment, the cardinality or both two criteria. Here we adopt a different standpoint in which a specific focal element can be removed if there exists a number of other focal elements compatible with it, i.e., its degree of incompatibility is small. Now the focus is how to define the degree of incompatibility of a focal element. We define the incompatibility degree for a focal element at first.

##### A. Degree of Incompatibility of Focal Elements

As mentioned before, the distance between two focal elements is given by Eq.(4). The compatible focal element  $A_i^C$ <sup>1</sup>

<sup>1</sup>We use the notation "C" as the upper index because it is the first letter of word "Compatible".

of a given focal element  $A_i \subseteq \Theta$  for a BBA  $m(\cdot)$  (with  $l$  focal elements) is defined by

$$A_i^C \triangleq \arg \min_{A_j} \delta_{\cap}(A_i, A_j) \quad (12)$$

$$s.t. \begin{cases} A_j \subseteq \Theta \\ j = 1, 2, \dots, l, j \neq i \end{cases}$$

$A_i^C$  has the smallest distance with the focal element  $A_i$ , i.e., among all focal elements,  $A_i^C$  is the most compatible with  $A_i$ . It should be noted that  $A_i^C$  can be replaced by  $A_i$ , but the reverse may not be true.

We define the degree of incompatibility of the focal element  $A_i$  by

$$ICP(A_i) \triangleq \begin{cases} \frac{m(A_i)}{|M_i^C|}, & M_i^C \neq \emptyset \\ \infty, & M_i^C = \emptyset \end{cases} \quad (13)$$

where

$$M_i^C = \{A_j | A_j^C = A_i, j = 1, 2, \dots, l, j \neq i\} \quad (14)$$

The set  $M_i^C$  contains all the focal elements which can replace  $A_i$ . The  $ICP(A_i)$  value describes the average effect on the  $|M_i^C|$  ( $M_i^C \neq \emptyset$ ) focal elements after removing  $A_i$ . The smaller  $ICP(A_i)$  value, the smaller the effect, which is preferred. From another perspective, the effect can be explained as the incompatibility degree of  $A_i$ . The smaller the effect, the smaller the incompatibility degree and the more it can be removed.  $M_i^C = \emptyset$  means that no focal elements can replace  $A_i$ , so its degree of incompatibility is infinite.

Here we provide a simple example to show how  $M_i^C$  and  $ICP(A_i)$  are computed.

**Example 1:** Consider the BBA  $m(\cdot)$  defined over the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The mass assignments of focal elements  $A_1 = \{\theta_1\}$ ,  $A_2 = \{\theta_2\}$ ,  $A_3 = \{\theta_2, \theta_3\}$  and  $A_4 = \{\theta_1, \theta_2, \theta_3\}$  are as follows.

$$m(A_1) = 0.5, m(A_2) = 0.28$$

$$m(A_3) = 0.17, m(A_4) = 0.05$$

- 1) We calculate the distance between any two focal elements and find the compatible focal element for each focal element.

$$\delta_{\cap}(A_1, A_2) = 0.78, \delta_{\cap}(A_1, A_3) = 0.84$$

$$\delta_{\cap}(A_1, A_4) = 0.1, \delta_{\cap}(A_2, A_3) = 0.17$$

$$\delta_{\cap}(A_2, A_4) = 0.1, \delta_{\cap}(A_3, A_4) = 0.05$$

$$A_1^C = A_2^C = A_3^C = A_4, \quad A_4^C = A_3$$

- 2) We compute  $M_i^C$  for each focal element.

$$M_1^C = M_2^C = \emptyset$$

$$M_3^C = \{A_4\}, M_4^C = \{A_1, A_2, A_3\}$$

- 3) We compute  $ICP(A_i)$  for each focal element.

$$ICP(A_1) = ICP(A_2) = \infty$$

$$ICP(A_3) = \frac{m(A_3)}{|M_3^C|} = \frac{0.17}{1} = 0.17$$

$$ICP(A_4) = \frac{m(A_4)}{|M_4^C|} = \frac{0.05}{3} = \mathbf{0.0167}$$

So,  $A_4 = \{\theta_1, \theta_2, \theta_3\}$  should be removed at first when approximating the original BBA  $m(\cdot)$ .

### B. New Mass Assignment Strategy

Here, we propose a new mass assignment strategy based on distance of focal elements. Let  $m(\cdot)$  denote the original BBA with  $l$  focal elements and  $\hat{m}(\cdot)$  denote the remaining BBA after removing the focal element  $A_r$ , where  $A'_i, i = 1, 2, \dots, l-1$  are the focal elements of  $\hat{m}(\cdot)$ . Then  $\hat{m}(\cdot)$  is obtained by

$$\hat{m}(A'_i) = \begin{cases} m(A'_i) + \frac{m(A_r)}{D \cdot \delta_{\cap}(A'_i, A_r)}, & A'_i \neq \emptyset \\ 0, & A'_i = \emptyset \end{cases} \quad (15)$$

where

$$D = \sum_{i=1}^{l-1} \frac{1}{\delta_{\cap}(A'_i, A_r)}, \quad A'_i \neq \emptyset \quad (16)$$

The proof that  $\hat{m}(\cdot)$  is a true normalized BBA is given in Appendix.

From Eq.(15) and (16), we can see that the mass of each removed focal element  $A_r$  is redistributed to remaining focal elements  $A_j$  according to their distances to  $A_r$ . The smaller the distance, the more mass is committed to  $A_j$ . Based on the compatibility of the focal elements and the new mass assignment strategy, we propose a novel BBA approximation approach described in the next subsection.

### C. New BBA Approximation Algorithm

Let  $m(\cdot)$  denote the original BBA with  $l$  focal elements. In the approximation, we want to keep  $k$  ( $k < l$ ) focal elements and remove the focal elements one by one iteratively. The detailed steps of this new BBA approximation method are as follows.

- **Step 1:** Calculate  $ICP(A_i)$  for each remaining focal element;
- **Step 2:** Sort all the focal elements in descending order according to their incompatibility degree to obtain the sorted list of focal elements;
- **Step 3:** Remove the last focal element  $A_r$  of the sorted list of focal elements, and redistribute its mass value to the mass of focal elements upper it in the sorted list to generate an approximated BBA  $\hat{m}$  according to our new mass assignment strategy. Reduce the number of focal elements by one, i.e.,  $l \leftarrow l - 1$ ;
- **Step 4:** Assign  $m = \hat{m}$ . If the number of removed focal elements is not reached, go to Step 1, otherwise output  $m$  as the final approximated BBA.

The whole procedure is illustrated in Fig.1.

Here we provide an illustrative example to show how our approximation method works and we compare it with other methods.

**Example 2:** Consider the BBA  $m(\cdot)$  defined over the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5\}$  listed in Table I.

1)  $k - l - x$  approximation. Here  $k$  and  $l$  are set to 5.  $x$  is set to 0.2. The focal elements  $A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$  and  $A_7 = \{\theta_2, \theta_5\}$  are removed without violating the constraints

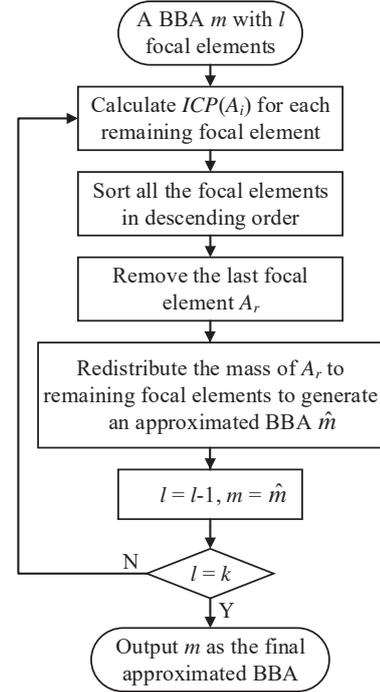


Fig. 1. Scheme of the new BBA approximation.

TABLE I  
FOCAL ELEMENTS AND MASS VALUES OF  $m(\cdot)$ .

Focal Elements	Mass Values
$A_1 = \{\theta_1\}$	0.13
$A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$	0.06
$A_3 = \{\theta_4, \theta_5\}$	0.3
$A_4 = \{\theta_3, \theta_5\}$	0.15
$A_5 = \{\theta_1, \theta_2\}$	0.14
$A_6 = \{\theta_2, \theta_4, \theta_5\}$	0.12
$A_7 = \{\theta_2, \theta_5\}$	0.1

in  $k - l - x$ . The remaining total mass value is  $1 - 0.06 - 0.1 = 0.84$ . Then, all the focal elements' mass values are divided by 0.84 to accomplish the normalization. The approximated BBA  $\hat{m}^{klx}(\cdot)$  is listed in Table II, where  $A'_i, i = 1, 2, 3, 4, 5$  are the focal elements of  $\hat{m}^{klx}(\cdot)$ .

TABLE II  
 $\hat{m}^{klx}(\cdot)$  OBTAINED USING  $k - l - x$ .

Focal Elements	Mass Values
$A'_1 = \{\theta_1\}$	0.1548
$A'_2 = \{\theta_4, \theta_5\}$	0.357
$A'_3 = \{\theta_3, \theta_5\}$	0.1786
$A'_4 = \{\theta_1, \theta_2\}$	0.1667
$A'_5 = \{\theta_2, \theta_4, \theta_5\}$	0.1429

2) Summarization approximation. Here  $k$  is set to 5. According to the summarization method, the focal elements

$A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$ ,  $A_7 = \{\theta_2, \theta_5\}$  and  $A_6 = \{\theta_2, \theta_4, \theta_5\}$  are removed and their union set  $\{\theta_2, \theta_3, \theta_4, \theta_5\}$  is generated as a new focal element (existed already) with mass value  $m(A_2) + m(A_7) + m(A_6) = 0.28$ . The approximated BBA  $\hat{m}^{Sum}(\cdot)$  is listed in Table III.

TABLE III  
 $\hat{m}^{Sum}(\cdot)$  OBTAINED USING SUMMARIZATION.

Focal Elements	Mass Values
$A'_1 = \{\theta_1\}$	0.13
$A'_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$	0.28
$A'_3 = \{\theta_4, \theta_5\}$	0.3
$A'_4 = \{\theta_3, \theta_5\}$	0.15
$A'_5 = \{\theta_1, \theta_2\}$	0.14

3) D1 approximation. Here  $k$  is set to 5. It can be obtained that  $A_3, A_4, A_5, A_1$  belong to  $M$ , and  $A_6, A_7, A_2$  belong to  $M^-$ . For  $A_6$  and  $A_2$ , there are no supersets of them in  $M$ , i.e.,  $M_A = \emptyset$ , and we can not construct the set  $M'_A$ , i.e.,  $M'_A = \emptyset$ . So the mass values of  $A_6$  and  $A_2$  are assigned to the total set  $\Theta$ . For  $A_7$ , we can construct the set  $M'_A = \{A_3, A_4, A_5\}$ . The parameter *ratio* and *number* are calculated to be 1 and 3. Therefore,  $m(A_7)/3 = 0.0333$  is added to the mass value of  $A_3, A_4$  and  $A_5$  respectively. The approximated BBA  $\hat{m}^{D1}(\cdot)$  is listed in Table IV.

TABLE IV  
 $\hat{m}^{D1}(\cdot)$  OBTAINED USING D1.

Focal Elements	Mass Values
$A'_1 = \{\theta_1\}$	0.13
$A'_2 = \{\theta_4, \theta_5\}$	0.3334
$A'_3 = \{\theta_3, \theta_5\}$	0.1833
$A'_4 = \{\theta_1, \theta_2\}$	0.1733
$A'_5 = \Theta$	0.18

4) Rank-level fusion approximation. Here  $k$  and  $l$  are set to 5 and  $x$  is 0.2. The parameter  $\alpha$  is set to 0.5. At the first iteration, we calculate the comprehensive vector  $r_f = [r_f(A_1), r_f(A_2), \dots, r_f(A_7)] = [5.5, 1, 5, 4.5, 4, 2.5, 2.5]$ . Then we remove  $A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$  at first and do the normalization of remaining focal elements. At the second iteration, we obtain the comprehensive vector  $r_f = [r_f(A_1), r_f(A_3), r_f(A_4), r_f(A_5), r_f(A_6), r_f(A_7)] = [4.5, 4, 3.5, 3, 1.5, 1.5]$ . Then, we remove  $A_6 = \{\theta_2, \theta_4, \theta_5\}$  (or  $A_7$ ) and normalize the remaining focal elements to obtain the final approximated BBA  $\hat{m}^{Rank}(\cdot)$  listed in Table V.

TABLE V  
 $\hat{m}^{Rank}(\cdot)$  OBTAINED USING RANK-LEVEL FUSION.

Focal Elements	Mass Values
$A'_1 = \{\theta_1\}$	0.1585
$A'_2 = \{\theta_4, \theta_5\}$	0.3659
$A'_3 = \{\theta_3, \theta_5\}$	0.1829
$A'_4 = \{\theta_1, \theta_2\}$	0.1707
$A'_5 = \{\theta_2, \theta_5\}$	0.122

5) Correlation coefficient approximation. Here  $k$  is set to 2, i.e., we have to remove two focal elements. The correlation

coefficients between the remaining BBA  $\hat{m}_i(\cdot), i = 1, 2, \dots, 7$  and the original BBA  $m(\cdot)$  are 0.9805, 0.9981, 0.9274, 0.9778, 0.9842, 0.9946 and 0.9927. We sort all the focal elements in ascending order according to the correlation coefficient and remove the two bottom focal elements  $A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$  and  $A_6 = \{\theta_2, \theta_4, \theta_5\}$  from the original BBA. Then, we redistribute the removed mass to remaining focal elements to obtain the final approximated BBA  $\hat{m}^{CC}(\cdot)$  listed in Table VI.

TABLE VI  
 $\hat{m}^{CC}(\cdot)$  OBTAINED USING CORRELATION COEFFICIENT.

Focal Elements	Mass Values
$A'_1 = \{\theta_1\}$	0.13
$A'_2 = \{\theta_4, \theta_5\}$	0.3718
$A'_3 = \{\theta_3, \theta_5\}$	0.1839
$A'_4 = \{\theta_1, \theta_2\}$	0.1677
$A'_5 = \{\theta_2, \theta_5\}$	0.1466

6) Iterative approximation based on distance of evidence. Here  $k$  is set to 2, i.e., we have to remove two focal elements. At the first iteration, Jousselme's distances between the remaining BBA  $\hat{m}_i(\cdot), i = 1, 2, \dots, 7$  and the original BBA  $m(\cdot)$  are 0.1053, 0.0315, 0.1932, 0.1049, 0.105, 0.05981 and 0.05982. We remove  $A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$  at first. Then, we normalize the remaining focal elements and assign  $m = \hat{m}_2$  to execute the next iteration. At the second iteration, Jousselme's distances between the remaining BBA  $\hat{m}_i(\cdot), i = 1, 3, 4, 5, 6, 7$  and  $m(\cdot)$  are 0.1113, 0.2101, 0.114, 0.1118, 0.0644 and 0.0663. So we remove  $A_6 = \{\theta_2, \theta_4, \theta_5\}$  and normalize the remaining focal elements to obtain the final approximated BBA  $\hat{m}^{Dis}(\cdot)$  listed in Table VII.

TABLE VII  
 $\hat{m}^{Dis}(\cdot)$  OBTAINED USING DISTANCE OF EVIDENCE.

Focal Elements	Mass Values
$A'_1 = \{\theta_1\}$	0.1585
$A'_2 = \{\theta_4, \theta_5\}$	0.3659
$A'_3 = \{\theta_3, \theta_5\}$	0.1829
$A'_4 = \{\theta_1, \theta_2\}$	0.1707
$A'_5 = \{\theta_2, \theta_5\}$	0.122

7) ICP method (Our approximation method). The desired remaining focal elements is set to  $k = 5$  and we obtain the final approximated BBA in two iterations as follows.

- The first iteration: We first calculate  $ICP(A_i), i = 1, 2, \dots, 7$  and sort all the focal elements in descending order according to  $ICP(A_i)$  value. The result of the first iteration is listed in Table VIII. Because  $ICP(A_2)$  is the smallest and the focal element  $A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$  is removed at first, then we redistribute the mass of  $A_2$  to remaining focal elements to proceed the next iteration.
- The second iteration: We recalculate  $ICP(A_i), i = 1, 3, 4, 5, 6, 7$  and sort all the remaining focal elements. The result of the second iteration is listed in Table VIII. Because  $ICP(A_7)$  is the smallest value, the focal

element  $A_7 = \{\theta_2, \theta_5\}$  is removed at this iteration. Now the number of remaining focal elements is five and we redistribute the mass of  $A_7$  to remaining focal elements to obtain the final approximated BBA  $\hat{m}^{ICP}(\cdot)$  listed in Table IX.

TABLE VIII  
THE RESULTS OF TWO ITERATIONS USING ICP.

The First Iteration			
Focal Elements	Mass Values	$ M_i^C $	$ICP(A_i)$
$A_3 = \{\theta_4, \theta_5\}$	0.3	$M_3^C = \emptyset$	$\infty$
$A_4 = \{\theta_3, \theta_5\}$	0.15	$M_4^C = \emptyset$	$\infty$
$A_7 = \{\theta_2, \theta_5\}$	0.1	$M_7^C = \emptyset$	$\infty$
$A_5 = \{\theta_1, \theta_2\}$	0.14	1	0.14
$A_1 = \{\theta_1\}$	0.13	1	0.13
$A_6 = \{\theta_2, \theta_4, \theta_5\}$	0.12	1	0.12
$A_2 = \{\theta_2, \theta_3, \theta_4, \theta_5\}$	0.06	4	<b>0.015</b>

The Second Iteration			
Focal Elements	Mass Values	$ M_i^C $	$ICP(A_i)$
$A_3 = \{\theta_4, \theta_5\}$	0.3105	$M_3^C = \emptyset$	$\infty$
$A_4 = \{\theta_3, \theta_5\}$	0.1605	$M_4^C = \emptyset$	$\infty$
$A_5 = \{\theta_1, \theta_2\}$	0.1439	1	0.1439
$A_1 = \{\theta_1\}$	0.1334	1	0.1334
$A_6 = \{\theta_2, \theta_4, \theta_5\}$	0.1411	2	0.0705
$A_7 = \{\theta_2, \theta_5\}$	0.1106	2	<b>0.0553</b>

TABLE IX  
 $\hat{m}^{ICP}(\cdot)$  OBTAINED USING ICP.

Focal Elements	Mass Values
$A'_1 = \{\theta_1\}$	0.1491
$A'_2 = \{\theta_4, \theta_5\}$	0.3237
$A'_3 = \{\theta_3, \theta_5\}$	0.181
$A'_4 = \{\theta_1, \theta_2\}$	0.1658
$A'_5 = \{\theta_2, \theta_4, \theta_5\}$	0.1804

V. EXPERIMENTS AND ANALYSIS

In this section, we compare all the aforementioned BBA approximation methods to demonstrate the effectiveness and interest of our method in terms of three Measures of Performance (MoP): 1) closeness, 2) computational efficiency, and 3) decision-making.

A. MoP of Closeness and Computational Efficiency

The smaller the distance between the new approximated BBA and the original BBA, the less information is lost, which is preferred. We use  $d_{BI}^E$  distance [20] to describe the degree of closeness between two pieces of evidence, which is defined as

$$d_{BI}^E(m_1, m_2) = \sqrt{N_c \cdot \sum_{i=1}^{2^n-1} [d^I(BI_1(A_i), BI_2(A_i))]^2} \quad (17)$$

Here  $N_c = 1/2^{n-1}$  is the normalization factor.  $BI_1(A_i)$  and  $BI_2(A_i)$  are belief intervals of  $A_i$  for  $m_1(\cdot)$  and  $m_2(\cdot)$ , which are denoted by  $[Bel_1(A_i), Pl_1(A_i)]$  and  $[Bel_2(A_i), Pl_2(A_i)]$ .

TABLE X  
ALGORITHM 1: RANDOM GENERATION OF BBA.

---

**Input:**  $\Theta$ : Frame of Discernment;  
 $N_{max}$ : Maximum number of focal elements  
**Output:**  $m(\cdot)$ : BBA  
 Generate  $\mathcal{P}(\Theta)$ , which is the power set of  $\Theta$ ;  
 Generate a random permutation of  $\mathcal{P}(\Theta) \rightarrow \mathcal{R}(\Theta)$ ;  
 Generate an integer between 1 and  $N_{max} \rightarrow l$ ;  
**FOReach** First  $k$  elements of  $\mathcal{R}(\Theta)$  do  
 Generate a value within  $[0, 1] \rightarrow m_i, i = 1, 2, \dots, l$ ;  
**END**  
 Normalize the vector  $m = [m_1, m_2, \dots, m_l] \rightarrow m'$ ;

---

The strict distance between interval numbers  $[a_1, b_1]$  and  $[a_2, b_2]$  ( $b_i \geq a_i, i = 1, 2$ ) is defined by

$$d^I([a_1, b_1], [a_2, b_2]) = \sqrt{\left[\frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2}\right]^2 + \frac{1}{3} \left[\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2}\right]^2} \quad (18)$$

Our comparative analysis is based on a Monte Carlo simulation using  $M = 200$  random runs. The cardinality of the FoD is  $|\Theta| = 5$ . In the  $j$ -th simulation run, a BBA  $m^j(\cdot)$  is randomly generated according to Algorithm 1 [21] of Table X. The number  $j$  of remaining focal elements for all the approaches are set to from 2 to 30 and then the different approximation results  $\hat{m}_i^j(\cdot)$  can be obtained using different methods, where  $i$  denotes the  $i$ -th approximation approach. We record the computational time of the original BBA combination of  $m^j(\cdot) \oplus m^j(\cdot)$  with Dempster's rule and the computational time of using Dempster's rule for each approximated BBA  $\hat{m}_i^j(\cdot) \oplus \hat{m}_i^j(\cdot)$ . The average (over 200 runs) computational time for the original and approximated combination are shown in Fig.2. The average (over 200 runs) distance between the original BBAs and the approximated BBAs obtained using different approaches given different remaining focal elements' number are shown in Fig.3.

As we can see in Fig.2, all the BBA approximation approaches permit to reduce the computational time with respect to the original computational time due to the removal of focal elements. Besides, from Fig.3 we observe that, the approximated BBAs using our new proposed approach are globally closer to the original one when compared with other approaches, which represents the least loss of information. Note that when the number of remaining focal elements is small, there are no data points for the curve of  $k - l - x$  and rank-level fusion methods because they can not remove a certain number of focal elements like other methods due to the constraint that the removed masses are no greater than  $x = 0.2$ .

B. MoP of Decision-making

In this work we use the DSMT Transformation [18] to make the final decision by selecting the  $\theta_i$  with the maxi-

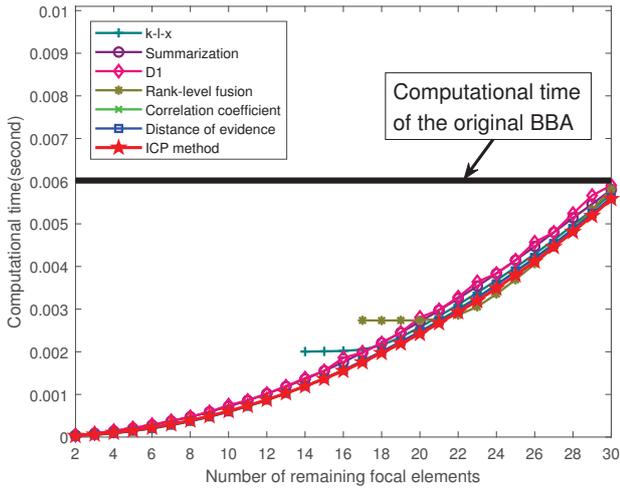


Fig. 2. Computational time comparisons.

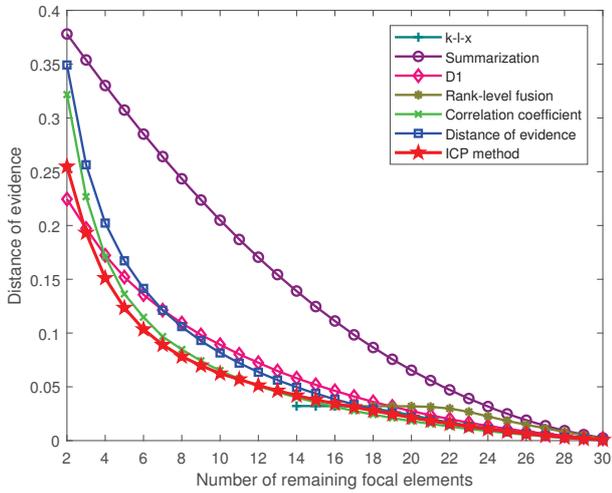


Fig. 3. Closeness comparisons.

mum  $DSmP_\epsilon(\theta_i)$  value. The  $DSmP_\epsilon(\theta_i)$  probability of any elements  $\theta_i, i = 1, 2, \dots, |\Theta|$  of the FoD  $\Theta$  can be obtained by

$$DS m P_\epsilon(\theta_i) = m(\theta_i) + \frac{[m(\theta_i) + \epsilon] \sum_{\substack{X \in 2^\Theta \\ X \supset \theta_i \\ |X| \geq 2}} \frac{m(X)}{\sum_{\substack{Y \in 2^\Theta \\ X \supset Y \\ |Y|=1}} m(Y) + \epsilon \cdot |X|} \quad (19)$$

where  $\epsilon \geq 0$  is a tuning parameter.

In our simulations, all the approximation approaches are compared from the aspect of the accuracy of decision-making. The cardinality of the FoD is  $|\Theta| = 5$  and the parameter  $\epsilon$  has been set to 0.001. Firstly, 1000 BBAs are randomly generated according to Algorithm 1 [21] of Table X. Then, use the DSMP Transformation to make the final decision for the original BBAs. After that, 1000 approximated BBAs are generated and 1000 decisions are made for each approximation method. Finally, the accuracy of decision-making is counted for each

method and the results with different number of remaining focal elements are shown in Fig.4.

As we can see in Fig.4, although ICP method is not the best, it presents a stable and good performance, especially when the number of remaining focal elements is small, which represents the less loss of information from our standpoint. It should be noted that there are no data points for the curve of  $k-l-x$  and rank-level fusion methods due to the constraint mentioned before, when the number of remaining focal elements is small.

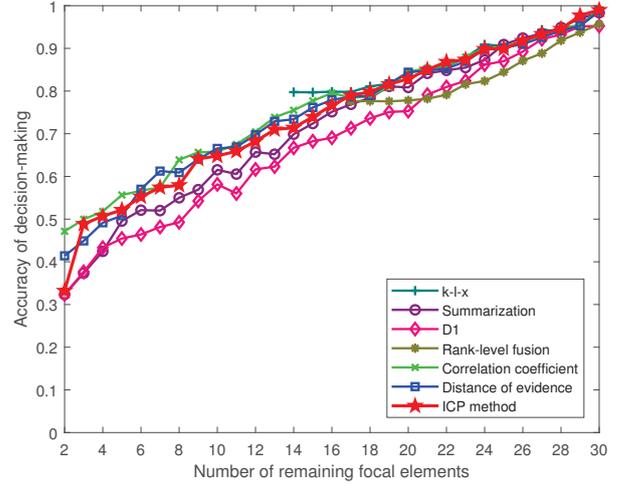


Fig. 4. Accuracy of decision-making comparisons.

## VI. CONCLUSION

With the increase of cardinality of the FoD, evidence combination exhibits a large computational cost. In this paper, a novel BBA approximation approach based on focal element compatibility is proposed based on a new mass assignment strategy. This new method offers a good balance between the computational time and the loss of information. Simulations and comparative analyses show the interest and efficiency of our new method. In future, we will consider other BBA approximation approaches based on the removal of focal elements to solve the bottleneck of BBA combination for different rules of combination.

## APPENDIX

The proof that  $\hat{m}(\cdot)$  which is obtained by the new mass assignment strategy is a true normalized BBA is as follows.

**Proof:**

- 1)  $\hat{m}(\emptyset) = 0$ .
- 2)  $\delta_\cap(A'_i, A_r) > 0$  for any focal element  $A'_i \neq \emptyset$ .

$$\begin{aligned}
 \delta_\cap(A'_i, A_r) &= m(A'_i)|A'_i| + m(A_r)|A_r| \\
 &\quad - [m(A'_i) + m(A_r)]|A'_i \cap A_r| \\
 &\geq m(A'_i)|A'_i| + m(A_r)|A_r| \\
 &\quad - [m(A'_i) + m(A_r)]\min\{|A'_i|, |A_r|\}
 \end{aligned}$$

Suppose that  $\min\{|A'_i|, |A_r|\} = |A_r|$ .

$$\begin{aligned}\delta_{\cap}(A'_i, A_r) &\geq m(A'_i)|A'_i| + m(A_r)|A_r| \\ &\quad - [m(A'_i) + m(A_r)]|A_r| \\ &= m(A'_i)(|A'_i| - |A_r|) > 0\end{aligned}$$

$$3) \sum_{i=1}^{l-1} \hat{m}(A'_i) = 1.$$

$$\begin{aligned}\sum_{i=1}^{l-1} \hat{m}(A'_i) &= \sum_{i=1}^{l-1} \left[ m(A'_i) + \frac{m(A_r)}{D \cdot \delta_{\cap}(A'_i, A_r)} \right] \\ &= \sum_{i=1}^{l-1} m(A'_i) + \frac{m(A_r)}{D} \sum_{i=1}^{l-1} \frac{1}{\delta_{\cap}(A'_i, A_r)} \\ &= \sum_{i=1}^{l-1} m(A'_i) + \frac{m(A_r)}{D} D \\ &= \sum_{i=1}^{l-1} m(A'_i) + m(A_r) = 1 \quad \square\end{aligned}$$

#### ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 61573097 and 91748106, in part by Key Laboratory of Integrated Automation of Process Industry (PAL-N201704), the Fundamental Research Funds for the Central Universities (3208008401), the Qing Lan Project and Six Major Top-talent Plan, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

#### REFERENCES

- [1] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pp. 57–72, Springer, 2008.
- [2] G. Shafer, *A mathematical theory of evidence*, vol. 42. Princeton university press, 1976.
- [3] C. Pohl and J. L. Van Genderen, "Multisensor image fusion in remote sensing: concepts, methods and applications," *International journal of remote sensing*, vol. 19, no. 5, pp. 823–854, 1998.
- [4] F. Yang and H. Wei, "Fusion of infrared polarization and intensity images using support value transform and fuzzy combination rules," *Infrared Physics & Technology*, vol. 60, pp. 235–243, 2013.
- [5] X. Shen and P. K. Varshney, "Sensor selection based on generalized information gain for target tracking in large sensor networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 2, pp. 363–375, 2014.
- [6] K. Yuan, F. Xiao, L. Fei, B. Kang, and Y. Deng, "Modeling sensor reliability in fault diagnosis based on evidence theory," *Sensors*, vol. 16, no. 1, p. 113, 2016.
- [7] W. Jiang, B. Wei, C. Xie, and D. Zhou, "An evidential sensor fusion method in fault diagnosis," *Advances in Mechanical Engineering*, vol. 8, no. 3, pp. 1–7, 2016.
- [8] B. Tessem *et al.*, "Approximations for efficient computation in the theory of evidence," *Artificial Intelligence*, vol. 61, no. 2, pp. 315–329, 1993.
- [9] J. D. Lowrance, T. D. Garvey, and T. M. Strat, "A framework for evidential-reasoning systems," in *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pp. 419–434, Springer, 2008.
- [10] M. Bauer, "Approximations for decision making in the Dempster-Shafer theory of evidence," in *Proc of 20th Int. Conf. on Uncertainty in artificial intelligence*, pp. 73–80, Morgan Kaufmann Publishers Inc., 1996.
- [11] M. Grabisch, "Upper approximation of non-additive measures by k-additive measures—the case of belief functions.," in *ISIPTA*, pp. 158–164, 1999.
- [12] J. Dezert, D. Han, Z. Liu, and J.-M. Tacnet, "Hierarchical proportional redistribution for BBA approximation," in *Belief functions: theory and applications*, pp. 275–283, Springer, 2012.
- [13] T. Denœux, "Inner and outer approximation of belief structures using a hierarchical clustering approach," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 04, pp. 437–460, 2001.
- [14] Y. Yang, D. Han, C. Han, and F. Cao, "A novel approximation of basic probability assignment based on rank-level fusion," *Chinese Journal of Aeronautics*, vol. 26, no. 4, pp. 993–999, 2013.
- [15] D. Han, J. Dezert, and Y. Yang, "Two novel methods for BBA approximation based on focal element redundancy," in *Proc of Int. Conf Fusion 2015, Washington D.C., USA, July 6-9, 2015*, pp. 428–434.
- [16] Y. Yang and Y. Liu, "Iterative approximation of basic belief assignment based on distance of evidence," *PLoS one*, vol. 11, no. 2, p. e0147799, 2016.
- [17] Y. Shou, X. Deng, X. Liu, H. Zheng, and W. Jiang, "Approximation of basic probability assignment in dempster-shafer theory based on correlation coefficient," in *Proc. of 20th International Conference on Information Fusion (Fusion 2017), Xi'an, China, July 10-13, 2017*, pp. 1–7.
- [18] F. Smarandache and J. Dezert, "Advances and applications of dsmt for information fusion-collected works-volume 3." [www.onera.fr/staff/jean-dezert/references](http://www.onera.fr/staff/jean-dezert/references), 2009.
- [19] W. Jiang, "A correlation coefficient for belief functions," *International Journal of Approximate Reasoning*, vol. 103, pp. 94–106, 2018.
- [20] D. Han, J. Dezert, and Y. Yang, "New distance measures of evidence based on belief intervals," in *International Conference on Belief Functions*, pp. 432–441, Springer, 2014.
- [21] T. Burger and S. Destercke, "Random generation of mass functions: A short howto," in *Belief Functions: Theory and Applications*, pp. 145–152, Springer, 2012.

# Evaluation of Probabilistic Transformations For Evidential Data Association

Mohammed Boumediene<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>Signals and Images Laboratory, USTO, Mohamed Boudiaf, Oran, Algeria.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

Emails: mboumediene@inttic.dz, jean.dezert@onera.fr

Originally published as: M. Boumediene, J. Dezert, *Evaluation of Probabilistic Transformations for Evidential Data Association*, in Proc. of IPMU 2020 Int. Conf., Lisbon, Portugal, June 15–19, 2020, and reprinted with permission.

**Abstract**—Data association is one of the main tasks to achieve in perception applications. Its aim is to match the sensor detections to the known objects. To treat such issue, recent research focus on the evidential approach using belief functions, which are interpreted as an extension of the probabilistic model for reasoning about uncertainty. The data fusion process begins by quantifying sensor data by belief masses. Thereafter, these masses are combined in order to provide more accurate information. Finally, a probabilistic approximation of these combined masses is done to make-decision on associations. Several probabilistic transformations have been proposed in the literature. However, to the best of our knowledge, these transformations have been evaluated only on simulated examples. For this reason, the objective of this paper is to benchmark most of interesting probabilistic transformations on real-data in order to evaluate their performances for the autonomous vehicle perception problematic.

**Keywords:** Data Association, Evidential Theory, Belief Functions, Probabilistic Transformation.

## I. INTRODUCTION

Multiple Target Tracking (MTT) is important in perception applications (autonomous vehicle, surveillance, etc.). The MTT system is usually based on two main steps: data association and tracking. The first step associates detected objects in the perceived scene, called *targets*, to known objects characterized by their predicted *tracks*. The second step estimates the *track* states over time typically thanks to Kalman Filters [1], or improved state estimation techniques (like particle filters, etc). Nevertheless, bad associations provide wrong *track* estimation and then leads to false perception results.

The data association problem is usually resolved by Bayesian theory [1], [2]. Several methods have been proposed as the Global Nearest Neighbor (GNN) method, the Probabilistic Data Association Filter (PDAF), and the Multiple Hypothesis Tracking (MHT) [3]–[5]. However, the Bayesian theory doesn't manage efficiently data imperfection due to the lack of knowledge we can have on sensor quality, reliability, etc. To circumvent this drawback, the Evidential theory [6], [7] appears as an interesting approach because of its ability to model and deal with epistemic uncertainty. Its provides a theoretical framework to manage ignorance and data imperfection.

Several evidential data association approaches have been proposed [8]–[11] in the framework of belief functions. Rombaut [11] uses the Evidential theory to measure the confidence of the association between perceived and known obstacles. To manage efficiently objects appearance and disappearance, Gruyer and Cherfaoui [12] propose the bi-directional data association. The first direction concerns the *target-to-track* pairings which provides a good way to manage the appearance of the new *tracks*. The second direction concerns the *track-to-target* pairings and then manage disappearance of *tracks*. This approach has been extended by Mercier *et al.* [10] to track vehicles by using a global optimization to make assignment decisions. To reduce the complexity for real-time applications, a local optimization has been used [8], [13]. For all these methods, the data fusion process begins by defining belief masses from sensor information and prior knowledge. These masses represent the belief and ignorance on the assignment hypotheses. Thereafter, the masses are combined in order to provide a complete information of the considered problem. Finally, to make a decision, the belief masses are classically approximated by a probability measure thanks to a chosen probabilistic transformation.

For data association applications, the widely used probabilistic transformation (i.e. approximation) is the pignistic transformation [8], [10], [13], [14]. This transformation is based on a simple mapping process from belief to probability domain. However, several published works criticize the pignistic transformation and propose generalized and/or alternative transformations [16]–[21]. To our knowledge, the proposed transformations have been evaluated by their authors only on simulated examples. The main objective of this paper is to compare these transformations on real-data in order to determine which one is well-suited for assignment problems.

The rest of the paper is structured as follows. Section II recalls the basics of belief functions and their uses in data association problems. In Section III, the most appealing probabilistic transformations are presented and compared on the well-known KITTI public database in Section IV. Finally, Section V concludes the paper.

## II. BELIEF FUNCTIONS FOR DATA ASSOCIATION

To select “best” associations, the data fusion process consists in four steps: modeling, estimation, combination and decision-making. This section presents their definitions and principles.

## A. Basic Fundamentals

The Belief Functions (BF) have been introduced by Shafer [7] based on Dempster’s researches [6]. They offer a theoretical framework for reasoning about uncertainty. Let’s consider a problem where we have an exhaustive list of hypotheses ( $H_j$ ) which are mutually exclusive. They define a so-called *frame of discernment*  $\Theta$ :

$$\Theta = \bigcup_{j=1}^k \{H_j\} \text{ with } H_i \cap H_j = \emptyset \quad (1)$$

The power set  $2^\Theta$  is the set of all subsets of  $\Theta$ , that is:

$$2^\Theta = \{\emptyset, H_1, \dots, H_k, \dots, \{H_1, H_2, H_3\}, \dots, \Theta\} \quad (2)$$

The proposition  $A = \{H_1, H_2, H_3\}$  represents the disjunction meaning that either  $H_1$  or  $H_2$  or  $H_3$  can be the solution to the problem under concern. In other words,  $A$  represents a partial ignorance if  $A$  is the disjunction of several elements of  $\Theta$ . The union of all hypotheses  $\Theta$  represents the total ignorance and  $\emptyset$  is the empty set that represents the impossible solution (interpreted usually as the conflicting information).

The truthfulness of each proposition  $A \in 2^\Theta$  issued from source  $j$  is modeled by a *basic belief assignment* (bba)  $m_j^\Theta(A)$ :

$$m_j^\Theta : 2^\Theta \rightarrow [0, 1], \sum_{A \in 2^\Theta} m_j^\Theta(A) = 1 \quad (3)$$

Thereafter, the different *bbas* ( $m_j^\Theta$ ) are combined which provides a global knowledge of the considered problem. Several rules of combination have been proposed [22], the conjunctive operator is widely used in many rules proposed in the literature for the combination of sources of evidence. For instance, Shafer [7] did propose Dempster’s rule of combination below which is nothing but the normalized version of the conjunctive rule [23]:

$$\begin{cases} m_{DS}^\Theta(A) &= \frac{1}{1-K} \sum_{A_1 \cap \dots \cap A_p = A} \prod_{j=1}^p m_j^\Theta(A_j) \\ m_{DS}^\Theta(\emptyset) &= 0, \end{cases} \quad (4)$$

where  $K$  is a normalized coefficient:

$$K = \sum_{A_1 \cap \dots \cap A_p = \emptyset} \prod_{j=1}^p m_j^\Theta(A_j). \quad (5)$$

Finally, in order to make decisions in  $\Theta$ , a probabilistic approximation of the combined *bbas* ( $m_{DS}^\Theta(A)$ ) is usually done. The upper and the lower bounds of the unknown

probability  $P(A)$  are defined by the belief  $Bel(A)$  and the plausibility  $Pl(A)$  functions given respectively by:

$$\begin{cases} Bel(A) &= \sum_{B \subseteq A} m_{DS}^\Theta(B) \\ Pl(A) &= \sum_{B \cap A \neq \emptyset} m_{DS}^\Theta(B) \end{cases} \quad (6)$$

## B. Belief Modeling

The data association problem can be analyzed from two points of view: *target-to-track* and *track-to-target* association. Consequently, two frames of discernment are defined:  $\Theta_{i,\cdot}$  and  $\Theta_{\cdot,j}$ ,  $i = 1, \dots, n$ , with  $n$  the number of targets, and  $j = 1, \dots, m$ , with  $m$  the number of tracks:

$$\begin{aligned} \Theta_{i,\cdot} &= \{Y_{(i,1)}, Y_{(i,2)}, \dots, Y_{(i,m)}, Y_{(i,*)}\} \\ \Theta_{\cdot,j} &= \{X_{(1,j)}, X_{(2,j)}, \dots, X_{(n,j)}, X_{(*,j)}\} \end{aligned} \quad (7)$$

where  $\Theta_{i,\cdot}$  is composed of the  $m$  possible target( $i$ )-to-track( $j$ ) associations denoted  $Y_{(i,j)}$ . The hypothesis of appearance is represented by  $Y_{(i,*)}$ <sup>1</sup>.  $\Theta_{\cdot,j}$  contains the  $n$  possible track( $j$ )-to-target( $i$ ) associations denoted  $X_{(i,j)}$ , and  $X_{(*,j)}$  is the track disappearance.

## C. Basic Belief Assignment

For *target-to-track* assignment, three *bba*’s are used to answer the question “Is target  $X_i$  associated with track  $Y_j$ ?”:

- $m_j^{\Theta_{i,\cdot}}(Y_{(i,j)})$ : belief in “ $X_i$  is associated with  $Y_j$ ”,
- $m_j^{\Theta_{i,\cdot}}(\overline{Y_{(i,j)}})$ : belief in “ $X_i$  is not associated with  $Y_j$ ”<sup>2</sup>,
- $m_j^{\Theta_{i,\cdot}}(\Theta_{i,\cdot})$ : the degree of ignorance.

The recent benchmark [24] on huge real data shows that the most suited model is the non-antagonist model [11], [25] which is defined as follows:

$$m_j^{\Theta_{i,\cdot}}(Y_{(i,j)}) = \begin{cases} 0 & , I_{i,j} \in [0, \tau] \\ \Phi_1(I_{i,j}) & , I_{i,j} \in [\tau, 1] \end{cases} \quad (8)$$

$$m_j^{\Theta_{i,\cdot}}(\overline{Y_{(i,j)}}) = \begin{cases} \Phi_2(I_{i,j}) & , I_{i,j} \in [0, \tau] \\ 0 & , I_{i,j} \in [\tau, 1] \end{cases} \quad (9)$$

$$m_j^{\Theta_{i,\cdot}}(\Theta_{i,\cdot}) = 1 - m_j^{\Theta_{i,\cdot}}(Y_{(i,j)}) - m_j^{\Theta_{i,\cdot}}(\overline{Y_{(i,j)}}), \quad (10)$$

where  $0 < \tau < 1$  represents the impartiality of the association process and  $I_{i,j} \in [0, 1]$  is an index of similarity between  $X_i$  and  $Y_j$ .  $\Phi_1(\cdot)$  and  $\Phi_2(\cdot)$  are two cosine functions defined by:

$$\begin{cases} \Phi_1(I_{i,j}) &= \frac{\alpha}{2} \left[ 1 - \cos\left(\pi \frac{I_{i,j} - \tau}{\tau}\right) \right] \\ \Phi_2(I_{i,j}) &= \frac{\alpha}{2} \left[ 1 + \cos\left(\pi \frac{I_{i,j}}{\tau}\right) \right], \end{cases} \quad (11)$$

where  $0 < \alpha < 1$  is the reliability factor of the data source. In the same manner, belief masses are generated for the *track-to-target* assignment.

<sup>1</sup> $Y_{(i,*)}$  refers to the fact that no track is assigned to the target( $i$ ).  
<sup>2</sup> $\overline{Y_{(i,j)}}$  defines the complementary hypothesis of  $Y_{(i,j)}$ ,  
 $\overline{Y_{(i,j)}} = \{Y_{(i,1)}, \dots, Y_{(i,j-1)}, Y_{(i,j+1)}, \dots, Y_{(i,m)}, Y_{(i,*)}\}$ .

Table I  
PROBABILITIES OF TARGET-TO-TRACK ASSOCIATIONS

$P_{i..}(.)$	$Y_1$	...	$Y_m$	$Y_*$
$X_1$	$P_{1..}(Y_{(1,1)})$	...	$P_{1..}(Y_{(1,m)})$	$P_{1..}(Y_{(1,*)})$
$X_2$	$P_{2..}(Y_{(2,1)})$	...	$P_{2..}(Y_{(2,m)})$	$P_{2..}(Y_{(2,*)})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_n$	$P_{n..}(Y_{(n,1)})$	...	$P_{n..}(Y_{(n,m)})$	$P_{n..}(Y_{(n,*)})$

D. Belief Combination

Based on Dempster’s rule (4), the combined masses  $m^{\Theta_{i..}}$  (and  $m^{\Theta_{i..j}}$ ) over  $2^{\Theta_{i..}}$  (and  $2^{\Theta_{i..j}}$ ) can be computed as follows [26]:

$$\begin{aligned}
 m^{\Theta_{i..}}(Y_{(i,j)}) &= K \cdot m_j^{\Theta_{i..}}(Y_{(i,j)}) \prod_{\substack{a=1 \\ a \neq j}}^m \alpha_{(i,a)} \\
 m^{\Theta_{i..}}(\{Y_{(i,j)}, \dots, Y_{(i,l)}, Y_{(i,*)}\}) &= K \cdot \gamma_{(i,(j,\dots,l))} \prod_{\substack{a=1 \\ a \neq j \\ \dots \\ a \neq l}}^m \beta_{(i,a)} \\
 m^{\Theta_{i..}}(Y_{(i,*)}) &= K \cdot \prod_{a=1}^m \beta_{(i,a)} \\
 m^{\Theta_{i..}}(\Theta_{i..}) &= K \cdot \prod_{a=1}^m m_a^{\Theta_{i..}}(\Theta_{i..})
 \end{aligned} \tag{12}$$

with:

$$\left\{ \begin{aligned}
 \alpha_{(i,a)} &= 1 - m_a^{\Theta_{i..}}(Y_{(i,a)}) \\
 \beta_{(i,a)} &= m_a^{\Theta_{i..}}(Y_{(i,a)}) \\
 \gamma_{(i,(j,\dots,l))} &= m_j^{\Theta_{i..}}(\Theta_{i..}) \dots m_l^{\Theta_{i..}}(\Theta_{i..}) \\
 K &= \left[ \prod_{a=1}^m \alpha_{(i,a)} + \sum_{a=1}^m m_a^{\Theta_{i..}}(Y_{(i,a)}) \prod_{\substack{b=1 \\ b \neq a}}^m \alpha_{(i,b)} \right]^{-1}
 \end{aligned} \right.$$

E. Decision-Making

Finally, the probabilities matrix  $P_{i..}$  ( $P_{i..j}$ ) is obtained by using a probabilistic transformation. Table I presents the  $P_{i..}$  matrix where each line defines the association probabilities of the target  $X_i$  with all tracks  $Y_j$ .  $P_{i..}(Y_{(i,*)})$  represents the appearance probability of  $X_i$ .

The association decisions are made by using a global or a local optimization strategy. The Joint Pignistic Probability (JPP) [10] selects associations that maximize the probability product. However, this global optimization is time-consuming and can select doubtful local associations. To cope these drawbacks, local optimizations have been proposed as the Local Pignistic Probability (LPP). Interested readers in the benchmark of these algorithms can refer to [14], [15].

III. PROBABILISTIC TRANSFORMATIONS

The generalized formula of the probabilistic transformation can be defined as follows:

$$\begin{aligned}
 P_{i..}(Y_{(i,j)}) &= m^{\Theta_{i..}}(Y_{(i,j)}) \\
 &+ \sum_{\substack{A \in 2^{\Theta_{i..}} \\ Y_{(i,j)} \subset A}} T(Y_{(i,j)}, A) \cdot m^{\Theta_{i..}}(A), \tag{13}
 \end{aligned}$$

where  $A$  represents the partial/global ignorance about the association of target  $X_i$  and  $T(Y_{(i,j)}, A)$  represents the rate of the ignorance mass  $m^{\Theta_{i..}}(A)$  which is transferred to singleton  $Y_{(i,j)}$ .

Several probabilistic transformations have been proposed in the literature. In this section, only the most interesting ones are presented.

A. Pignistic Probability

The pignistic transformation denoted by  $BetP$  and proposed by Smets [27], [28] is still widely used for evidential data association applications [8], [10], [25], [29]. This transformation redistributes equitably the mass of ignorance on singletons as follows:

$$T_{BetP_{i..}}(Y_{(i,j)}, A) = \frac{1}{|A|}, \tag{14}$$

where  $|A|$  represents the cardinality of the subset  $A$ . However, the pignistic transformation (14) ignores the *bbas* of singletons which can be considered as a crude commitment.  $BetP$  is easy to implement because it has a low complexity due to its simple redistribution process.

B. Dezert-Smarandache Probability

Besides of the cardinality, Dezert-Smarandache Probability ( $DSmP$ ) transformation [18] considers the values of masses when transferring ignorance on singletons:

$$T_{DSmP_{i..}}(Y_{(i,j)}, A) = \frac{m^{\Theta_{i..}}(Y_{(i,j)}) + \epsilon}{\sum_{Y_{(i,k)} \subset A} m^{\Theta_{i..}}(Y_{(i,k)}) + \epsilon \cdot |A|} \tag{15}$$

The value of the tuning parameter  $\epsilon \geq 0$  is used to adjust the effect of focal element’s cardinality in the proportional redistribution, and to make  $DSmP$  defined and computable when encountering zero masses. Typically, one takes  $\epsilon = 0.001$ . The smaller  $\epsilon$ , the better approximation of probability measure we get [18].  $DSmP$  allows to obtain in general a higher Probabilistic Information Content (PIC) [30] than  $BetP$  because it uses more information than  $BetP$  for its establishment. The PIC indicates the level of the available knowledge to make a correct decision.  $PIC = 0$  indicates that no knowledge exists to take a correct decision.

C. MultiScale Probability

The Multiscale Probability ( $MulP$ ) transformation [19] highlights the proportion of each hypothesis in the frame of discernment by using a difference function between belief and plausibility:

$$T_{MulP_{i..}}(Y_{(i,j)}, A) = \frac{(Pl^{\Theta_{i..}}(Y_{(i,j)}) - Bel^{\Theta_{i..}}(Y_{(i,j)}))^q}{\sum_{Y_{(i,k)} \subset A} (Pl^{\Theta_{i..}}(Y_{(i,k)}) - Bel^{\Theta_{i..}}(Y_{(i,k)}))^q}, \quad (16)$$

where  $q \geq 0$  is a factor used to amend the proportion of the difference  $(Pl(\cdot) - Bel(\cdot))$ . However, the  $T_{MulP_{i..}}$  is not defined ( $\frac{0}{0}$ ) when  $m(\cdot)$  is a Bayesian mass ( $Pl(\cdot) = Bel(\cdot)$ ).

#### D. Sudano's Probabilities

Sudano proposes several alternatives to *BetP* as the Proportional Plausibility (*PrPl*) and the Proportional Belief (*PrBel*) transformations [18], [21]. Those latter redistribute respectively the ignorance mass according to the normalized plausibility and belief functions:

$$T_{PrPl_{i..}}(Y_{(i,j)}, A) = \frac{Pl^{\Theta_{i..}}(Y_{(i,j)})}{\sum_{Y_{(i,k)} \subset A} Pl^{\Theta_{i..}}(Y_{(i,k)})} \quad (17)$$

$$T_{PrBel_{i..}}(Y_{(i,j)}, A) = \frac{Bel^{\Theta_{i..}}(Y_{(i,j)})}{\sum_{Y_{(i,k)} \subset A} Bel^{\Theta_{i..}}(Y_{(i,k)})} \quad (18)$$

#### E. Pan's Probabilities

Other proportional transformations have been proposed in [20]. Those transformations assume that the *bba* are proportional to a function  $S(\cdot)$  which is based on the belief and the plausibility:

$$T_{PrBP_{i..}}(Y_{(i,j)}, A) = \frac{S(i,j)}{\sum_{Y_{(i,k)} \subset A} S(i,k)}, \quad (19)$$

where different definitions of  $S$  have been proposed:

$$\begin{cases} PrBP1_{i..} : S(i,j) = Pl^{\Theta_{i..}}(Y_{(i,j)}) \cdot Bel^{\Theta_{i..}}(Y_{(i,j)}) \\ PrBP2_{i..} : S(i,j) = Bel^{\Theta_{i..}}(Y_{(i,j)}) \cdot (1 - Pl^{\Theta_{i..}}(Y_{(i,j)}))^{-1} \\ PrBP3_{i..} : S(i,j) = Pl^{\Theta_{i..}}(Y_{(i,j)}) \cdot (1 - Bel^{\Theta_{i..}}(Y_{(i,j)}))^{-1} \end{cases} \quad (20)$$

## IV. RESULTS

This section presents a benchmark of the probabilistic transformations in the framework of the object association system for autonomous vehicles. The aim is to assign detected objects in the scene (*targets*) to known ones (*tracks*). The transformations have been evaluated on real data.

The KITTI dataset<sup>3</sup> provides 21 sequences recorded from cameras mounted on a moving vehicle on urban roads [31]. To our knowledge, no comparison of probabilistic transformations has been done on real data where more than 30000 associations have been observed. These latter cover different road scenarii as shown in Fig. 1. For this work, detections are defined only by 2D bounding box in the image plane as presented in Fig. 1.

<sup>3</sup>[http://www.cvlibs.net/datasets/kitti/eval\\_tracking.php](http://www.cvlibs.net/datasets/kitti/eval_tracking.php)

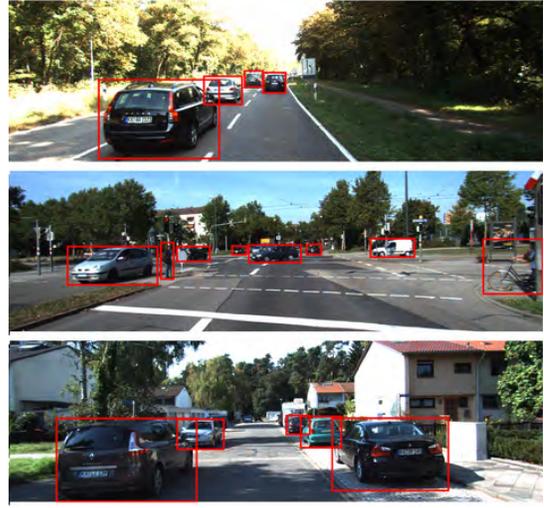


Figure 1. Examples of images provided by KITTI [24].

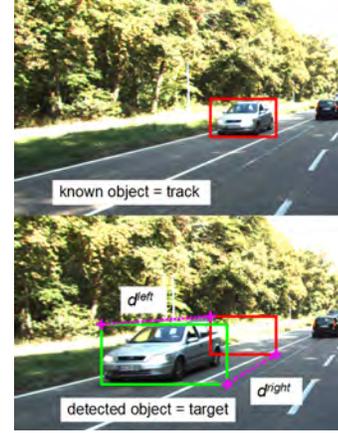


Figure 2. The illustration of the distances  $d_{i,j}^{right}$  and  $d_{i,j}^{left}$  [24].

#### A. Experimental Setting

The assignment information are based on the distance between objects in the image plane. For that, the distance  $d_{i,j}$  is defined as follows:

$$d_{i,j} = \frac{1}{2}(d_{i,j}^{right} + d_{i,j}^{left}), \quad (21)$$

where  $d_{i,j}^{right}$  (resp.  $d_{i,j}^{left}$ ) is the Euclidean distance between bottom-right (resp. top-left) corners of the bounding boxes of target  $X_i$  (detected object) and track  $Y_j$  (known object) as presented in Fig. 2.

The parameters of the *bba* model (11) are:  $\alpha = 0.9$  and  $\tau = 0.5$ . The index of similarity is defined as follows:

$$I_{i,j} = \begin{cases} 1 - \frac{d_{i,j}}{D} & , \text{if } d_{i,j} < D \\ 0 & , \text{otherwise,} \end{cases} \quad (22)$$

where  $D$  is the limit distance for association which is determined heuristically, e.g.  $D = 210$  in this work.

The tuning parameters  $\epsilon = 0.001$  and  $q = 5$  for *DSmP* and *MulP* transformations respectively. The LPP algorithm

has been used as optimization strategy in the decision-making step.

### B. Comparison of probabilistic transformations

All discussed transformations are characterized by an equivalent complexity except the pignistic transformation. *BetP* is computed directly from combined masses which leads to a lower computational time.

To compare the performance of the probabilistic transformations presented previously, the object association system is evaluated by the True Associations Rate (*TAR*):

$$TAR = \frac{\sum_t True\ Association_t}{\sum_t Ground\ Truth_t}, \quad (23)$$

where  $t$  is the frame index.

Table II compares association outcomes of the system based on different probabilistic transformations. Only *target-to-track* association results have been presented in Table II due to the lack of space. However, from *track-to-target* association results, similar comments/conclusions hold. The penultimate row of Table II shows the weighted average of *TAR* value based on all sequences which is given by:

$$TAR_{avg} = \sum_{i=0}^{20} w_i TAR_i \quad (24)$$

where  $TAR_i$  is the *TAR* value of the  $i$ -th sequence, and where the weight  $w_i$  is  $w_i = n_i / \sum_{i=0}^{20} n_i$  and  $n_i$  being the number of associations of the  $i$ -th sequence. For instance,  $TAR_{aver} = 0.9852$  (or 98.52%) for the *BetP* transformation, etc. The last row of Table II represents the weighted standard deviation ( $\sigma_w$ ) of association scores defined as follows:

$$\sigma_w = \sqrt{\sum_{n=0}^{20} w_i (TAR_i - TAR_{avg.})^2} \quad (25)$$

The obtained results show that *PrBel*, *PrBP1*, and *PrBP2* provide the worst mean associations scores ( $\leq 97.40\%$ ) with the largest standard deviation (1.36%) for *PrBP2*. It can be explained by the fact that these transformations are based on the *Bel* function which is a pessimistic measurement. The rest of the transformations provide rates of correct association (i.e. scores)  $> 98.40\%$  which represents a gain of +1%. The best mean score  $\approx 98.50\%$  is given by *BetP*, *PrPl*, and *MultP* transformations. Based only on the mean score criterion, *BetP* seems more interesting because it provides better scores on 15 sequences from 21 as illustrated in Fig. 3. In addition, *BetP* is based on a very simple transferring process of uncertainty which makes *BetP* a good choice for real-time applications. However, this apparent advantage of *BetP* needs to be seen in relative terms because *BetP* also generates a quite large standard deviation of 1.38%, which clearly indicate that *BetP* is not very precise. *PrPl* and *MultP* are also characterized by a relatively high standard deviation (1.22% and 1.39%). On the other hand, the lower standard deviation 1.05% is given by *DSmP* transformation with a good association score = 97.85%. This

transformation performs well in term of PCI criteria which leads to make correct decisions [18]. Consequently, *DSmP* is an interesting alternative to *BetP* for the data association process in autonomous vehicle perception system.

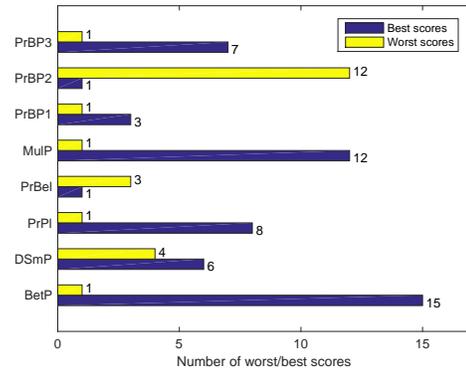


Figure 3. The number of worst/best scores obtained by each probabilistic transformation on 21 sequences; e.g. *PrBel* provides three worst scores (sequences 0, 10, and 17) and only one best score on sequence 12.

## V. CONCLUSION

An evaluation of several probabilistic transformations for evidential data association has been presented in this paper. These transformations approximate the belief masses by a probability measure in order to make association decisions. The widely used probabilistic approximation is the pignistic transformation. However, several published studies criticize the choice of this method of approximation and propose generalized transformations.

We did compare the performances of these probabilistic transformations on real-data in order to determine which one is more suited for assignment problems in the context of autonomous vehicle navigation based on real datasets. The obtained results based on the well-known KITTI dataset show that the pignistic transformation provides one of the better scores. However, it provides a quite large standard deviation contrary to *DSmP* transformation which provides the lowest standard deviation. In addition, *DSmP* procures a nearly similar association score to that given by *BetP*. Consequently, *DSmP* can be a good alternative to *BetP* for the autonomous vehicle perception problematic requiring a bit more computational power with respect to *BetP*.

## REFERENCES

- [1] Y. Bar-Shalom, X. Li, *Multitarget-multisensor tracking: principles and techniques*, YBS Publishing, Bradford, U.K., 1995.
- [2] S. Blackman, R. Popoli, *Design and analysis of modern tracking systems*, Artech House, Norwood, MA, USA, 1999.
- [3] S. Blackman, *Multiple-target Tracking with Radar Applications*, Radar Library, Artech House, 1986.
- [4] T. E. Fortmann, Y. Bar-Shalom, M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association", *IEEE J. of Ocean. Eng.*, Vol. 8, pp. 173–184, 1983.
- [5] D.B. Reid, "An algorithm for tracking multiple targets", *IEEE Trans. on Automatic Control*, Vol. 24(6), pp. 843–854, 1979.
- [6] A.P. Dempster, "Upper and lower probabilities induced by a multiple valued mapping", *Ann. Math. Statistics*, Vol. 38, pp. 325–339, 1967.

Table II  
*Target-to-track* ASSOCIATIONS SCORE (IN %) OBTAINED BY DIFFERENT PROBABILISTIC TRANSFORMATIONS.

Seq. $n^\circ$	nb of Frame	nb of Ass.	<i>BetP</i>	<i>DSmP</i>	<i>PrPl</i>	<i>PrBel</i>	<i>MulP</i>	<i>PrBP1</i>	<i>PrBP2</i>	<i>PrBP3</i>
Seq. 0	154	675	<b>99.41</b>	99.11	<b>99.41</b>	99.26	<b>99.41</b>	<b>99.41</b>	99.26	99.11
Seq. 1	447	2643	<b>97.50</b>	96.71	97.43	96.03	97.47	95.88	95.50	97.39
Seq. 2	233	668	<b>99.70</b>	<b>99.70</b>	<b>99.70</b>	97.75	<b>99.70</b>	98.65	97.46	<b>99.70</b>
Seq. 3	144	337	<b>99.41</b>	<b>99.41</b>	<b>99.41</b>	98.81	<b>99.41</b>	99.11	98.81	<b>99.41</b>
Seq. 4	314	545	89.72	93.39	93.21	92.29	90.09	92.29	91.56	<b>93.76</b>
Seq. 5	297	925	98.59	97.51	98.16	96.00	<b>99.46</b>	96.32	95.24	97.95
Seq. 6	270	474	<b>100</b>	<b>100</b>	<b>100</b>	98.95	<b>100</b>	98.95	98.73	<b>100</b>
Seq. 7	800	2084	<b>97.60</b>	96.55	97.17	95.11	97.55	95.25	94.63	97.02
Seq. 8	390	492	99.19	98.78	99.19	97.76	<b>99.39</b>	97.76	97.56	99.19
Seq. 9	802	2888	<b>98.44</b>	97.33	98.10	97.09	98.37	97.13	96.92	97.82
Seq. 10	294	541	98.71	98.34	98.71	97.78	<b>99.26</b>	98.15	98.89	98.34
Seq. 11	373	3001	<b>99.37</b>	98.77	99.30	99.30	99.33	99.27	99.27	99.23
Seq. 12	78	67	<b>100</b>							
Seq. 13	340	617	93.35	<b>95.62</b>	94.00	93.35	93.19	93.35	92.06	94.00
Seq. 14	106	374	89.04	89.57	88.50	88.77	88.50	<b>89.84</b>	89.04	88.77
Seq. 15	376	1249	<b>99.28</b>	<b>99.28</b>	<b>99.28</b>	99.04	<b>99.28</b>	98.80	98.80	<b>99.28</b>
Seq. 16	209	1872	<b>97.54</b>	96.63	97.44	96.69	<b>97.54</b>	96.85	96.90	97.38
Seq. 17	145	486	<b>99.18</b>	98.35	<b>99.18</b>	96.71	<b>99.18</b>	97.33	96.91	<b>99.18</b>
Seq. 18	339	1130	<b>99.82</b>	98.41	99.65	98.94	<b>99.82</b>	99.03	98.94	99.29
Seq. 19	1059	4968	<b>99.42</b>	98.73	<b>99.42</b>	97.83	99.36	97.89	97.36	99.34
Seq. 20	837	4673	<b>99.68</b>	98.35	99.59	98.35	99.64	98.20	98.10	99.42
<i>All Seq.</i>	8007	30709	<b>98.52</b>	97.85	<b>98.47</b>	97.35	<b>98.52</b>	97.40	97.10	98.37
	<i>std. dev.</i>	$\sigma_w$	1.38	<b>1.05</b>	1.22	1.26	1.39	1.21	1.36	1.18

- [7] G. Shafer, "A mathematical theory of evidence", *Princeton University Press*, Princeton, NJ, USA, 1976.
- [8] M. Boumediene, J.-P. Lauffenburger, J. Daniel, C. Cudel, A. Ouamri, "Multi-roi association and tracking with belief functions: Application to traffic sign recognition", *IEEE Trans. on Intell. Transp. Syst.*, Vol. 15, pp. 2470–2479, 2014.
- [9] T. Dencoux, N. El Zoghby, V. Cherfaoui, A. Jouglet, "Optimal object association in the Dempster-Shafer framework", *IEEE Trans. on Cybern.*, Vol. 44, pp. 2521–2531, 2014.
- [10] D. Mercier, E. Lefèvre, D. Jolly, "Object association with belief functions, an application with vehicles", *Inf. Sci.*, Vol. 181, pp. 5485–5500, 2011.
- [11] M. Rombaut, "Decision in multi-obstacle matching process using Dempster-Shafer's theory", in *Int. Conf. on Advances in Vehicle Control and Safety*, pp. 63–68, Amiens, France, 1998.
- [12] D. Gruyer, V. Berge-Cherfaoui, "Multi-objects association in perception of dynamical situation", in *Proc. of the Fifteenth Conf. on Uncertainty in Artificial Intelligence*, pp. 255–262, San Francisco, CA, USA, 1999.
- [13] M. Boumediene, J.-P. Lauffenburger, J. Daniel, C. Cudel, "Coupled detection, association and tracking for traffic sign recognition", in *IEEE Intell. Vehicle Symposium*, Dearborn, Michigan, USA, pp. 1402–1407, 2014.
- [14] J.-P. Lauffenburger, M. Boumediene, "Adaptive Credal Multi-Target Assignment for Conflict Resolving", in *Proc. of the International Conference on Information Fusion (FUSION)*, pp. 1578–1584, Heidelberg, Germany, July 2016.
- [15] J.-P. Lauffenburger, J. Daniel, M. Boumediene, "Traffic Sign Recognition: Benchmark of credal object associatio algorithms", in *Proc. of Int. Conf. on Information Fusion (FUSION 2014)*, Salamanca, Spain, July 2014.
- [16] B.R. Cobb, P. Shenoy, "On the plausibility transformation method for translating belief function models to probability models", *IJAR*, 41(3), 2006.
- [17] F. Cuzzolin, "On the properties of the Intersection probability", Preprint submitted to Elsevier Science on 21 February 2007 (available on ResearchGate web site).
- [18] J. Dezert, F. Smarandache, "A new probabilistic transformation of belief mass assignment", in *Proc. of Int. Conf. on Information Fusion (FUSION)*, pp. 1410–1417, Cologne, Germany, 2008
- [19] M. Li, X. Lu, Q. Zhang, Y. Deng, "Multiscale probability transformation of basic probability assignment", *Mathematical Problems in Engineering*, 2014.
- [20] W. Pan, H. Yang, "New Methods of Transforming Belief Functions to Pignistic Probability Functions in Evidence Theory", in *Proc. of Int. Workshop on Intelligent Systems and Applications*, May 2009, China
- [21] J. Sudano, "Pignistic Probability Transforms for Mixes of Low- and High-Probability Events", in *Proc. of Int. Conf. on Information Fusion (FUSION 2001)*, August 2001, Montreal, Canada.
- [22] P. Smets, "Analyzing the combination of conflicting belief functions", *Information Fusion*, Vol. 8(4), pp. 387–412, 2007.
- [23] P. Smets, "The Combination of Evidence in the Transferable Belief Model", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 12, pp. 447–458, 1990.
- [24] M. Boumediene, "Evidential Data Association: Benchmark of Belief Assignment Models", in *Proc. of International Conference on Advanced Electrical Engineering*, Algeria, November 2019.
- [25] D. Gruyer, S. Demmel, V. Magnier, R. Belaroussi, "Multi-hypotheses tracking using the Dempster-Shafer theory, application to ambiguous road context", *Inf. Fus.*, Vol. 29, pp. 40–56, 2016.
- [26] C. Royère, D. Gruyer, V. Cherfaoui, "Data association with believe theory", in *Proc. of Int. Conf. on Information Fusion (FUSION 2000)*, Paris, France, 2000.
- [27] P. Smets, R. Kennes, "The transferable belief model", *Artif. Intell.*, Vol. 66, pp. 191–234, 1994.
- [28] P. Smets, "Decision Making in the TBM: the Necessity of the Pignistic Transformation", *IJAR*, Vol. 38, pp. 133–147, 2005
- [29] S. Hachour, F. Delmotte, M. David, "Comparison of Credal Assignment Algorithms in Kinematic Data Tracking Context", in *Proc. of Information Processing and Management of Uncertainty (IPMU 2014)*, Montpellier, France, 2014.
- [30] J. Sudano, "The system probability information content (PIC) relationship to contributing components, combining independent multi-source beliefs, hybrid and pedigree pignistic probabilities", in *Proc. of Int. Conf. on Information Fusion*, Annapolis, USA, July 2002
- [31] A. Geiger, P. Lenz, R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite", in *Proc. of CVPR 2012 Int. Conf.*, Rhode Island, USA, 2012.

# Generation of Fuzzy Evidence Numbers for the Evaluation of Uncertainty Measures

Samia Barhoumi<sup>a</sup>, Imene Khanfir Kallel<sup>a,b</sup>, Sonda Ammar Bouhamed<sup>a,b</sup>,  
Eloi Bosse<sup>b,c</sup>, Basel Solaiman<sup>b</sup>

<sup>a</sup>Control and Energy Management Lab, Cybernetics Team, ENIS/ISBS, University of Sfax, Tunisia.

<sup>b</sup>Image & Information Processing, Department (iTi), IMT-Atlantique, Technopôle Brest, Brest, France.

<sup>c</sup>Expertises Parafuse Inc., 1006 Blvd Pie XII, Québec, QC, G1W 4N1, Québec, Canada.

Emails: semya.barhoumi@gmail.com, imen.khanfir.kallel@gmail.com, sonda.ammar@gmail.com,  
ebosse861@gmail.com, basel.solaiman@imt-atlantique.fr

Originally published as: S. Barhoumi, I.K. Kallel, S.A. Bouhamed, E. Bosse, B. Solaiman, *Generation of Fuzzy Evidence Numbers for the Evaluation of Uncertainty Measures*, in Proc. of 5th Int. Conf. on Advanced Technologies for Signal and Image Processing (ATSIP 2020), 02–05 September 2020, Sousse, Tunisia, and reprinted with permission.

**Abstract**— Uncertainty is an important dimension to consider to evaluate the quality of information. In real world, information tends, usually, to be uncertain, vague and imprecise leading to different types of uncertainty, such as randomness, ambiguity and imprecision. Methods to quantify uncertainty, will help to quantify information quality. This paper presents a general measure of uncertainty framed into the fuzzy evidence theory named GM, quantifying in an aggregate way the three basic types of uncertainty: non-specificity, fuzziness and discord considered within the framework of Generalized Information Theory (GIT). Monte-Carlo simulations are used to study the behavior of GM with respect to the up-cited uncertainty types. Results show that the total uncertainty GM behave properly as we increase and decrease the various types of uncertainty.

**Keywords**—fuzzy evidence theory, uncertainty measures, fuzziness, non-specificity, discord, ambiguity, imprecision, fuzzy randomness

## I. INTRODUCTION

Uncertainty is a primary importance in evaluating information quality. In the field of Generalized Information Theory (GIT), Klir and Yan [1] defined three basic types of uncertainty: non-specificity, fuzziness and discord (or randomness). These types are considered as the three main types of uncertainty that covered all the aspects. Indeed, fuzziness is due to the non-crisp boundaries of a set. Non-specificity is due to the numerous elements of a set. Discord appears when different occurrences are possible. The typology, presented in Fig. 1, is an extension of Klir and Yan's typology that was originally initiated in Liu [2].

The typology exhibits three types of uncertainty, resulting from the combination of each pair of basic types of uncertainty. In fact, imprecision is introduced as a general concept for both fuzziness and non-specificity. Ambiguity is a combination between non-specificity and discord. For the third combination, the concept of discord used here, is equivalent to the randomness one, used by Pal et al. [3], but clearly differs from the

inconsistency concept introduced by Smets [4], or of conflict concept, in the Dempster's sense [5]. No term has been proposed so far for combining fuzziness and discord. Hence, we propose here fuzzy randomness to designate the total uncertainty in the fuzzy-probability theory. Hence, three general terms are used to designate the combinations of (1) non-specificity and discord, i.e. ambiguity, (2) non-specificity and fuzziness, i.e. imprecision and (3) fuzziness and confusion, i.e. fuzzy randomness. The term "total uncertainty" is kept for the combination of the three basic kinds of uncertainty.

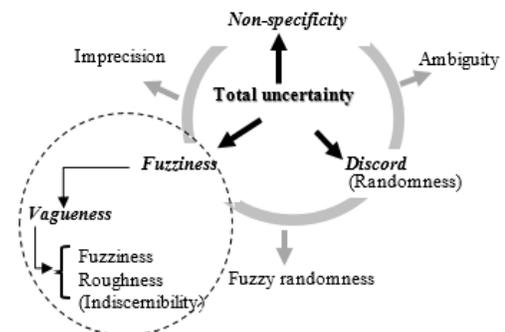


Fig. 1. Circular typology of uncertainty

This paper proposes to pursue the analysis of the properties and behaviors of a general measure of uncertainty named GM, framed into the fuzzy evidence theory [6] that was originally defined in Liu [2]. The analysis has been further developed by Jusselme et al. [7] that resulted in a new measure of ambiguity, AM. Finally, Burkov et al. [8] presented preliminary results from an empirical study about the behavior of GM and AM.

The paper is organized as follows: Section II introduces the fuzzy evidence theory that will be used in the following sections. It also presents existing measures of uncertainty as well as proposed measures, namely imprecision measure and GM

measure that quantifies all kinds of uncertainty. In section III, a consistency analysis is performed on the mentioned measures, to illustrate GM consistency behavior under the three main types of uncertainty: discord, non-specificity and fuzziness. The data used for experiments is provided by a fuzzy evidence number generator. We present our conclusions on the fourth section.

## II. UNCERTAINTY IN FUZZY EVIDENCE THEORY

### A. Basic concepts

**Definition 1.** Let  $X = \{x_1, x_2, \dots, x_N\}$ , classically referred to the universe of discourse in the probability theory and the frame of discernment in the evidence theory, be a frame of discernment (FoD). It is a finite discrete set with  $N$  exclusive and exhaustive hypotheses. It defines the working space for the application as it entails all the propositions, for which information sources can provide evidence. We denote by:

- $\mathcal{P}(X) = 2^X$  the power set of  $X$  (all possible sub-sets of  $X$ ),
- $\tilde{\mathcal{P}}(X) = [0,1]^X$  the set of fuzzy parts of  $X$ ,
- $A$  and  $B$  are two crisp sets from  $\mathcal{P}(X)$ ,
- $\tilde{A}$  and  $\tilde{B}$  are two fuzzy sets from  $\tilde{\mathcal{P}}(X)$  with fuzzy membership degree  $\mu_A(x_i), \forall x_i \in A$  and  $\mu_B(x_j), \forall x_j \in B$ .

### B. Fuzzy evidence theory

#### 1) Fuzzy sets theory

**Definition 2.** A fuzzy set  $\tilde{A}$  [9] is a generalization of a classical set allowing each one of its elements to have a degree of membership to the set. The membership function is defined on the FoD  $X$  by:

$$\mu_{\tilde{A}}: X \rightarrow [0,1], x \mapsto \mu_{\tilde{A}}(x) \quad (1)$$

Where  $\mu_{\tilde{A}}(x)$  is the membership degree of the element  $x$  to the fuzzy set  $\tilde{A}$ . A crisp set is a special case of a fuzzy set where  $\mu_A(x) = 1$  if  $x \in A$  and  $\mu_A(x) = 0$  if  $x \notin A$ .

#### 2) Evidence theory

The evidence theory started with Glenn Shafer [10]. He formalized the field of belief functions based on the work of Arthur Dempster on upper and lower bounds of probability [5]. For that, it is also called Dempster-Shafer theory (DST). DST is often defined as an extension of probability theory. Indeed, a probability distribution is a belief function whose focal elements are singletons. The advantage of the DST is that it provides important tools to handle both random and epistemic uncertainty. It is based on two dual non-additive measures, i.e. belief measure and plausible measure and it assigns a mass to every subset of FoD. In the following, we define the different notions.

**Definition 3.** A mass assignment  $m$  is mapping function from  $\mathcal{P}(X)$  to  $[0,1]$ . It satisfies the following conditions:

$$a) \quad m(\emptyset) = 0, \quad (2)$$

$$b) \quad \sum_{A \subseteq \mathcal{P}(X)} m(A) = 1 \quad (3)$$

The value  $m(A)$  expresses the degree of support of the evidential claim that the true alternative is in the set  $A$ , but not in any specific subset. Any additional evidence, supporting the

claim that the true alternative is in a subset of  $A$ , let's be  $B \subset A$ , must be expressed by another nonzero value  $m(B)$ . Condition (b) of the equation 3, is called a normalization condition of DST. A subset  $A$  is called a focal element if  $m(A) > 0$ .  $\mathcal{F} \equiv \{A \subseteq \mathcal{P}(X) | m(A) > 0\}$  is the set of focal elements.

The belief function can then be deduced from  $m$  as:

$$\text{Bel}(A) = \sum_{B/B \subseteq A} m(B) \quad (4)$$

The plausibility function also can be deduced from  $m$  as:

$$\text{Pl}(A) = \sum_{B/B \cap A \neq \emptyset} m(B) \quad (5)$$

The pignistic probability  $BetP_m$  called as such by Smets [40], corresponds to a classical probabilistic transformation of a belief function:

$$BetP_m(A) = \sum_{B \subseteq X} \frac{m(B)}{|B|} |A \cap B|, \forall A \subseteq X \quad (6)$$

Where  $|A|$  is the cardinality of  $A$ . If  $A$  reduces to a singleton  $\{x\}$ :  $BetP_m(x) = \sum_{x \in B} \frac{m(B)}{|B|}$ .

**Definition 4.** A piece of evidence is an information that supports different hypotheses with different probabilities. It can contain variety of uncertainties due to the diversity of the information sources.

**Definition 5.** A body of evidence (BoE), which is also called basic probability assignment (BPA) or basic belief assignment (BBA), is defined as the focal sets and their corresponding mass functions:

$$BoE \equiv \{ \langle A_i, m(A_i) \rangle : A_i \in \mathcal{F}, m(A_i) > 0 \}_{i=1:f} \quad (7)$$

Where  $f \equiv |\mathcal{F}|$  is the cardinality of  $\mathcal{F}$ , called also the number of focal elements.

In DST, given two pieces of information, represented in the form of two different bodies of evidence, Dempster's combination rule for combining them, is defined as follows:

$$m(C) \equiv \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A) m_2(B)} \quad (8)$$

#### 3) Fuzzy evidence theory

DST is a powerful and flexible mathematical tool for handling uncertainty, impreciseness, and incomplete information. Even though, it represents appropriately non-specificity and discord, it exists some types of uncertainty that cannot be represented: for instance, fuzziness. Fuzzy evidence theory [6] is built to solve this problem. In fact, it combines the concepts of DST with fuzzy sets in order to represent the three types of uncertainty within one framework (fig. 2).

**Definition 6.** A fuzzy mass assignment  $m$  is mapping function from  $\tilde{\mathcal{P}}(X)$  to  $[0,1]$ . It satisfies the following conditions [12]:

$$a) \quad \text{The set of focal element } \tilde{\mathcal{F}} \equiv \{ \tilde{A} \in \tilde{\mathcal{P}}(X), m(\tilde{A}) > 0 \} \text{ is finite.}$$

$$b) \quad m(\emptyset) = 0,$$

$$c) \quad \sum_{A \subseteq \tilde{\mathcal{P}}(X)} m(\tilde{A}) = 1$$

A fuzzy subset  $\tilde{A}$  is called a focal element if  $m(\tilde{A}) > 0$ .

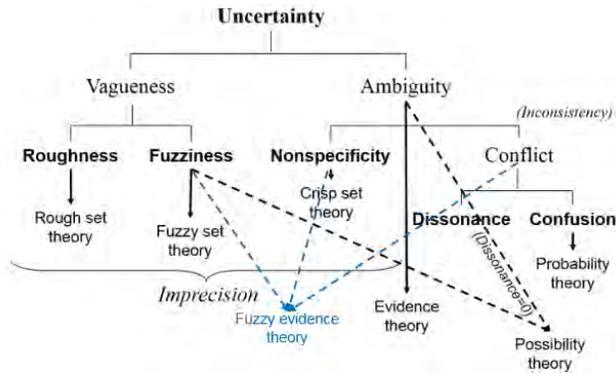


Fig. 2. Extended Klir et al.'s typology, adopted from [11]

**Definition 7.** In the framework of the fuzzy evidence theory, a piece of evidence is characterized by a portion of weight called mass function  $m$ . The generation of the mass, with taking into account uncertainties, is an essential step to generate a fuzzy evidence number. A Fuzzy evidence number is assumed to a fuzzy subset to which is granted a mass.

**Definition 8.** A fuzzy body of evidence (BoE) [8], known as a fuzzy basic probability assignment (FBPA) or fuzzy basic belief assignment (FBBA) [13-15], is defined as the fuzzy focal sets and their corresponding mass functions:

$$FBoE \equiv \{ \langle \tilde{A}_i, m(\tilde{A}_i), \mu_{\tilde{A}_i} \rangle : \tilde{A}_i \in \tilde{\mathcal{P}}(X), m(\tilde{A}_i) > 0 \}_{i=1:f} \quad (9)$$

Where  $f \equiv |\mathcal{F}|$  is the number of focal elements.

### C. Uncertainty measures

For Harmanec [16], “measuring uncertainty or information means assigning a number or value from some ordinal scale to a given model of an epistemic state”.

#### 1) Non-specificity

Non-specificity exists when numerous alternatives exist in a given set  $A$ . In a fuzzy set is obligatory linked to fuzziness since based on the fuzzy cardinality. It also exists in evidence theory, since the BPAs are defined over the crisp subsets of  $X$ . The Hartley measure is proposed in [15] that quantify the non-specificity by the formulation:

$$(A) = \log_2 |A| \quad (10)$$

where  $|A|$  is the cardinality of  $A$ .

A natural generalization of the Hartley measure of non-specificity to the fuzzy-set interpretation of possibility theory was developed by Higashi and Klir [17] under the name U-uncertainty:

$$U(A) = \int_0^1 \log_2 |^\alpha A| d\alpha \quad (11)$$

Dubois and Prade [18] proposed another generalization of the non-specificity measure in Dempster–Shafer’s theory:

$$N(m) = \sum_{A \in \mathcal{P}(X)} m(A) \log_2 |A| \quad (12)$$

Many other authors gave other generalization types of Hartley measure as in [19] by Abellan and Moral and by Klir and Yuan in [20].

#### 2) Discord

Discord represents a feature that expresses the fact that conflictual alternatives are considered as potentially occurring, and commonly represented by an additional measure. The basic measure of discord has been established by Shannon in the probability theory [21]:

$$S(p) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (13)$$

Where  $p$  is a probability distribution on  $X$ .

This measure has been used as the starting point for many theories to quantify uncertainty in situations where the probabilistic representation is inadequate.

#### 3) Fuzziness

Fuzziness is the type of uncertainty represented by fuzzy sets theory, which is clearly distinct from discord. Two main approaches exist for measuring fuzziness, namely either “entropy-like” measures when the membership function is related to a probability distribution [22], [23], [24], [25], or “non-specificity-like” measures when an extension to the classical measure of cardinality is involved [26], [27].

The degenerated measure of fuzziness proposed by De Luca and Termini, [25] and called the entropy of a fuzzy set is given by the following equation:

$$F_{DTe}(\tilde{A}) = -K \sum_{x \in \tilde{A}} [\mu(x) \log_2 \mu(x) + (1 - \mu(x)) \log_2 (1 - \mu(x))] \quad (14)$$

where  $K$  is a normalizing constant and  $\tilde{A}$  is the fuzzy set.

A wide literature survey of the different measures of fuzziness is presented in [28], where 15 measures are reviewed.

#### 4) Ambiguity

Following Klir and Yuan [1], ambiguity is the sum of non-specificity and discord. It is called the total uncertainty in a BPA. Other terms are used, such as total uncertainty, aggregate uncertainty or general uncertainty to designate this type of uncertainty. A measure of ambiguity AM is proposed by Jousselme et al [4]:

$$AM(m) = - \sum_{x \in X} BetP_m(x) \log_2 BetP_m(x) \quad (15)$$

where  $BetP_m(A) = \sum_{B \subseteq X} \frac{m(B)}{|B|} |A \cap B|$ ,  $\forall A \subseteq X$

Shahpari and Seyedin presented a modified version of AM named Modified Ambiguity Measure (MAM) [29] to resolve the issues raised in [30] about the subadditivity property of AM. Abellán and Bossé [31, 32] have analyzed the drawbacks of these measures defined around the pignistic transformation and belief intervals.

#### 5) Imprecision

Imprecision is the total uncertainty of a fuzzy set that accounts for both fuzziness and non-specificity. We name this measure, IM, standing for imprecision measure. This measure was slightly discussed in Liu [2] and with the authors of [7]. IM is composed of two parts as in Eq.16: 1) the non-specificity generally quantified by  $H$ , the Hartley measure, and; 2) the fuzziness quantified by  $F_{DTe}$ .

$$IM(\tilde{A}) = \frac{N_A}{|\tilde{A}|} [F_{DTe}(\tilde{A}) + U(\tilde{A})] \quad (16)$$

where  $F_{DTe}(\tilde{A})$  is the entropy of a fuzzy set,  $U(\tilde{A})$  is U-uncertainty representing the Hartley measure of a fuzzy set  $\tilde{A}$  and  $N$  the number of hypothesis in the frame of discernment.

6) *Fuzzy randomness*

Fuzzy randomness is the proposed term to designate the total uncertainty of a fuzzy probability measure, since it contains both fuzziness and randomness. The first term concerns the prediction of the result, i.e., the event that will occur, and the second term is related to the interpretation of the result as 1 or 0. Fuzzy randomness can be modeled by a fuzzy random (stochastic) variable, which is a mathematical description of a fuzzy stochastic phenomenon. The first measure of uncertainty in fuzzy probability theory was proposed by Zadeh [33] as the weighted entropy for quantifying uncertainty. De Luca and Termini in [25] proposed a measure of uncertainty in the fuzzy probability theory framework that can represent the fuzzy randomness with the following formulation:

$$H_{DT}^{tot}(\tilde{p}) = -\sum_{x \in X} [p(x) \log_2 p(x) + p(x)[\mu(x) \log_2 \mu(x) + (1 - \mu(x)) \log_2 (1 - \mu(x))]] \quad (17)$$

where  $\tilde{p}$  is a fuzzy probability distribution on  $X$  and  $\mu$  is the membership degree of each element  $\mu$  to  $X$ .

D. *Total uncertainty measure*

To evaluate the performance of fusion systems, it is necessary to evaluate the information that processes it, i.e. to quantify the different types of uncertainty related to this information. GM is a measure proposed by Liu [2] in the framework of the fuzzy evidence theory, to gather the different types of uncertainty.

**Definition 10.** For a given  $FBoE = \{\tilde{A}_i, m(\tilde{A}_i), \mu_{\tilde{A}_i}(x)\}$ , the formulation of GM is :

$$GM(FBoE) \equiv -\sum_{x \in X} [BetP(x) \log_2 BetP(x) + \overline{BetP}(x) \log_2 \overline{BetP}(x)] \quad (18)$$

Where :

$$BetP(x_i) \equiv \sum_{i=1}^f \frac{m_X(\tilde{A}_i) \mu_{\tilde{A}_i}(x)}{\sum_{x' \in S_{\tilde{A}_i}} \mu_{\tilde{A}_i}(x')} \quad (19)$$

$$BetP(x_i) \equiv \sum_{i=1}^f \frac{m_X(\tilde{A}_i) (1 - \mu_{\tilde{A}_i}(x))}{\sum_{x' \in S_{\tilde{A}_i}} \mu_{\tilde{A}_i}(x')} \quad (20)$$

Here  $S_{\tilde{A}}$  is the Support  $S_{\tilde{A}} = \{x \in X, \mu_{\tilde{A}}(x) > 0\}$  and  $f$  is the number of focal elements in the FBoE.

E. *Basic Scenarios to study the behavior of GM*

This section attempts to illustrate the behavior of GM which is an aggregate measure including all kinds of uncertainty. By varying these kinds of uncertainty using three basic operations on a fuzzy BPA:

- 1) **Defuzzification:** this operation gives more precision to the information. It transforms a fuzzy BPA into a

crisp one. When it's applied to a fuzzy set, it gives a crisp set and when it's applied to a fuzzy probability distribution, defuzzification gives a classical probability distribution.

- 2) **Specification:** this operation removes the ambiguity part. It transforms a fuzzy BPA into a fuzzy probability distribution. When it's applied to a fuzzy set, specification gives a nonspecific fuzzy set, while applied to a crisp set, specification gives a singleton.
- 3) **Accordance:** this operation reduces randomness. It transforms a fuzzy BPA into a fuzzy set. When it is applied to a fuzzy probability distribution, accordance gives a nonspecific fuzzy set, while applied to a classical probability distribution accordance gives a singleton.

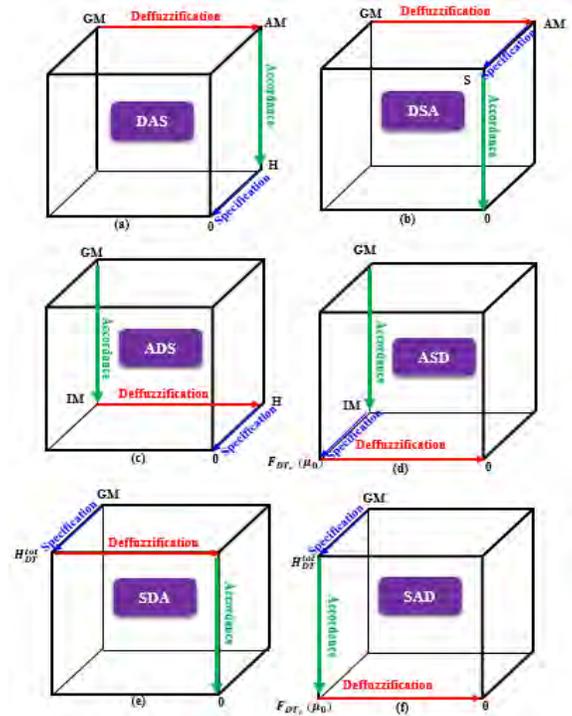


Fig. 3 Schemes to vary the amount and type of uncertainty to study the GM behavior

Six ways to vary the amount and type of uncertainty on a fuzzy BPA, using different combinations of these three operations. Hence, we can see the behavior of GM according to these variations.

F. *GM behavior*

Each time one of the operations, mentioned above, is applied to a FBPA, GM must behave accordingly. For instance, if we reduce a type of uncertainty the GM measure must decrease. Fig.3 shows the six ways to vary the uncertainty and the associated measures generated along the path. For instance, there are six (6) different ways to make GM decreasing. GM will correspond to the different measures quantifying the remaining types of uncertainty as we progress along the path such as  $AM, IM, H_{DT}^{tot}, F_{DTe}, S$  or  $H$ .

For example, in Fig. 3(a), the defuzzification operation is first applied to the initial fuzzy BPA, which leads to a crisp BPA. Hence, the framework of fuzzy evidence theory turns into evidence theory framework and the uncertainty measure is then be quantified by  $AM$ . Then the crisp BPA that includes both discord and non-specificity is accorded, leading to a single classical set, whose uncertainty is the Hartley measure  $H$ . Finally, this classical set is specified, leading to a singleton, thus a null uncertainty. And so on for all the other five figures for  $GM$ .

### III. A FUZZY EVIDENCE NUMBER GENERATOR

We need simulations to study the behavior of GM. To this end, fuzzy bodies of evidence (FBPA) have to be generated to implement the schemes depicted in Fig. 3 for the computation of GM. Firstly, we have to generate, on the frame of discernment  $X$ , a collection  $\tilde{\mathcal{F}} = \{\tilde{A}_1, \dots, \tilde{A}_i, \dots, \tilde{A}_f\}$  of fuzzy numbers of evidence that are fuzzy subsets. These sets define the focal elements of a FBPA. To each focal element, a mass function is associated. We generate the mass function as in Burkov [8] presented in Algorithms 1 & 2 :

#### Algorithm 1

1. **input:**  $\tilde{\mathcal{P}}$ , the set of size  $f$ ;
2.  $rest \leftarrow 1$ ;
3. **For**  $i \leftarrow 1$  to  $f - 1$
4.   **do** generate an exponentially distributed random value  $y$ ;
5.    $m_X(\tilde{A}_i) \leftarrow P(Y \leq y).rest$ ;
6.    $rest \leftarrow rest - m_X(\tilde{A}_i)$ ;
7.    $m(\tilde{A}_f) \leftarrow rest$ ;
8. **return**  $\{m(\tilde{A}_i)\}_{1,f}$

#### Algorithm 2

1. **input:**  $X$ , the frame of discernment;
2.  $type = 'trapezoidal'$ ;
3.  $fun_1 = 'quadratic', fun_2 = 'quadratic'$ ;
4. **Select four random (uniformly distributed) points**  $A, B, C$  and  $D$  such that  $inf_X \leq A < B < C < D \leq sup_X$ ;
5. **define**  $f_1$  according to  $fun_1$  and values of  $A, B$ ;
6. **define**  $f_2$  according to  $fun_2$  and values of  $C, D$ ;
7. **Define**  $\tilde{A}$  as  $\langle A, f_1, B, C, f_2, D \rangle$ .
8. **return**  $\tilde{A}$

### IV. EVALUATION OF GM

The generated FBPA is represent by two components:

- A vector  $m$  that contains the masses of all focal elements,
- A matrix  $FB$  that contains the membership degrees of elements in each focal element of the FBPA.

One focal element is a fuzzy subset of FoD with non-zero mass values.

$$FB = \begin{bmatrix} \mu_{11} & \dots & \mu_{1N} \\ \mu_{21} & \dots & \mu_{2N} \\ \vdots & \vdots & \vdots \\ \mu_{f1} & \dots & \mu_{fN} \end{bmatrix} \text{ and } m = [m_1 \ m_2 \ \dots \ m_f]^T$$

where  $f$  is the number of focal elements and  $N$  is the number of elements in a FoD.

An initial fuzzy BPA, say  $FBPA_0$ , is first randomly generated using a uniform distribution between 0 and 1 for the matrix  $FB$  and the vector  $m$ . This latter must satisfy moreover

the normalization condition of DST for a BPA. Then, the uncertainty in  $FBPA_0$  is successively reduced using the three basic operations described above.

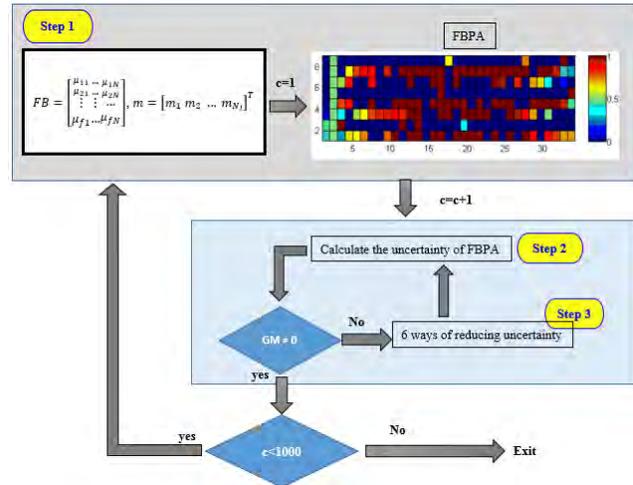
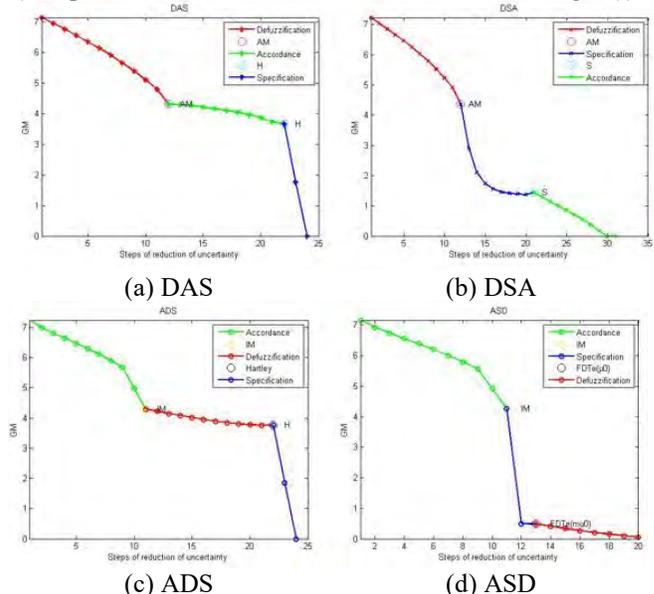


Fig.4 Proposed approach to compute GM

Technically, to perform **defuzzification**, we change step by step the degrees of membership of  $FB$ , whose greater than 0.5 rising to 1 and whose lower than 0.5 decreasing to 0. To perform **accordance**, we change step by step the values in  $m$ , the maximum rising to 1 and all the other values decreasing to 0. And the **specification** is performed by randomly pruning elements of  $FB$  (i.e. changing them for 0). After each basic operation, GM is computed using the corresponding consistent expression (see Fig.4).

Fig. 5 shows the result of a Monte Carlo simulation using 1000 runs and a reducing scheme of uncertainty according to six different ways for the 1000 randomly selected fuzzy BPAs:

- 1) Defuzzification, Accordance, Specification in Fig. 5(a);
- 2) Defuzzification, Specification, Accordance in Fig. 5(b);
- 3) Accordance, Defuzzification, Specification in Fig. 5(c);
- 4) Accordance, Specification, Defuzzification in Fig. 5(d);
- 5) Specification, Defuzzification, Accordance in Fig. 5(e);
- 6) Specification, Accordance, Defuzzification in Fig. 5(f).



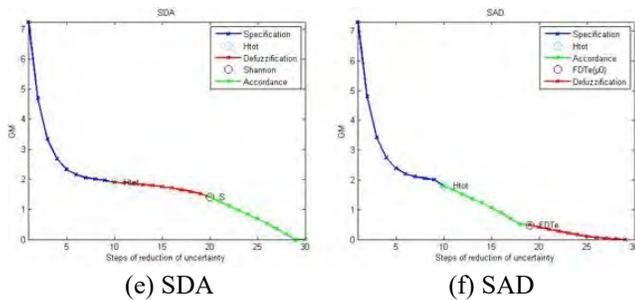


Fig. 5 GM computation following 6 combinations of operations: defuzzification (D), accordance (A) and specification (S).

The results of Fig.5 from simulations show that GM behaves as expected according to the variation of amount and type of uncertainty: GM decreases in a monotonous way when we decrease the amount of uncertainty and obviously increases as we increase uncertainty. The method to generate FBPA's is being validated so that it results in a simulation tool that is appropriate to evaluate and select approaches such as fusion rules that are being used to reduce uncertainty.

#### V. CONCLUSION

The problem of measuring uncertainty within the general framework of fuzzy evidence theory has been discussed. This paper contributed with a simulation tool to support studies on the three main types of uncertainty and their associated measures: nonspecificity, fuzziness and discord. The tool that is centralized around the simulation of FBPA's can be used to evaluate the performance of systems and techniques that have the objective of reducing uncertainty. For instance, it can be used to evaluate and select fusion rules that are framed in a general theory of uncertainty such as fuzzy evidence.

#### REFERENCES

[1] G. J. Klir and B. Yuan, "Fuzzy sets and fuzzy logic: theory and applications," Upper Saddle River, p. 563, 1995.

[2] C. Liu, "A general measure of uncertainty-based information," Ph.D. Ph.D., Laval, QC, Quebec, 2005.

[3] N. R. Pal, J. C. Bezdek, and R. Hemasinha, "Uncertainty measures for evidential reasoning I: A review," *International Journal of Approximate Reasoning*, vol. 7, pp. 165-183, 1992.

[4] P. Smets, "Imperfect Information: Imprecision-Uncertainty en Uncertainty Management in Information Systems: from Needs to Solutions. Amuhai Motro and Phillip Smets," ed: Kluwer Academics Publisher, 1999.

[5] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *Classic Works of the Dempster-Shafer Theory of Belief Functions*, ed: Springer, 2008, pp. 57-72.

[6] S. Petit-Renaud and T. Denoeux, "Regression analysis based on fuzzy evidence theory," in *FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No. 99CH36315)*, 1999, pp. 1229-1234.

[7] A.-L. Jousselme, C. Liu, D. Grenier, and É. Bossé, "Measuring ambiguity in the evidence theory," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 36, pp. 890-903, 2006.

[8] A. Burkov, S. Paquet, G. Michaud, and P. Valin, "An empirical study of uncertainty measures in the fuzzy evidence theory," in *14th International Conference on Information Fusion*, 2011, pp. 1-8.

[9] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, pp. 338-353, 1965.

[10] G. Shafer, *A mathematical theory of evidence* vol. 42: Princeton university press, 1976.

[11] E. Bossé, A. Jousselme, and P. Maupin, "Knowledge, uncertainty and belief in information fusion and situation analysis," *NATO Sciences Series in Computer Sciences*, vol. 198, p. 61, 2005.

[12] G. J. Klir and R. M. Smith, "On measuring uncertainty and uncertainty-based information: recent developments," *Annals of Mathematics and Artificial Intelligence*, vol. 32, pp. 5-33, 2001.

[13] A. Sarabi-Jamab and B. N. Araabi, "An information-based approach to handle various types of uncertainty in fuzzy bodies of evidence," *PLoS one*, vol. 15, p. e0227495, 2020.

[14] Y. Tang, D. Zhou, and F. T. Chan, "An extension to Deng's entropy in the open world assumption with an application in sensor data fusion," *Sensors*, vol. 18, p. 1902, 2018.

[15] D. Zhou, Y. Tang, and W. Jiang, "A modified belief entropy in Dempster-Shafer framework," *PLoS one*, vol. 12, p. e0176832, 2017.

[16] D. Harmanec and G. J. Klir, "Measuring total uncertainty in Dempster-Shafer theory: A novel approach," *International journal of general system*, vol. 22, pp. 405-419, 1994.

[17] M. Higashi and G. J. Klir, "Measures of uncertainty and information based on possibility distributions," *International Journal of General Systems*, vol. 9, pp. 43-58, 1982.

[18] D. Dubois and H. Prade, "Properties of measures of information in evidence and possibility theories," *Fuzzy sets and systems*, vol. 24, pp. 161-182, 1987.

[19] J. Abellan and S. Moral, "A non-specificity measure for convex sets of probability distributions," *International journal of uncertainty, fuzziness and knowledge-based systems*, vol. 8, pp. 357-367, 2000.

[20] G. J. Klir and B. Yuan, "On nonspecificity of fuzzy sets with continuous membership functions," in *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, 1995, pp.25-29.

[21] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, pp. 379-423, 1948.

[22] B. Kosko, "Fuzzy entropy and conditioning," *Information sciences*, vol. 40, pp. 165-174, 1986.

[23] D. Bhandari and N. R. Pal, "Some new information measures for fuzzy sets," *Information Sciences*, vol. 67, pp. 209-228, 1993.

[24] N. Pal and S. K. Pal, "Object-background segmentation using new definitions of entropy," *IEE Proceedings E (Computers and Digital Techniques)*, vol. 136, pp. 284-295, 1989.

[25] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," *Information and control*, vol. 20, pp. 301-312, 1972.

[26] A. Kaufman, "Theory of fuzzy subsets," New York, San Francisco, London, 1975.

[27] R. R. Yager, "Generalized probabilities of fuzzy events from fuzzy belief structures," *Information sciences*, vol. 28, pp. 45-62, 1982.

[28] N. R. Pal and J. C. Bezdek, "Measuring fuzzy uncertainty," *IEEE Transactions on Fuzzy Systems*, vol. 2, pp. 107-118, 1994.

[29] A. Shahpari and S. A. Seyedin, "Using mutual aggregate uncertainty measures in a threat assessment problem constructed by Dempster-Shafer network," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, pp. 877-886, 2014.

[30] G. J. Klir and H. W. Lewis, "Remarks on "Measuring ambiguity in the evidence theory",," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 38, pp. 995-999, 2008.

[31] J. Abellán and É. Bossé, "Critique of recent uncertainty measures developed under the evidence theory and belief intervals," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017.

[32] J. Abellán and É. Bossé, "Drawbacks of uncertainty measures based on the pignistic transformation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, pp. 382-388, 2016.

[33] L. A. Zadeh, "Probability measures of fuzzy events," *Journal of mathematical analysis and applications*, vol. 23, pp. 421-427, 1968.

# Critique of Recent Uncertainty Measures Developed Under the Evidence Theory and Belief Intervals

Joaquín Abellán<sup>a</sup>, Eloi Bossé<sup>b</sup>

<sup>a</sup>Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain.

<sup>b</sup>Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada.

Emails: jabellan@decsai.ugr.es, eboss861@gmail.com

Originally published as: J. Abellán, E. Bossé, *Critique of Recent Uncertainty Measures Developed Under the Evidence Theory and Belief Intervals*, IEEE Trans. On Systems, Man, and Cybernetics: Systems, Vol. 50(3), pp. 1186-1192, March 2020. Reprinted with permission.

**Abstract**—The theory of evidence has been largely used for many applications. This theory is a generalization of probability distribution and offers a mathematical representation for two types of uncertainty-based information: discord and non-specificity. Several measures have already been developed to quantify these two types of uncertainty. They have been called total uncertainty measures since they quantify both types of uncertainty. The generalized Hartley measure and the maximum entropy have been the only measures so far that satisfy a list of properties very desirable for practical applications. Recently, two new measures of non-specificity and total uncertainty based on belief intervals have been proposed. These two measures do not satisfy the properties of additivity, superadditivity and subadditivity in the theory of evidence. The present critique is about these shortcomings and provides a more complete analysis of those uncertainty measures with respect to a list of desired properties. A potential consequence of an ill-characterized measure may yield selecting an inappropriate rule for decision-making in the processing chain from data to information to decisions.

**Keywords:** Imprecise probabilities, theory of evidence, uncertainty measures, non-specificity, additivity, subadditivity.

## I. INTRODUCTION

The main goal of information is to enable adequate decisions and actions. Uncertainty and information are two sides of the same coin. Uncertainty-based information is a major dimension of information quality that is paramount to decision quality. The representation of uncertainty is a crucial issue in applications in many areas of science and engineering that support the transformation of information along the processing chain: data to information to knowledge to decisions and actions.

Classical set theory and probability theory (PT) have been regarded as reference frameworks for centuries to represent uncertainty. However, these two frameworks cannot easily represent all types of uncertainty. Numerous other theories have been developed by expanding the conceptual frameworks on which those classical theories are based.

Amongst them, the *theory of evidence* (TE), also known as Dempster-Shafer's theory [8], [32], has been presented as an important extension of the classical probability theory (PT). In TE, the available information is represented via a new concept called *basic probability assignment* (BPA), which is introduced to generalize the probability distribution concept of PT. This theory has been widely used in several areas and applications such as in [34], [35], [25], [41], [13], [44], [19].

Characterizing information is a crucial step in the development of applications framed within evidential theory as well as within any other uncertainty representation framework. The development of measures of uncertainty is part of that step that impacts on the subsequent calculus processes such as updating (conditioning), aggregation, combination, and decision rules. If a property of a measure is ill-defined or misinterpreted then it could mislead the rest of the processing chain. As an example, a non-additive measure that is characterized as being additive may be combined with an improper rule [36]. The main motivation behind the critique presented in this paper is to insist upon the need to best characterize the information and its imperfections upstream of the processing chain to prevent misinterpretation at decision level.

Uncertainty measures within TE have been exploited successfully in several applications namely in pattern classification [2], [5], [27], [28], [29] to reference a few.

In TE, more types of uncertainty can be represented by a BPA than in PT [23]. Yager [38] makes the distinction between two types of uncertainty called: *discord* (or *randomness* or *conflict*) and *non-specificity*. The first one has been related to entropy and the second to imprecision.

The majority of measures presented in the literature have Shannon's entropy [33] as a starting point. The motivation has been because Shannon's entropy satisfies a large set of properties in PT. For example, the maximum entropy measure in [15] satisfies a similar list of properties in TE than in PT. Hence, it has been considered as the best established measure in TE quantifying jointly both types of uncertainty [4], [5], [6]. This type of measures is called *total uncertainty measures* (TU). Several other measures have been recently proposed to quantify, jointly or separately, both types of uncertainty [31], [9], [7], [39], [40], [43] within the TE framework.

Non-specificity is associated with cases where the information is focused on sets with cardinality greater than one. A non-specificity measure is then based on the way to quantify imprecision in a BPA. This type of uncertainty does not appear in the PT and can be considered as a major difference between PT and other theories that claim to generalize it. The majority of these theories are based on imprecise probabilities [42], [21].

A measure of non-specificity must satisfy a list of properties [12], [23]. So far, only the called *generalized Hartley measure*

satisfies all the required properties for a non-specificity measure in TE [21]. However, there is no study in the literature that establishes the relative importance of that list of properties and its exhaustivity for the use of that measure in practical applications.

Very recently, Yang et al. [39] have presented a new measure of non-specificity in TE. This measure is based on beliefs intervals that a BPA represents about single elements of a finite set. The authors make simplifications of that measure to facilitate its use in applications. In addition, the authors claim that the new measure satisfies the following list of properties: range, monotonicity, symmetry, additivity and subadditivity.

By the means of a counter-example, this paper shows that the properties of additivity and subadditivity are not always satisfied with the proposed measure. In their mathematical derivations, the authors of [39] have not taken into account some considerations related to the cardinal of the focal sets. The contribution offered by our critique here is to provide a more complete analysis of the listed properties and to point out where the shortcomings are.

Another contribution of this paper is to complete the analysis of the Total Uncertainty (TU) measure presented by Yang and Han [40] based on belief intervals. This measure uses the distance between intervals numbers, presented in [18], [37] and applied to belief intervals of an evidence. This TU measure satisfies some properties in the list, but additivity and subadditivity properties have not been discussed. This paper completes the discussion by showing that TU does not satisfy both additivity and subadditivity properties.

In theory of evidence (TE), the complexity of a problem can be reduced by using the principle of decomposition. This is achieved through projections of an evidence. However, for those measures based on this decomposition, it can exhibit a very conflicting behavior: sometimes an increase of information and sometimes a decrease in information. This behavior makes no sense when the same functional projection has been used.

The paper is organized as follows. Section 2 reviews briefly the necessary background about theory of evidence, uncertainty measures and the definitions of some of their desired properties. Section 3 presents an analysis of the additivity and subadditivity of the new TU measures based on belief intervals. Finally, section 4 presents conclusions.

## II. BRIEF BACKGROUND

### A. The theory of evidence

The theory of evidence (TE) [8], [32], is a type of mathematical theory based on imprecise probabilities (see Walley [42]). Its principal characteristics and concepts can be described as follows.

Let  $X$  be a finite set, considered as a set of possible situations,  $|X| = n$ ,  $\wp(X)$  the power set of  $X$  and  $x$  any element in  $X$ . Evidential theory used the concept of a *basic probability assignment* (BPA), also called *mass assignment*. A BPA is a mapping  $m : \wp(X) \rightarrow [0, 1]$ , such that  $m(\emptyset) = 0$

and  $\sum_{A \subseteq X} m(A) = 1$ . A set  $A$  such that  $m(A) > 0$  is called a *focal element*<sup>1</sup> of  $m$ .

Let  $X, Y$  be finite sets. Considering the product space of possible situations  $X \times Y$  and  $m$  a BPA on  $X \times Y$ . The marginal BPA on  $X$ ,  $m^{\downarrow X}$  (and similarly on  $Y$ ,  $m^{\downarrow Y}$ ), is defined as follows:<sup>2</sup>

$$m^{\downarrow X}(A) = \sum_{R|A=R_X} m(R), \quad \forall A \subseteq X, \quad (1)$$

where  $R_X$  is the set projection of  $R$  on  $X$ .

There are two functions associated with each BPA, a belief function,  $Bel$ ; and a plausibility function,  $Pl$ :

$$Bel(A) = \sum_{B \subseteq A} m(B); \quad Pl(A) = \sum_{A \cap B \neq \emptyset} m(B).$$

They can be seen as the belief bounds of  $A$  (lower and upper belief of the set  $A$ , respectively).

We note that belief and plausibility are interrelated for all  $A \in \wp(X)$ , by  $Pl(A) = 1 - Bel(A^c)$ , where  $A^c$  denotes the set complement of  $A$ . Furthermore,  $Bel(A) \leq Pl(A)$ . Hence, the interval  $[Bel(A), Pl(A)]$  is called the *belief interval* for the set  $A$ .

### B. Measures of uncertainty in the theory of evidence (TE)

Shannon [33] presented a measure of entropy on probability theory (PT) defined as follows:

$$S(p) = - \sum_{x \in X} p(x) \log_2(p(x)), \quad (2)$$

where  $p = (p(x))_{x \in X}$  is a probability distribution on  $X$ , a finite set;  $p(x)$  is the probability of the value  $x$ . Here,  $\log_2$  is normally used to measure the value in bits.  $S(p)$  measures the only type of uncertainty that can be represented in probability theory and it satisfies a list of desirable properties [33], [23].

There exist two types of uncertainty in evidence theory (Yager [38]): “one associated with cases where the information is focused on sets with empty intersections; and one associated with cases where the information is focused on sets with cardinality greater than one”. The first concept is known as *discord* (also as *randomness* or *conflict*); and the second one is known as *non-specificity*.

A significant effort has been allocated to quantify the degree of discord in evidence theory [23]. In this paper, the discussion will be focussed on measures of non-specificity and total uncertainty within the framework of TE.

Dubois and Prade [10] have introduced in TE a function based on the Hartley measure [14], which was defined in the classical set theory. It represents a measure of non-specificity associated with a BPA and it is expressed as follows:

<sup>1</sup>The focal elements can be noted as  $A \in \wp(X)$  or  $A \subseteq X$ , with  $m(A) > 0$ . The empty set is never considered here because always  $m(\emptyset) = 0$ .

<sup>2</sup>The expression of the marginal BPAs from Eq. (21) in the paper of Yang et al. [39] has an erratum: the set “R” must be “S”.

$$I(m) = \sum_{A \subseteq X} m(A) \log_2(|A|). \quad (3)$$

$I(m)$  attains its minimum value (zero) when  $m$  is a probability distribution. Its maximum value ( $\log_2(|X|)$ ), is obtained for a BPA,  $m$ , with  $m(X) = 1$  and  $m(A) = 0, \forall A \subset X$ .  $I$  satisfies a large set of desired properties [10], [23]. In the literature, to our knowledge, there is no other measure of non-specificity that satisfies a similar list of properties.

Several composed measures have appeared within the theory of evidence to jointly quantify both parts of uncertainty. They are called *total uncertainty measures* (TU) in TE. The most established one seems to be the measure proposed by Harmanec and Klir [15], [16]:  $S^*(m)$  equal to the maximum of the entropy (upper entropy) of the probability distributions satisfying  $Bel(A) \leq \sum_{x \in A} p(x) \leq Pl(A), \forall A \subseteq X$ . This measure quantifies discord and non-specificity but it does not exhibit the respective separate parts. Some years later, Abellán, Klir and Moral [4], have proposed upper entropy as an aggregate measure framed in more general theories than TE, coherently separating discord and non-specificity. Similar splitting can be done within TE:

$$S^*(m) = S_*(m) + (S^* - S_*)(m), \quad (4)$$

where  $S^*(m)$  represents maximum entropy and  $S_*(m)$  represents minimum entropy on the credal set<sup>3</sup>  $K_m$  associated with a BPA  $m$ , which can be defined the following way:

$$K_m = \{p \mid Bel(A) \leq p(A), \forall A \in \wp(X)\} \quad (5)$$

Here  $S_*(m)$  measures respectively the discord part and  $(S^* - S_*)(m)$  the non-specificity part of the BPA  $m$  [3], [6].

### C. Additivity and subadditivity of non-specificity measures in the theory of evidence

Klir and Wierman [23] define a list of desired properties for an uncertainty measure in TE. The additivity and subadditivity properties belong to that list. A measure of uncertainty (MU), can then be defined as follows [7]:

**Additivity:** “Let  $m$  be a BPA on the space  $X \times Y$ ,  $m^{\downarrow X}$  and  $m^{\downarrow Y}$  its marginal BPAs on  $X$  and  $Y$  respectively such that these marginals are not interactive ( $m(A \times B) = m^{\downarrow X}(A) m^{\downarrow Y}(B)$ , with  $A \subseteq X, B \subseteq Y$  and  $m(C) = 0$  if  $C \neq A \times B$ ). Then  $MU$  satisfies the additivity property iff it satisfies the equality:

$$MU(m) = MU(m^{\downarrow X}) + MU(m^{\downarrow Y})” \quad (6)$$

The same property can be expressed in a reverse manner. If we build a BPA  $m$  on  $X \times Y$  from two independent BPAs on  $X$  and  $Y$ ,  $m_1$  and  $m_2$  respectively ( $m = m_1 \cdot m_2$ ), the total amount of information should be preserved. In this case the marginals of  $m$  on  $X$  and  $Y$ , are  $m_1$  and  $m_2$  respectively (the marginals are not interactive).

<sup>3</sup>A closed and convex set of probability distributions [4].

**Subadditivity:** “Let  $m$  be a BPA on the space  $X \times Y$ ,  $m^{\downarrow X}$  and  $m^{\downarrow Y}$  its marginal BPAs on  $X$  and  $Y$  respectively. Then  $MU$  satisfies the subadditivity property iff it satisfies the inequality:

$$MU(m) \leq MU(m^{\downarrow X}) + MU(m^{\downarrow Y})” \quad (7)$$

The expression indicates that the amount of information must not be increased through a disaggregation process to respect the principle of information conservation.

Additivity and subadditivity represent important properties that a measure of non-specificity must satisfy in TE [12].

To further analyze the meaning of these properties, let us examine the following practical example.

**Example 1:** Three riders and two horses are participating agents of an obstacle race. The competition is that each rider with each horse performs a circuit of obstacles in the shortest possible time. The final score depends on the time taken and the number of overturned obstacles. There are 3 different competitions as follows: (i) the best binomial set (rider + horse); (ii) the best rider; and (iii) the best horse. One can bet money on each type of competition, but the greatest reward is for type (i) since there are more alternatives and it is more difficult to win. The reward for (i) can be considered as an aggregation value of the rewards for (ii) and (iii).

Two cases are analyzed. Case “D” where there is a priori knowledge about riders and horses that can provide an advantage to win the race. Pairing knowledge between riders and horses does exist (no independence between rider and horse). The second case denoted as “I” represents the case where riders do not possess any knowledge about any horse.

In Case D, experts can assign a numerical value to a binomial set (rider + horse) that would be the winning one, as given: 1) the circuit done with different types of obstacles, and 2) the experts’ a priori knowledge obtained from past competitions. In Case I, there are horse experts and rider experts. There is no cross-knowledge (pairing knowledge) and experts are consulted separately on horses and on riders.

For both cases, a set of possible riders is defined as  $RI = \{r_1, r_2, r_3\}$ ; a set of horses as  $H = \{h_1, h_2\}$ ; and a set of binomials as  $B = \{b_{ij} \mid i = 1, 2, 3; j = 1, 2\}$ , where  $b_{ij}$  expresses the binomial  $r_i + h_j$  and  $B = RI \times H$ .

All experts use belief functions to express their knowledge on riders, horses and binomial sets.

These two cases, Case D and Case I, are analyzed below with respect to the additivity and subadditivity properties.

- **Case I.** The BPA  $m_{RI}$  is given by an expert on riders while BPA  $m_H$  is from an expert on horses. Each quantified knowledge is independent on each other, i.e. if

we build  $m = m_{RI} \cdot m_H$ , the marginals are not interactive since  $m_{RI} = m^{\downarrow RI}$  and  $m_H = m^{\downarrow H}$ .

The information obtained from  $m$ , built from the marginal BPAs, is used to bet on the binomial sets. The product space creates more alternatives, then the uncertainty must be greater (or at least not less) than with each marginal taken separately. The question raised here is the following: is it coherent that the sum of information from the original independent BPAs be less or greater than the one of the joint BPA via an information based-uncertainty measure?

Let then examine the following situation: (a) a bet on the binomial sets OR (b) bets on riders and horses separately. Here, (a) and (b) are exclusive. Rewards for (a) and (b) are equivalent according to the previous definitions. Our common sense would say to choose (a) or (b) with the same level of credibility than obtained with the available information. This is what the definition of the additivity property means. A joint case is being built from independent sources so as a result there will be neither more neither less information: **the available amount of information must be preserved**. Consequently, the additivity property is then an essential property for a measure of information (uncertainty) in TE.

- **Case D.** An expert expresses his knowledge via the BPA  $m$  on  $B = RI \times H$ , and we want to make the three possible bets. To this end, we use: 1) the marginal BPAs  $m^{\downarrow RI}$ ,  $m^{\downarrow H}$  and, 2) an uncertainty measure  $UM$ . Let then examine the following statement:

$$UM(m) > UM(m^{\downarrow RI}) + UM(m^{\downarrow H})$$

The expression above would mean that via the measure  $UM$ , one can gain information with a simple mathematical procedure on BPAs. Moreover, if we build the BPA<sup>4</sup>  $m' = m^{\downarrow RI} \cdot m^{\downarrow H}$  on  $B$ , it could result in providing more information via  $UM$  than obtained from the original expert on binomial sets. That means improving the original information source. Based upon the above additivity property, to be coherent, we can have the following situation:

$$UM(m) > UM(m^{\downarrow RI}) + UM(m^{\downarrow H}) = UM(m')$$

In this case, it results in having more information to the binomial sets than the one obtained from the original expert (!). The subadditivity property indicates, in accordance with our common sense, that is not coherent. The amount of information must not be increased with only calculations of the marginal BPAs.

<sup>4</sup> $m \neq m'$  might be possible.

### III. MEASURES OF UNCERTAINTY BASED ON BELIEF INTERVALS

#### A. A measure of non-specificity based on belief intervals

Very recently, Yang et al. [39] have presented a new non-specificity measure in TE based on belief intervals. This measure takes into account the maximum difference of belief of each possible state of a finite set  $X$ . If we consider a BPA  $m$  on a finite set  $X$  with states  $\{x_1, \dots, x_n\}$ , the measure is defined using the values  $Pl(\{x_i\}) - Bel(\{x_i\})$ ,  $i \in \{1, \dots, n\}$ . The measure is the average of those values on the belief intervals, and can be expressed as follows:

$$NE^{BI}(m) = \frac{1}{n} \sum_i (Pl(\{x_i\}) - Bel(\{x_i\})). \quad (8)$$

This definition makes sense in relation to the non-specificity concept and its coherence. Non-specificity is focused on the degree of imprecision of a BPA. Then, it is related to the values of the belief intervals used in the definition of the measure.

Yang et al. [39] have shown that  $NE^{BI}$  can be reduced<sup>5</sup> to the following expression:

$$NE^{BI}(m) = \sum_{A \subseteq X, |A| > 1} m(A) \frac{|A|}{n}. \quad (9)$$

1) *Properties* : The measure in Yang et al. [39], has been shown to satisfy the properties of range, monotonicity and symmetry. These three properties will not be discussed here. In addition, Yang et al. have shown that their measure satisfies a *multiplicativity* and a *submultiplicativity* property. These properties are equivalent to the additivity and subadditivity properties<sup>6</sup>, as it is remarked in Yang et al. [39]. These properties can be described in a similar way as for additivity and subadditivity.

**Multiplicativity**: “Let  $m$  be a BPA on the space  $X \times Y$ ,  $m^{\downarrow X}$  and  $m^{\downarrow Y}$  its marginal BPAs on  $X$  and  $Y$  respectively such that these marginals are not interactive ( $m(A \times B) = m^{\downarrow X}(A) m^{\downarrow Y}(B)$ , with  $A \subseteq X$ ,  $B \subseteq Y$  and  $m(C) = 0$  if  $C \neq A \times B$ ). Then  $NE^{BI}$  satisfies the equality:

$$NE^{BI}(m) = NE^{BI}(m^{\downarrow X}) \cdot NE^{BI}(m^{\downarrow Y})”. \quad (10)$$

**Submultiplicativity**: “Let  $m$  be a BPA on the space  $X \times Y$ ,  $m^{\downarrow X}$  and  $m^{\downarrow Y}$  its marginal BPAs on  $X$  and  $Y$  respectively. Then  $NE^{BI}$  satisfies the inequality:

<sup>5</sup>The expression of the summations from Eq. (15) in the paper of Yang et al. [39] brings some sort of confusion. For example, to express the value of the  $Pl(\{\theta_i\})$ ,  $\sum_{j=1, i < j}^n m(\{\theta_i, \theta_j\})$  is used ; and for  $Pl(\{\theta_n\})$  one could interpret that the summation does not include any term, because there is no  $j > n$ . One should use  $\sum_{j=1, i \neq j}^n m(\{\theta_i, \theta_j\})$  instead, which represents the correct form.

<sup>6</sup>They represent their counter-part definitions [12]. The additivity and subadditivity are used when the measure has a range of  $[0, \log(n)]$ , whereas the multiplicativity and submultiplicativity with a range of  $[0, 1]$ ; but they represent the same concept.

$$NE^{BI}(m) \leq NE^{BI}(m^{\downarrow X}) \cdot NE^{BI}(m^{\downarrow Y}). \quad (11)$$

At this point, we would like to cite Yang et al. [39] about the importance of these two properties:

”Note that the physical meaning of submultiplicativity is essential in the conservation of information, i.e., the amount of uncertainty in a joint BPA is no greater than the total amount of uncertainty of its corresponding marginal BPAs. The equation holds if and only if the corresponding marginal BPAs are independent, i.e., there is not correlated part. If two marginal BPAs are dependent, then the double counting uncertainty amount should be removed, therefore, the total amount of uncertainty in the joint BPA is larger than the total amount in marginal BPAs.”

Let then consider the following example based on the Example 1:

**Example 2:** Using the case expressed in Example 1. Let  $RI \times H$  be the product space of the sets  $RI = \{r_1, r_2, r_3\}$  and  $H = \{h_1, h_2\}$ , and  $m^1$  and  $m^2$  the following BPAs on  $RI$  and  $H$  expressed by two experts on riders and horses, respectively:

$$m^1(\{r_1, r_2\}) = 0.45, m^1(\{r_3\}) = 0.45, m^1(RI) = 0.1; \\ m^2(H) = 1.$$

Hence, the BPA  $m = m^1 \times m^2$  on  $RI \times H$  has the following masses:

$$m(\{b_{11}, b_{12}, b_{21}, b_{22}\}) = 0.45, m(\{b_{31}, b_{32}\}) = 0.45, \\ m(RI \times H) = 0.1,$$

where we note  $b_{ij} = (r_i, h_j)$ . Then  $m^1 = m^{\downarrow RI}$  and  $m^2 = m^{\downarrow H}$ , and they are not interactive by definition.

The values of uncertainty via  $NE^{BI}$  measure are the following ones:

$$NE^{BI}(m) = 0.3 + 0.15 + 0.1 = 0.55; \\ NE^{BI}(m^1) \cdot NE^{BI}(m^2) = (0.3 + 0.1) \cdot 1 = 0.4.$$

Then  $NE^{BI}(m') \neq NE^{BI}(m^1) \cdot NE^{BI}(m^2)$ , and the multiplicativity property is not satisfied by  $NE^{BI}$ . □

The above Example 2 can also be used to prove that  $NE^{BI}$  does not satisfy the submultiplicativity property. Let use the same BPA  $m$  on  $RI \times H$  and then its marginal BPAs are  $m_1$  and  $m_2$  on  $RI$  and  $H$ , respectively. We have that:

$$NE^{BI}(m) > NE^{BI}(m^{\downarrow RI}) \cdot NE^{BI}(m^{\downarrow H}),$$

and this implies that the submultiplicativity property is not satisfied for  $NE^{BI}$ .

The results above say that the new measure, as opposed of what has been said in Yang et al. [39], does not satisfy the

multiplicativity and submultiplicativity properties. The main error in Yang et al. [39] mathematical proofs is located at the cardinal of the sets. For example, if  $A \subseteq RI$  and  $B \subseteq H$ , then  $A \times B \subseteq RI \times H$  and it is possible to find focal sets of each BPA such that,  $|A \times B| > 1$  and  $|A| = 1, |B| > 1$ , as it is happening in the Example 2.

To analyze the proof about the multiplicativity, we apply the values of the Example 2 on the penultimate step in the Equation (23) of Yang et al. [39]. We can observe the following situation<sup>7</sup>, (detailed calculations of Example 2):

$$NS^{BI}(m) = \dots = \sum_{A, B; |A \times B| > 1} m_X(A)m_Y(B) \frac{|A||B|}{n_X n_Y} = \\ = (0.45 \cdot 1) \frac{(2 \cdot 2)}{3 \cdot 2} + (0.45 \cdot 1) \frac{(1 \cdot 2)}{3 \cdot 2} + (0.1 \cdot 1) \frac{(1 \cdot 3)}{3 \cdot 2};$$

$$NS^{BI}(m^1) = \sum_{A; |A| > 1} m_X(A) \frac{|A|}{n_X} = 0.45 \frac{2}{3} + 0.1 \frac{3}{3};$$

$$NS^{BI}(m^2) = \sum_{B; |B| > 1} m_Y(B) \frac{|B|}{n_Y} = 1 \frac{2}{2}.$$

It is easy to see that

$$\sum_{A, B; |A \times B| > 1} m_X(A)m_Y(B) \frac{|A||B|}{n_X n_Y} \neq (0.45 \frac{2}{3} + 0.1 \frac{3}{3}) \cdot (1 \frac{2}{2}) = \\ = \sum_{A; |A| > 1} m_X(A) \frac{|A|}{n_X} \cdot \sum_{B; |B| > 1} m_Y(B) \frac{|B|}{n_Y}$$

Hence, the penultimate step in Eq. (23) of Yang et al. [39] is shown to be incorrect.

Focal elements with cardinal 1 in the set  $X$ , i.e. sets that do not produce any imprecision, can be components of sets in the product space  $X \times Y$  that produce imprecision. This case has not been considered in the proof of multiplicativity in Yang et al. [39].

Now, let us look at the proof of the submultiplicativity property and here again there is a problem with the last step. That last step expresses the following equality:

$$NS^{BI}(m^{\downarrow X} \times m^{\downarrow Y}) = \sum_{R \subseteq X \times Y; |R_X| > 1, |R_Y| > 1} m(R) \frac{|R|}{n_X n_Y}$$

But, this equality is not always correct. Again, if we consider the values of the Example 2, the left term of the above equality is  $NS^{BI}(m)$  and contains the following addend:

<sup>7</sup>Here,  $X$  is  $RI$  and  $Y$  is  $H$  to follow the notation of [39].

$$NS^{BI}(m) = \dots + m(\{b_{31}, b_{32}\})\frac{2}{6} + \dots$$

Now, considering  $R = \{b_{31}, b_{32}\}$ , we have that  $R_{RI} = \{x_3\}$  and  $|R_{RI}| = 1$ . Hence, that addend can not be in the right part of the equality. In this specific case, the equality is shown to be incorrect.

That could lead us to think that this measure always produces an uncertainty decrease when using decomposition with the marginals in TE. That might not be a good operation for this measure and implies to add more coherence. The property associated with that situation is known as **supermultiplicativity**.<sup>8</sup> With the above notation, it can be defined, for a  $UM$  measure, as:

$$UM(m) \geq UM(m^{\downarrow X}) \cdot UM(m^{\downarrow Y}), \quad (12)$$

but with the example below we show that  $NE^{BI}$  does not satisfy that property.

**Example 3:**

Let  $X \times Y$  be the product space of the sets  $X = \{x_1, x_2, x_3\}$  and  $Y = \{y_1, y_2\}$ , and  $m$  the following BPA on  $X \times Y$ :

$$m(\{z_{11}, z_{12}, z_{21}, z_{22}, z_{31}\}) = 0.4, m(\{z_{21}, z_{22}, z_{31}, z_{32}\}) = 0.4, \\ m(X \times Y) = 0.2,$$

where we note  $z_{ij} = (x_i, y_j)$ .

Now, the marginals on  $X$  and  $Y$  have the following values:

$$m^{\downarrow X}(\{x_2, x_3\}) = 0.4, m^{\downarrow X}(X) = 0.6$$

$$m^{\downarrow Y}(Y) = 1$$

The values of uncertainty via  $NE^{BI}$  measure are the following ones:

$$NE^{BI}(m) = 0.8;$$

$$NE^{BI}(m^{\downarrow X}) \cdot NE^{BI}(m^{\downarrow Y}) = 0.866 \cdot 1 = 0.866.$$

Hence,  $NE^{BI}(m) < NE^{BI}(m^{\downarrow X}) \cdot NE^{BI}(m^{\downarrow Y})$ , and the supermultiplicativity property is not satisfied by  $NE^{BI}$ .  $\square$

The Examples 2 and 3 represents a very conflictive situation for the new measure of Yang et al. [39]. If we have a complex set of information represented within TE that can be decomposed using the marginals to make projections on two (2) less complex sets, then in some situations the information available can be decreased; and in others ones it can be increased. This is a very undesirable behaviour for such a measure.

*B. A measure of total uncertainty based on belief intervals*

Let  $m$  be a BPA on a finite set  $X$  with  $n$  elements  $\{x_1, \dots, x_n\}$ . Each element has the following belief intervals  $[Bel(\{x_i\}), Pl(\{x_i\})]$ ,  $i \in \{1, \dots, n\}$ , that we simplify as

$[Bel_i, Pl_i]$ . Yang and Han [40] have recently proposed the following measure of total uncertainty in TE:

$$TU'(m) = 1 - \frac{1}{n} \sqrt{3} \sum_i d'([Bel_i, Pl_i]; [0, 1]) \quad (13)$$

where  $d'$  is a distance function between intervals obtained from [18], [37], and it has the following expression:

$$d'([a, b_1]; [a_2, b_2]) = \\ = \sqrt{\left[\frac{a_1+b_1}{2} - \frac{a_2+b_2}{2}\right]^2 + \frac{1}{3} \left[\frac{b_1-a_1}{2} - \frac{b_2-a_2}{2}\right]^2} \quad (14)$$

1) *Properties*: Yang and Han [40] prove that  $TU'$  measure has some desirable properties for a TU measure in TE: range, monotonicity, probability consistency and set consistency. Unfortunately they do not show whether this TU measure satisfies additivity and subadditivity properties; in this case, since the range of TU is  $[0, 1]$ , the equivalent properties of multiplicativity and submultiplicativity. These two properties will be verified below.

The following Example 4 shows that  $TU'$  measure does not satisfy the multiplicativity property.

**Example 4:** Let  $X \times Y$  be the product space of the sets  $X = \{x_1, x_2, x_3\}$  and  $Y = \{y_1, y_2\}$ . Let the following BPAs  $m^1$  and  $m^2$  be on  $X$  and  $Y$  respectively:

$$m^1(\{x_1, x_2\}) = 0.5, m^1(\{x_3\}) = 0.5;$$

$$m^2(\{y_1\}) = 0.5, m^2(Y) = 0.5.$$

We build the BPA  $m = m^1 \times m^2$  on  $X \times Y$  (the marginals of  $m$  are not interactive). Then,  $m$  has the following values, where again we note  $z_{ij} = (x_i, y_j)$ :

$$m(\{z_{11}, z_{21}\}) = 0.25, m(\{z_{11}, z_{12}, z_{21}, z_{22}\}) = 0.25,$$

$$m(\{z_{31}\}) = 0.25, m(\{z_{31}, z_{32}\}) = 0.25,$$

Now, the set of belief intervals of each  $z_{ij}$  are the following ones:

$$\{[0, 0.5]; [0, 0.25]; [0, 0.5]; [0, 0.25]; [0.25, 0.5]; [0, 0.25]\}$$

and we obtain:

$$TU'(m) = 0.386$$

Similarly, for  $m^1$  and  $m^2$  we have the following belief intervals, respectively:

$$\{[0, 0.5]; [0, 0.5]; [0.5, 0.5]\},$$

$$\{[0.5, 1]; [0, 0.5]\},$$

and  $TU'$  values:

$$TU'(m^1) = 0.5, \quad TU'(m^2) = 0.5$$

<sup>8</sup>It is equivalent to the super-additivity property [11].

Hence, we have that

$$0.386 = TU'(m) \neq TU'(m^1) \cdot TU'(m^2) = 0.5 \cdot 0.5 = 0.25$$

and the multiplicativity property is not satisfied.  $\square$

To prove that  $TU'$  also does not satisfy the submultiplicativity property we only need the Example 4. Considering the BPA  $m$  on  $X \times Y$ , we have that its marginal BPAs are  $m^1$  and  $m^2$  on  $X$  and  $Y$ , respectively so that:

$$0.386 = TU'(m) > TU'(m^{\downarrow X}) \cdot TU'(m^{\downarrow Y}) = 0.5 \cdot 0.5 = 0.25$$

implying that  $TU'$  does not satisfy the submultiplicativity property.

As happens with the  $NE^{BI}$  measure,  $TU'$  also does not satisfy the supermultiplicativity property, expressed in Eq. (12). To prove it, we only need to consider the Example 5.

**Example 5:**

Let  $X \times Y$  be the product space of the sets  $X = \{x_1, x_2, x_3\}$  and  $Y = \{y_1, y_2\}$ , and  $m$  the following BPA on  $X \times Y$ :

$$m(\{z_{11}, z_{12}, z_{21}, \}) = 0.6, m(\{z_{31}\}) = 0.1, \\ m(X \times Y) = 0.3,$$

where we note  $z_{ij} = (x_i, y_j)$ .

Now the marginals on  $X$  and  $Y$  have the following values:

$$m^{\downarrow X}(\{x_1, x_2\}) = 0.6, m^{\downarrow X}(\{x_3\}) = 0.1, m^{\downarrow X}(X) = 0.3$$

$$m^{\downarrow Y}(\{y_1\}) = 0.1, m^{\downarrow Y}(Y) = 0.9$$

These evidences produces the following sets of belief intervals on  $X \times Y$ ,  $X$  and  $Y$  respectively:

$$\{[0, 0.9]; [0, 0.9]; [0, 0.9]; [0, 0.3]; [0.1, 0.4]; [0, 0.3]\} \\ \{[0, 0.9]; [0, 0.9]; [0.1, 0.4]\}, \\ \{[0.1, 1]; [0, 0.9]\},$$

The values of uncertainty via  $TU'$  measure are the following ones:

$$TU'(m) = 0.624;$$

$$TU'(m^{\downarrow X}) \cdot TU'(m^{\downarrow Y}) = 0.748 \cdot 0.9 = 0.673.$$

Now  $TU'(m) < TU'(m^{\downarrow X}) \cdot TU'(m^{\downarrow Y})$ , and the supermultiplicativity property is not satisfied by  $TU'$ .  $\square$

We see that the new total uncertainty measure,  $TU'$ , in TE has a similar undesirable behaviour than the  $NE^{BI}$  measure. In some situations the use of the projections produces an increase of information, and a decrease in other ones. This behaviour is incoherent. It could impact negatively a subsequent decision-making process (not analyzed in this current paper).

#### IV. CONCLUSIONS

In this paper, we have analyzed the properties of additivity and subadditivity of new measures of uncertainty in TE based on belief intervals. The definitions of these measures makes sense and they satisfy a list of interesting and important properties but they have shortcomings: they do not satisfy the properties of additivity and subadditivity. These two properties belongs to a list of required properties for such types of measures in TE. It has been also shown that these new measures present incoherent results when a decomposition is done using the same functional projections: in some situation that decomposition presents an increase in information, but in others it presents a decrease of information. The importance of those shortcomings in real life applications cannot be appreciated without more analysis and extensive empirical studies. However, considering the results presented in this paper, more work is required to adjust those measures or to develop new ones that possess all required properties for exploitation along the complete processing chain from data to decisions and actions.

#### REFERENCES

- [1] J. Abellán, S. Moral, *A Non-specificity Measure for Convex Sets of Probability Distributions*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 8, pp. 357–367, 2000.
- [2] J. Abellán, S. Moral, *Upper Entropy of Credal Sets. Applications to Credal Classification*, Int. J. of Approximate Reasoning, Vol. 39, pp. 235–255, 2005.
- [3] J. Abellán, S. Moral, *Difference of entropies as a non-specificity measure for credal sets*, Int. J. of General Systems, Vol. 34(3), pp. 201–214, 2005.
- [4] J. Abellán, G.J. Klir, S. Moral, *Disaggregated total uncertainty measure for credal sets*, Int. J. of General Systems, Vol. 35(1), pp. 29–44, 2006.
- [5] J. Abellán, A. Masegosa, *Requirements for total uncertainty measures in Dempster-Shafer theory of evidence*, Int. J. of General Systems, Vol. 37(6), pp. 733–747, 2008.
- [6] J. Abellán, *Combining nonspecificity measures in Dempster-Shafer theory of evidence*, Int. J. of General Systems, Vol. 40(6), pp. 611–622, 2011.
- [7] J. Abellán, E. Bossé, *Drawbacks of uncertainty measures based on the pignistic transformation*, IEEE Trans. on Systems, Man and Cybernetics: Systems, Vol. 48(3), pp. 382–388, March 2018.
- [8] A.P. Dempster, *Upper and Lower Probabilities Induced by a Multivaluated Mapping*, Ann. Math. Statistic, Vol. 38, pp. 325–339, 1967.
- [9] Y. Deng, *Deng entropy*, Chaos, Solitons and Fractals, Vol. 91, pp. 549–553, 2016.
- [10] D. Dubois, H. Prade, *A Note on Measure of Specificity for Fuzzy Sets*, BUSEFAL, Vol. 19, pp. 83–89, 1984.
- [11] D. Dubois, H. Prade, *Additivity and monotonicity of measures of information defined in the setting of Shafer's evidence theory*, BUSEFAL, Vol. 24, pp. 64–76, 1985.
- [12] D. Dubois, H. Prade, *Properties of measures of information in evidence and possibility theories*, Fuzzy sets and systems, Vol. 24(2), pp. 161–182, 1987.
- [13] D. Han, W. Liu, J. Dezert, Y. Yang, *A novel approach to pre-extracting support vectors based on the theory of belief functions*, Knowledge-Based Systems, Vol. 110, pp. 210–223, 2016.
- [14] R.V.L. Hartley, *Transmission of information*, The Bell Systems Technical Journal, Vol. 7, pp. 535–563, 1928.
- [15] D. Harmanec, G.J. Klir, *Measuring total uncertainty in Dempster-Shafer Theory: a novel approach*, Int. J. General System, Vol. 22, pp. 405–419, 1994.
- [16] D. Harmanec, G.J. Klir, *Principle of uncertainty revisited*, in Proc. of 4th Intern. Fuzzy Systems and Intelligent Control Conf., Maui, Hawaii, pp. 331–339, 1996.
- [17] M. Higashi, G.J. Klir, *Measures of uncertainty and information based on possibility distributions*, Int. J. General System, Vol. 9, pp. 43–58, 1983.

- [18] A. Iripino, R. Verde, *Dynamic clustering of interval data using a Wasserstein-based distance*, Pattern Recognition Letters, Vol. 29(11), pp. 1648–1658, 2008.
- [19] M.B. Farah, D. Mercier, F. Delmotte, E. Lefevre, *Methods using belief functions to manage imperfect information concerning events on the road in VANETs*, Transportation Research Part C: Emerging Technologies, Vol. 67, pp. 299–320, 2016.
- [20] A.-L. Jusselme, C. Liu, D. Grenier, E. Bossé, *Measuring Ambiguity in the Evidence Theory*, IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 36(5), pp. 890–903, 2006.
- [21] G.J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*, John Wiley, Hoboken, NJ, 2006.
- [22] G.J. Klir, R.M. Smith, *Recent developments in generalized information theory*, Int. J. Fuzzy Systems of Mathematics and Artificial Intelligence, Vol. 1(1), pp. 1–13, 1999.
- [23] G.J. Klir, M.J. Wierman, *Uncertainty-Based Information*, Phisica-Verlag, 1998.
- [24] G.J. Klir, H.W. Lewis, *Remarks on “Measuring Ambiguity in the Evidence Theory”*, IEEE Trans. on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol. 38(4), pp. 995–999, 2008.
- [25] D. Li, H. Wang, R. Wang, Y. Xiong, *Professional Competence Evaluation of Information Management Undergraduates Based on Rough Set and DS Evidence Theory*, International Journal of Database Theory and Application, Vol. 9(5), pp. 111–120, 2016.
- [26] C. Liu, D. Grenier, A.-L. Jusselme, E. Bossé, *Reducing Algorithm Complexity for Computing an Aggregate Uncertainty Measure*, IEEE Trans. on Systems, Man and Cybernetics, Part A, Vol. 37(5), pp. 669–679, 2007.
- [27] Z. Liu, Q. Pan, J. Dezert, G. Mercier, *Credal classification rule for uncertain data based on belief functions*, Pattern Recognition, Vol. 47(7), pp. 2532–2541, 2014.
- [28] Z. Liu, Q. Pan, J. Dezert, *Credal classification rule for uncertain data based on belief functions*, Pattern Recognition, Vol. 46(3), pp. 834–844, 2013.
- [29] Z. Liu, Q. Pan, J. Dezert, G. Mercier, *Hybrid Classification System for Uncertain Data*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 47(10), pp. 2783–2790, October 2017.
- [30] A. Meyerowitz, F. Richman, E.A. Walker, *Calculating maximum-entropy probabilities densities for belief functions*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 2, pp. 377–389, 1994.
- [31] A. Shahpari, S.A. Seyedin, *Using mutual aggregate uncertainty measures in a threat assessment problem constructed by Dempster-Shafer network*, IEEE Trans. on Systems, Man and Cybernetics: Systems, Vol. 45(6), pp. 877–886, 2015.
- [32] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
- [33] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.
- [34] X. Su, S. Mahadevan, P. Xu, Y. Deng, *Dependence Assessment in Human Reliability Analysis Using Evidence Theory and AHP*, Risk Analysis, Vol. 35(7), pp. 1296–1316, 2015.
- [35] J. Tang, Z. Wu, Ch. Yang, *Epistemic uncertainty quantification in flutter analysis using evidence theory*, Chinese Journal of Aeronautics, Vol. 28(1), pp. 164–171, 2015.
- [36] V. Torra, Y. Narukawa, M. Sugeno, *Non-Additive Measures*, Springer, 2014.
- [37] L. Tran, L. Duckstein, *Comparison of fuzzy numbers using a fuzzy distance measure*, Fuzzy Sets and Systems, Vol. 130(3), pp. 331–341, 2002.
- [38] R.R. Yager, *Entropy and Specificity in a Mathematical Theory of Evidence*, Int. J. of General Systems, Vol. 9, pp. 249–260, 1983.
- [39] Y. Yang, D. Han, J. Dezert, *A new non-specificity measure in evidence theory based on belief intervals*, Chinese Journal of Aeronautics, Vol. 29(3), pp. 704–713, 2016.
- [40] Y. Yang, D. Han, *A new distance-based total uncertainty measure in the theory of belief functions*, Knowledge-Based System, Vol. 94, pp. 114–123, 2016.
- [41] K. Yang, S. Liu, X. Li, X.A. Wang, *DS Evidence Theory Based Trust Detection Scheme in Wireless Sensor Networks*, International Journal of Technology and Human Interaction, Vol. 12(2), pp. 48–59, 2016.
- [42] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, New York, 1991.
- [43] J. Wang, Y. Hu, F. Xiao, X. Deng, Y. Deng, *A novel method to use fuzzy soft sets in decision making based on ambiguity measure and Dempster-Shafer theory of evidence: An application in medical diagnosis*, Artificial intelligence in medicine, Vol. 69, pp. 1–11, 2016.
- [44] Q. Wang, W. Li, Y. Wu, Y. Pei, M. Xing, D. Yang, *A comparative study on the landslide susceptibility mapping using evidential belief function and weights of evidence models*, Journal of Earth System Science, Vol. 125(3), pp. 645–662, 2016.

# On Effectiveness of Measures of Uncertainty of Basic Belief Assignments

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA - DTIS/MIDL, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.  
Emails: jean.dezert@onera.fr, tchamova@bas.bg

Originally published as: J. Dezert, A. Tchamova, *On the Effectiveness of Measures of Uncertainty of Basic Belief Assignments*, Information & Security Journal, Vol.N52, pp. 9–36, 2022, and reprinted with permission.

**Abstract**—In this paper we examine many existing measures of uncertainty (MoU) of basic belief assignments proposed in the literature related with the theory of belief functions. Some measures capture only a particular aspect of the uncertainty, others propose a total measure of uncertainty to characterize the information quality of a source of information. We discuss the effectiveness of these measures with respect to four main important desiderata that we consider essential for the definition of a satisfactory MoU (i.e. effective entropy of basic belief assignment).

**Keywords:** Measure of Uncertainty, MoU, belief functions, Shannon entropy.

## I. INTRODUCTION

In the classical framework of belief functions, a source of evidence expresses its belief on the possible solutions of a given problem defined with respect to a chosen (finite) frame of discernment (FoD)  $\Theta$ . This belief is usually characterized by a basic belief assignment (BBA), referred also as a belief mass denoted by  $m(\cdot)$ . One of the major concern related with belief function is how to measure/quantify the uncertainty encompassed by a source of evidence and inherent to any BBA. This problem is challenging and of crucial importance because its effective solution would allow to well characterize any BBA, to make fair comparisons of sources of evidence, to compare fusion results in term of uncertainty reduction, to achieve a BBA complexity reduction by new approximations methods, etc.

In this paper we make a state-of-the-art survey of most of existing MoUs available in the literature, and point out their theoretical drawbacks to warn the reader about their misuses and irrelevances in applications. This work justifies the requirement for better effective MoUs to make a step-ahead in the understanding and characterization of uncertainty in the belief functions framework. There exist several survey papers covering different proposals for measures of uncertainty, among them we must cite by chronological order [1]–[14], and more recently in [15], [16]. These papers however do not consider the effectiveness of MoU as we propose in this paper.

In the sequel, we suppose the reader familiar with the classical (i.e. Shannon) information theory [17]–[22], and specially with Shannon entropy measure, and with the theory of belief functions introduced by Shafer in [23]. Some of

these basics are recalled in appendix for convenience and for recalling the classical notations.

This paper is organized as follows. In section II we present and justify the four essential desiderata that a MoU should satisfy in order to be considered as effective. In section III we examine many existing MoUs proposed in the literature over 40 years, and check if they pass the effectiveness test, or not. For those that pass successfully the test, we examine in details in section IV if they are sufficiently well justified for considering them as serious candidate for effective MoU to be used in applications. Section V concludes this survey and gives some perspectives for future research works.

## II. DESIDERATA FOR AN EFFECTIVE MOU

Our analysis of many existing works on Measures of Uncertainty (MoU) of belief functions reveals that most of MoUs suffer of serious problems, and we explain why in the next section. Here we introduce several very essential desiderata that a satisfactory MoU, denoted by  $U(m)$ , should satisfy. Some of these desiderata have already been identified in the past by some researchers working towards axiomatic approaches of MoUs, for instance by Klir [8] and Abellán [12], [13], [15]. Here we keep only the four desiderata that we consider as really important and indispensable, and we justify our choice for these desiderata. We also explain why we consider the other desiderata not fundamental, and why we decide to discard them. The four essential and indispensable desiderata we consider for a satisfactory MoU are mathematically expressed as follows

- **Desideratum D1:** (zero min value of  $U(m)$ )

$$U(m) = 0 \quad (1)$$

if the BBA  $m$  defined on the power set  $2^\Theta$  of the frame of discernment  $\Theta$  is focused on a singleton, that is if  $m(X) = 1$  for some  $X$  of  $2^\Theta$  with  $|X| = 1$ .

**Justification of D1:** This desideratum is very natural and intuitive because any particular BBA for which  $m(X) = 1$  with  $|X| = 1$  characterizes the certainty of a singleton  $X$ , which is one of most specific element of  $2^\Theta$ . There is no uncertainty about the choice of this element  $X$  characterized by  $m(X) = 1$  since this element  $X$  (a smallest information granule) does not include other

smaller element in it. So, the measure of uncertainty must be minimal, and it can always be arbitrarily set to zero reflecting well such non-uncertainty case.

- **Desideratum D2:** (increasing of MoU of vacuous BBA)

$$U(m_v^\Theta) < U(m_v^{\Theta'}), \quad \text{if } |\Theta| < |\Theta'|. \quad (2)$$

where  $m_v^\Theta$  and  $m_v^{\Theta'}$  are the vacuous BBAs defined respectively on the frames of discernment (FoDs)  $\Theta$  and  $\Theta'$  of cardinalities  $|\Theta|$  and  $|\Theta'|$ .

**Justification of D2:** This desideratum stipulates that the measure of uncertainty of a total ignorant source of evidence represented by the vacuous BBA must increase with the cardinality of the frame of discernment. This desideratum makes perfect sense because the total ignorant source of evidence on  $\Theta = \{\theta_1, \dots, \theta_N\}$  for which  $m_v^\Theta(\Theta) = 1$  means that one knows absolutely nothing about only  $N$  elements, whereas the total ignorant source of evidence on  $\Theta' = \{\theta_1, \dots, \theta_N, \theta_{N+1}, \dots, \theta_{N'}\}$  for which  $m_v^{\Theta'}(\Theta') = 1$  means that one knows absolutely nothing about more elements because  $N' > N$ . This clearly indicates that  $m_v^{\Theta'}$  must be considered in fact as more ignorant than  $m_v^\Theta$ , and the condition (2) reflects this necessity.

- **Desideratum D3:** (compatibility with Shannon entropy)

$$U(m) = - \sum_{X \in \Theta} m(X) \log(m(X)) \quad (3)$$

if the BBA  $m(\cdot)$  is a Bayesian BBA defined on the FoD  $\Theta$ . We recall that any Bayesian BBA commits zero belief mass for all elements of the power set of  $\Theta$  having their cardinality greater than one [23].

**Justification of D3:** This desideratum D3 seems also very natural because Shannon entropy is the most well-known (and justified [20], [24]–[27]) measure used so far to quantify the uncertainty (i.e. the randomness, or variability, also called conflict by some authors) of a probability mass function (pmf). Because any Bayesian BBA induces belief and plausibility functions that coincide with a probability measure, one must have a total coherence of  $U(m)$  with Shannon entropy when the BBA is Bayesian if one admits, as we do here, that Shannon entropy is an effective measure the uncertainty (or randomness) of a pmf. Under the acceptance of Shannon entropy as MoU for pmf, the desideratum D3 makes perfect sense. Of course, this desideratum D3 could be disputed (and eventually rejected) if one can cast in doubt (based on very strong justification) the use of Shannon entropy as MoU for pmf. For alternatives of Shannon entropy, see the non-exhaustive list of alternatives given in [28]–[30], and discussions in [9], [31]–[33] for instance.

- **Desideratum D4:** (unicity of max value of  $U(m)$ )

$$\forall m \neq m_v, \quad U(m) < U(m_v), \quad (4)$$

where  $m$  is any BBA different of the vacuous BBA  $m_v$  defined with respect to the same FoD.

**Justification of D4:** This fourth desideratum is very important and it makes perfect sense also because the total ignorant source of evidence is characterized by the vacuous BBA  $m_v(\cdot)$ , and no source of evidence can be more uncertain than the total ignorant source, so the unique maximum value of  $U(m)$  must be obtained for  $U(m_v)$ . As it will be shown next, many existing MoUs fail to satisfy this important and essential desideratum.

**Effectiveness of a measure of uncertainty:** A measure of uncertainty  $U(m)$  is said effective if and only if it satisfies desiderata D1, D2, D3, and D4 and if it is strongly well justified. Any MoU that fails to satisfy at least one of these desiderata is said non-effective, and in this case it cannot be considered seriously as a satisfactory measure of uncertainty for characterizing a basic belief assignment of a source of evidence. Consequently, all non-effective MoUs should be discarded in all applications that necessitate some MoU evaluation.

**Remark 1:** It is worth noting that we do not specify a priori what should be the range of an effective MoU in contrary to some axiomatic attempts made by different authors as reported, for instance, in [15], [34], [35]. We consider that the choice of the range must not be chosen a priori. The maximum range must result of the effective MoU mathematical definition. We only request the satisfaction of the desideratum D4, which is much more general, natural and essential.

**Remark 2:** We voluntarily do not include the subadditivity desideratum in our list of our desiderata for the search of an effective MoU in the belief function framework because this desideratum appears in general (i.e. for non-Bayesian non-vacuous BBAs) to be incompatible with essential desideratum D4, and thus it is illusory and vain to ask for a sub-additive MoU for non-Bayesian non-vacuous BBAs. We recall that the subadditivity condition is defined by  $U(m^{\Theta \times \Theta'}) \leq U(m^{\downarrow \Theta}) + U(m^{\downarrow \Theta'})$  or any joint BBA defined on the cartesian product  $\Theta \times \Theta'$  of FoDs  $\Theta$  and  $\Theta'$ , where  $m^{\downarrow \Theta}$  is the marginal (i.e. projection) of  $m^{\Theta \times \Theta'}(\cdot)$  on the power-set  $2^\Theta$ , and  $m^{\downarrow \Theta'}$  is the marginal (i.e. projection, see [36], [37] for definition) of  $m^{\Theta \times \Theta'}(\cdot)$  on the power-set  $2^{\Theta'}$ . This impossibility comes from the fact that there exist in general  $2^{|\Theta \times \Theta'|} - 2^{|\Theta|} \cdot 2^{|\Theta'|} > 0$  elements of the power set  $2^{\Theta \times \Theta'}$  (including some disjunctions of elements of  $\Theta \times \Theta'$ ) whose mass of belief cannot be obtained from the masses of elements of  $2^\Theta$  and of  $2^{\Theta'}$ , and which contribute in the uncertainty measure of the joint BBA  $m^{\Theta \times \Theta'}$ . Indeed, if  $|\Theta| = N$  and  $|\Theta'| = N'$  the cartesian product space  $\Theta \times \Theta'$  has  $N \cdot N'$  elements and its power set  $2^{\Theta \times \Theta'}$  has  $2^{N \cdot N'}$  elements which is always bigger than the cartesian product space of power sets  $2^\Theta \times 2^{\Theta'}$  because  $2^N \cdot 2^{N'} (= 2^{N+N'}) < 2^{N \cdot N'}$  as soon as  $N > 2$  and  $N' > 2$ . It is worth mentioning also that most of elements of  $2^\Theta \times 2^{\Theta'}$  do not have the same structure as the elements of the power set  $2^{\Theta \times \Theta'}$ . This means that we cannot recover the joint BBA

$m^{\Theta \times \Theta'}$  from the product, or combination, of its marginal  $m^{\downarrow \Theta}$  and  $m^{\downarrow \Theta'}$  in general, but if the joint BBA is totally vacuous or if the joint BBA is Bayesian and if it is equal to the product of two so-called non-interacting (or independent) probability measures [8]. To be more clear, consider two FoDs  $\Theta$  and  $\Theta'$  with  $|\Theta| = 2$  and  $|\Theta'| = 3$ . Hence the cartesian product space  $\Theta \times \Theta'$  has  $2 \cdot 3 = 6$  elements<sup>1</sup>, and its power set  $2^{\Theta \times \Theta'}$  has  $2^6 = 64$  elements (couples, and unions of couples). If we consider the vacuous BBA  $m_v^{\Theta \times \Theta'}$  on  $2^{\Theta \times \Theta'}$  defined by  $m_v^{\Theta \times \Theta'}(\Theta \times \Theta') = 1$ , then its projection on  $\Theta$  is the vacuous BBA  $m_v^{\Theta}(\Theta) = 1$  defined on the FoD  $\Theta = \{\theta_1, \theta_2\}$  having only two elements, and its projection on  $\Theta'$  is the vacuous BBA  $m_v^{\Theta'}(\Theta') = 1$  defined on the FoD  $\Theta' = \{\theta'_1, \theta'_2, \theta'_3\}$  having only three elements. Why the MoU of  $m_v^{\Theta \times \Theta'}$  (i.e. full ignorant source) related to 6 elements of  $\Theta \times \Theta'$  should be less (or equal) to the sum of MoU of  $m_v^{\Theta}$  related to only the two elements of  $\Theta$  and the MoU of  $m_v^{\Theta'}$  only related to the three elements of  $\Theta'$ ? To amplify this point, if we consider  $|\Theta| = 5$  and  $|\Theta'| = 8$  then  $|\Theta \times \Theta'| = 40$ . Why the MoU of the vacuous BBA  $m_v^{\Theta \times \Theta'}$  related to 40 elements of  $\Theta \times \Theta'$  should be less (or equal) to the sum of MoU of vacuous BBA  $m_v^{\Theta}$  related to only 5 elements of  $\Theta$  and the MoU of the vacuous BBA  $m_v^{\Theta'}$  only related to the 8 elements of  $\Theta'$ ? We do not see any solid theoretical reason, nor intuitive reason, for justifying and requiring the subadditivity desideratum in the general framework of belief functions, and put it as a property to satisfy in general listed in [15]. Unlike Vejnarova and Klir opinions [38] (p.28) and many authors, we do not consider that the meaningful measure of uncertainty of basic belief assignment must satisfy the subadditivity property. The proposal of adding the desiderata of subadditivity, additivity, and monotonicity for a search of a MoU of belief functions had been explored and defended by Klir in [2] at the end of 1980s. It is however worth mentioning that if a MoU satisfies the desideratum D3 (when the BBA is Bayesian), its subadditivity property is always guaranteed because Shannon entropy is subadditive [8], [20].

### III. EXISTING MEASURES OF UNCERTAINTY

In this section we analyze most of existing measures of uncertainty available in the open literature related to belief functions. We verify if these measure pass, or not, the effectiveness test. We say that a MoU fails the effectiveness test if at least one of the desiderata D1, D2, D3 or D4 is not satisfied by the MoU under test. If necessary, we explain what is the problem with this MoU and when necessary we give a counter-example for it.

The Tables I and II show the formulas of all the MoUs analyzed in this work. Some existing MoUs capture only some aspects of uncertainty<sup>2</sup> and have specific names given by their authors (e.g. conflict, dissonance, discord, strife, etc)

<sup>1</sup>Each element is a couple of the form  $(\theta_i, \theta'_j)$ ,  $i = 1, 2$  and  $j = 1, 2, 3$ .

<sup>2</sup>referred to as entropy-like uncertainty, nonspecificity (or imprecision), and fuzziness which is uniquely connected with fuzzy sets [10].

listed in the third column of these tables<sup>3</sup>. For convenience, the MoUs have been indexed and listed by the year of their publication in the tables I and II. We have also included in Tables I and II the name of authors of the MoUs, the names of the MoU when it exists (and eventually new names if needed for clarity), and the formulas of the MoUs. For convenience, we have used the natural log in the mathematical expressions of MoUs for the homogeneity of the presentation. Some authors prefer  $\log_2$  instead, but this preference does not really matter because the values of an expression will differ only from the constant multiplicative factor  $1/\log(2)$ , and the unity will just change from nats to bits.

The Table III indicates if each MoU satisfies, or not, the desiderata D1, D2, D3 and D4, and thus if it passes the effectiveness test, or not. Most of results listed in Table III are easy to verify directly from the mathematical definition of each MoU of Tables I and II, and are left as exercises for the reader. Some results however of Table III, specially those related to the failure of D4 desideratum, may appear less obvious to verify and that is why we give some numerical counter-examples for them in the Tables IV and V for convenience<sup>4</sup>. These counter-examples have been obtained from Monte-Carlo simulation of randomly generated BBAs for testing the desiderata. Of course, many more counter-examples can be found by Monte-Carlo simulation, but of course only one is sufficient to prove the failure of a MoU for a desideratum, specially for D4. Extra justifications about violation of desiderata by some MoUs are presented next.

The MoU<sub>1984</sub>( $m$ ) =  $-\sum_{X \subseteq \Theta} m(X) \log(m(X))$  does not satisfy D2 desideratum because MoU<sub>1984</sub>( $m_v^{\Theta}$ ) = 0 whatever is the size of the FoD  $\Theta$ . Consequently, MoU<sub>1984</sub>( $m$ ) > MoU<sub>1984</sub>( $m_v$ ) if  $m \neq m_v$ , hence D4 desideratum is violated. That is why MoU<sub>1984</sub>( $m$ ) cannot be recommended as an effective measure of uncertainty.

The MoU<sub>1990b</sub>( $m$ ) =  $T(m)$  does not satisfy D4 desideratum because we can have  $m \neq m_v$  such that  $T(m) = T(m_v)$  as shown in the counter-example given in [52] (p.165). See also our simpler counter-example given in Table IV.

The MoU<sub>1992</sub>( $m$ ) =  $S(m)$  (i.e. the strife) does not satisfy D2 desideratum because one can easily verify that one has always<sup>5</sup>  $S(m_v^{\Theta}) = S(m_v^{\Theta'}) = 0$  when  $|\Theta| \neq |\Theta'|$ . The strife does not satisfy D4 either because if  $m$  is the uniform Bayesian BBA on (non-empty) FoD  $\Theta$ , one has  $S(m) = \log(|\Theta|)$  which is greater than zero, proving that  $S(m)$  violates D4.

The MoU<sub>1992b</sub>( $m$ ) =  $NS(m)$  does not satisfy D4 desideratum because we can have  $m \neq m_v$  but such that  $NS(m) = NS(m_v)$ , as shown in the counter-example of Table IV,

<sup>3</sup>The names and notations are not always homogeneous from one author to another, for instance U-uncertainty is also called nonspecificity and denoted by  $N(m)$  in [39]–[41].

<sup>4</sup>The numerical values have been truncated to their 3rd digit.

<sup>5</sup>It is worth noting that Klir's statement, at the bottom of page 86 of [8], saying (using our notation)  $S(m_v) = \log(|\Theta|)$  is clearly wrong.

Table I  
LIST OF EXISTING MOUS FOR THE PERIOD 1980–2000.

Measures of Uncertainty	Author(s) & Ref.	Name	Mathematical expression
MoU <sub>1981</sub> ( $m$ )	Höhle [42], [43]	confusion	$C(m) = - \sum_{X \subseteq \Theta} m(X) \log(Bel(X))$
MoU <sub>1983</sub> ( $m$ )	Yager [44]	dissonance	$E(m) = - \sum_{X \subseteq \Theta} m(X) \log(Pl(X))$
MoU <sub>1983b</sub> ( $m$ )	Yager [44], [45]	nonspecificity	$N(m) = 1 - \sum_{X \subseteq \Theta} m(X)/ X $
MoU <sub>1983c</sub> ( $m$ )	Dubois et al. [46]–[48]	U-uncertainty	$U(m) = \sum_{X \subseteq \Theta} m(X) \log( X )$
MoU <sub>1984</sub> ( $m$ )	Höhle [1], [49], [50]	entropy of discernibleness	$E_m(m) = - \sum_{X \subseteq \Theta} m(X) \log(m(X))$
MoU <sub>1987</sub> ( $m$ )	Dubois et al. [1]	entropy-like index	$C'(m) = - \sum_{X \subseteq \Theta} m(X) \log(q(X))$
MoU <sub>1987b</sub> ( $m$ )	Dubois et al. [1]	index of fuzziness	$d(m) = - \log(\sum_{X \subseteq \Theta} m(X) Bel(X))$
MoU <sub>1988</sub> ( $m$ )	Dubois et al. [51]	imprecision	$l(m) = \sum_{X \subseteq \Theta} m(X) X $
MoU <sub>1988b</sub> ( $m$ )	Lamata et al. [37]	lower entropy	$L_{ent}(m) = E(m) + U(m) = - \sum_{X \subseteq \Theta} m(X) \log(Pl(X)/ X )$
MoU <sub>1988c</sub> ( $m$ )	Lamata et al. [37]	upper entropy	$U_{ent}(m) = - \sum_{X \subseteq \Theta} m(X) \sup\{\log(Pl(\theta_i))   \theta_i \in X\} + \log(\sum_{X \subseteq \Theta} m(X) X )$
MoU <sub>1990</sub> ( $m$ )	Klir et al. [52]–[54]	discord	$D(m) = - \sum_{X \subseteq \Theta} m(X) \log(\sum_{Y \subseteq \Theta} m(Y) \frac{ X \cap Y }{ Y })$
MoU <sub>1990b</sub> ( $m$ )	Klir et al. [52], [53]	total uncertainty	$T(m) = U(m) + D(m)$
MoU <sub>1992</sub> ( $m$ )	Klir et al. [38], [53]	strife	$S(m) = - \sum_{X \subseteq \Theta} m(X) \log(\sum_{Y \subseteq \Theta} m(Y) \frac{ X \cap Y }{ X })$
MoU <sub>1992b</sub> ( $m$ )	Klir et al. [53]		$NS(m) = U(m) + S(m)$
MoU <sub>1993</sub> ( $m$ )	Pal et al. [6]	average total uncertainty	$ATU(m) = - \sum_{X \subseteq \Theta} m(X) \log(m(X)) + U(m)$
MoU <sub>1993b</sub> ( $m$ )	Maeda et al. [55]	Maeda extended entropy	$M(m) = AU(m) + U(m)$
MoU <sub>1994</sub> ( $m$ )	Harmanec et al. [39], [56]	amount of uncertainty	$AU(m) = - \sum_{\theta_i \in \Theta} P^*(\theta_i) \log(P^*(\theta_i))$
MoU <sub>1996</sub> ( $m$ )	George et al. [9], [45]	total conflict	$TC(m) = \sum_{X \subseteq \Theta} m(X) (\sum_{Y \subseteq \Theta} m(Y) [1 - \frac{ X \cap Y }{ X \cup Y }])$
MoU <sub>1997</sub> ( $m$ )	Maluf [57]	Maluf entropy	$H_{ds}(m) = - \sum_{X \subseteq \Theta   m(X) > 0} Pl(X) \log(Bel(X))$
MoU <sub>1999</sub> ( $m$ )	Klir [58]	Shannon-like measure	$SL(m) = - \sum_{\theta_i \in \Theta} \frac{Bel(\theta_i) \log(Bel(\theta_i)) + Pl(\theta_i) \log(Pl(\theta_i))}{\sum_{\theta_j \in \Theta} Bel(\theta_j) + Pl(\theta_j)}$
MoU <sub>2000</sub> ( $m$ )	Yager [59], [60]	Shapley entropy	$H_S(m) = - \sum_{\theta_i \in \Theta} [\sum_{X \subseteq \Theta   \theta_i \in X} \frac{m(X)}{ X }] \log(\sum_{X \subseteq \Theta   \theta_i \in X} \frac{m(X)}{ X })$

where<sup>6</sup>  $U(m) = \log(2)$  and  $S(m) = \log(3) - \log(2)$ , so that  $NS(m) = U(m) + S(m) = \log(3)$ , and we have  $U(m_v) = \log(3)$  and  $S(m_v) = 0$  yielding  $NS(m_v) = \log(3)$ , and hence proving  $NS(m) = NS(m_v)$ .

The MoU<sub>1994</sub>( $m$ ) =  $AU(m)$  proposed by Harmanec and Klir [39], [40] is nothing but the maximal Shannon entropy value obtained by analyzing all the pmfs  $P(\cdot)$  compatible with  $Bel(\cdot)$  and  $Pl(\cdot)$  functions of the BBA  $m(\cdot)$  such that for all  $X \subseteq \Theta$ ,  $Bel(X) \leq \sum_{\theta_i \in X} P(\theta_i) \leq Pl(X)$ . More precisely,

$$P^*(\cdot) = \arg \max_{\substack{\text{All compatible} \\ \text{pmf } P(\cdot)}} - \sum_{\theta_i \in \Theta} P(\theta_i) \log(P(\theta_i))$$

This max-entropy pmf  $P^*(\cdot)$  is obtained by solving a non linear optimization problem, see [86]–[88]. It is clear that this

<sup>6</sup>The easy verification from  $U(m)$  and  $S(m)$  formulas is left to the reader.

MoU, as well as all other Shannon-like entropy measures based on different probabilistic approximations techniques<sup>7</sup> (as BetP-entropy, PIPr-entropy, or DSMP-entropy, etc) of (non-bayesian) BBA  $m$  to a bayesian BBA fail to satisfy D4 desideratum. Indeed, the vacuous BBA  $m_v$  will always be approximated by the uniform pmf  $P^{unif}(\cdot)$  defined on the FoD  $\Theta$ , and there will be no difference between the Shannon-like entropy value for  $m_v$  (for the total ignorant source of evidence) and the Shannon-like entropy value of the Bayesian uniform BBA. This explains why  $AU(m)$  and all other Shannon-like entropies violate the D4 desideratum.

The MoU<sub>1996</sub>( $m$ ) =  $TC(m)$  violates D2 because  $TC(m_v) = 0$  whatever is the dimension of the (non-

<sup>7</sup>BetP<sub>m</sub>, DSMP<sub>m</sub> and PIPr<sub>m</sub> are different probabilistic transformations of a non-Bayesian BBA into a Bayesian one. They have been proposed by different authors, see in [89]–[91] for details.

Table II  
LIST OF EXISTING MOUS FOR THE PERIOD 2001–2021.

Measures of Uncertainty	Author(s) & Ref.	Name	Mathematical expression
MoU <sub>2003</sub> ( <i>m</i> )	Dezert et al. [61], Jusselme et al. [11]	Pignistic entropy or BetP-entropy or Ambiguity measure	$AM(m) = - \sum_{\theta_i \in \Theta} BetP_m(\theta_i) \log(BetP_m(\theta_i))$
MoU <sub>2016</sub> ( <i>m</i> )	Deng [62]	Deng entropy	$E_d(m) = - \sum_{X \subseteq \Theta} m(X) \log \left( \frac{m(X)}{2^{ X -1}} \right)$
MoU <sub>2016b</sub> ( <i>m</i> )	Yang et al. [63]	total uncertainty, $d^I(\cdot, \cdot)$ is Wasserstein distance	$TU^I(m) = 1 - \frac{\sqrt{3}}{ \Theta } \sum_{\theta_i \in \Theta} d^I([Bel(\theta_i), Pl(\theta_i)], [0, 1])$
MoU <sub>2017</sub> ( <i>m</i> )	Deng et al. [64]	improved $TU^I(m)$ with Euclidean distance $d_E^I$	$iTU^I(m) = \sum_{\theta_i \in \Theta} [1 - d_E^I([Bel(\theta_i), Pl(\theta_i)], [0, 1])]$
MoU <sub>2017b</sub> ( <i>m</i> )	Zhou et al. [65]–[67]	improved Deng entropy	$E_{Id}(m) = - \sum_{X \subseteq \Theta} m(X) \log \left( \frac{m(X)}{2^{ X -1}} e^{\frac{ X -1}{ \Theta }} \right)$
MoU <sub>2017c</sub> ( <i>m</i> )	Tang et al. [68]	Tang weighted belief entropy	$E_{Wd}(m) = - \sum_{X \subseteq \Theta} \frac{ X }{ \Theta } m(X) \log \left( \frac{m(X)}{2^{ X -1}} \right)$
MoU <sub>2018</sub> ( <i>m</i> )	Jiroušek et al. [69]	Extended PlPr-entropy	$H_{PlPr}^{ext}(m) = - \sum_{\theta_i \in \Theta} PlPr_m(\theta_i) \log(PlPr_m(\theta_i)) + U(m)$
MoU <sub>2018b</sub> ( <i>m</i> )	Jiroušek et al. [70], [71]	<i>q</i> -entropy	$H_q(m) = \sum_{X \subseteq \Theta} (-1)^{ X } q(X) \log(q(X))$
MoU <sub>2018c</sub> ( <i>m</i> )	Mambé et al. [72]	Mambé entropy	$E_{Nm}(m) = - \sum_{X \subseteq \Theta} m(X) \log \left( \frac{m(X)}{2^{ X -1}} e^{\frac{ X -1}{2 \Theta }} \right)$
MoU <sub>2018d</sub> ( <i>m</i> )	Pan et al. [73]	Pan 1st entropy	$H_{bel}(m) = - \sum_{X \subseteq \Theta} \frac{Bel(X)+Pl(X)}{2} \log \left( \frac{Bel(X)+Pl(X)}{2(2^{ X -1})} \right)$
MoU <sub>2018e</sub> ( <i>m</i> )	Wang et al. [74]	Wang entropy	$SU(m) = \sum_{\theta_i \in \Theta} \left[ -\frac{Bel(\theta_i)+Pl(\theta_i)}{2} \log_2 \left( \frac{Bel(\theta_i)+Pl(\theta_i)}{2} \right) + \frac{Pl(\theta_i)-Bel(\theta_i)}{2} \right]$
MoU <sub>2019</sub> ( <i>m</i> )	Li et al. [75]	Li entropy	$IQ(m) = \sum_{X \subseteq \Theta} \left( \frac{m(X)}{2^{ X -1}} \right)^2$
MoU <sub>2019b</sub> ( <i>m</i> )	Cui et al. [76]	Cui entropy	$E_{Cui}(m) = - \sum_{X \subseteq \Theta} m(X) \log \left( \frac{m(X)}{2^{ X -1}} \cdot e^{\sum_{Y \subseteq \Theta, Y \neq X \& m(Y) > 0} \frac{ X \cap Y }{2^{ \Theta -1}}} \right)$
MoU <sub>2019c</sub> ( <i>m</i> )	Pan et al. [77]	Pan 2nd entropy	$H_{PQ}(m) = - \sum_{X \subseteq \Theta} m(X) \log \left( \sum_{\theta_i \in X} PlPr_m(\theta_i) \right) + U(m)$
MoU <sub>2019d</sub> ( <i>m</i> )	Chen et al. [78]	Chen entropy	$E_i(m) = - \sum_{X \subseteq \Theta} m(X) \log \left( \frac{m(X)}{2^{ X -1}} \cdot \prod_{\substack{Y \subseteq \Theta \\ m(Y) > 0}} \frac{ X }{ Y } \right)$
MoU <sub>2019e</sub> ( <i>m</i> )	Zhao et al. [79]	Zhao entropy	$H_{inter}(m) = - \sum_{\theta_i \in \Theta} \frac{Bel(\theta_i)+Pl(\theta_i)}{2} \log \left( \frac{Bel(\theta_i)+Pl(\theta_i)}{2} e^{-(Pl(\theta_i)-Bel(\theta_i))} \right) - \sum_{X \subseteq \Theta,  X  > 1} m(X) \log \left( \frac{m(X)}{2^{ X -1}} e^{-(Pl(X)-Bel(X))} \right)$
MoU <sub>2020</sub> ( <i>m</i> )	Li et al. [80]	Li improved entropy	$IQ_{Li}(m) = \sum_{X \subseteq \Theta} \left( \frac{m(X)}{2^{ X -1}} \right)^2 \cdot e^{\sum_{Y \subseteq \Theta, Y \neq X} \frac{ X \cap Y }{ \Theta }}$
MoU <sub>2020b</sub> ( <i>m</i> )	Wen et al. [81]	Wen entropy	$U_{exp}(m) = \frac{1}{e^{-\frac{1}{ \Theta ^2}} e^{\frac{1}{ \Theta ^2}}} \left[ e - \sum_{\substack{X \subseteq \Theta \\  X =1}} m(X) e^{m(X)} - \sum_{\substack{X \subseteq \Theta \\  X  \neq 1}} \frac{m(X)}{ X } \prod_{\substack{Y \subseteq \Theta \\ m(Y) > 0}} \frac{ X }{ Y } e^{\frac{m(X)}{m(Y)}} \right]$
MoU <sub>2020c</sub> ( <i>m</i> )	Chen et al. [82]	Chen improved entropy	$E_{Wd}^C(m) = - \sum_{X \subseteq \Theta} \frac{ X }{ \Theta } \frac{1-m(X)}{ F_{\Theta}(m) -1} \log \left( \frac{1-m(X)}{ F_{\Theta}(m) -1} \cdot \frac{1}{2^{ X -1}} \right)$
MoU <sub>2020d</sub> ( <i>m</i> )	Qin et al. [83]	Qin entropy	$Q(m) = E_m(m) + \sum_{X \subseteq \Theta} \frac{ X }{ \Theta } m(X) \log( X )$
MoU <sub>2020e</sub> ( <i>m</i> )	Yan et al. [84]	Yan entropy	$H_n(m) = - \sum_{X \subseteq \Theta} m(X) \log \left( \frac{m(X)+Bel(X)}{2} \cdot \frac{1}{2^{ X -1}} \cdot e^{\frac{ X -1}{ C(m) }} \right)$
MoU <sub>2020f</sub> ( <i>m</i> )	Li et al. [85]	Li-Pan entropy	$H_{BF}(m) = E_m(m) +  \Theta  \cdot U(m)$
MoU <sub>2021</sub> ( <i>m</i> )	This paper	Extended BetP-entropy	$H_{BetP}^{ext}(m) = - \sum_{\theta_i \in \Theta} BetP_m(\theta_i) \log(BetP_m(\theta_i)) + U(m)$
MoU <sub>2021b</sub> ( <i>m</i> )	This paper	Extended DSMP-entropy	$H_{DSMP}^{ext}(m) = - \sum_{\theta_i \in \Theta} DSMP_m(\theta_i) \log(DSMP_m(\theta_i)) + U(m)$

Table III  
DESIDERATA VERIFICATION, AND EFFECTIVENESS TEST RESULTS.

Measures of Uncertainty	D1 $\min U(m) = 0$ for $m(\theta_i) = 1$	D2 $U(m_v^\Theta) < U(m_v^{\Theta'})$ if $ \Theta  <  \Theta' $	D3 $U(m)$ is Shannon entropy for Bayesian BBA	D4 $U(m) < U(m_v)$ if $m \neq m_v$	Effectiveness test result
MoU <sub>1981</sub> ( $m$ ) = $C(m)$	yes	no	yes	no	failed
MoU <sub>1983</sub> ( $m$ ) = $E(m)$	yes	no	yes	no	failed
MoU <sub>1983b</sub> ( $m$ ) = $N(m)$	yes	yes	no	yes	failed
MoU <sub>1983c</sub> ( $m$ ) = $U(m)$	yes	yes	no	yes	failed
MoU <sub>1984</sub> ( $m$ ) = $E_m(m)$	yes	no	yes	no	failed
MoU <sub>1987</sub> ( $m$ ) = $C'(m)$	yes	no	yes	no	failed
MoU <sub>1987b</sub> ( $m$ ) = $d(m)$	yes	no	no	no	failed
MoU <sub>1988</sub> ( $m$ ) = $l(m)$	no	yes	yes	yes	failed
MoU <sub>1988b</sub> ( $m$ ) = $L_{ent}(m)$	yes	yes	yes	no	failed
MoU <sub>1988c</sub> ( $m$ ) = $U_{ent}(m)$	yes	yes	yes	no	failed
MoU <sub>1990</sub> ( $m$ ) = $D(m)$	yes	no	yes	no	failed
MoU <sub>1990b</sub> ( $m$ ) = $T(m)$	yes	yes	yes	no	failed
MoU <sub>1992</sub> ( $m$ ) = $S(m)$	yes	no	yes	no	failed
MoU <sub>1992b</sub> ( $m$ ) = $NS(m)$	yes	yes	yes	no	failed
MoU <sub>1993</sub> ( $m$ ) = $ATU(m)$	yes	yes	yes	no	failed
MoU <sub>1993b</sub> ( $m$ ) = $M(m)$	yes	yes	yes	yes	okay
MoU <sub>1994</sub> ( $m$ ) = $AU(m)$	yes	yes	yes	no	failed
MoU <sub>1996</sub> ( $m$ ) = $TC(m)$	yes	no	no	no	failed
MoU <sub>1997</sub> ( $m$ ) = $H_{ds}(m)$	yes	no	yes	no	failed
MoU <sub>1999</sub> ( $m$ ) = $SL(m)$	yes	no	yes	no	failed
MoU <sub>2000</sub> ( $m$ ) = $H_S(m)$	yes	yes	yes	no	failed
MoU <sub>2003</sub> ( $m$ ) = $AM(m)$	yes	yes	yes	no	failed
MoU <sub>2016</sub> ( $m$ ) = $E_d(m)$	yes	yes	yes	no	failed
MoU <sub>2016b</sub> ( $m$ ) = $TU^I(m)$	yes	no	yes	yes	failed
MoU <sub>2017</sub> ( $m$ ) = $iTU^I(m)$	yes	yes	no	yes	failed
MoU <sub>2017b</sub> ( $m$ ) = $E_{Id}(m)$	yes	yes	yes	no	failed
MoU <sub>2017c</sub> ( $m$ ) = $E_{Wd}(m)$	yes	yes	no	no	failed
MoU <sub>2018</sub> ( $m$ ) = $H_{PIPr}^{ext}(m)$	yes	yes	yes	yes	okay
MoU <sub>2018b</sub> ( $m$ ) = $H_q(m)$	no	no	yes	no	failed
MoU <sub>2018c</sub> ( $m$ ) = $E_{Nm}(m)$	yes	yes	yes	no	failed
MoU <sub>2018d</sub> ( $m$ ) = $H_{bel}(m)$	no	yes	no	no	failed
MoU <sub>2018e</sub> ( $m$ ) = $SU(m)$	yes	yes	yes	yes	okay
MoU <sub>2019</sub> ( $m$ ) = $IQ(m)$	yes	no	no	no	failed
MoU <sub>2019b</sub> ( $m$ ) = $EC_{ui}(m)$	yes	yes	yes	no	failed
MoU <sub>2019c</sub> ( $m$ ) = $HPQ(m)$	yes	yes	yes	no	failed
MoU <sub>2019d</sub> ( $m$ ) = $E_i(m)$	yes	yes	no	no	failed
MoU <sub>2019e</sub> ( $m$ ) = $H_{inter}(m)$	yes	yes	yes	no	failed
MoU <sub>2020</sub> ( $m$ ) = $IQ_{Li}(m)$	no	yes	no	no	failed
MoU <sub>2020b</sub> ( $m$ ) = $U_{exp}(m)$	yes	no	no	yes	failed
MoU <sub>2020c</sub> ( $m$ ) = $E_{Wd}^C(m)$	no (NaN)	no (NaN)	no	no (NaN)	failed
MoU <sub>2020d</sub> ( $m$ ) = $Q(m)$	yes	yes	yes	no	failed
MoU <sub>2020e</sub> ( $m$ ) = $H_n(m)$	yes	yes	no	no	failed
MoU <sub>2020f</sub> ( $m$ ) = $H_{BF}(m)$	yes	yes	yes	no	failed
MoU <sub>2021</sub> ( $m$ ) = $H_{BetP}^{ext}(m)$	yes	yes	yes	yes	okay
MoU <sub>2021b</sub> ( $m$ ) = $H_{DSMP}^{ext}(m)$	yes	yes	yes	yes	okay

Table IV  
COUNTER-EXAMPLES FOR SOME MoUs ON  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ .

Elem. of $2^\Theta$	$m$	$m$	$m$	$m$	$m$	$m$
$\emptyset$	0	0	0	0	0	0
$\theta_1$	1/3	0.18	0	0	0	0.22
$\theta_2$	1/3	0.17	0	0	0	0.09
$\theta_1 \cup \theta_2$	0	0.32	1/3	1/3	1/3	0.02
$\theta_3$	1/3	0.31	0	0	0	0.01
$\theta_1 \cup \theta_3$	0	0	1/3	1/3	1/3	0.11
$\theta_2 \cup \theta_3$	0	0	1/3	1/3	1/3	0.17
$\theta_1 \cup \theta_2 \cup \theta_3$	0	0.02	0	0	0	0.38
MoU( $m$ )	$L_{ent}(m) = 1.098$	$U_{ent}(m) = 1.105$	$T(m) = \log(3)$	$NS(m) = \log(3)$	$ATU(m) = 1.791$	$H_{BF}(m) = 3.462$
MoU( $m_v$ )	$L_{ent}(m_v) = 1.098$	$U_{ent}(m_v) = 1.098$	$T(m_v) = \log(3)$	$NS(m_v) = \log(3)$	$ATU(m_v) = 1.098$	$H_{BF}(m_v) = 3.295$

Table V  
COUNTER-EXAMPLES FOR SOME MOUS ON  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ .

Elem. of $2^\Theta$	$m$	$m$	$m$	$m$	$m$	$m$
$\emptyset$	0	0	0	0	0	0
$\theta_1$	0	0	0.10	0	0.11	0.12
$\theta_2$	0	0	0.10	0	0.26	0.09
$\theta_1 \cup \theta_2$	1/3	1/3	0.16	1/3	0.24	0.17
$\theta_3$	0	0	0.03	0	0.01	0.04
$\theta_1 \cup \theta_3$	1/3	1/3	0.06	1/3	0.04	0.15
$\theta_2 \cup \theta_3$	1/3	1/3	0.21	1/3	0.15	0.23
$\theta_1 \cup \theta_2 \cup \theta_3$	0	0	0.34	0	0.19	0.20
MoU( $m$ )	$E_d(m) = 2.197$	$E_{Id}(m) = 1.863$	$E_{Wd}(m) = 2.058$	$E_{Nm}(m) = 2.072$	$H_n(m) = 1.795$	$E_{Cui}(m) = 2.003$
MoU( $m_v$ )	$E_d(m_v) = 1.945$	$E_{Id}(m_v) = 1.279$	$E_{Wd}(m_v) = 1.945$	$E_{Nm}(m_v) = 1.695$	$H_n(m_v) = 1.279$	$E_{Cui}(m_v) = 1.945$

empty) FoD  $\Theta$ . It violates D3, because for Bayesian BBA one gets  $TC(m) = \sum_{i=1}^n P(\theta_i)(1 - P(\theta_i))$  as reported in [45]. It also violates D4 in general because for Bayesian BBA one has  $TC(m) > 0$ , except in the particular Bayesian case where the BBA is entirely focused on a singleton  $\theta_i$  (i.e.  $m(\theta_i) = 1$ ). In this particular case we obtain  $TC(m) = TC(m_v) = 0$ . So for all Bayesian BBAs  $m$  we will always have  $TC(m) \geq TC(m_v)$ , which clearly violates D4 desideratum.

The original formula of  $MoU_{1997}(m) = H_{ds}(m)$  proposed by Maluf in [57] was actually  $H_{ds}(m) = -\sum_{X \subseteq \Theta} Pl(X) \log(Bel(X))$  which is obviously ill-defined when  $Pl(X) > 0$  and  $Bel(X) = 0$  because  $\log(0) = -\infty$ . That is why we did consider only focal elements of the BBAs  $m$  in the modified formula  $H_{ds}(m)$  given in Table I. For any cardinality of non-empty FoD  $\Theta$  we have always  $H_{ds}(m_v) = 0$  because for the vacuous BBA  $m_v$ , the only focal element is  $\Theta$  for which  $Bel(\Theta) = Pl(\Theta) = 1$  so that  $H_{ds}(m_v) = -Pl(\Theta) \log(Bel(\Theta)) = -1 \log(1) = 0$ . So,  $H_{ds}(m)$  violates D2. This MoU violates also D4 because for Bayesian BBA  $H_{ds}(m)$  is the same as Shannon entropy, and Shannon entropy is greater than zero in general.

The  $MoU_{2000}(m) = H_S(m)$  (Shapley entropy) coincides with Shannon entropy for Bayesian BBAs, and one can easily verify that  $H_S(m_v) = \log(|\Theta|)$  which is also the same maximum value of Shannon entropy for the uniform Bayesian BBA. Hence  $H_S(m_v)$  is not the unique maximum measure of uncertainty value when we use Shapley entropy. Also it can be verified that this maximum value can be also obtained by non-Bayesian BBA. For instance, if  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and  $m(\theta_1 \cup \theta_2) = m(\theta_1 \cup \theta_3) = m(\theta_2 \cup \theta_3) = 1/3$ , then  $H_S(m) = \log(3)$ , which is also the same value as for  $H_S(m_v)$ . Because Shapley entropy proposed by Yager violates D4 desideratum, we cannot recommend it as an effective MoU.

The  $MoU_{2016}(m) = E_d(m)$  (Deng entropy) has recently aroused the interest and enthusiasm of some researchers because it was highly publicized by Deng during the last five years [14]. We really wonder about such strong interest of this MoU because Deng entropy is obviously not effective,

as proved by our simple counter-example given in Table V. Abellán has already pointed out the problem of Deng entropy in [92]. Nevertheless, some researchers try to use it, publicize it or improve it unsuccessfully as shown in our analysis summarized in Table III. So, it is clear that Deng Entropy is not recommended for applications, as well as other generalizations (modifications or extensions) of it, as those recently proposed by the same author (Rényi-Deng (R-D) entropy, Tsallis-Deng (T-D) entropy, Rényi-Tsallis-Deng (R-T-D) entropy, Interval-valued Deng entropy, Fractal-based belief Deng entropy, Deng entropy for orderable set, etc), see for instance [93], [94] because they do not have interest since they are non-effective. We emphasize that even if a MoU collapses with Shannon entropy (as Deng entropy does) when a BBA is Bayesian, it can be non-effective and useless if it violates D4 desideratum. That is why Deng entropy (and all its recent variants based on it) is non-effective as most of other MoUs actually reported in Table III.

The  $MoU_{2018b}(m) = H_q(m)$  ( $q$ -entropy alike) violates D1 because  $H_q(m)$  can be negative so its minimum value is not zero. For instance if  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and  $m(\theta_1 \cup \theta_2) = m(\theta_1 \cup \theta_3) = m(\theta_2 \cup \theta_3) = 1/3$ , then  $H_q(m) \approx -0.2877$ . This MoU also violates D2 because  $H_q(m_v) = 0$  whatever is the dimension of the (non-empty) FoD  $\Theta$ . This MoU collapses with Shannon entropy because if  $m$  is a Bayesian BBA one has  $q(X) = m(X)$  for all  $X \subseteq \Theta$ , and the focal elements of  $m$  are necessarily singletons  $X \subseteq \Theta$  for which  $|X| = 1$ , so that  $(-1)^{|X|} = -1$ , and consequently the mathematical definition of  $H_q(m)$  given in Table I is same as Shannon entropy. This MoU violates D4 because for Bayesian BBA  $H_q(m)$  is the same as Shannon entropy, and Shannon entropy is greater than zero in general<sup>8</sup>. For instance if  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and  $m(\theta_1) = m(\theta_2) = m(\theta_3) = 1/3$ , then  $H_q(m) = \log(|\Theta|) = \log(3) > 0$ . Hence  $H_q(m) > H_q(m_v)$ .

The  $MoU_{2018d}(m) = H_{bel}(m)$  (Pan 1st entropy) violates D1 because if we consider the simplest case of FoD with  $\Theta = \{\theta_1, \theta_2\}$ , and the specific BBA  $m(\theta_1) = 1$ , we have  $[Bel(\theta_1), Pl(\theta_1)] = [1, 1]$ ,  $[Bel(\theta_2), Pl(\theta_2)] = [0, 0]$  and  $[Bel(\theta_1 \cup \theta_2), Pl(\theta_1 \cup \theta_2)] = [1, 1]$ , so we have

<sup>8</sup>except in the case where  $m(\theta_i) = 1$  for some  $\theta_i \in \Theta$ .

$(Bel(\theta_1) + Pl(\theta_1))/2 = 1$ ,  $(Bel(\theta_2) + Pl(\theta_2))/2 = 0$  and  $(Bel(\theta_1 \cup \theta_2) + Pl(\theta_1 \cup \theta_2))/2 = 1$ . Hence  $H_{bel}(m) = -1 \log(1/(2^1 - 1)) - 0 \log(1/(2^1 - 1)) - 1 \log(1/(2^2 - 1)) = \log(3) > 0$ . Pan 1st entropy violates D3 (Shannon entropy consistency) too because if  $m$  is the uniform Bayesian BBA given by  $m(\theta_1) = m(\theta_2) = 0.5$ , then  $H_{bel}(m) = -0.5 \log(0.5) - 0.5 \log(0.5) - 1 \log(1/3) = \log(2) + \log(3)$  which is greater than Shannon entropy which is equal to  $-0.5 \log(0.5) - 0.5 \log(0.5) = \log(2)$ . Pan 1st entropy violates D4 also because for the vacuous BBA  $m_v(\theta_1 \cup \theta_2) = 1$ , one has  $[Bel(\theta_1), Pl(\theta_1)] = [0, 1]$ ,  $[Bel(\theta_2), Pl(\theta_2)] = [0, 1]$  and  $[Bel(\theta_1 \cup \theta_2), Pl(\theta_1 \cup \theta_2)] = [1, 1]$ , and  $(Bel(\theta_1) + Pl(\theta_1))/2 = 0.5$ ,  $(Bel(\theta_2) + Pl(\theta_2))/2 = 0.5$  and  $(Bel(\theta_1 \cup \theta_2) + Pl(\theta_1 \cup \theta_2))/2 = 1$ , so that  $H_{bel}(m_v) = -0.5 \log(0.5) - 0.5 \log(0.5) - 1 \log(1/3) = \log(2) + \log(3)$  which is the same value as for uniform Bayesian BBA, so  $H_{bel}(m_v)$  is not strictly greater than other Pan 1st entropy values.

The formula of  $MoU_{2018e}(m) = SU(m)$  (Wang entropy) has been kept with its original formulation (with  $\log_2(\cdot)$  function) in Table II, so it is expressed in bits. If one wants to express  $SU(m)$  in nats we must replace  $\log_2(\cdot)$  function by the natural logarithm function  $\log(\cdot)$  and the second terms  $(Pl(\theta_i) - Bel(\theta_i))/2$  must be multiplied by  $\log(2)$  in the mathematical definition of  $SU(m)$ .

For the  $MoU_{2019b}(m) = E_{Cui}(m)$  (Cui entropy) proposed in [76], it is clear that the original mathematical definition of this entropy does not fit with the derivations of what the authors have in mind when making their numerical examples in their paper because of a mistake in their exponential term. That is why we have to correct this term by replacing  $\sum_{\substack{Y \subseteq \Theta \\ Y \neq X}}^{Y \subseteq \Theta}$  by  $\sum_{\substack{Y \subseteq \Theta \\ Y \neq X \& m(Y) > 0}}$  in the original formula. Cui entropy violates D4 desideratum as shown in the example of Table V.

The  $MoU_{2019c}(m) = H_{PQ}(m)$  (Pan 2nd entropy) is not effective because  $H_{PQ}(m_v)$  coincides with  $H_{PQ}(m)$  when  $m$  is the uniform Bayesian BBA, so it violates D4 desideratum.

The  $MoU_{2019d}(m) = E_i(m)$  (Chen entropy) is not effective because one can have  $E_i(m) > E_i(m_v)$ . For instance, consider the vacuous BBA  $m_v$  on FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ , then  $E_i(m_v) = \log(2^{|\Theta|} - 1) = \log(7) = 1.9459$ , and if one considers the uniform Bayesian BBA for which  $m(\theta_1) = m(\theta_2) = m(\theta_3) = 1/3$  one gets  $E_i(m) = -\log(\frac{1}{3} \cdot \frac{1}{3}) = 2 \log(3) = 2.1972 > E_i(m_v)$ . So, Chen entropy violates D4 desideratum.

The  $MoU_{2019e}(m) = H_{inter}(m)$  (Zhao entropy) is not effective because it violates D4 desideratum. As simple counter-example, consider  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  with the BBA  $m(\theta_1 \cup \theta_2) = m(\theta_1 \cup \theta_3) = m(\theta_2 \cup \theta_3) = 1/3$ , then  $H_{inter}(m) = 4.6291$  nats, where as for vacuous BBA  $m_v(\Theta) = 1$  we get  $H_{inter}(m_v) = 4.4856$  nats. Clearly  $H_{inter}(m) > H_{inter}(m_v)$ , which does not make sense because the vacuous BBA  $m_v$  characterizes the most ignorant source of evidence.

It is worth mentioning that the numerical examples given by Li and Cui in their paper are incorrect because they are inconsistent with their original new entropy formula (12) for  $IQ_{mi}$ , see [80]. If we admit that the original Li's definition of entropy is correct then we get the effectiveness test results listed for this entropy in Table III, and we conclude that the  $MoU_{2020}(m) = IQ_{Li}(m)$  (Li improved entropy) is not effective. If we consider that numerical examples by Li and Cui are correct, then we need to modify the exponent term in the original Li's definition (12) of  $IQ_{mi}$  as  $\sum_{\substack{Y \subseteq \Theta \\ Y \neq X \& m(Y) > 0}} |X \cap Y|/|\Theta|$ . In this case the effectiveness test result is worse because this modified Li improved entropy will fail to pass the four desiderata, and it is still non-effective.

The  $MoU_{2020b}(m) = U_{exp}(m)$  (Wen entropy) violates clearly Shannon entropy compatibility desideratum D2, and for the vacuous BBA  $m_v$  one has always  $U_{exp}(m_v) = 1$  whatever is the dimension of the FoD  $\Theta$ . Therefore Wen entropy does not verify desideratum D2. It is not certain that  $U_{exp}(m)$  satisfies, or not, D4 desideratum, but we did thousands of Monte Carlo tests with random BBAs for different size of FoD  $\Theta$ , and  $U_{exp}(m)$  did always pass successfully the D4 test, so we conjecture that Wen entropy satisfies D4. Even if our conjecture about satisfaction of D4 for  $U_{exp}(m)$  is wrong, it does not change our conclusion that Wen entropy is not effective because it fails to verify D2 and D3.

$MoU_{2020c}(m) = E_{Wd}^C(m)$  (Chen improved entropy) is not mathematically well-defined because when the BBA  $m$  has only one focal element (i.e.  $|\mathcal{F}_\Theta(m)| = 1$ ), then one has a division by  $|\mathcal{F}_\Theta(m)| - 1 = 0$  which yields a NaN (Not a Number) indeterminate value in Table III. Even if  $|\mathcal{F}_\Theta(m)| > 1$  this entropy is not compatible with Shannon entropy for Bayesian BBAs. So, Chen improved entropy is not effective.

$MoU_{2020d}(m) = Q(m)$  (Qin entropy) violates D4 desideratum because Qin entropy takes same value  $\log(|\Theta|)$  for the vacuous BBA and for the uniform Bayesian BBA.

$MoU_{2020e}(m) = H_n(m)$  (Yan entropy) is non-effective. A counter-example for D4 desideratum is given in Table V expressed in nats. To express them in bits we have of course to divide our results by  $\log(2)$ . It is worth noting that in Section III.B of [84], the numerical results given by Yan and Deng for  $H_n(m_3)$  and  $H_n(m_4)$  for their example 5 are wrong.

$MoU_{2020f}(m) = H_{BF}(m)$  (Li-Pan entropy) is also non-effective. A counter-example for D4 desideratum is given in Table IV.

#### IV. DISCUSSION

Our analysis of forty-five measures of uncertainty listed in Tables I and II covering 40 years of research in this field reveals that almost 89 % of them are non-effective because they violate at least one of the four very essential desiderata D1, D2, D3 or D4. In our analysis only five MoUs ( $M(m)$  1993,  $H_{PIPr}^{ext}(m)$  2018,  $SU(m)$  2018,  $H_{BetP}^{ext}(m)$  2021,  $H_{DSmP}^{ext}(m)$  2021) pass successfully the effectiveness

test as we can observe in Table III. We see that all these effective MoUs share two basic principles: 1) approximate the BBA  $m$  by a probability measure (i.e. a Bayesian BBA)  $P_m$  based on some method and evaluate its Shannon entropy to estimate the randomness (or conflict) inherent to the BBA, and 2) add a term to Shannon entropy value that estimates the level of ambiguity (or non specificity) inherent of the BBA (usually thanks to Dubois & Prade  $U$ -uncertainty). This general principle is simple and quite intuitive but it lacks seriously of theoretical justification. We consider such type of effective MoU construction is unfortunately conceptually flawed and not very satisfactory for the two following reasons.

- **1st reason:** These effective MoUs highly depend on the choice of the method of approximation. This mechanism appears quite arbitrary, and we do not see any strong justification for preferring one of them, either  $P^*$  in  $M(m)$  MoU,  $BetP$  in  $H_{BetP}^{ext}(m)$  MoU,  $DSmP$  in  $H_{DSmP}^{ext}(m)$  MoU, mid belief-interval value  $(Bel(\theta_i) + Pl(\theta_i))/2$  in  $SU(m)$  MoU, etc. Worse, a method of approximation can be totally misleading as for instance Cobb-Shenoy  $PlPr_m$  transformation [90] because the evaluation of probabilities can be inconsistent with belief interval values. More precisely, one can have  $PlPr_m(\theta_i) \notin [Bel(\theta_i), Pl(\theta_i)]$  with Cobb-Shenoy method, which is obviously not reasonable, nor acceptable at all. As a simple counter-example of Cobb-Shenoy transformation, just consider  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  with  $m(\theta_1) = 0.2$  and  $m(\theta_2 \cup \theta_3) = 0.8$ . Then,  $[Bel(\theta_1), Pl(\theta_1)] = [0.2, 0.2]$ ,  $[Bel(\theta_2), Pl(\theta_2)] = [0, 0.8]$  and  $[Bel(\theta_3), Pl(\theta_3)] = [0, 0.8]$ . Applying  $PlPr_m$  transformation, we get  $PlPr_m(\theta_1) = 0.2/(0.2 + 0.8 + 0.8) \approx 0.112$ . Therefore  $PlPr_m(\theta_1) < Bel(\theta_1)$  which shows that  $PlPr_m(\theta_1) \notin [Bel(\theta_1), Pl(\theta_1)]$ . We emphasize the fact that if a method of approximation of a BBA  $m$  by a probability measure  $P_m$  is chosen, it must be at least consistent with belief interval values generated by the BBA  $m$  under concern. Clearly, we cannot recommend Cobb-Shenoy  $PlPr_m$  transformation for building an effective MoU based on aforementioned principles 1) and 2) as  $H_{PIP_r}^{ext}(m)$  MoU proposed recently by Jiroušek and Shenoy based on questionable Shafer semantics and fallacious Dempster's rule arguments.
- **2nd reason:** More fundamentally, we do not see any serious reason which necessitates the arbitrary use of an approximation of any (non-Bayesian) BBA by a Bayesian BBA at first for using Shannon entropy measure as 1st valid principle. Also why do we need, or request, to make the distinction of the two aspects of uncertainty (conflict and non-specificity) in additive manner? This is conceptually very disputable because the randomness (or conflict) and ambiguity (or nonspecificity) are actually interwoven in a subtle way that needs to be explored in deep for a better understanding of the mechanism governing the uncertainty with a better description of the (probably non-additive) link between them.

Very recently however Zhang et al. in [104] did propose three new innovant effective MoUs not based on arbitrary approximation of the BBA by a probability as in the aforementioned effective MoUs. These measures are denoted by  $H^1(m)$ ,  $H^2(m)$  and  $H^3(m)$  and respectively defined by<sup>9</sup>

$$H^1(m) = - \sum_{X \subseteq \Theta} m(X) \log_2(Pl(X)) + \sum_{X \subseteq \Theta} m(X) 2 \log_2(|X|) \quad (5)$$

$$H^2(m) = - \sum_{X \subseteq \Theta} m(X) \log_2(Pl(X)) + \sum_{X \subseteq \Theta} m(X) \log_2(2^{|X|} - 1) \quad (6)$$

$$H^3(m) = - \sum_{X \subseteq \Theta} m(X) \log_2(Pl(X)) + \sum_{\substack{X \subseteq \Theta \\ |X| > 1}} m(X) |X| \quad (7)$$

These new effective MoUs differ conceptually from the previous effective MoUs  $M(m)$ ,  $H_{PIP_r}^{ext}(m)$ ,  $SU(m)$ ,  $H_{BetP}^{ext}(m)$  and  $H_{DSmP}^{ext}(m)$  but the authors fail to capture well the interwoven link between conflict and non-specificity (or imprecision). Actually the authors set arbitrarily the range of their MoU as a simple parameter, either taken as  $[0, 2 \log_2(|\Theta|)]$ ,  $[0, \log_2(2^{|\Theta|} - 1)]$  or  $[0, |\Theta|]$ , to define their  $H^1(m)$ ,  $H^2(m)$  and  $H^3(m)$  measures of uncertainty. This approach is rather ad-hoc and very questionable, and possibly other ranges could have been chosen instead. The authors do not identify (or propose) the best MoU to select between  $H^1(m)$ ,  $H^2(m)$  and  $H^3(m)$  which is a serious problem for using them in applications. Which one to choose? The other serious problem in this approach is the lack of solid justification for using the plausibility function in the summation  $-\sum_{X \subseteq \Theta} m(X) \log_2(Pl(X))$ . Although effective, these three new MoUs are actually ill-justified and heuristically defined, and somehow they can be considered as conceptually flawed.

## V. CONCLUSION

In this paper we have clearly proved that most of existing measures of uncertainty proposed during the last forty years are actually non-effective, and we consider that the effective ones are conceptually defective. We emphasize the fact that in this jungle of non-effective measures, many of them have bloomed like mushrooms since 2016 with the publication of Deng's paper because of its high publicity. Most of papers since 2016 do not pay attention to the four essential properties that an effective MoU must satisfy, which is a serious problem. We regret this matter of fact, and we hope that this paper have pointed out clearly this concern, and also that it will help to reduce the proliferation of useless publications about non-effective MoUs. We encourage the future authors working on

<sup>9</sup>We correct here the definition of  $H^3(m)$  which is mathematically badly formulated in [104].

new MoUs to verify the effectiveness of their MoU as done recently by Zhang et al. in [104]. We agree with Abellán, Mantas and E. Bossé vision that an (effective) MoU should not be too complicate to calculate (with direct simple explicit mathematical formula), must obviously incorporate the two aspects of uncertainty (in a subtle and efficient interwoven manner), and must be sensitive to changes of evidence. Recently, we have developed in [105] a better conceptual effective measure of uncertainty for the basic belief assignments not based on the additive decomposition of conflict and non-specificity which, we hope, will attract the attention of all readers interested by this topic for their own applications.

## APPENDIX

### Shannon entropy

Consider a random variable represented by a probability mass function (pmf)  $P_N = (p_1, p_2, \dots, p_N)$ , where  $p_i = P(\theta_i)$  is the probability of the  $i$ -th state  $\theta_i$  (i.e. outcome) of  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ . Shannon was interested in communication systems where the various events were the carriers of coded messages, and he did propose (and justify) his entropy measure as appropriate measure of average uncertainty (or measure of randomness) of a random variable [17], [18], [21], [22]. In the classical information theory, the entropy of a random variable is the average level of *surprisal*, or *uncertainty* inherent in the variable's possible outcomes [95]. It is worth noting that Shannon theory does not concern the semantic aspects of the content of a message [46], [96], [97], but only its transmission through communication systems. Shannon entropy formula is defined by<sup>10</sup>

$$H(P_N) \triangleq - \sum_{i=1}^{|\Theta|} P(\theta_i) \log(P(\theta_i)) \quad (8)$$

By convention, we take  $P(\theta_i) \log(P(\theta_i)) \triangleq 0$  if  $P(\theta_i) = 0$  which is easily justified by continuity since  $x \log(x) \rightarrow 0$  as  $x \rightarrow 0$ . Adding terms of zero probability does not change the entropy. In (8) we use the natural logarithm (i.e. base  $e$  logarithm) and in this case the Shannon entropy value is expressed in *nats* unity. We can also use the base 2 logarithm ( $\log_2$ ) function instead of the natural logarithm, and if so the Shannon entropy value will be expressed in *bits*. In this case, the entropy is the number of bits on average required to describe the random variable, or equivalently the minimum expected number of binary questions required to determine the value of the random variable.

Shannon entropy can be interpreted as a generalization of Hartley entropy (1928) [98], [99] when presuming the pmf of equally probable states (i.e. uniform pmf  $P_N^{\text{unif}}$  for which  $P(\theta_i) = 1/N$  for  $i = 1, 2, \dots, N$ ), hence getting  $H(P_N^{\text{unif}}) = \log(|\Theta|) = \log(N)$ . Note that if we have a uniform pmf  $P_N^{\text{unif}}$  defined on  $\Theta$  with  $|\Theta| = N$  and another uniform pmf  $P_{N'}^{\text{unif}}$  defined on  $\Theta'$  with  $|\Theta'| = N'$ , and if  $|\Theta| < |\Theta'|$  then  $H(P_N^{\text{unif}}) < H(P_{N'}^{\text{unif}})$  because  $\log(|\Theta|) <$

$\log(|\Theta'|)$  since  $\log(x)$  is an increasing function. The minimum value of Shannon entropy is zero, which characterizes a *non-random* (or sure) event  $\theta_j$  for which  $P(\theta_j) = 1$ , because  $-\sum_{i=1}^{|\Theta|} P(\theta_i) \log(P(\theta_i)) = -P(\theta_j) \log(P(\theta_j)) = 0$ .

In fact, Shannon rarely used the term information (nor information content) in his works, and he preferred the term entropy to describe the scattering of symbols in the communication system. As reported in [100], in 1961 Shannon explained to Tribus his choice for naming the measure of uncertainty as *entropy*, instead of *information* as follows: "My greatest concern was what to call it. I thought of calling it 'information,' but the word was overly used, so I decided to call it 'uncertainty'". When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one really knows what entropy really is, so in a debate you will always have the advantage.'" Shannon did not prove that his entropy formula is the best measure of uncertainty, and even if it is a measure for information. He only stated a set of reasonable criteria [101] to describe a measure that would serve the requirements of his signal transmission theory, and he found that the entropy formula meets those criteria. We prefer to interpret Shannon entropy as a measure of uncertainty (or randomness) of a pmf, rather than a measure of information content [101], because of multiple possible interpretations and definitions of information.

The main algebraic properties of the Shannon entropy are, see [20] p. 30 for details: the symmetry, the normality<sup>11</sup>, expansibility, decisivity, additivity and recursivity. We recall that Shannon entropy value  $H(P_N)$  is always smaller than  $H(P_N^{\text{unif}})$  if  $P_N \neq P_N^{\text{unif}}$ , expressing the fact that the uniform pmf is the only pmf giving the maximal Shannon entropy value, and characterizing the maximum of uncertainty (or randomness), which is called the maximality property. Another important property of Shannon entropy is its subadditivity property when considering two (not necessarily independent) events, see [20] p. 36, which can be formulated by the following inequality

$$H(P_{N \cdot N'}) \leq H(P_N) + H(P_{N'}) \quad (9)$$

where  $P_{N \cdot N'}$  is the joint pmf defined on cartesian product space  $\Theta \times \Theta' = \{(\theta_i, \theta'_j), i = 1, 2, \dots, N, j = 1, 2, \dots, N'\}$ .  $P_N$  and  $P_{N'}$  are marginal pmfs (i.e. the projections) of the joint pmf  $P_{N \cdot N'}$  on spaces (i.e. frames of discernements)  $\Theta$  and  $\Theta'$  respectively.

### Belief functions

The belief functions (BF) have been introduced by Shafer [23] to model epistemic uncertainty to reason about uncertainty. We assume that the answer of the problem under concern belongs to a known finite discrete frame of discernement (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , with  $n > 1$ , and

<sup>10</sup>The symbol  $\triangleq$  means *equal by definition*.

<sup>11</sup>This stipulates that  $H(P_2^{\text{unif}}) = 1$  using base 2 logarithm function in (8).

where all elements of  $\Theta$  are exhaustive and exclusive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of  $2^\Theta$  is  $2^{|\Theta|}$ . A (normal) basic belief assignment (BBA) associated with a given source of evidence is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  satisfying  $m(\emptyset) = 0$  and  $\sum_{A \in 2^\Theta} m(A) = 1$ . The number  $m(A)$  is called the mass of  $A$  committed by the source of evidence. The subset  $A \in 2^\Theta$  is called a focal element (FE) of the BBA  $m(\cdot)$  if and only if  $m(A) > 0$ . The set of all the focal elements of the BBA  $m(\cdot)$  is noted by  $\mathcal{F}_\Theta(m) = \{X \in 2^\Theta | m(X) > 0\}$ , or just  $\mathcal{F}$  for shorthand notation when there is no ambiguity on the FoD  $\Theta$  and the BBA  $m$  we are using. The core  $\mathcal{C}(m)$  of a BBA  $m$  is the union of all its focal elements, i.e.  $\mathcal{C}(m) = \bigcup_{X \in \mathcal{F}_\Theta(m)} X$ .

The belief of  $A$  denoted  $Bel(A)$  and the plausibility of  $A$  denoted  $Pl(A)$  are usually interpreted respectively as lower and upper bounds of an unknown (subjective) probability measure  $P(A)$ . They are respectively defined for any  $A \in 2^\Theta$  from the BBA  $m(\cdot)$  by

$$Bel(A) = \sum_{X \in 2^\Theta | X \subseteq A} m(X) \quad (10)$$

and

$$Pl(A) = \sum_{X \in 2^\Theta | A \cap X \neq \emptyset} m(X) = 1 - Bel(\bar{A}). \quad (11)$$

where  $\bar{A}$  represents the complement of  $A$  in  $\Theta$ , that is  $\bar{A} \triangleq \Theta \setminus \{A\} = \{X | X \in \Theta \text{ and } X \notin A\}$ . The symbol  $\setminus$  denotes the set difference operator. Also, the commonality function  $q(\cdot)$  defined for all  $A \subseteq \Theta$  by  $q(A) = \sum_{X \subseteq \Theta | A \subseteq X} m(X)$  is involved in the some derivations, for instance in the definition of  $MoU_{1987}(m)$  (cf Table I). The vacuous BBA (VBBA for short) representing a totally ignorant source is defined by  $m_v(\Theta) = 1$ . In this short presentation, we implicitly work on the FoD  $\Theta$  and so we did omit to refer to it in our previous notations. If we have to work with BBAs defined on different FoDs, say  $\Theta$  and  $\Theta'$ , then we will explicitly indicate these FoDs in the BBA notations as  $m^\Theta(\cdot)$  and  $m^{\Theta'}(\cdot)$ . In the classical theory of belief functions the combination of several distinct sources of evidence characterized by their BBAs defined on the same FoD is done with Dempster's rule of combination, see [23]. To circumvent the problems of Dempster's rule (e.g. its dictatorial behavior, its possible insensitivity to conflict level, its counter-intuitive results in high and low conflicting situations, etc), other rules have been developed in particular those based on proportional conflict redistribution (PCR) principles, see [102], [103].

#### REFERENCES

- [1] D. Dubois, H. Prade, *Properties of measures of information in evidence and possibility theories*, Fuzzy Sets Syst., Vol. 24, no. 2, pp. 161–182, 1987. (reprinted in Fuzzy Sets and Systems, Vol. 100 Supplement, pp. 35–49, 1999).
- [2] G.J. Klir, *Where do we stand on measures of uncertainty, ambiguity, fuzziness, and the like?*, Fuzzy Sets and Systems, Vol. 24, pp. 141–160, 1987.
- [3] G.J. Klir, T. Folger, *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall, Englewood Cliffs, NJ, 1988.
- [4] B. Bouchon, *Entropic models: A general framework for measures of uncertainty and information*, in Fuzzy Logic in Knowledge-Based Systems. Decision and Control (M.M Gupta, T. Yamakawa, Eds.), North-Holland, New York, pp. 93–105, 1988.
- [5] N.R. Pal, J.C. Bezdek, R. Hemasinha, *Uncertainty measures for evidential reasoning I: a review*, Int. J. Approx. Reasoning, 7(3/4), 165–183, 1992.
- [6] N.R. Pal, J.C. Bezdek, R. Hemasinha, *Uncertainty measures for evidential reasoning II: A new measure of total uncertainty*, Int. J. Approx. Reason., pp. 1–16, 1993.
- [7] G.J. Klir, *Developments in uncertainty-based information*, in Advances in Computers (M.C. Yovits, Ed.), Academic Press, San Diego, pp. 255–332, 1993.
- [8] G.J. Klir, M.J. Wierman, *Uncertainty-based Information: Elements of Generalized Information Theory* (2nd corrected edition), Berlin: Springer, 1999.
- [9] N.R. Pal, *On quantification of different facets of uncertainty*, Fuzzy Sets and Systems, Vol. 107, pp. 81–91, 1999.
- [10] G. Klir, *Measures of uncertainty and information*, in Fundamentals of fuzzy sets (D. Dubois et al. Eds.), Chap. 8, pp. 439–457, Kluwer Academic Publishers, 2000.
- [11] A.L. Jousselme, C. Liu, D. Grenier, E. Bossé, *Measuring ambiguity in the evidence theory*, IEEE Trans. Syst. Man Cybern., Part A: Syst. Hum., Vol. 36, pp. 890–903, 2006.
- [12] J. Abellán, A. Masegosa, *Requirements for total uncertainty measures in Dempster-Shafer theory of evidence*, Int. J. Gen.Syst., Vol. 37(6), 2008.
- [13] J. Abellán, C.J. Mantas, E. Bossé, *Basic Properties for Total Uncertainty Measures in the Theory of Evidence*, pp. 99–108, Chap. 5 of É. Bossé, G. L. Rogova (eds.), Information Quality in Information Fusion and Decision Making, Information Fusion and Data Science, Springer 2019.
- [14] Y. Deng, *Uncertainty measure in evidence theory*, Information Science, Science China Press and Springer-Verlag GmbH Germany, Vol. 63, 210201:1–210201:19, November 2020.
- [15] S. Moral-García, J. Abellán, *Required mathematical properties and behaviors of uncertainty measures on belief intervals*, Int. J. Intell. Syst., pp. 1–24, 2021.
- [16] A.-L. Jousselme, F. Pichon, N. Ben Abdallah, *A Note About Entropy and Inconsistency in Evidence Theory*, in Proc. of Belief 2021 Int. Conf, Shanghai, China, Oct. 15–17, 2021.
- [17] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379–423 & 623–656, July & October 1948 (reprinted in [21]).
- [18] C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: The University of Illinois Press, 1–117, 1949.
- [19] R.B. Ash, *Information Theory* (1990 reedition), Dover Publications, 345 pages, 1965.
- [20] J. Aczél, Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, 1975.
- [21] N.J.A. Sloane, A.D. Wyner (Editors), *Claude Elwood Shannon - Collected Papers*, IEEE Press, 924 pages, 1993.
- [22] T.M. Cover, J.A. Thomas, *Elements of Information Theory* (2nd edition), John Wiley & Sons, 2006.
- [23] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [24] J. Aczél, B. Forte, C.T. Ng, *Why the Shannon and Hartley Entropies Are 'Natural'*, Advances in Applied Probability, Vol. 6, No. 1, pp. 131–146, Mar., 1974.
- [25] B. Forte, *Why Shannon entropy*, Symposia Mathematica, XV, Academic Press, New York, 1975.
- [26] S. Guiasu, *Information Theory and Applications*, McGraw-Hill, New York, 1977.
- [27] J.C. Baez, T. Fritz, T. Leinster, *A Characterization of Entropy in Terms of Information Loss*, Entropy, Vol. 13(11), pp. 1945–1957, 2011.
- [28] <http://www.mtm.ufsc.br/~taneja/book/node43.html>
- [29] I.J. Taneja, *New Developments in Generalized Information Measure*, Advances in imaging and electron physics, Vol. 91, 1995
- [30] V.M. Ilić, M.S. Stanković, *A unified characterization of generalized information and certainty measures*, Physica A 415, pp. 229–239, 2014.
- [31] C. Tsallis, *Nonadditive entropy: The concept and its use*, The European Physical Journal A, 40(3), 2009.
- [32] T.O. Kvålseth, *On the Measurement of Randomness (Uncertainty): A More Informative Entropy*, Entropy 2016, 18, 159.

- [33] G.W. Petty, *On Some Shortcomings of Shannon Entropy as a Measure of Information Content in Indirect Measurements of Continuous Variables*, Journal of Atmospheric and Oceanic Technology, Vol. 35, No. 5, pp. 1011-1021, 2018.
- [34] A. Bronevich, G. Klir, *Measures of uncertainty for imprecise probabilities: an axiomatic approach*, Int. J. Approx. Reason., Vol. 5 (4), pp. 365-390, 2010.
- [35] G.J. Klir, M.J. Wierman, *Uncertainty-Based Information: Elements of Generalized Information Theory*, (Vol. 15), Springer: Berlin, Germany, 2013.
- [36] D. Dubois, H. Prade, *Additivity and monotonicity of measures of information defined in the setting of Shafer's evidence theory*, BUSEFAL 24, pp. 64-76, 1985.
- [37] M.T. Lamata, S. Moral, *Measures of entropy in the theory of evidence*, Int. J. Gen. Systems, Vol. 14(4), pp. 297-305, 1988.
- [38] J. Vejnarova, G.J. Klir, *Measure of strife in Dempster-Shafer theory*, Int. J. of General Systems, Vol. 22(1), pp. 25-42, 1993.
- [39] D. Harmanec, G.J. Klir, *Measuring total uncertainty in Dempster-Shafer theory: a novel approach*, Int J General Syst., Vol. 22(4), pp. 405-419, 1994.
- [40] D. Harmanec, *Toward a characterization of uncertainty measure for the Dempster-Shafer theory*, in Proc. of the 11th Conf. on Uncertainty in Artificial Intelligence (UAI), pp. 255-261, San Mateo, CA, 1995.
- [41] D. Harmanec, *Measures of uncertainty and information*, Documentation Section on the website of the Society for Imprecise Probability Theory and Applications (SIPTA), 1999. [https://www.sipta.org/documentation/summary\\_measures/ippinfom.pdf](https://www.sipta.org/documentation/summary_measures/ippinfom.pdf).
- [42] U.Höhle, *Fuzzy plausibility measures*, in Proc. Third Intern. Seminar on Fuzzy Set Theory at Johannes Kepler Univ. (E.P. Klement, Ed.), Linz, Austria, pp. 7-30, 1981.
- [43] U. Höhle, *Entropy with respect to plausibility measures*, in Proc. of the 12th IEEE Int. Symp. on Multiple Valued Logic, Paris, 1982.
- [44] R.R. Yager, *Entropy and specificity in a mathematical theory of evidence*, Int. J. Gen. Syst., Vol. 9, pp. 249-260, 1983.
- [45] T. George, N.R. Pal, *Quantification of conflict in Dempster-Shafer framework: A new approach*, Int. J. Gen. Syst., Vol. 24(4), pp. 407-423, 1996.
- [46] M. Higashi, G.J. Klir, *Measures of uncertainty and information based on possibility distributions*, Int. J. General Systems, Vol. 9, pp. 43-58, 1983.
- [47] D. Dubois, H. Prade, *A note on measures of specificity for fuzzy sets*, (E)BUSENFAL, Vol. 19, paper # 8, May, 1984.
- [48] D. Dubois, H. Prade, *A note on measures of specificity for fuzzy sets*, Int. J. Gen. Syst., Vol. 10, no. 4, pp. 279-283, 1985.
- [49] U. Höhle, E.P. Klement, *Plausibility measures - A general framework for possibility and fuzzy probability measures*, in Aspects of vagueness (H.J. Skala, S. Termini, and E. Trillas, Eds.), pp.31-50, D.Reidel Publishing Company, 1984.
- [50] H.T. Nguyen, *On entropy of random sets and possibility distributions*, Anal. Fuzzy Inf., Vol. 1, pp. 45-156, 1987.
- [51] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Comput. Intell., Vol. 4, pp. 244-264, 1988.
- [52] G.J. Klir, A. Ramer, *Uncertainty in the Dempster-Shafer theory: a critical re-examination*, Int. J. Gen. Syst., Vol. 18(2), pp. 155-166, 1990.
- [53] G.J. Klir, B. Parviz, *A note on the measure of discord*, Proc. of 8th UAI Conf., pp. 138-141, Stanford Univ., USA, July 17-19, 1992.
- [54] A. Ramer, G.J. Klir, *Measures of discord in the Dempster-Shafer theory*, Information Sciences, Vol. 67, pp.35-50, 1993
- [55] Y. Maeda, H. Ichihashi, *An uncertainty measure with monotonicity under the random set inclusion*, Int. J. of General Systems, Vol. 21(4), pp. 379-392, 1993.
- [56] D. Harmanec, G. Resconi, G.J. Klir, Y. Pan, *emphOn the computation of the uncertainty measure for the Dempster-Shafer theory*, Int. J. of General Systems, Vol. 25(2), pp. 153-163, 1996.
- [57] D.A. Maluf, *Monotonicity of Entropy Computations in Belief Functions*, Intelligent Data Analysis, Vol. 1, No. 1, pp. 207-213, 1997.
- [58] G.J. Klir, *Uncertainty and information measures for imprecise probabilities: An overview*, in Proc. of 1st Int. Symposium on Imprecise Probabilities and Their Applications (ISIPTA'99), pp. 234-240, Ghent, Belgium, June 29-July 2, 1999.
- [59] R. Yager, *On the entropy of fuzzy measures*, IEEE Trans. Fuzzy Syst., vol. 8, pp. 453-461, 2000.
- [60] R. Yager, *Uncertainty Representation Using Fuzzy Measures*, IEEE Trans on SMC - Part B: Cybernetics, Vol. 32, No. 1, pp. 13-20, Feb. 2002
- [61] J. Dezert, F. Smarandache, A. Tchamova, *On the Blackman's association problem*, in Proc of Fusion 2003 Conf., pp. 1349-1356, Cairns, Queensland, Australia, 2003.
- [62] Y. Deng, *Deng entropy*, Chaos, Solitons & Fractals, Vol. 91, pp. 549-553, 2016.
- [63] Y. Yang, D. Han, *A new distance-based total uncertainty measure in the theory of belief functions*, Knowledge-Based Systems, Vol. 94, pp. 114-123, Feb. 2016.
- [64] X. Deng, F. Xiao, Y. Deng, *An improved distance-based total uncertainty measure in belief function theory*, Appl. Intell., 46, pp. 898-915, 2017.
- [65] D. Zhou, Y. Tang, W. Jiang, *A modified belief entropy in Dempster-Shafer framework*, Plos One, Vol. 12, no. 5, May, 2017.
- [66] D. Zhou, Y. Tang, W. Jiang, *An Improved Belief Entropy and Its Application in Decision-Making*, Complexity, Volume 2017, 2017.
- [67] Y. Tang, D. Zhou, Z. He, S. Xu, *An improved belief entropy-based uncertainty management approach for sensor data fusion*, International Journal of Distributed Sensor Networks, Vol. 13(7), 2017.
- [68] Y. Tang, D. Zhou, S. Xu, Z. He, *A Weighted Belief Entropy-Based Uncertainty Measure for Multi-Sensor Data Fusion*, Sensors, Vol. 17, 928, 2017.
- [69] R. Jiroušek, P.P. Shenoy, *A new definition of entropy of belief functions in the Dempster-Shafer theory*, Int. J. Approx. Reasoning, Vol. 92(1), pp. 49-65, Jan. 2018.
- [70] R. Jiroušek, P.P. Shenoy, *A decomposable entropy of belief functions in the Dempster-Shafer theory*, in: S. Destercke, T. Denoeux, F. Cuzzolin, A. Martin (Eds.), Belief Functions: Theory and Applications, in: Lecture Notes in Artificial Intelligence, vol. 11069, Springer Nature, Switzerland AG, 2018, pp. 146-154.
- [71] R. Jiroušek, P.P. Shenoy, *On properties of a new decomposable entropy of Dempster-Shafer belief functions*, Int. J. Approx. Reason., Vol. 119(4), pp. 260-279, 2020.
- [72] M.D. Mambé, et al., *A New Uncertainty Measure in Belief Entropy Framework*, (IJACSA) International Journal of Advanced Computer Science and Application, Vol. 9, No. 11, pp. 600-606, 2018.
- [73] L. Pan, Y. Deng, *A new belief entropy to measure uncertainty of basic probability assignments based on belief function and plausibility function*, Entropy, Vol. 20, no. 11, p. 842, 2018.
- [74] X. Wang, Y. Song, *Uncertainty measure in evidence theory with its application*, Applied Intelligence, Vol. 48, pp. 1672-1688, 2018.
- [75] D. Li, X. Gao, Y. Deng, *A Generalized Expression for Information Quality of Basic Probability Assignment*, IEEE Access, Vol. 7, pp. 174734-174739, December, 2019.
- [76] H.Cui, K. Liu, J. Zhang, B. Kang, *An Improved Deng Entropy and Its Application in Pattern Recognition*, IEEE Access, Vol. 7, pp. 18284-18292, 2019.
- [77] Q. Pan, D. Zhou, Y. Tang, X. Li, J. Huang *A Novel Belief Entropy for Measuring Uncertainty in Dempster-Shafer Evidence Theory Framework Based on Plausibility Transformation and Weighted Hartley Entropy*, Entropy 2019, 21, 163, 2019.
- [78] L. Chen, L. Diao, J. Sang, *A novel weighted evidence combination rule based on improved entropy function with a diagnosis application*, Int. Journal of Distributed Sensor Networks, Vol. 15(1), 2019.
- [79] Y. Zhao, D. Ji, X. Yang, L. Fei, C. Zhai, *An Improved Belief Entropy to Measure Uncertainty of Basic Probability Assignments Based on Deng Entropy and Belief Interval*, Entropy 2019, 21, 1122, 2019.
- [80] H. Li, R. Cai, *An Improved Expression for Information Quality of Basic Probability Assignment and Its Application in Fault Diagnosis*, techrxiv.org, Feb. 2020. DOI:10.36227/techrxiv.11725980.v2
- [81] K. Wen, Y. Song, C. Wu, I. Li, *A Novel Measure of Uncertainty in the Dempster-Shafer Theory*, IEEE Access, Vol. 8, pp. 51550-51559, 2020.
- [82] Y. Chen, Y. Tang, Y. Lei, *An Improved Data Fusion Method Based on Weighted Belief Entropy considering the Negation of Basic Probability Assignment*, Hindawi Journal of Mathematics, Volume 2020.
- [83] M. Qin, Y. Tang, J. Wen, *An Improved Total Uncertainty Measure in the Evidence Theory and Its Application in Decision Making*, Entropy 2020, 22, 487, 2020.
- [84] H. Yan, Y. Deng, *An Improved Belief Entropy in Evidence Theory*, IEEE Access, Volume 8, pp. 57505-57516, 2020.
- [85] J. Li, Q. Pan, *A New Belief Entropy in Dempster-Shafer Theory Based on Basic Probability Assignment and the Frame of Discernment*, Entropy 2020, 22, 691, 2021.

- [86] Y. Maeda, H.T. Nguyen, H. Ichihashi, *Maximum entropy algorithms for uncertainty measures*, Int. J. of Uncertainty, Fuzziness and Knowledge-based Systems, Vol. 1, No. 1, pp. 69–93, 1993 .
- [87] A. Meyerowitz, F. Richman, E. Walker, *Calculating maximum-entropy densities for belief functions*, Int. J. Uncertainty Fuzziness Knowledge-Based Syst., Vol. 2(4), pp. 377–389, 1994.
- [88] D. Harmanec, *A note on uncertainty, Dempster rule of combination, and conflict*, Int. J. of General Systems, Vol. 26 (1–2), pp. 63–72, 1997.
- [89] P. Smets, *Decision making in the TBM: the necessity of the pignistic transformation*, Int. Journal of Approximate Reasoning, Vol. 38, pp. 133–147, 2005.
- [90] B. Cobb, P.P. Shenoy, *On the plausibility transformation method for translating belief function models to probability models*, IJAR, Vol. 41, 2006.
- [91] J. Dezert, F. Smarandache, *A new probabilistic transformation of belief mass assignment*, in Int. Conf. on Information Fusion 2008, Germany, Cologne, pp. 1–8, June 30th–July 3rd, 2008.
- [92] J. Abellán, *Analyzing properties of Deng entropy in the theory of evidence*, Chaos Solitons Fractals, Vol. 95, pp. 195–199, 2017.
- [93] F. Liu, X. Gao, J. Zhao, Y. Deng, *Generalized Belief Entropy and Its Application in Identifying Conflict Evidence*, IEEE Access, Volume7, 2019.
- [94] Y. Xue, Y. Deng, *Interval-valued belief entropies for Dempster-Shafer structures*, Soft Computing, June 2021.
- [95] [https://en.wikipedia.org/wiki/Entropy\\_\(information\\_theory\)](https://en.wikipedia.org/wiki/Entropy_(information_theory))
- [96] O. Lombardi, F. Holik, L. Vanni, *What is Shannon information?*, Synthese 193, pp. 1983–2012, 2016.
- [97] O. Rioul, *This is IT: A Primer on Shannon's Entropy and Information*, L'Information, Séminaire Poincaré XXIII, 43–77, 2018.
- [98] R. Hartley, *Transmission of information*, Bell System Tech. J., Vol. 7, pp. 535–563, 1928.
- [99] [https://fr.wikipedia.org/wiki/Entropie\\_de\\_Hartley](https://fr.wikipedia.org/wiki/Entropie_de_Hartley)
- [100] M. Tribus, E.C. McIrvine, *Energy and Information*, Scientific American, Vol. 225, No. 3, pp. 179–190, Sept. 1971
- [101] [https://en.wikipedia.org/wiki/Information\\_content](https://en.wikipedia.org/wiki/Information_content)
- [102] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information Fusion (Collected works)*, Vol. 1–4, American Research Press, 2004–2015.
- [103] T. Dezert, J. Dezert, F. Smarandache, *Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments*, in Journal of Advances in Information Fusion (JAIF), Vol. 16, No. 1, June 2021.
- [104] Y. Zhang, F. Huang, X. Deng, W. Jiang, *A New Total Uncertainty Measure from A Perspective of Maximum Entropy Requirement*, Entropy 2021, 23, 1061, 2021.
- [105] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, In proc. of Int. Conf. on Information Fusion (Fusion 2022), Linköping, Sweden, pp. 1–10, July, 2022.



# An Effective Measure of Uncertainty of Basic Belief Assignments

Jean Dezert

Department of Information Processing and Systems  
The French Aerospace Lab - ONERA  
Palaiseau, France.  
Email: jean.dezert@onera.fr

Originally published as: J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, in Proc. of Int. Conf. on Information Fusion (Fusion 2022), Linköping, Sweden, July 4–7, 2022, and reprinted with permission.

**Abstract**—This paper presents a new effective measure of uncertainty (MoU) of basic belief assignments. This new continuous measure is effective in the sense that it satisfies a small number of very natural and essential desiderata. Our new simple mathematical definition of MoU captures well the interwoven link of randomness and imprecision inherent to basic belief assignments. Its numerical value is easy to calculate. This new effective MoU characterizes efficiently any source of evidence used in the belief functions framework. Because this MoU coincides with Shannon entropy for any Bayesian basic belief assignment, it can be also interpreted as an effective generalization of Shannon entropy. We also provide several examples to show how this new MoU works.

**Keywords:** Measure of Uncertainty, MoU, belief functions, Shannon entropy.

## I. INTRODUCTION

In the classical probabilistic framework of the theory of communication developed by Shannon in 1948 [1], [2], the measure of uncertainty (MoU), also called entropy, for characterizing a source of information (from signal transmission standpoint) is represented by Shannon entropy. This entropy measures the randomness of a probability mass function. Shannon entropy has played a very important role in the development of modern communication systems during the second half of the 20th century, and in signal and image coding, data compression, and cryptography [3] until today. Shannon theory does not concern the semantic aspects of the content of a message but only its transmission.

From 1980s and until now, many research works have been proposed to try to extend Shannon measure of uncertainty (i.e. entropy) in the belief functions framework since their introduction by Shafer in the mid of 1970s [5]. In parallel, other research works have been done on the characterization of particular aspects of the uncertainty which are related to the set consistency (or non-specificity) of basic belief assignments (BBAs). Recently Jousselme et al. [6] proposed an interesting attempt of mathematical unification of existing MoU formulations. In our recent survey paper [7], we did analyze in details 40 years of research works on MoUs. Our deep analysis of forty-eight MoUs reveals that only very few of them can be considered as effective in the mathematical sense defined in Section III. Unfortunately, these existing effective MoUs are conceptually flawed. The main contribution of this

paper is to provide a clear positive answer with a new well-justified mathematical solution to the fundamental challenging question stated in the conclusion of [7]:

*Is there a better conceptual effective measure of uncertainty for the basic belief assignments?*

This paper is organized as follows. Section II presents the basics of belief functions. Section III presents and justifies the four essential desiderata that a MoU must satisfy in order to be effective. In the section IV we list the existing effective MoUs and we explain their conceptual flaws. Section V presents the new effective MoU for BBA (i.e. generalized Shannon entropy) with some examples in the section VI. Concluding remarks and perspectives are given in the section VII.

## II. BELIEF FUNCTIONS

The belief functions (BF) were introduced by Shafer [5] for modeling epistemic uncertainty, reasoning about uncertainty and combining distinct sources of evidence. The answer of the problem under concern is assumed to belong to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \dots, \theta_N\}$  where all elements (i.e. members) of  $\Theta$  are exhaustive and exclusive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of the power-set is  $2^{|\Theta|}$ . A (normalized) basic belief assignment (BBA) associated with a given source of evidence is a mapping  $m^\Theta(\cdot) : 2^\Theta \rightarrow [0, 1]$  such that  $m^\Theta(\emptyset) = 0$  and  $\sum_{X \in 2^\Theta} m^\Theta(X) = 1$ . A BBA  $m^\Theta(\cdot)$  characterizes a source of evidence related with a FoD  $\Theta$ . For notation shorthand, we can omit the superscript  $\Theta$  in  $m^\Theta(\cdot)$  notation if there is no ambiguity on the FoD we work with. The quantity  $m(X)$  is called the mass of belief of  $X$ .  $X \in 2^\Theta$  is called a focal element (FE) of  $m(\cdot)$  if  $m(X) > 0$ . The set of all focal elements of  $m(\cdot)$  is denoted<sup>1</sup> by  $\mathcal{F}_\Theta(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$ . The belief and the plausibility of  $X$  are respectively defined for any  $X \in 2^\Theta$  by [5]

$$Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y) \quad (1)$$

$$Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) = 1 - Bel(\bar{X}). \quad (2)$$

where  $\bar{X} \triangleq \Theta \setminus \{X\}$  is the complement of  $X$  in  $\Theta$ .

<sup>1</sup> $\triangleq$  means equal by definition.

One has always  $0 \leq Bel(X) \leq Pl(X) \leq 1$ , see [5]. For  $X = \emptyset$ ,  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and for  $X = \Theta$  one has  $Bel(\Theta) = Pl(\Theta) = 1$ .  $Bel(X)$  and  $Pl(X)$  are often interpreted as the lower and upper bounds of unknown probability  $P(X)$  of  $X$ , that is  $Bel(X) \leq P(X) \leq Pl(X)$ . To quantify the uncertainty (i.e. the imprecision) of  $P(X) \in [Bel(X), Pl(X)]$ , we use  $u(X) \in [0, 1]$  defined by

$$u(X) \triangleq Pl(X) - Bel(X) \quad (3)$$

The quantity  $u(X) = 0$  if  $Bel(X) = Pl(X)$  which means that  $P(X)$  is known precisely, and one has  $P(X) = Bel(X) = Pl(X)$ . One has  $u(\emptyset) = 0$  because  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and one has  $u(\Theta) = 0$  because  $Bel(\Theta) = Pl(\Theta) = 1$ . If all focal elements of  $m(\cdot)$  are singletons of  $2^\Theta$  the BBA  $m(\cdot)$  is a Bayesian BBA because  $\forall X \in 2^\Theta$  one has  $Bel(X) = Pl(X) = P(X)$  and  $u(X) = 0$ . Hence the belief and plausibility of  $X$  coincide with a probability measure  $P(X)$  defined on the FoD  $\Theta$ . The vacuous BBA characterizing a totally ignorant source of evidence is defined by  $m_v(X) = 1$  for  $X = \Theta$ , and  $m_v(X) = 0$  for all  $X \in 2^\Theta$  different of  $\Theta$ . This very particular BBA plays a major role in the establishment of a new effective measure of uncertainty for BBA.

### III. ESSENTIAL DESIDERATA FOR A MOU

Before defining our new effective measure of uncertainty, denoted by  $U(m)$ , for any basic belief assignment  $m(\cdot)$  related to a (non-empty) FoD  $\Theta$ , we present the four essential and very natural desiderata that an effective MoU must satisfy [7].

**Desideratum D1:** For any non-empty frame of discernment  $\Theta$  and for any BBA  $m(\cdot)$  focused on a singleton  $X$  of  $2^\Theta$  one must have

$$U(m) = 0 \quad (4)$$

*Justification of D1:* this desideratum is natural and intuitive because any particular BBA for which a singleton  $X$  has  $m(X) = 1$  characterizes its certainty, which means that there is no uncertainty about the choice of this element since it does not include other smaller element in it. So, in this case  $U(m)$  must take zero value.

**Desideratum D2:** The measure of uncertainty of a total ignorant source of evidence must increase with the cardinality of the frame of discernment. That is

$$U(m_v^\Theta) < U(m_v^{\Theta'}), \quad \text{if } |\Theta| < |\Theta'|. \quad (5)$$

*Justification of D2:* this second desideratum makes perfect sense because the total ignorant source of evidence on  $\Theta = \{\theta_1, \dots, \theta_N\}$  for which  $m_v^\Theta(\Theta) = 1$  knows absolutely nothing about only  $N$  elements, whereas the total ignorant source of evidence on  $\Theta' = \{\theta_1, \dots, \theta_N, \theta_{N+1}, \dots, \theta_{N'}\}$  with  $m_v^{\Theta'}(\Theta') = 1$  knows absolutely nothing about more elements because  $N' > N$ . This clearly indicates that  $m_v^{\Theta'}$  must be in fact more ignorant than  $m_v^\Theta$ .

**Desideratum D3:** The measure of uncertainty  $U(m)$  must coincide with Shannon entropy [1]–[3] if the BBA  $m(\cdot)$  is a Bayesian BBA. This desideratum is mathematically expressed

for any Bayesian BBA  $m(\cdot)$  defined on the FoD  $\Theta$  by the condition<sup>2</sup>

$$U(m) = - \sum_{X \in \Theta} m(X) \log(m(X)) \quad (6)$$

*Justification of D3:* this third desideratum is also very natural because Shannon entropy is the most well-known and justified [9] measure used to characterize the uncertainty (the randomness, or variability) of a probability mass function. Because any Bayesian BBA induces belief and plausibility functions that coincide with a probability measure, one must have a coherence of  $U(m)$  with Shannon entropy when the BBA is Bayesian.

**Desideratum D4:** For any non-vacuous BBA  $m(\cdot)$  and for the vacuous BBA  $m_v(\cdot)$  defined with respect to the same FoD one must have

$$U(m) < U(m_v) \quad (7)$$

*Justification of D4:* this last desideratum is also a very important one and it makes perfect sense because the total ignorant source is always characterized by the vacuous BBA  $m_v(\cdot)$ , and obviously no source of evidence can be more uncertain than the total ignorant source.

**Effectiveness of a measure of uncertainty:** A measure of uncertainty (MoU) is said effective if and only if it satisfies the four essential desiderata D1, D2, D3, and D4.

Any MoU that fails to satisfy at least one of these four desiderata is said non-effective, and in this case it cannot be considered seriously as a good measure of uncertainty for characterizing a basic belief assignment of a source of evidence. Consequently, a non-effective MoU should not be used in applications involving MoU.

As justified in [7], we voluntarily do not include the sub-additivity desideratum in the list of our desiderata for the search of an effective MoU in the belief function framework because this desideratum does not make sense when working with general (i.e. non-Bayesian) BBAs, and it is incompatible with the essential desideratum D4. We recall that the sub-additivity condition is defined by  $U(m^{\Theta \times \Theta'}) \leq U(m^{\downarrow \Theta}) + U(m^{\downarrow \Theta'})$  or any joint BBA defined on the cartesian product  $\Theta \times \Theta'$  of FoDs  $\Theta$  and  $\Theta'$ , where  $m^{\downarrow \Theta}$  is the marginal (i.e. projection) of  $m^{\Theta \times \Theta'}(\cdot)$  on the power-set  $2^\Theta$ , and  $m^{\downarrow \Theta'}$  is the marginal (i.e. projection, see [10], [11]) of  $m^{\Theta \times \Theta'}(\cdot)$  on the power-set  $2^{\Theta'}$ . To justify our choice, just consider a simple example with  $|\Theta| = 5$  and  $|\Theta'| = 8$ , which means that the cartesian product space  $\Theta \times \Theta'$  has  $|\Theta \times \Theta'| = 40$  elements. Why the MoU of the vacuous BBA  $m_v^{\Theta \times \Theta'}$  related to 40 elements of  $\Theta \times \Theta'$  should be less (or equal) to the sum of MoU of vacuous BBA  $m_v^\Theta$  related to only 5 elements of  $\Theta$  and the MoU of the vacuous BBA  $m_v^{\Theta'}$  only related to the 8 elements of  $\Theta'$ ? We do not see any solid theoretical reason, nor intuitive reason, for justifying and requiring the

<sup>2</sup>Shannon entropy [1] is given here in *nats*, and we take  $0 \log(0) = 0$  because  $\lim_{x \rightarrow 0^+} x \log(x) = 0$  which is proved using L'Hôpital's rule [4].

subadditivity desideratum in the general framework of belief functions, and to select it as an axiom to satisfy in general as done in [12]. Unlike Vejnarova and Klir opinions [15] (p.28) (and some authors following them), we do not consider that the meaningful (or effective) measure of uncertainty of basic belief assignment must satisfy the sub-additivity desideratum in general.

#### IV. EXISTING EFFECTIVE MOUS

Before presenting our new effective MoU (or generalized entropy) in the next section, we must discuss a bit of the few existing effective measures of uncertainty proposed in the literature. As shown in [7], most<sup>3</sup> of existing MoUs are actually non-effective, and only eight MoUs can be considered as effective in the mathematical sense defined in the previous section. Most of effective MoUs share two basic principles: 1) approximate the BBA  $m$  by a probability measure (i.e. a Bayesian BBA)  $P_m$  based on some method of approximation and evaluate its Shannon entropy to estimate the randomness (or conflict) inherent to the BBA, and 2) add a term to Shannon entropy that characterizes the level of ambiguity (or non-specificity) inherent of the BBA (usually thanks to Dubois & Prade  $U$ -uncertainty [16]). For instance in [7] the BetP and DSMT transformations are used, in [17] the Cobb-Shenoy transformation [18] is used, and in [19] the authors suggest to use<sup>4</sup> the Bayesian BBA compatible with belief intervals drawn from  $m(\cdot)$  that maximizes Shannon entropy. This general 2-steps principle is rather simple and quite intuitive but it seriously lacks of theoretical justification. We consider that such type of effective MoU construction is conceptually flawed and not very satisfactory for two main reasons:

*Reason 1:* these effective MoUs highly depend on the method of approximation whose choice is quite arbitrary. Worse, a method of approximation of a BBA  $m(\cdot)$  to a Bayesian BBA can be totally misleading as for instance Cobb-Shenoy  $PlPr_m$  transformation [18] because for this transformation the evaluation of probabilities can be inconsistent with belief interval values. More precisely, one can have  $PlPr_m(\theta_i) \notin [Bel(\theta_i), Pl(\theta_i)]$  with Cobb-Shenoy method, which is obviously not reasonable, nor acceptable at all, see discussion in [7] with example. We emphasize the fact that if a method of approximation of a BBA  $m$  by a probability measure  $P_m$  is chosen, it must be at least consistent with belief interval values generated by the BBA  $m$  under concern. Clearly, we cannot recommend Cobb-Shenoy transformation for building an effective MoU based on aforementioned principles 1) and 2) as proposed recently by Jiroušek and Shenoy in [17].

*Reason 2:* In fact, there is no solid reason or evidence that necessitates to approximate any (non-Bayesian) BBA by a Bayesian BBA (for using Shannon entropy measure) in the construction of MoU. Also, there is no reason why

we need (or request) to make the distinction of the two aspects of uncertainty (conflict and non-specificity), and to consider them as additively separable. This is conceptually very disputable because the randomness (or conflict) and ambiguity (or non-specificity) are actually interwoven through the mass value of the focal elements of the BBA and their belief intervals.

Very recently, Zhang et al. in [22] did propose three new effective MoUs not directly based on the aforementioned 2-steps principle approach, and that is why they have attracted our attention. These MoUs are denoted by  $H^1(m)$ ,  $H^2(m)$  and  $H^3(m)$  and they are respectively defined by<sup>5</sup>

$$\begin{aligned} H^1(m) &= - \sum_{X \subseteq \Theta} m(X) \log_2(Pl(X)) + \sum_{X \subseteq \Theta} m(X) 2 \log_2(|X|) \\ H^2(m) &= - \sum_{X \subseteq \Theta} m(X) \log_2(Pl(X)) + \sum_{X \subseteq \Theta} m(X) \log_2(2^{|X|} - 1) \\ H^3(m) &= - \sum_{X \subseteq \Theta} m(X) \log_2(Pl(X)) + \sum_{\substack{X \subseteq \Theta \\ |X| > 1}} m(X) |X| \end{aligned}$$

Unfortunately, Zhang et al. fail to capture well the interwoven link between conflict and non-specificity (or imprecision). Actually the authors set arbitrarily the range of their MoU as a simple parameter, either taken arbitrarily as  $[0, 2 \log_2(|\Theta|)]$ ,  $[0, \log_2(2^{|\Theta|} - 1)]$  or  $[0, |\Theta|]$ , to define their  $H^1(m)$ ,  $H^2(m)$  and  $H^3(m)$  measures of uncertainty. Zhang's approach is very questionable, and actually other ranges could have been chosen instead. Moreover Zhang et al. do not identify (nor propose) the best MoU to use between  $H^1(m)$ ,  $H^2(m)$  and  $H^3(m)$ . The other serious problem with Zhang's approach is its lack of solid justification for using the plausibility function in the summation  $-\sum_{X \subseteq \Theta} m(X) \log_2(Pl(X))$ . Although effective in the mathematical sense defined in section III, Zhang's new MoUs are ill-justified and they can also be considered as conceptually flawed. That is why we present a better conceptual effective measure of uncertainty for BBA in the next section.

#### V. A NEW EFFECTIVE MEASURE OF UNCERTAINTY

##### A. Mathematical definition

The new effective measure of uncertainty we propose is given by the following simple formula

$$U(m) = \sum_{X \in 2^\Theta} s(X) \quad (8)$$

with

$$\begin{aligned} s(X) \triangleq & -(1 - u(X))m(X) \log(m(X)) \\ & + u(X)(1 - m(X)) \quad (9) \end{aligned}$$

$s(X)$  is the uncertainty contribution of  $X$  in the MoU  $U(m)$ . We call  $s(X)$  the *entropiece* of  $X$ . Because  $u(X) \in [0, 1]$  and  $m(X) \in [0, 1]$  one has  $s(X) \geq 0$ , and  $U(m) \geq 0$ . The

<sup>3</sup>Forty-eight MoUs have been analyzed in [7].

<sup>4</sup>found using a complicate optimization method, see [20], [21] for details.

<sup>5</sup>We have corrected here the definition of  $H^3(m)$  which is mathematically ill-formulated in [22].

entropiece  $s(X)$  takes into account the belief mass  $m(X)$ , and the uncertainty (or imprecision)  $u(X) = Pl(X) - Bel(X)$  about the unknown probability of  $X$  in a subtle interwoven manner. The cardinality of  $X$  enters indirectly (i.e. not explicitly) in the derivations of  $Bel(X)$  and  $Pl(X)$ , and thus in the calculation of  $u(X)$  and in the entropiece  $s(X)$ . The quantity  $-(1-u(X))\log(m(X)) = (1-u(X))\log(1/m(X))$  entering in  $s(X)$  in (9) is the surprisal [8]  $\log(1/m(X))$  of  $X$  discounted by the confidence  $(1-u(X))$  one has about the precision of  $P(X)$ . The term  $-m(X)(1-u(X))\log(m(X))$  is the weighted discounted surprisal of  $X$ . The second term  $u(X)(1-m(X))$  corresponds to the imprecision of  $P(X)$  discounted by  $(1-m(X))$  because the greater  $m(X)$  the less one should take into account the imprecision  $u(X)$  in the MoU. As we will prove next, this new very simple MoU  $U(m)$  satisfies the four essential desiderata, and thus it is effective and conceptually well justified, and it presents several advantages over existing effective MoUs given in Section VII.

Because for  $X = \emptyset$ , one has  $m(\emptyset) = 0$  and  $u(\emptyset) = 0$  the entropiece of the empty set  $\emptyset$  is  $s(\emptyset) = 0$ . Hence the expression of  $U(m)$  can be written equivalently as

$$U(m) = s(\emptyset) + \sum_{X \in 2^\Theta | X \neq \emptyset} s(X) = \sum_{X \in 2^\Theta | X \neq \emptyset} s(X) \quad (10)$$

It is worth noting that for any BBA focused on  $X \neq \emptyset$  with  $m(X) = 1$ , we have  $m(X) = Bel(X) = Pl(X) = 1$ , and thus  $u(X) = 0$ . In this case, the entropiece of  $X$  is<sup>6</sup>

$$\begin{aligned} s(X) &= -(1-u(X))m(X)\log(m(X)) + u(X)(1-m(X)) \\ &= -(1-0)1\log(1) + 0(1-1) = 0 \end{aligned}$$

In particular, if  $m(\Theta) = 1$  (which corresponds to the vacuous BBA) we have the entropiece  $s(\Theta) = 0$ .

$U(m)$  is expressed in *nats* because we use the natural logarithm which makes derivations simpler, specially for making some proofs in the sequel.  $U(m)$  can be expressed in *bits* by dividing the  $U(m)$  value in *nats* by  $\log(2) = 0.69314718\dots$ . This measure of uncertainty  $U(m)$  is a continuous function in its basic belief mass arguments because it is a summation of continuous functions.

### B. Entropy of the vacuous BBA

Consider the FoD  $\Theta$  of cardinality  $|\Theta| = N$  greater than zero, and the vacuous BBA  $m_v$  defined on this FoD for which  $m_v(\Theta) = 1$  and  $m_v(X) = 0$  for any  $X \neq \Theta$  in  $2^\Theta$ . For this vacuous BBA one always has  $Bel(\Theta) = Pl(\Theta) = 1$  and thus  $u(\Theta) = Pl(\Theta) - Bel(\Theta) = 0$ , and one has also  $u(\emptyset) = 0$ . For all elements  $X \neq \Theta$  with  $X \in 2^\Theta \setminus \{\emptyset\}$  one has also necessarily  $Bel(X) = 0$ ,  $Pl(X) = 1$  and thus

$u(X) = Pl(X) - Bel(X) = 1$ . Consequently, the expression (10) with the BBA  $m_v$  becomes<sup>7</sup>

$$\begin{aligned} U(m_v) &= - \sum_{X \in 2^\Theta | X \neq \emptyset} (1-u(X))m_v(X)\log(m_v(X)) \\ &\quad + \sum_{X \in 2^\Theta | X \neq \emptyset} u(X)(1-m_v(X)) \\ &= -(1-u(\Theta))m_v(\Theta)\log(m_v(\Theta)) \\ &\quad - \sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (X \neq \Theta)} (1-u(X))m_v(X)\log(m_v(X)) \\ &\quad + u(\Theta)(1-m_v(\Theta)) \\ &\quad + \left[ \sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (X \neq \Theta)} u(X)(1-m_v(X)) \right] \end{aligned}$$

In this expression of  $U(m_v)$  we have<sup>8</sup>

$$\begin{cases} -(1-u(\Theta))m_v(\Theta)\log(m_v(\Theta)) = -(1-0)1\log(1) = 0 \\ -\sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (X \neq \Theta)} (1-u(X))m_v(X)\log(m_v(X)) = 0 \\ u(\Theta)(1-m_v(\Theta)) = 0(1-1) = 0 \\ \sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (X \neq \Theta)} u(X)(1-m_v(X)) = 2^N - 2 \end{cases}$$

Therefore, it comes finally for the vacuous BBA  $m_v$  defined on a FoD of size  $N > 0$  the following MoU value

$$U(m_v) = 2^N - 2 \quad (11)$$

The entropy  $U(m)$  makes perfect sense because for the vacuous BBA  $m_v(\cdot)$  there is no information about the conflicts between the elements of the FoD. One has  $u(\emptyset) = 0$  because  $[Bel(\emptyset), Pl(\emptyset)] = [0, 0]$ ,  $u(\Theta) = 0$  because  $[Bel(\Theta), Pl(\Theta)] = [1, 1]$ , and for all  $X \in 2^\Theta \setminus \{\emptyset, \Theta\}$  one has  $u(X) = 1$  because  $[Bel(X), Pl(X)] = [0, 1]$ . Hence, the sum of all imprecisions of  $P(X)$  for all  $X \in 2^\Theta$  is exactly equal to  $2^N - 2$  when  $|\Theta| = N$ . In the degenerate case where  $|\Theta| = N = 1$ , one has  $U(m_v) = 2^1 - 2 = 0$  which indicates that there is absolutely no uncertainty in this very particular case. This result makes perfect sense also. For non-degenerate FoD (i.e. when  $|\Theta| > 1$ ) one has always  $U(m_v) > \log(N)$  which means that the vacuous BBA representing the totally ignorant source of evidence has an entropy greater than the maximum of Shannon entropy  $\log(N)$  obtained with the uniform probability mass function distributed on  $\Theta$ . This is an expected result because no BBA can represent the total ignorance, but the vacuous BBA.

### C. Effectiveness of $U(m)$

In this subsection we establish the effectiveness of our new generalized entropy  $U(m)$  defined in (8). For this, we prove the following four lemmas.

**Lemma 1:**  $U(m)$  satisfies the desideratum D1.

**Proof:** Consider at first the very special case where  $\Theta$  includes only one element  $\theta$ , that is  $\Theta = \{\theta\}$  and  $|\Theta| = 1$ . In this case there exists only one possible

<sup>6</sup>because  $\log(1) = 0$ .

<sup>7</sup>The notation  $a \wedge b$  means that the conditions  $a$  and  $b$  are both satisfied.

<sup>8</sup>For  $X \neq \Theta$ ,  $m_v(X) = 0$  and  $m_v(X)\log(m_v(X)) = 0\log(0) = 0$ .

BBA over  $2^\Theta = \{\emptyset, \theta\}$  defined by  $m(\emptyset) = 0$  and  $m(\theta) = 1$ . Hence  $Bel(\theta) = Pl(\theta) = 1$ ,  $u(\theta) = 0$ , and  $s(\theta) = (1 - u(\theta))m(\theta) \log(m(\theta)) + u(\theta)(1 - m(\theta)) = 0$ . Therefore  $U(m) = s(\emptyset) + s(\theta) = 0$ . In more general (i.e. when  $|\Theta| > 1$ ) if  $X$  is a singleton of  $2^\Theta$  (i.e.  $|X| = 1$ ) and if  $m(X) = 1$  then  $Bel(X) = Pl(X) = 1$  and  $u(X) = 0$ . For the elements  $Y$  of  $2^\Theta \setminus \{\emptyset\}$  containing  $X$  one has also  $Bel(Y) = Pl(Y) = 1$  and therefore  $u(Y) = 0$ . For all elements  $Y$  of  $2^\Theta \setminus \{\emptyset\}$  not containing  $X$  one has always  $Bel(Y) = Pl(Y) = 0$  and therefore  $u(Y) = 0$ . In summary, one has: 1)  $m(X) = 1$ ,  $u(X) = 0$ ,  $s(X) = 0$ , 2)  $m(Y) = 0$ ,  $u(Y) = 0$ ,  $s(Y) = 0$  for all  $Y \neq X$ ,  $Y \in 2^\Theta \setminus \{\emptyset\}$ , and 3)  $s(\emptyset) = 0$ . Applying formula (8) (or (10)) we obtain  $U(m) = 0$ , which completes the proof of lemma 1.

**Lemma 2:**  $U(m)$  satisfies the desideratum D2.

**Proof:** Consider two FoD  $\Theta$  et  $\Theta'$  with  $|\Theta| = N$  and  $|\Theta'| = N'$  greater than zero, and suppose  $N < N'$ . For the vacuous BBA  $m_v^\Theta$  defined on the FoD  $\Theta$ , one has  $U(m_v^\Theta) = 2^N - 2$ . Similarly, for the vacuous BBA  $m_v^{\Theta'}$  defined on the FoD  $\Theta'$ , one has  $U(m_v^{\Theta'}) = 2^{N'} - 2$ . Because the exponential function is an increasing function, one has always  $2^N < 2^{N'}$ , and also  $2^N - 2 < 2^{N'} - 2$ . Therefore  $U(m_v^\Theta) < U(m_v^{\Theta'})$  when  $|\Theta| < |\Theta'|$ , which completes the proof of lemma 2.

**Lemma 3:**  $U(m)$  satisfies the desideratum D3.

**Proof:** When the BBA  $m$  is Bayesian, its focal elements are only singletons of  $2^\Theta$  and  $Bel(X) = Pl(X)$  for all  $X \in 2^\Theta$ . Hence  $u(X) = 0$  for all  $X \in 2^\Theta$ . Thus, in the expression (9) of  $s(X)$  one has always  $-(1 - u(X))m(X) \log(m(X)) = -m(X) \log(m(X))$  and  $u(X)(1 - m(X)) = 0(1 - m(X)) = 0$ , so that  $s(X) = -m(X) \log(m(X))$ . Therefore  $U(m) = \sum_{X \in 2^\Theta} s(X) = -\sum_{X \in 2^\Theta} m(X) \log(m(X))$ . Because the masses of all non-singleton elements of  $2^\Theta$  are zero, we finally obtain  $U(m) = -\sum_{X \in 2^\Theta, |X|=1} m(X) \log(m(X)) = -\sum_{X \in \Theta} m(X) \log(m(X))$ , and this is Shannon entropy. This completes the proof of lemma 3.

**Lemma 4:**  $U(m)$  satisfies the desideratum D4.

**Proof:** see the appendix.

**Theorem:**  $U(m)$  is an effective measure of uncertainty of a basic belief assignment.

**Proof:** Because  $U(m)$  satisfies all desiderata D1, D2, D3, and D4 as proved in lemmas 1–4, the measure of uncertainty  $U(m)$  defined in (8) is effective.

#### D. Remarks about $U(m)$

**Remark 1:** It is worth noting that we do not have specified a priori what should be the range of an effective MoU in contrary to some axiomatic attempts made by different authors as reported, for instance, in [12]. We consider that the choice of the range must not be chosen a priori. The maximum range must result of the effective MoU mathematical definition. We

only request the satisfaction of the desideratum D4, which is much more general, natural and essential.

**Remark 2:** The choice of the desideratum D3 (compatibility with Shannon probabilistic entropy) could be disputed because other entropy definitions and generalizations exist in the probabilistic framework (as those defined by Rényi [13], Tsallis [14], etc). We think however that Shannon entropy is still the most used and preferred one for engineers working in information fusion. The measure of uncertainty  $U(m)$  presented in this paper could be (hopefully) generalized by replacing the desideratum D3 by another one using another choice of generalized entropy definition, which would obviously necessitate a modification of the definition of  $U(m)$ . This theoretical question has not yet been explored, and is left for future research.

**Remark 3:** It can be proved<sup>9</sup> that  $U(m)$  verifies the monotonicity property. More precisely, if  $m_Y$  and  $m_Z$  are two distinct BBAs defined on the same FoD  $\Theta$  and respectively focused on  $Y$  and on  $Z$  in  $2^\Theta$ , then one has always  $U(m_Y) < U(m_Z)$  if  $|Y| < |Z|$ . As a special case, one has  $U(m_Y) < U(m_Z)$  if  $Y \subset Z$ .

**Remark 4:** Consider a BBA  $m^\Theta$  defined on a FoD  $\Theta$ . Its zero-extension  $m^{\Theta'}$  on a FoD  $\Theta'$  including  $\Theta$  (i.e.  $\Theta \subseteq \Theta'$ ) is defined by  $m^{\Theta'}(X) = 0$  for all  $X \in 2^{\Theta'}$  not included in  $2^\Theta$ , and  $m^{\Theta'}(X) = m^\Theta(X)$  for all  $X \in 2^\Theta$ . It means that  $[Bel(\theta_i), Pl(\theta_i)] = [0, 0]$  for all  $\theta_i \in \Theta' \setminus \Theta$ . Under this condition, one has always  $U(m^\Theta) \leq U(m^{\Theta'})$  because  $u^{\Theta'}(X) \geq 0$  if  $X \cap Y \neq \emptyset$  for some  $Y \in 2^\Theta$ . Hence there exists at least an extra term  $s^{\Theta'}(X) > 0$  entering in  $U(m^{\Theta'})$  calculation (w.r.t.  $U(m^\Theta)$ ) if  $m^\Theta \neq m_v^{\Theta'}$ . Therefore, the extendability property of Shannon entropy for probability measures must be extended as  $U(m^\Theta) \leq U(m^{\Theta'})$  for (non-Bayesian) basic belief assignments. The equality  $U(m^\Theta) = U(m^{\Theta'})$  holds if  $m^\Theta$  is a Bayesian BBA because  $U(m^\Theta)$  coincides with Shannon entropy in this case.

## VI. EXAMPLES

In this section we give several simple numerical examples of the value of the measure of uncertainty  $U(m)$  expressed in nats. The examples are given in Table I and they correspond to different BBAs  $m_i$  ( $i = 1, 2, \dots, 6$ ), and to the vacuous BBA  $m_v$  defined on a FoD  $\Theta$ . For  $|\Theta| = 2$ , we have only one possible union/disjunction  $\theta_1 \cup \theta_2$  in  $2^\Theta$  which makes the examples too simple and not very interesting. Because for  $|\Theta| \geq 4$  we have  $2^4 = 16$  elements of  $2^\Theta$  to list, and due to paper length restriction we just give here some examples for  $|\Theta| = 3$  with  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ .

The numerical values of  $U(m)$  have been truncated to their third decimal.  $m_1$  and  $m_2$  are Bayesian BBAs, and  $m_2$  is the uniform Bayesian BBA. Hence we have  $U(m_2) = \log(|\Theta|) = \log(3) \approx 1.098$  which is the maximum of Shannon entropy for this FoD. The BBAs  $m_3, \dots, m_6$  and  $m_v$  are non-Bayesian

<sup>9</sup>Sketch of proof: prove that  $U(m_Y) = 2^{|\Theta|} - 1 - |\{X \in 2^\Theta | Y \subseteq X\}| - |\{X \in 2^\Theta | X \cap Y = \emptyset\}|$  and  $U(m_Z) = 2^{|\Theta|} - 1 - |\{X \in 2^\Theta | Z \subseteq X\}| - |\{X \in 2^\Theta | X \cap Z = \emptyset\}|$ , and compare  $U(m_Y)$  and  $U(m_Z)$  when  $|Y| < |Z|$  to complete the proof.

BBA, and  $U(m_v) = 2^3 - 2 = 6$  is the maximum value of the new proposed generalized entropy.

$X \in 2^\Theta$	$m_1$	$m_2$	$m_3$	$m_4$	$m_5$	$m_6$	$m_v$
$\emptyset$	0	0	0	0	0	0	0
$\theta_1$	0.2	1/3	0.1	0.1	1/7	0	0
$\theta_2$	0.3	1/3	0.2	0.2	1/7	0	0
$\theta_1 \cup \theta_2$	0	0	0.7	0.05	1/7	1	0
$\theta_3$	0.5	1/3	0	0.3	1/7	0	0
$\theta_1 \cup \theta_3$	0	0	0	0.03	1/7	0	0
$\theta_2 \cup \theta_3$	0	0	0	0.02	1/7	0	0
$\Theta$	0	0	0	0.3	1/7	0	1
$U(m_i)$	1.029	1.098	3.005	3.100	3.435	4	6

Table I  
EXAMPLES FOR  $U(m_i)$ ,  $i = 1, 2, \dots, 6$  AND  $U(m_v)$ .

It is worth noting that a non-Bayesian BBA  $m$  can have an entropy value  $U(m)$  smaller than the maximum of Shannon entropy, which is normal and not surprising. For instance, if we consider  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the BBA  $m(\theta_1) = 0.1$ ,  $m(\theta_2) = 0.8$  and  $m(\theta_1 \cup \theta_2) = 0.1$ , we get  $U(m) \approx 0.909$  which is smaller than  $\log(|\Theta|) = \log(3) \approx 1.098$ . Therefore, the condition  $U(m) < \log(|\Theta|)$  does not imply that the BBA  $m$  is necessarily a Bayesian BBA, but if  $U(m) > \log(|\Theta|)$  we are sure that  $m$  is a non-Bayesian BBA. We recall also that any BBA focused on a singleton has always zero uncertainty because lemma 1 holds.

#### Abellán and Moral's example revisited

We revisit Abellán and Moral's example [23] with the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the BBAs  $m(\cdot)$  and  $m'(\cdot)$  defined by

$$\begin{cases} m(\theta_1) = m(\theta_2) = m(\theta_3) = 0.2 \\ m(\Theta) = 0.4 \\ m'(\theta_1) = m'(\theta_2) = m'(\theta_3) = 0.161 \\ m'(\theta_2 \cup \theta_3) = 0.317 \\ m'(\Theta) = 0.2 \end{cases}$$

Abellán and Moral's intuitively think it is reasonable that  $m$  should represent more uncertainty than  $m'$  as  $m$  is completely symmetrical and  $m'$  points to  $\theta_2 \cup \theta_3$ . We disagree with this intuition because the authors did not take into account the changes of masses values between  $m$  and  $m'$ , nor the imprecisions of all unknown probabilities  $P(X)$  generated by  $m$ , and the imprecisions of  $P'(X)$  generated by  $m'$ .

If we analyze more carefully these two basic belief assignments we get the belief intervals  $[Bel(X), Pl(X)]$  based on  $m$ , and the belief intervals  $[Bel'(X), Pl'(X)]$  based on  $m'$  listed in Table II. Based on the belief interval values listed in Table II, it is clear that  $m'$  generates in fact globally more uncertainty (imprecisions on probabilities of elements of the power set of  $\Theta$ ) than  $m$  if we compare  $u(X)$  and  $u'(X)$  values. If we apply our new effective MoU definition, we obtain  $U(m) = 3.1059$  nats, and  $U(m') = 3.3384$  nats. One sees that  $U(m) < U(m')$ , which well reflects that  $m'$  is actually a bit more uncertain than  $m$ , contrary to what one would expect based on an incorrect intuition. This simple example is very interesting because it shows clearly how a simplistic intuition can easily fail.

$X \in 2^\Theta$	$[Bel(X), Pl(X)]$	$u(X)$	$[Bel'(X), Pl'(X)]$	$u'(X)$
$\emptyset$	[0,0]	0	[0,0]	0
$\theta_1$	[0.2,0.6]	0.4	[0.161,0.361]	0.200
$\theta_2$	[0.2,0.6]	0.4	[0.161,0.678]	0.517
$\theta_1 \cup \theta_2$	[0.4,0.8]	0.4	[0.322,0.839]	0.517
$\theta_3$	[0.2,0.6]	0.4	[0.161,0.678]	0.517
$\theta_1 \cup \theta_3$	[0.4,0.8]	0.4	[0.322,0.839]	0.517
$\theta_2 \cup \theta_3$	[0.4,0.8]	0.4	[0.639,0.839]	0.200
$\Theta$	[1,1]	0	[1,1]	0

Table II  
BELIEF INTERVALS DRAWN FROM  $m$  AND  $m'$ .

Entropic surface for all BBAs  $m(\cdot)$  defined on  $\Theta = \{\theta_1, \theta_2\}$

The figure 1 shows the entropic surface corresponding to  $U(m)$  when  $m(\theta_1) \in [0, 1]$ ,  $m(\theta_2) \in [0, 1]$  such that  $m(\theta_1) + m(\theta_2) \leq 1$ , and with  $m(\theta_1 \cup \theta_2) = 1 - m(\theta_1) - m(\theta_2)$ .

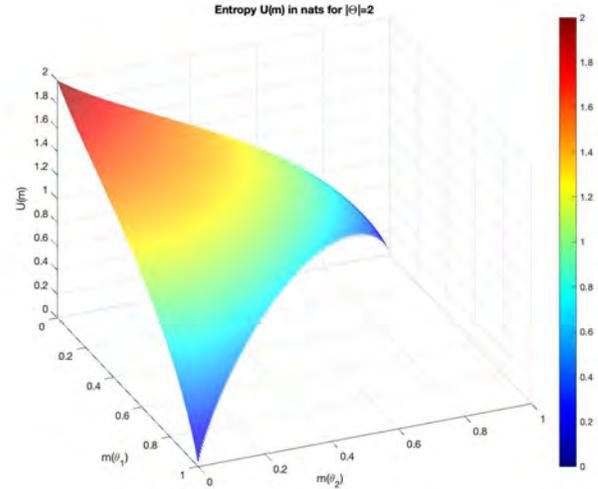


Figure 1. Entropy value  $U(m)$  for all  $m(\cdot)$  defined on  $\Theta = \{\theta_1, \theta_2\}$ .

One verifies visually that  $U(m)$  surface is smooth. Its border in the vertical plane passing through the points  $(m(\theta_1) = 1, m(\theta_2) = 0)$  and  $(m(\theta_1) = 0, m(\theta_2) = 1)$  corresponds to Shannon entropy curve whose maximum value is  $\log(2) \approx 0.6931$ , which is what we naturally expect. The unique maximum value of  $U(m)$  is for the vacuous BBA  $m_v$ , and it is  $U(m_v) = 2^{|\Theta|} - 2 = 2$ .

## VII. CONCLUSION

In this paper we have presented a new effective measure of uncertainty for basic belief assignments which is conceptually better justified than the few existing effective measures defined so far. This new generalized entropy measure verifies all the four very natural and essential desiderata, and presents the main advantages of simplicity, continuity, monotonicity and it also responds to the change of dimension of the frame of discernment. It is based on the interwoven link between the randomness and the imprecision of unknown probabilities of

all elements of the power set of the frame of discernment which is inherent to any basic belief assignment.

This new entropy measure makes a clear distinction between the maximum uncertainty of the vacuous BBA, and the uncertainties related to all non-vacuous BBAs, in particular with respect to Bayesian BBAs. Hence, we have answered positively to the challenging question about the existence of a better conceptual effective measure of uncertainty for BBAs. We hope that this new effective entropy measure will arouse the interest of users of belief functions who need an effective entropy measure in their own applications. It is worth mentioning that a dual of this new measure of entropy can be defined to characterize the information content of any BBA, as well as the notion of information gain and information loss between two BBAs. This will be reported in a future publication.

As a first perspective of this theoretical work, this new entropy measure could be useful to develop advanced methods for performance evaluation of information fusion techniques, and for reasoning under uncertainty using the belief functions. As a second perspective, this new entropy could also serve to measure the uncertainty of quantitative possibility measures in the possibility theory because any quantitative possibility measure is a special case of a plausibility function which is one-to-one with a consonant belief mass function (i.e. a BBA having nested focal elements).

#### ACKNOWLEDGMENTS

I wish to express my warmest thanks to Prof. Christine Marois and Albena Tchamova, and to Dr. Théo Dezert for their interesting remarks and discussions during the preparation of this paper that help me to clarify some technical points.

#### APPENDIX

##### *Proof of Lemma 4*

We first note from the expression (9) of  $s(X)$  that we always have  $s(X) = u(X)$  for  $X \in 2^\Theta$  and  $X \neq \emptyset$  if  $m(X) = 0$ . We have also  $s(X) = 0$  for  $X \in 2^\Theta$  and  $X \neq \emptyset$  if  $m(X) = 1$ . For  $X \in 2^\Theta$  and  $X \neq \emptyset$ , if  $0 < m(X) < 1$  one has

$$\begin{aligned} s(X) &= -(1 - u(X))m(X) \log(m(X)) + u(X)(1 - m(X)) \\ &= (1 - u(X))m(X) \log\left(\frac{1}{m(X)}\right) + u(X)(1 - m(X)) \\ &< (1 - u(X))m(X)\left(\frac{1}{m(X)} - 1\right) + u(X)(1 - m(X)) \end{aligned}$$

This strict inequality comes from the fact that for any real number  $x > 0$  with  $x \neq 1$ , the strict inequality  $\log(x) < x - 1$  holds<sup>10</sup> (see [24], p. 68). Because  $(1 - u(X))m(X)\left(\frac{1}{m(X)} - 1\right) + u(X)(1 - m(X)) = 1 - m(X)$ , one has finally the following inequality

$$s(X) < 1 - m(X) \quad (12)$$

<sup>10</sup>because the derivative  $f'(x)$  of  $f(x) = x - 1 - \log(x)$  is always positive for  $x > 0$  except for  $x = 1$  where  $f'(1) = 0$ .

To prove that  $U(m) < U(m_v)$ , we consider all the cases for the distribution of the belief masses in the BBA  $m \neq m_v$  as follows:

**Case 1:**  $0 < m(X) < 1$  for all  $X \neq \emptyset$  of  $2^\Theta$ .

In this (most general) case we have

$$U(m) = \sum_{X \in 2^\Theta | X \neq \emptyset} s(X) < \sum_{X \in 2^\Theta | X \neq \emptyset} (1 - m(X))$$

The majorant  $\sum_{X \in 2^\Theta | X \neq \emptyset} (1 - m(X))$  can be written as

$$\sum_{X \in 2^\Theta | X \neq \emptyset} 1 - m(X) = \sum_{X \in 2^\Theta | X \neq \emptyset} 1 - \sum_{X \in 2^\Theta | X \neq \emptyset} m(X)$$

Because one has  $\sum_{X \in 2^\Theta | X \neq \emptyset} 1 = 2^N - 1$ , and  $\sum_{X \in 2^\Theta | X \neq \emptyset} m(X) = 1$ , the majorant is given by

$$\sum_{X \in 2^\Theta | X \neq \emptyset} 1 - m(X) = 2^N - 1 - 1 = 2^N - 2$$

This majorant corresponds exactly to  $U(m_v)$ , therefore we have proved that

$$U(m) < U(m_v) \quad (13)$$

when  $0 < m(X) < 1$  for all  $X \neq \emptyset$  of  $2^\Theta$ .

**Case 2:** Consider the particular BBA for which  $m(X) = 1$  for some  $X \neq \emptyset$  and  $X \neq \Theta$  in  $2^\Theta$ .

- If  $X$  is a singleton of  $2^\Theta$  then  $Bel(X) = Pl(X) = 1$  and  $u(X) = 0$ . For the elements  $Y$  of  $2^\Theta$  including  $X$  one has  $Bel(Y) = Pl(Y) = 1$  and thus  $u(Y) = 0$ . for the elements  $Y$  of  $2^\Theta$  not including  $X$  one always has  $Bel(Y) = Pl(Y) = 0$  and thus  $u(Y) = 0$ . Hence,  $m(X) = 1$ ,  $u(X) = 0$ ,  $s(X) = 0$ , and also  $m(Y) = 0$ ,  $u(Y) = 0$ ,  $s(Y) = 0$  for all  $Y \neq X$ . Therefore we get  $U(m) = 0$  which is smaller than  $U(m_v) = 2^N - 2$ , i.e.  $U(m) < U(m_v)$  in this case.
- If  $X$  is not a singleton of  $2^\Theta$  and if  $m(X) = 1$  then  $Bel(X) = Pl(X) = 1$ ,  $u(X) = 0$  and  $s(X) = 0$ . We have also  $s(\Theta) = 0$  because  $m(\Theta) = 0$ , and we have  $u(\Theta) = 0$  because  $Bel(\Theta) = Pl(\Theta) = 1$ . For all  $Y \neq \emptyset$ ,  $Y \neq X$  and  $Y \neq \Theta$  such that  $X \cap Y = \emptyset$ , we always have  $u(Y) = 0$  because  $Bel(Y) = 0$  and  $Pl(Y) = 0$ . For all  $Y \neq \emptyset$ ,  $Y \neq X$  and  $Y \neq \Theta$  such that  $X \cap Y \neq \emptyset$ , we always have  $u(Y) = 1$  because  $Bel(Y) = 0$  and  $Pl(Y) = m(X) = 1$  because  $X$  has a non-empty intersection with  $Y$ . Consequently, the expression of  $U(m)$  can be reformulated as

$$\begin{aligned} U(m) &= s(\emptyset) + s(X) + s(\Theta) \\ &\quad + \sum_{Y \in 2^\Theta \setminus \{\emptyset, X, \Theta\} | Y \cap X = \emptyset} s(Y) \\ &\quad + \sum_{Y \in 2^\Theta \setminus \{\emptyset, X, \Theta\} | Y \cap X \neq \emptyset} s(Y) \quad (14) \end{aligned}$$

We have  $s(\emptyset) + s(X) + s(\Theta) = 0$  because  $s(\emptyset) = 0$ ,  $s(\Theta) = 0$  and  $s(X) = 0$  when  $m(X) = 1$ . For  $Y \in 2^\Theta \setminus \{\emptyset, X, \Theta\}$  such that  $Y \cap X = \emptyset$ , we have

$u(Y) = 0$  and  $m(Y) = 0$ , hence  $s(Y) = -(1 - u(Y))m(Y) \log(m(Y)) + u(Y)(1 - m(Y)) = (1 - 0)0 \log(0) + 0(1 - 0) = 0$ . Consequently

$$\sum_{Y \in 2^\Theta \setminus \{\emptyset, X, \Theta\} | Y \cap X = \emptyset} s(Y) = 0$$

For  $Y \in 2^\Theta \setminus \{\emptyset, X, \Theta\}$  such that  $Y \cap X \neq \emptyset$ , we have  $u(Y) = 1$  and  $m(Y) = 0$ , hence  $s(Y) = -(1 - u(Y))m(Y) \log(m(Y)) + u(Y)(1 - m(Y)) = (1 - 1)0 \log(0) + 1(1 - 0) = 1$ . Consequently,

$$\sum_{Y \in 2^\Theta \setminus \{\emptyset, X, \Theta\} | Y \cap X \neq \emptyset} 1 < 2^N - 2$$

Therefore, if a BBA is focused on any element  $X \neq \Theta$  (singleton, or not), that is if  $m(X) = 1$ , we have proved that the strict inequality  $U(m) < U(m_v)$  always holds.

**Case 3:** Some elements of the BBA have at least a zero mass value, and others have some strictly positive mass values strictly smaller than 1.

The measure of uncertainty  $U(m)$  defined in (10) requires  $2^N - 1$  terms  $s(X)$  to calculate in general (i.e. when all  $X \in 2^\Theta \setminus \{\emptyset\}$  are focal elements of  $m$ ). If some elements  $X$  have zero mass value, this measure  $U(m)$  can always be decomposed as

$$U(m) = \sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (m(X)=0)} s(X) + \sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (0 < m(X) < 1)} s(X) \quad (15)$$

Because one has  $s(X) = u(X)$  when  $m(X) = 0$ , the first summation of (15) is equal to  $\sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (m(X)=0)} u(X)$ . Because  $u(X) \leq 1$ , and  $s(X) < 1 - m(X)$  when  $m(X) < 1$ , one has the following strict inequality that holds

$$U(m) < \sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (m(X)=0)} 1 + \sum_{X \in 2^\Theta | (X \neq \emptyset) \wedge (0 < m(X) < 1)} (1 - m(X))$$

We can have at most  $2^N - 3$  elements of  $2^\Theta \setminus \{\emptyset\}$  having a mass equal to zero because we must have at least  $(2^N - 1) - (2^N - 3) = 2$  elements  $X_1$  and  $X_2$  of  $2^\Theta$  for which  $0 < m(X_1) < 1$ ,  $0 < m(X_2) < 1$  with  $m(X_1) + m(X_2) = 1$ . If we assume that there are  $1 < M \leq 2^N - 3$  elements of  $2^\Theta \setminus \{\emptyset\}$  that have zero mass value, then there exist  $K = 2^N - 1 - M$  elements  $X_1, X_2, \dots, X_K$  of  $2^\Theta \setminus \{\emptyset\}$  for which  $0 < m(X_k) < 1$ ,  $k = 1, \dots, K$  and with  $\sum_{k=1}^{(2^N-1)-M} m(X_k) = 1$ . Hence,

$$U(m) < M + \sum_{k=1}^{(2^N-1)-M} (1 - m(X_k))$$

or equivalently,

$$U(m) < \underbrace{M + (2^N - 1) - M}_{2^N - 1} - \underbrace{\sum_{k=1}^{(2^N-1)-M} m(X_k)}_1$$

Hence,  $U(m) < 2^N - 2$ , and consequently we have  $U(m) < U(m_v)$  because  $U(m_v) = 2^N - 2$ .

In summary, we have examined all possible cases for the distribution of the belief masses, and we have proved that we always have the strict inequality  $U(m) < U(m_v)$  satisfied. This completes the proof of the Lemma 4.

## REFERENCES

- [1] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379-423 & 623-656, July & October 1948 (reprinted in [3]).
- [2] C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: The University of Illinois Press, 1-117, 1949.
- [3] N.J.A. Sloane, A.D. Wyner (Editors), *Claude Elwood Shannon - Collected Papers*, IEEE Press, 924 pages, 1993.
- [4] R.E. Bradley, S.J. Petrioli, C.E. Sandifer, *L'Hôpital's analyse des infiniments petits (An annotated translation with source material by Johann Bernoulli)*, Birkhäuser, 311 pages, 2015.
- [5] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [6] A.-L. Jousselme, F. Pichon, N. Ben Abdallah, *A Note About Entropy and Inconsistency in Evidence Theory*, in Proc. of Belief 2021 Int. Conf, Shanghai, China, Oct. 15-17, 2021.
- [7] J. Dezert, A. Tchamova, *On Effectiveness of Measures of Uncertainty of Basic Belief Assignments*, Information & Security Journal: An International Journal (ISIJ), Vol. 52, Feb. 2022.
- [8] M. Tribus, *Rational Descriptions, Decisions and Designs*, Pergamon Press, New York, pp. 26-28, 1969.
- [9] J. Aczél, Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, 1975.
- [10] D. Dubois, H. Prade, *Additivity and monotonicity of measures of information defined in the setting of Shafer's evidence theory*, BUSEFAL 24, pp. 64-76, 1985.
- [11] M.T. Lamata, S. Moral, *Measures of entropy in the theory of evidence*, Int. J. Gen. Systems, Vol. 14(4), pp. 297-305, 1988.
- [12] S. Moral-García, J. Abellán, *Required mathematical properties and behaviors of uncertainty measures on belief intervals*, Int. J. Intell. Syst., pp. 1-24, 2021.
- [13] A. Rényi, *On measures of entropy and information*, in Proc. of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, Vol. 1, pp. 547-561, 1961.
- [14] C. Tsallis, *Nonadditive entropy: The concept and its use*, The European Physical Journal A, 40(3), 2009.
- [15] J. Vejnarova, G.J. Klir, *Measure of strife in Dempster-Shafer theory*, Int. J. of General Systems, Vol. 22(1), pp. 25-42, 1993.
- [16] D. Dubois, H. Prade, *A note on measures of specificity for fuzzy sets*, Int. J. Gen. Syst., Vol. 10, no. 4, pp. 279-283, 1985.
- [17] R. Jiroušek, P.P. Shenoy, *A new definition of entropy of belief functions in the Dempster-Shafer theory*, Int. J. Approx. Reasoning, Vol. 92(1), pp. 49-65, Jan. 2018.
- [18] B. Cobb, P.P. Shenoy, *On the plausibility transformation method for translating belief function models to probability models*, IJAR, Vol. 41, 2006.
- [19] Y. Maeda, H. Ichihashi, *An uncertainty measure with monotonicity under the random set inclusion*, Int. J. of General Systems, Vol. 21(4), pp. 379-392, 1993.
- [20] D. Harmanec, G.J. Klir, *Measuring total uncertainty in Dempster-Shafer theory: a novel approach*, Int J General Syst., Vol. 22(4), pp. 405-419, 1994.
- [21] D. Harmanec, G. Resconi, G.J. Klir, Y. Pan, *On the computation of the uncertainty measure for the Dempster-Shafer theory*, Int. J. of General Systems, Vol. 25(2), pp. 153-163, 1996.
- [22] Y. Zhang, F. Huang, X. Deng, W. Jiang, *A New Total Uncertainty Measure from A Perspective of Maximum Entropy Requirement*, Entropy 2021, 23, 1061, 2021.
- [23] J. Abellán, S. Moral, *Completing a total uncertainty measure in Dempster-Shafer theory*, Int. J. of General Systems, Vol. 28(4-5), pp. 299-314, 1999.
- [24] M. Abramowitz, I. Stegun, *Handbook of mathematical functions*, (9th Edition), Dover, 1970.

# Analytical Solution of the Simplest Entropiece Inversion Problem

Jean Dezert<sup>a</sup>, Florentin Smarandache<sup>b</sup>, Albena Tchamova<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>b</sup>Department of Mathematics, University of New Mexico, Gallup, NM, USA.

<sup>c</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.  
Emails: jean.dezert@onera.fr, smarand@unm.edu, tchamova@bas.bg

Originally published as: J. Dezert, F. Smarandache, A. Tchamova, *Analytical Solution of the Simplest Entropiece Inversion Problem*, in Proc. of 8th Int. Conf. of Modelling and development of Intelligent Systems (MDIS 2022), Sibiu, Romania, October 28–30, 2022, and reprinted with permission.

**Abstract**—In this paper, we present a method to solve analytically the simplest Entropiece Inversion Problem (EIP). This theoretical problem consists in finding a method to calculate a Basic Belief Assignment (BBA) from the knowledge of a given entropiece vector which quantifies effectively the measure of uncertainty of a BBA in the framework of the theory of belief functions. We give an example of the calculation of EIP solution for a simple EIP case, and we show the difficulty to establish the explicit general solution of this theoretical problem that involves transcendental Lambert's functions.

**Keywords:** belief functions, entropy, measure of uncertainty.

## I. INTRODUCTION

In this paper, we suppose the reader to be familiar with the theory of Belief Functions (BF) introduced by Shafer in [1], and we do not present in details the basics of BF. We just recall that a frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$  is a finite exhaustive set of  $N > 1$  mutually exclusive elements  $\theta_i$  ( $i = 1, \dots, N$ ), and its power set (i.e. the set of all subsets) is denoted by  $2^\Theta$ . A FoD represents a set of potential solutions of a decision-making problem under consideration. A Basic Belief Assignment (BBA)<sup>1</sup> is a mapping  $m : 2^\Theta \rightarrow [0, 1]$  with  $m(\emptyset) = 0$ , and  $\sum_{X \in 2^\Theta} m(X) = 1$ .

A new effective entropy measure  $U(m)$  for any BBA  $m(\cdot)$  defined on a FoD  $\Theta$  has been defined as follows [2]:

$$U(m) = \sum_{X \in 2^\Theta} s(X), \quad (1)$$

where  $s(X)$  is named the *entropiece* of  $X$ , which is defined by

$$s(X) = -m(X)(1 - u(X)) \log(m(X)) + u(X)(1 - m(X)), \quad (2)$$

with

$$\begin{aligned} u(X) &= Pl(X) - Bel(X) \\ &= \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) - \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y). \end{aligned} \quad (3)$$

<sup>1</sup>For notation convenience, we denote by  $m$  or  $m(\cdot)$  any BBA defined implicitly on the FoD  $\Theta$ , and we also denote it as  $m^\Theta$  to explicitly refer to the FoD when necessary.

$Pl(X)$  and  $Bel(X)$  are respectively the plausibility and the belief of the element  $X$  of the power set of  $\Theta$ , see [1] for details.  $u(X)$  quantifies the imprecision of the unknown probability of  $X$ . The vacuous BBA characterizing the total ignorant source of evidence is denoted by  $m_v$ , and it is such that  $m_v(\Theta) = 1$  and  $m_v(X) = 0$  for any  $X \subset \Theta$ .

This measure of uncertainty  $U(m)$  (i.e. entropy measure) is effective because it satisfies the following four essential properties [2]:

- 1)  $U(m) = 0$  for any BBA  $m(\cdot)$  focused on a singleton  $X$  of  $2^\Theta$ ;
- 2)  $U(m_v^\Theta) < U(m_v^{\Theta'})$  if  $|\Theta| < |\Theta'|$ ;
- 3)  $U(m) = -\sum_{X \in \Theta} m(X) \log(m(X))$  if  $m(\cdot)$  is a Bayesian<sup>2</sup> BBA. Hence,  $U(m)$  reduces to Shannon entropy [7] in this case;
- 4)  $U(m) < U(m_v)$  for any non-vacuous BBA  $m(\cdot)$  and for the vacuous BBA  $m_v(\cdot)$  defined with respect to the same FoD.

The proof of the three first properties is quite simple to make, whereas the proof of  $U(m) < U(m_v)$  is much more difficult, see [2] for proofs and examples. A detailed analysis of other (non-effective) entropy measures proposed in the literature during the last four decades is done in [3].

The entropiece  $s(X)$  given by (2) corresponds to the contribution of  $X$  to the whole uncertainty measure  $U(m)$ . The entropiece  $s(X)$  involves  $m(X)$  and the imprecision  $u(X) = Pl(X) - Bel(X)$  about the unknown probability of  $X$  in a subtle interwoven manner named *epistemic entanglement*. The cardinality of  $X$  is indirectly taken into account in the derivation of  $s(X)$  thanks to  $u(X)$  which requires the derivation of  $Pl(X)$  and  $Bel(X)$  functions that depend on the cardinality of  $X$ . Because  $u(X) \in [0, 1]$  and  $m(X) \in [0, 1]$  one has  $s(X) \geq 0$ , and  $U(m) \geq 0$ . The quantity  $U(m)$  is expressed in *nats* because we use the natural logarithm.  $U(m)$  can be expressed in *bits* by dividing the  $U(m)$  value in *nats* by  $\log(2) = 0.69314718\dots$ . This measure of uncertainty  $U(m)$  is a continuous function in its basic belief mass arguments because it is a summation of continuous func-

<sup>2</sup> $m$  is Bayesian BBA if it has only singletons as focal elements, i.e.  $m(\theta_i) > 0$  for some  $\theta_i \in \Theta$  and  $m(X) = 0$  for all non-singletons  $X$  of  $2^\Theta$ .

tions. In formula (2), we always take  $m(X) \log(m(X)) = 0$  when  $m(X) = 0$  because  $\lim_{m(X) \rightarrow 0^+} m(X) \log(m(X)) = 0$  which can be proved using L'Hôpital rule [4]. Note that for any BBA  $m$ , one has always  $s(\emptyset) = 0$  because  $m(\emptyset) = 0$  and  $u(\emptyset) = Pl(\emptyset) - Bel(\emptyset) = 0 - 0 = 0$ . For the vacuous BBA, one has  $s(\Theta) = 0$  because  $m_v(\Theta) = 1$  and  $u(\Theta) = Pl(\Theta) - Bel(\Theta) = 1 - 1 = 0$ .

As proved in [2], the entropy of the vacuous BBA on the FoD  $\Theta$  is equal to

$$U(m_v) = 2^{|\Theta|} - 2. \quad (4)$$

This maximum entropy value  $2^{|\Theta|} - 2$  makes perfectly sense because for the vacuous BBA there is no information at all about the conflicts between the elements of the FoD. Actually for all  $X \in 2^\Theta \setminus \{\emptyset, \Theta\}$  one has  $u(X) = 1$  because  $[Bel(X), Pl(X)] = [0, 1]$ , and one has  $u(\emptyset) = 0$  and  $u(\Theta) = 0$ . Hence, the sum of all imprecisions of  $P(X)$  for all  $X \in 2^\Theta$  is exactly equal to  $2^{|\Theta|} - 2$  which corresponds to  $U(m_v)$  as expected. Moreover, one has always  $U(m_v) > \log(|\Theta|)$  which means that the vacuous BBA has always an entropy greater than the maximum of Shannon entropy  $\log(|\Theta|)$  obtained with the uniform probability mass function distributed on  $\Theta$ . As a dual concept of this entropy measure  $U(m)$ , we have defined in [8] the measure of information content of any BBA by

$$IC(m) = U(m_v) - U(m) = (2^{|\Theta|} - 2) - \sum_{X \in 2^\Theta} s(X). \quad (5)$$

From the definition (5), one sees that for  $m \neq m_v^\Theta$  one has  $IC(m) > 0$  because  $U(m) < U(m_v)$ , and for  $m = m_v$  one has  $IC(m_v) = 0$  (i.e. the vacuous BBA carries no information), which is what we naturally expect.

Note that the information content  $IC(m^\Theta)$  of a BBA depends not only of the BBA  $m(\cdot)$  itself but also on the cardinality of the frame of discernment  $\Theta$  because  $IC(m)$  requires the knowledge of  $|\Theta| = N$  to calculate the max entropy value  $U(m_v) = 2^{|\Theta|} - 2$  entering in (5). This remark is important to understand that even if two BBAs (defined on different FoDs) focus entirely on a same focal element, their information contents are necessarily different. This means that the information content depends on the context of the problem, i.e. the FoD. The notions of information gain and information loss between two BBAs are also mathematically defined in [8] for readers interested in this topic.

This paper is organized as follows. Section 2 defines the general entropiece inversion problem (EIP). Section 3 describes the simplest entropiece inversion problem (SEIP). An analytical solution of SEIP is proposed and it is applied on a simple example also in Section 3. The conclusion is made in Section 4.

## II. GENERAL ENTROPIECE INVERSION PROBLEM (EIP)

The set  $\{s(X), X \in 2^\Theta\}$  of the entropieces values  $s(X)$  given by (2) can be represented by an entropiece vector  $s(m) = [s(X), X \in 2^\Theta]^T$ , where any order of elements  $X$  of the power set  $2^\Theta$  can be chosen. For simplicity, we

suggest to use the classical  $N$ -bits representation if  $|\Theta| = N$ , with the increasing order (see example in Section 3). The general Entropiece Inversion Problem, or EIP for short, is an interesting theoretical problem which can be easily stated as follows:

Suppose that if the entropiece vector  $s(m)$  known (estimated or given), is it possible to calculate a BBA  $m(\cdot)$  corresponding to this entropiece vector  $s(m)$ ? and how?

Also we would like to know if the derivation of  $m(\cdot)$  from  $s(m)$  provides a unique BBA solution, or not?

This general entropiece inversion problem is a challenging mathematical problem, and we do not know if a general analytical solution of EIP is possible, or not. We leave it as an open mathematical question for future research. However, we present in this paper the analytical solution for the simplest case where the FoD  $\Theta$  has only two elements, i.e. when  $|\Theta| = N = 2$ . Even in this simplest case, the EIP solution is no so easy to calculate as it will be shown in the next section. This is the main contribution of this paper.

The mathematical EIP addressed in this paper is not related (for now) to any problem for the natural world and it cannot be confirmed experimentally using data from nature because the entropy concept is not directly measurable, but only computable from the estimation of probability  $p(\cdot)$  or belief mass functions  $m(\cdot)$ . So, why do we address this entropiece inversion problem? Because in advanced information fusion systems we can imagine to have potentially access to this type of information and it makes sense to assess the underlying BBA provided by a source of evidence to eventually modify it in some fusion systems for some aims. We could also imagine to make adjustments of entropieces values to voluntarily improve (or degrade)  $IC(m)$ , and to generate the proper modified BBA for some tasks. At this early stage of research work it is difficult to anticipate the practical interests of the calculation of solutions of the general EIP, but to present its mathematical interest for now.

## III. SIMPLEST ENTROPIECE INVERSION PROBLEM (SEIP)

### A. Example

We consider a FoD  $\Theta$  with only two elements, say  $\Theta = \{A, B\}$ , where  $A$  and  $B$  are mutually exclusive and exhaustive, and the following BBA

$$m(A) = 0.5, \quad m(B) = 0.3, \quad m(A \cup B) = 0.2.$$

Because  $[Bel(\emptyset), Pl(\emptyset)] = [0, 1]$  one has  $u(\emptyset) = 0$ . Because  $[Bel(A), Pl(A)] = [0.5, 0.7]$ ,  $[Bel(B), Pl(B)] = [0.3, 0.5]$ ,  $[Bel(\Theta), Pl(\Theta)] = [1, 1]$ , one has  $u(A) = 0.2$ ,  $u(B) = 0.2$ , and  $u(\Theta) = 0$ . Applying (2), one gets  $s(\emptyset) = 0$ ,  $s(A) \approx 0.377258$ ,  $s(B) \approx 0.428953$  and  $s(\Theta) \approx 0.321887$ . Using the 2-bits representation with increasing ordering<sup>3</sup>, we encode

<sup>3</sup>Once the binary values are converted into their digit value with the most significant bit on the left (i.e the least significant bit on the right).

the elements of the power set as  $\emptyset = 00$ ,  $A = 01$ ,  $B = 10$  and  $A \cup B = 11$ . The entropiece vector is

$$s(m^\Theta) = \begin{bmatrix} s(\emptyset) \\ s(A) \\ s(B) \\ s(A \cup B) \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0.3773 \\ 0.4290 \\ 0.3219 \end{bmatrix}. \quad (6)$$

If we use the classical 2-bits (here  $|\Theta| = 2$ ) representation with increasing ordering (as we recommend) the first component of entropiece vector  $s(m)$  will be  $s(\emptyset)$  which is always equal to zero for any BBA  $m$ , hence the first component of  $s(m)$  is always zero and it can be dropped (i.e. removed of the vector representation actually). By summing all the components of the entropiece vector  $s(m)$  we obtain the entropy  $U(m) \approx 1.128098$  nats of the BBA  $m(\cdot)$ . Note that the components  $s(X)$  (for  $X \neq \emptyset$ ) of the entropieces vector  $s(m)$  are not independent because they are linked to each other through the calculation of  $Bel(X)$  and  $Pl(X)$  values entering in  $u(X)$ .

### B. Analytical solution of SEIP

Because we suppose  $\Theta = \{A, B\}$ , the expression of three last components<sup>4</sup> of the entropiece vector  $s(m)$  are given by (2), and we have

$$s(A) = -m(A)(1 - u(A)) \log(m(A)) + u(A)(1 - m(A)),$$

$$s(B) = -m(B)(1 - u(B)) \log(m(B)) + u(B)(1 - m(B)),$$

$$s(A \cup B) = -m(A \cup B)(1 - u(A \cup B)) \log(m(A \cup B)) + u(A \cup B)(1 - m(A \cup B)).$$

Because  $u(A) = Pl(A) - Bel(A) = (m(A) + m(A \cup B)) - m(A) = m(A \cup B)$ ,  $u(B) = Pl(B) - Bel(B) = (m(B) + m(A \cup B)) - m(B) = m(A \cup B)$  and  $u(A \cup B) = Pl(A \cup B) - Bel(A \cup B) = 1 - 1 = 0$ , one gets the following system of equations to solve

$$s(A) = -m(A)(1 - m(A \cup B)) \log(m(A)) + m(A \cup B)(1 - m(A)), \quad (7)$$

$$s(B) = -m(B)(1 - m(A \cup B)) \log(m(B)) + m(A \cup B)(1 - m(B)), \quad (8)$$

$$s(A \cup B) = -m(A \cup B) \log(m(A \cup B)). \quad (9)$$

The set of equations (7), (8) and (9) is called the EIP *transcendental equation system* for the case  $|\Theta| = 2$ .

The plot of function  $s(A \cup B) = -m(A \cup B) \log(m(A \cup B))$  is given in Figure 1 for convenience. By derivating the function  $-m(A \cup B) \log(m(A \cup B))$  we see that its maximum value is obtained for  $m(A \cup B) = 1/e \approx 0.3679$  for which

$$s(A \cup B) = -\frac{1}{e} \log(1/e) = \frac{1}{e} \log(e) = \frac{1}{e}.$$

Therefore, the numerical value of  $s(A \cup B)$  always belongs to the interval  $[0, 1/e]$ .

<sup>4</sup>We always omit the 1st component  $s(\emptyset)$  of entropiece vector  $s(m)$  which is always equal to zero and not necessary in our analysis.

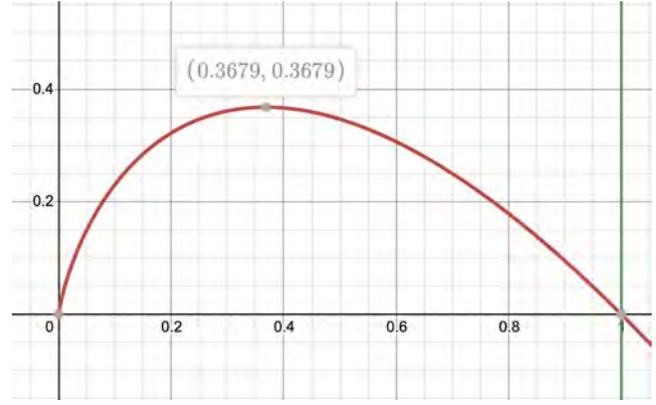


Figure 1. Plot of  $s(A \cup B) = -m(A \cup B) \log(m(A \cup B))$  (in red) with  $x$ -axis equals  $m(A \cup B) \in [0, 1]$ , and  $y$ -axis equals  $s(A \cup B)$  in nats.

Without loss of generality, we assume  $0 < s(A \cup B) \leq 1/e$  because if  $s(A \cup B) = 0$  then one deduces directly without ambiguity that either  $m(A \cup B) = 1$  (which means that the BBA  $m(\cdot)$  is the vacuous BBA) if  $s(A) = s(B) = 1$ , or  $m(A \cup B) = 0$  otherwise. With the assumption  $0 < s(A \cup B) \leq 1/e$ , the equation (9) is of the general transcendental form

$$ye^y = a \Leftrightarrow \log(m(A \cup B))m(A \cup B) = -s(A \cup B). \quad (10)$$

by considering the known value as  $a = -s(A \cup B)$  in  $[-\frac{1}{e}, 0)$ , and the unknown as  $y = \log(m(A \cup B))$ .

Unfortunately the solution of the transcendental equation (10) does not have an explicit expression involving simple functions. Actually, the solution of this equation is actually given by the Lambert's  $W$ -function which is a multivalued function (called also the omega function or product logarithm in mathematics) [6]. It can however be calculated<sup>5</sup> with a good precision by some numerical methods - see [5] for details. The equation  $ye^y = a$  admits real solution(s) only if  $a \geq -\frac{1}{e}$ . For  $a \geq 0$ , the solution of  $ye^y = a$  is  $y = W_0(a)$ , and for  $-\frac{1}{e} \leq a < 0$  there are two possible real values of  $W(a)$  - see Figure 1 of [5] which are denoted respectively  $y_1 = W_0(a)$  and  $y_2 = W_{-1}(a)$ . The principal branch of the Lambert's function  $W(x)$  satisfying  $-1 \leq W(x)$  is denoted  $W_0(x)$ , and the branch satisfying  $W(x) \leq -1$  is denoted by  $W_{-1}(x)$  by Corless et al. in [5]. In our context because we have  $a \in [-\frac{1}{e}, 0)$ , the solutions of  $ye^y = a$  are given by

$$y_1 = W_0(a) = W_0(-s(A \cup B)),$$

$$y_2 = W_{-1}(a) = W_{-1}(-s(A \cup B)).$$

Hence we get two possible solutions for the value of  $m(A \cup B)$ , which are

$$m_1(A \cup B) = e^{y_1} = e^{W_0(-s(A \cup B))}, \quad (11)$$

$$m_2(A \cup B) = e^{y_2} = e^{W_{-1}(-s(A \cup B))}. \quad (12)$$

Of course, at least one of these solutions is necessarily correct but we do not know which one. So, at this current

<sup>5</sup>Lambert's  $W$ -function is implemented in Matlab<sup>TM</sup> as *lambertw* function.

stage, we must consider<sup>6</sup> and the two solutions  $m_1(A \cup B)$  and  $m_2(A \cup B)$  for  $m(A \cup B)$  as acceptable, and we must continue to solve equations (7) and (8) to determine the mass values  $m(A)$  and  $m(B)$ .

Let's now determine  $m(A)$  at first by solving (7). Suppose we set the value of  $m(A \cup B)$  is known and taken either as  $m_1(A \cup B)$ , or as  $m_2(A \cup B)$ , then we can rearrange the equation (7) as

$$-\frac{s(A) - m(A \cup B)}{1 - m(A \cup B)} = m(A) \left[ \log(m(A)) + \frac{m(A \cup B)}{1 - m(A \cup B)} \right],$$

which can be rewritten as the general equation of the form

$$(y + a)e^y = b, \quad (13)$$

by taking

$$y = \log(m(A)), \quad (14)$$

$$a = \frac{m(A \cup B)}{1 - m(A \cup B)}, \quad (15)$$

$$b = -\frac{s(A) - m(A \cup B)}{1 - m(A \cup B)}. \quad (16)$$

The solution of (13) are given by [5]

$$y = W(be^a) - a. \quad (17)$$

Once  $y$  is calculated by formula (17) and since  $y = \log(m(A))$  we obtain the solution for  $m(A)$  given by

$$m(A) = e^y = e^{W(be^a) - a}. \quad (18)$$

Similarly, the solution for  $m(B)$  will be given by

$$m(B) = e^y = e^{W(be^a) - a}, \quad (19)$$

by solving the equation  $(y + a)e^y = b$  with

$$y = \log(m(B)), \quad (20)$$

$$a = \frac{m(A \cup B)}{1 - m(A \cup B)}, \quad (21)$$

$$b = -\frac{s(B) - m(A \cup B)}{1 - m(A \cup B)}. \quad (22)$$

We must however check if there is one solution only  $m(A) = e^{W_0(be^a) - a}$ , or in fact two solutions  $m_1(A) = e^{W_0(be^a) - a}$  and  $m_2(A) = e^{W_{-1}(be^a) - a}$ , and similarly for the solution for  $m(B)$ . This depends on the parameters  $a$  and  $b$  with respect to  $[-1/e, 0)$  interval and  $[0, \infty)$ .

We illustrate in the next subsection how to calculate the SEIP solution from these analytical formulas for the previous exemple.

<sup>6</sup>If the two masses values are admissible, that is if  $m_1(A \cup B) \in [0, 1]$  and if  $m_2(A \cup B) \in [0, 1]$ . If one of them is non-admissible it is eliminated.

### C. SEIP solution of the previous exemple

We recall that we have for this exemple  $s(\emptyset) = 0$ ,  $s(A) \approx 0.3773$ ,  $s(B) \approx 0.4290$  and  $s(\Theta) \approx 0.3219$ . If we apply formulas (11)-(12) for this exemple, we have  $a = -s(A \cup B) = -0.3219$  and therefore

$$y_1 = W_0(-0.3219) = -0.5681,$$

$$y_2 = W_{-1}(-0.3219) = -1.6094.$$

Hence the two potential solutions for the mass  $m(A \cup B)$  are

$$m_1(A \cup B) = e^{y_1} \approx 0.5666,$$

$$m_2(A \cup B) = e^{y_2} = 0.2000.$$

It can be easily verified that

$$-m_1(A \cup B) \log(m_1(A \cup B)) = 0.3219 = s(A \cup B),$$

$$-m_2(A \cup B) \log(m_2(A \cup B)) = 0.3219 = s(A \cup B).$$

We see that the second potential solution  $m_2(A \cup B) = 0.2000$  is the solution that corresponds to the original mass of  $A \cup B$  of the BBA  $m(A \cup B)$  of our exemple.

Now, we examine what would be the values of  $m(A)$  and  $m(B)$  given respectively by (18) and (19) by taking either  $m(A \cup B) = m_1(A \cup B) = 0.5666$  or  $m(A \cup B) = m_2(A \cup B) = 0.20$ .

- Let's examine the 1st possibility with the potential solution

$$m(A \cup B) = m_1(A \cup B) = 0.5666.$$

For determining  $m(A)$ , we have to solve  $(y + a)e^y = b$  with the unknown  $y = \log(m(A))$  and with

$$a = \frac{m(A \cup B)}{1 - m(A \cup B)} \approx \frac{0.5666}{1 - 0.5666} = 1.3073,$$

$$b = -\frac{s(A) - m(A \cup B)}{1 - m(A \cup B)} \approx -\frac{0.3773 - 0.5666}{1 - 0.5666} = 0.4369.$$

Hence,  $be^a = 0.4368 \cdot e^{1.3073} \approx 1.6148$ .

Applying formula (18), one gets<sup>7</sup>

$$m_1(A) = e^{W_0(be^a) - a} = 0.5769,$$

$$m_2(A) = e^{W_{-1}(be^a) - a} = -0.0216 + 0.0924i.$$

For determining  $m(B)$  we have to solve  $(y + a)e^y = b$  with the unknown  $y = \log(m(B))$  and with

$$a = \frac{m(A \cup B)}{1 - m(A \cup B)} \approx \frac{0.5666}{1 - 0.5666} = 1.3073,$$

$$b = -\frac{s(B) - m(A \cup B)}{1 - m(A \cup B)} \approx -\frac{0.4290 - 0.5666}{1 - 0.5666} = 0.3176.$$

Hence,  $be^a = 0.3176 \cdot e^{1.3073} \approx 1.1739$ .

Applying formula (19), one gets

$$m_1(B) = e^{W_0(be^a) - a} = 0.5065,$$

$$m_2(B) = e^{W_{-1}(be^a) - a} = -0.0204 + 0.0657i.$$

<sup>7</sup>Using *lambertw* Matlab™ function.

One sees that there is no effective choice for the values of  $m(A)$  and  $m(B)$  if we suppose  $m(A \cup B) = m_1(A \cup B) = 0.5666$  because if one takes as real values solutions  $m(A) = m_1(A) = 0.5769$  and  $m(B) = m_1(B) = 0.5065$  one would get

$$m(A) + m(B) + m(A \cup B) = 0.5769 + 0.5065 + 0.5666 = 1.65,$$

which is obviously greater than one. This generates an improper BBA.

- Let's consider the 2nd possibility with the potential solution

$$m(A \cup B) = m_2(A \cup B) = 0.20.$$

For determining  $m(A)$ , we have to solve  $(y + a)e^y = b$  with the unknown  $y = \log(m(A))$  and with

$$a = \frac{m(A \cup B)}{1 - m(A \cup B)} = \frac{0.20}{1 - 0.20} = 0.25,$$

$$b = -\frac{s(A) - m(A \cup B)}{1 - m(A \cup B)} \approx -\frac{0.3773 - 0.20}{1 - 0.20} = -0.2216.$$

Hence,  $be^a = -0.2216 \cdot e^{0.25} \approx -0.2845$ .

$$m_1(A) = e^{W_0(be^a) - a} = 0.5000,$$

$$m_2(A) = e^{W_{-1}(be^a) - a} = 0.1168.$$

For determining  $m(B)$  we have to solve  $(y + a)e^y = b$  with the unknown  $y = \log(m(B))$  and with

$$a = \frac{m(A \cup B)}{1 - m(A \cup B)} \approx \frac{0.20}{1 - 0.20} = 0.25,$$

$$b = -\frac{s(B) - m(A \cup B)}{1 - m(A \cup B)} \approx -\frac{0.4290 - 0.20}{1 - 0.20} = -0.2862.$$

Hence,  $be^a = -0.2862 \cdot e^{0.25} \approx -0.3675$ .

Applying formula (19), one gets

$$m_1(B) = e^{W_0(be^a) - a} = 0.3000,$$

$$m_2(B) = e^{W_{-1}(be^a) - a} = 0.2732.$$

Based on this 2nd possibility for potential solution  $m(A \cup B) = 0.20$ , one sees that the only possible effective choice of mass values  $m(A)$  and  $m(B)$  is to take  $m(A) = m_1(A) = 0.50$  and  $m(B) = m_1(B) = 0.30$  which gives the proper sought BBA such that  $m(A) + m(B) + m(A \cup B) = 1$  which exactly corresponds to the original BBA that has been used to generate the entropiece vector  $s(m)$  for this example.

In summary, for the case  $|\Theta| = 2$  it is always possible to calculate the BBA  $m(\cdot)$  from the knowledge of the entropiece vector, and the solution of SEIP is obtained by analytical formulas.

#### D. Remark

In the very particular case where  $s(A \cup B) = 0$  the equation (9) reduces to

$$-m(A \cup B) \log(m(A \cup B)) = 0, \quad (23)$$

which has two possible solutions  $m(A \cup B) = m_1(A \cup B) = 1$ , and  $m(A \cup B) = m_2(A \cup B) = 0$ .

If  $m(A \cup B) = 1$ , then it means that necessarily the BBA is the vacuous BBA, and so  $m(A) = m(B) = 0$ ,  $u(A) = Pl(A) - Bel(A) = 1$ ,  $u(B) = Pl(B) - Bel(B) = 1$ . Therefore<sup>8</sup>

$$\begin{aligned} s(A) &= -m(A)(1 - u(A)) \log(m(A)) + u(A)(1 - m(A)) \\ &= -m(A)(1 - m(A \cup B)) \log(m(A)) \\ &\quad + m(A \cup B)(1 - m(A)) \\ &= 0(1 - 1) \log(0) + 1(1 - 0) = 1, \\ s(B) &= -m(B)(1 - u(B)) \log(m(B)) + u(B)(1 - m(B)) \\ &= -m(B)(1 - m(A \cup B)) \log(m(B)) \\ &\quad + m(A \cup B)(1 - m(B)) \\ &= 0(1 - 1) \log(0) + 1(1 - 0) = 1. \end{aligned}$$

So the choice of  $m(A \cup B) = m_1(A \cup B) = 1$  is the only possible if the entropiece vector is  $s(m) = [110]^T$ .

If  $s(A) < 1$ , or if  $s(B) < 1$  (or both) then we must choose  $m(A \cup B) = m_2(A \cup B) = 0$ , and in this case we have to solve the equations

$$\begin{aligned} s(A) &= -m(A)(1 - u(A)) \log(m(A)) + u(A)(1 - m(A)) \\ &= -m(A)(1 - m(A \cup B)) \log(m(A)) \\ &\quad + m(A \cup B)(1 - m(A)) \\ &= -m(A) \log(m(A)), \\ s(B) &= -m(B)(1 - u(B)) \log(m(B)) + u(B)(1 - m(B)) \\ &= -m(B)(1 - m(A \cup B)) \log(m(B)) \\ &\quad + m(A \cup B)(1 - m(B)) \\ &= -m(B) \log(m(B)). \end{aligned}$$

The possible solutions of equation  $s(A) = -m(A) \log(m(A))$  are given by

$$m_1(A) = e^{W_0(-s(A))}, \quad (24)$$

$$m_2(A) = e^{W_{-1}(-s(A))}, \quad (25)$$

and the possible solutions of equation  $s(B) = -m(B) \log(m(B))$  are given by

$$m_1(B) = e^{W_0(-s(B))}, \quad (26)$$

$$m_2(B) = e^{W_{-1}(-s(B))}. \quad (27)$$

<sup>8</sup>We use the formal notation  $\log(0)$  even if  $\log(0)$  is  $-\infty$  because in our derivations we have always a  $0 \log(0)$  product which is equal to zero due to L'Hôpital's rule [4].

In this particular case where  $s(A \cup B) = 0$ , and  $s(A) < 1$  or  $s(B) < 1$ , we have to select the pair of possible solutions among the four possible choices

$$\begin{aligned} (m(A), m(B)) &= (m_1(A), m_1(B)), \\ (m(A), m(B)) &= (m_1(A), m_2(B)), \\ (m(A), m(B)) &= (m_2(A), m_1(B)), \\ (m(A), m(B)) &= (m_2(A), m_2(B)). \end{aligned}$$

The judicious choice of pair  $(m(A), m(B))$  must satisfy the proper BBA constraint  $m(A) + m(B) + m(A \cup B) = 1$ , where  $m(A \cup B) = 0$  because  $s(A \cup B) = 0$  in this particular case.

For instance, if we consider  $\Theta = \{A, B\}$  and the following (Bayesian) BBA

$$m(A) = 0.6, m(B) = 0.4, m(A \cup B) = 0.$$

The entropiece vector  $\mathbf{s}(m)$  is

$$\mathbf{s}(m) = \begin{bmatrix} s(A) \\ s(B) \\ s(A \cup B) \end{bmatrix} \approx \begin{bmatrix} 0.3065 \\ 0.3665 \\ 0 \end{bmatrix}. \quad (28)$$

Hence from  $\mathbf{s}(m)$  we can deduce  $m(A \cup B) = 0$  because we cannot consider  $m(A \cup B) = 1$  as a valid solution because  $s(A) < 1$  and  $s(B) < 1$ . The possible solutions of equation  $s(A) = -m(A) \log(m(A))$  are

$$\begin{aligned} m_1(A) &= e^{W_0(-s(A))} = e^{W_0(-0.3065)} = 0.6000, \\ m_2(A) &= e^{W_{-1}(-s(A))} = e^{W_{-1}(-0.3065)} = 0.1770, \end{aligned}$$

and the possible solutions of the equation  $s(B) = -m(B) \log(m(B))$  are

$$\begin{aligned} m_1(B) &= e^{W_0(-s(B))} = e^{W_0(-0.3665)} = 0.4000, \\ m_2(B) &= e^{W_{-1}(-s(B))} = e^{W_{-1}(-0.3665)} = 0.3367. \end{aligned}$$

One sees that the only effective (or judicious) choice for  $m(A)$  and  $m(B)$  is to take  $m(A) = m_1(A) = 0.60$  and  $m(B) = m_1(B) = 0.40$ , which coincides with the original Bayesian BBA that has been used to generate the entropiece vector  $\mathbf{s}(m) = [0.3065, 0.3665, 0]^T$ .

#### IV. CONCLUSION

In this paper we have introduced for the first time the entropiece inversion problem (EIP) which consists in calculating a basic belief assignment from the knowledge of a given entropiece vector which quantifies effectively the measure of uncertainty of a BBA in the framework of the theory of belief functions. The general analytical solution of this mathematical problem is a very challenging open problem because it involves transcendental equations. We have shown however how it is possible to obtain an analytical solution for the simplest EIP involving only two elements in the frame of discernment. Even in this simplest case the analytical solution of EIP is not easy to obtain because it requires a calculation of values of the transcendental Lambert's functions. Even if no general analytical formulas are found for the solution of general EIP, it would be interesting to develop numerical methods to approximate the general EIP solution, and to exploit it in future advanced information fusion systems.

#### REFERENCES

- [1] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [2] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, in Proc. of Fusion 2022 Int. Conf., ISIF Editor, Linköping, Sweden, July 2022.
- [3] J. Dezert, A. Tchamova, *On Effectiveness of Measures of Uncertainty of Basic Belief Assignments*, Information & Security Journal: An International Journal (ISIJ), Vol. 52, 2022.
- [4] R.E. Bradley, S.J. Petrilli, C.E. Sandifer, *L'Hôpital's analyse des infiniments petits (An annotated translation with source material by Johann Bernoulli)*, Birkhäuser, 311 pages, 2015.
- [5] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth, *On the Lambert W Function*, Advances in Computational Mathematics, Vol. 5, pp. 329–359, 1996.
- [6] Wikipedia, *Lambert W function* [https://en.wikipedia.org/wiki/Lambert\\_W\\_function](https://en.wikipedia.org/wiki/Lambert_W_function). Accessed 1st December 2022.
- [7] C.E. Shannon, *A mathematical theory of communication*, in [9] and in The Bell System Technical Journal, Vol. 27, pp. 379–423 & pp. 623–656, July, 1948.
- [8] J. Dezert, A. Tchamova, D. Han, *Measure of Information Content of Basic Belief Assignments*, in Proc. of Belief 2022 Int. Conf., Paris, France, Oct. 26–28, 2022.
- [9] N.J.A., Sloane, A.D. Wyner, *Claude Elwood Shannon - Collected Papers*, IEEE Press, 924 pages, 1993.

# Measure of Information Content of Basic Belief Assignments

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>, Deqiang Han<sup>c</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, F-91761 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

<sup>c</sup>Institute of Integrated Automation, Xi'an Jiaotong University, Xi'an, China.

Emails: jean.dezert@onera.fr, tchamova@bas.bg, deqhan@xjtu.edu.cn

Originally published as: J. Dezert, A. Tchamova, D. Han, *Measure of Information Content of Basic Belief Assignments*, in Proc. of Belief 2022 Int. Conf., Paris, France, Oct. 26–28, 2022, and reprinted with permission.

**Abstract**—In this paper, we present a measure of Information Content (IC) of Basic Belief Assignments (BBAs), and we show how it can be easily calculated. This new IC measure is interpreted as the dual of the effective measure of uncertainty (i.e. generalized entropy) of BBAs developed recently.

**Keywords:** belief functions, information content, generalized entropy.

## I. INTRODUCTION

Information quality (IQ) evaluation is of major importance for information processing and for helping the decision-making under uncertainty. In [1], the authors introduced the Accessibility, Interpretability, Relevance, and Integrity concepts as main attributes to describe the information quality in the context of assurance and belief networks, but unfortunately they present only general concepts without explicit formulas to evaluate quantitatively these attributes. In several recent books devoted to IQ [2]–[5], the authors proposed different models and methods of IQ evaluations. Recently in [6], Bouhamed et al. proposed a quantitative IQ evaluation using the possibility theory framework, which could be extended to the belief functions theory framework with further investigations. In this latter work, the information quantity component being necessary for the IQ evaluation is based on Gini's entropy rather than classical Shannon entropy. From the examination of these aforementioned references (and some references therein), it is far from obvious to make a clear justified choice among all these methods, especially when we model the uncertain information by belief functions (BF). What is clear however is that several distinct factors (or components) must be taken into account in the IQ evaluation mechanism. In this paper we focus on one of these components which is the Information Content (IC) component that we consider as the very (if not the most) essential component for IQ evaluation and indispensable for developing an effective IQ evaluation method in future research works.

It is worth noting that we do not address directly the whole IQ evaluation problem in this work but to provide a mathematical solution for measuring the IC of any Basic Belief Assignments (BBA) in the belief functions (BF) framework.

Our new IC measure is interpreted as the dual of an effective Measure of Uncertainty (MoU) developed recently [7]. We show how to calculate the IC of a BBA, and we also discuss the notion of information gain and information loss in the BF context. In our opinion, we cannot define a measure of Information Content independently of a Measure of Uncertainty (MoU) because they must be strongly related to each other. Actually these measures are two different sides of a same *abstract coin* we would say. On one side (the uncertainty side), more uncertainty content we have harder is the decision or choice to make, and on the other side (the information side) more information content we have easier and stronger is the decision or choice to make. This very simple and natural basic principle will be clarified mathematically next. So, the measure of information content of a BBA must reflect somehow the easiness and strength in the choice of an element of the frame of discernment drawn from the BBA (i.e. in the decision-making). This paper is organized as follows. After a brief recall of basics of belief functions in section II, we recall the effective MoU adopted in this work in section III. Section IV defines the measure of information content of a BBA and the information granules vector. Section V introduces the notions of information gain and information loss. Conclusions and perspectives appear in the last section.

## II. BELIEF FUNCTIONS

The belief functions (BF) were introduced by Shafer [8] for modeling epistemic uncertainty, reasoning about uncertainty and combining distinct sources of evidence. The answer of the problem under concern is assumed to belong to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \dots, \theta_N\}$  where all elements (i.e. members) of  $\Theta$  are exhaustive and mutually exclusive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of the power-set is  $2^{|\Theta|}$ . A (normalized) basic belief assignment (BBA) associated with a given source of evidence is a mapping  $m^\Theta(\cdot) : 2^\Theta \rightarrow [0, 1]$  such that  $m^\Theta(\emptyset) = 0$  and  $\sum_{X \in 2^\Theta} m^\Theta(X) = 1$ . A BBA  $m^\Theta(\cdot)$  characterizes a source of evidence related with a FoD  $\Theta$ . For notation shorthand, we can omit the superscript  $\Theta$  in  $m^\Theta(\cdot)$

notation if there is no ambiguity on the FoD we work with<sup>1</sup>. The quantity  $m(X)$  is called the mass of belief for  $X$ . The element  $X \in 2^\Theta$  is called a focal element (FE) of  $m(\cdot)$  if  $m(X) > 0$ . The set of all focal elements of  $m(\cdot)$  is denoted<sup>2</sup> by  $\mathcal{F}_\Theta(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$ .

The belief and the plausibility of  $X$  are defined for any  $X \in 2^\Theta$  by [8]

$$Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y), \quad (1)$$

$$Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) = 1 - Bel(\bar{X}), \quad (2)$$

where  $\bar{X} \triangleq \Theta \setminus \{X\}$  is the complement of  $X$  in  $\Theta$ .

One has always  $0 \leq Bel(X) \leq Pl(X) \leq 1$ , see [8]. For  $X = \emptyset$ ,  $Bel(\emptyset) = 0$  and  $Pl(\emptyset) = 0$ , and for  $X = \Theta$  one has  $Bel(\Theta) = 1$  and  $Pl(\Theta) = 1$ .  $Bel(X)$  and  $Pl(X)$  are often interpreted as the lower and upper bounds of unknown probability  $P(X)$  of  $X$ , that is  $Bel(X) \leq P(X) \leq Pl(X)$ . To quantify the uncertainty (i.e. the imprecision) of  $P(X) \in [Bel(X), Pl(X)]$ , we use the notation  $u(X) \in [0, 1]$  defined by

$$u(X) \triangleq Pl(X) - Bel(X). \quad (3)$$

The quantity  $u(X) = 0$  if  $Bel(X) = Pl(X)$  which means that  $P(X)$  is known precisely, and one has  $P(X) = Bel(X) = Pl(X)$ . One has  $u(\emptyset) = 0$  because  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and one has  $u(\Theta) = 0$  because  $Bel(\Theta) = Pl(\Theta) = 1$ . If all focal elements of  $m(\cdot)$  are singletons of  $2^\Theta$  the BBA  $m(\cdot)$  is a Bayesian BBA because  $\forall X \in 2^\Theta$  one has  $Bel(X) = Pl(X) = P(X)$  and  $u(X) = 0$ . Hence the belief and plausibility of  $X$  coincide with a probability measure  $P(X)$  defined on the FoD  $\Theta$ . The vacuous BBA characterizing a totally ignorant source of evidence is defined by  $m_v(X) = 1$  for  $X = \Theta$ , and  $m_v(X) = 0$  for all  $X \in 2^\Theta$  different from  $\Theta$ . This particular BBA has played a major role in the establishment of a new effective measure of uncertainty of BBA defined in [7].

### III. GENERALIZED ENTROPY OF A BBA

In [9] we did analyze in details forty-eight measures of uncertainty (MoU) of BBAs by covering 40 years of research works on this topic. Some of these MoUs capture only a particular aspect of the uncertainty inherent to a BBA (typically, the non-specificity and the conflict). Other MoUs propose a total uncertainty measure to capture jointly several aspects of the uncertainty. Unfortunately, most of these MoUs fail to satisfy four very simple reasonable and essential desiderata, and so they cannot be considered as really effective and useful. Actually only six MoUs can be considered as effective from the mathematical sense presented next, but unfortunately they appear as conceptually defective and disputable, see discussions in [9]. That is why, a better effective measure of uncertainty (MoU), i.e. generalized entropy of BBAs has been

developed and presented in [7]. The mathematical definition of this new effective entropy is given by

$$U(m) = \sum_{X \in 2^\Theta} s(X), \quad (4)$$

with

$$s(X) \triangleq -m(X)(1 - u(X)) \log(m(X)) + u(X)(1 - m(X)). \quad (5)$$

In (5), the term  $-(1 - u(X)) \log(m(X))$  is equal to  $(1 - u(X)) \log(1/m(X))$ , and  $\log(1/m(X))$  is called the surprisal<sup>3</sup> of  $X$ . Therefore  $(1 - u(X)) \log(1/m(X))$  represents the discounted surprisal of  $X$  by the confidence factor  $(1 - u(X))$  one has on the precision of the probability  $P(X)$ . The term  $u(X)(1 - m(X))$  entering in (5) corresponds to the imprecision  $u(X)$  about  $P(X)$  discounted by  $(1 - m(X))$ . The main idea is the greater  $m(X)$  the less one should take into account the imprecision  $u(X)$  in the MoU. The quantity  $s(X)$  is the uncertainty contribution related to element  $X$  (named the *entropiece* of  $X$ ) in the MoU  $U(m)$ . This entropiece  $s(X)$  involves  $m(X)$  and the imprecision  $u(X) = Pl(X) - Bel(X)$  about the unknown probability of  $X$  in a subtle interwoven manner. The cardinality of  $X$  is indirectly taken into account in the derivation of  $s(X)$  thanks to  $u(X)$  which requires the derivation of  $Pl(X)$  and  $Bel(X)$  functions depending on the cardinality of  $X$ . Because  $u(X) \in [0, 1]$  and  $m(X) \in [0, 1]$  one has  $s(X) \geq 0$ , and  $U(m) \geq 0$ . The quantity  $U(m)$  is expressed in *nats* because we use the natural logarithm.  $U(m)$  can be expressed in *bits* by dividing the  $U(m)$  value in *nats* by  $\log(2) = 0.69314718\dots$ . This measure of uncertainty  $U(m)$  is a continuous function in its basic belief mass arguments because it is a summation of continuous functions. In formula (5), we always take  $m(X) \log(m(X)) = 0$  when  $m(X) = 0$  because  $\lim_{m(X) \rightarrow 0^+} m(X) \log(m(X)) = 0$  which can be proved using L'Hôpital rule [11]. Note that for any BBA  $m$ , one has always  $s(\emptyset) = 0$  because  $m(\emptyset) = 0$  and  $u(\emptyset) = Pl(\emptyset) - Bel(\emptyset) = 0 - 0 = 0$ . For the vacuous BBA, one has  $s(\Theta) = 0$  because  $m_v(\Theta) = 1$  and  $u(\Theta) = Pl(\Theta) - Bel(\Theta) = 1 - 1 = 0$ .

The set  $\{s(X), X \in 2^\Theta\}$  of the entropieces values  $s(X)$  can be represented by an entropiece vector  $\mathbf{s}(m^\Theta) = [s(X), X \in 2^\Theta]^T$ , where any order of elements  $X$  of the power set  $2^\Theta$  can be chosen. For simplicity, we suggest to use the classical  $N$ -bits representation (if  $|\Theta| = N$ ) with the increasing order - see the next example.

This measure of uncertainty  $U(m)$  is effective because it can be proved (see proofs in [7]) that it satisfies the following four essential properties:

- 1)  $U(m) = 0$  for any BBA  $m(\cdot)$  focused on a singleton  $X$  of  $2^\Theta$ ;
- 2)  $U(m_v^\Theta) < U(m_v^{\Theta'})$  if  $|\Theta| < |\Theta'|$ ;

<sup>3</sup>This terminology is not used by Shannon in his original paper but it has been introduced by Tribus in [10] in the probabilistic context, and by analogy we adopt Tribus' terminology also for BBAs.

<sup>1</sup>However, we will keep  $m^\Theta(\cdot)$  notation when very necessary.

<sup>2</sup> $\triangleq$  means *equal by definition*.

- 3)  $U(m) = -\sum_{X \in \Theta} m(X) \log(m(X))$  if the BBA  $m(\cdot)$  is a Bayesian BBA. Hence,  $U(m)$  reduces to Shannon entropy [12] in this case;
- 4)  $U(m) < U(m_v)$  for any non-vacuous BBA  $m(\cdot)$  and for the vacuous BBA  $m_v(\cdot)$  defined with respect to the same FoD.

The proof of the three first properties is quite simple to make. The proof of the last property is much more difficult. As explained in [7], we do not consider that the sub-additivity property [13] of  $U(m)$  is a fundamental desideratum that an effective MoU must satisfy in general. In fact the sub-additivity desideratum is incompatible with the fourth important property  $U(m) < U(m_v)$  above which stipulates that none non-vacuous BBA can be more uncertain (i.e. more ignorant about the problem under consideration) than the vacuous BBA. Actually, it does not make sense to have the entropy  $U(m_v^{\Theta \times \Theta'})$  of the vacuous joint BBA  $m_v^{\Theta \times \Theta'}$  defined on the cartesian product space  $\Theta \times \Theta'$  smaller than (or equal to) the sum  $U(m_v^{\Theta}) + U(m_v^{\Theta'})$  of entropies of vacuous BBAs  $m_v^{\Theta}$  and  $m_v^{\Theta'}$  defined respectively on  $\Theta$  and  $\Theta'$ . There is no theoretical justification, nor intuitive reason for this sub-additivity desideratum in the context of non-Bayesian BBAs. Of course for Bayesian BBAs,  $U(m)$  is equivalent to Shannon entropy which is in this case sub-additive.

It can be also proved, see [7] for details, that the entropy of the vacuous BBA  $m_v$  related to a FoD  $\Theta$  is equal to

$$U(m_v^{\Theta}) = 2^{|\Theta|} - 2. \quad (6)$$

This maximum entropy value  $U(m_v)$  makes perfect sense because for this very particular BBA there is no information at all about the conflicts between the elements of the FoD. Actually for all  $X \in 2^{\Theta} \setminus \{\emptyset, \Theta\}$  one has  $u(X) = 1$  because  $[Bel(X), Pl(X)] = [0, 1]$ , and one has  $u(\emptyset) = 0$  and  $u(\Theta) = 0$ . Hence, the sum of all imprecisions of  $P(X)$  for all  $X \in 2^{\Theta}$  is exactly equal to  $2^{|\Theta|} - 2$  which corresponds to  $U(m_v^{\Theta})$  as expected. Moreover, one has always  $U(m_v^{\Theta}) > \log(|\Theta|)$  which means that the vacuous BBA has always an entropy greater than the maximum of Shannon entropy  $\log(|\Theta|)$  obtained with the uniform probability mass function distributed on  $\Theta$ .

**Example 1 of entropy calculation:** consider  $\Theta = \{\theta_1, \theta_2\}$  and the BBA  $m^{\Theta}(\theta_1) = 0.5$ ,  $m^{\Theta}(\theta_2) = 0.3$  and  $m^{\Theta}(\theta_1 \cup \theta_2) = 0.2$ , then one has  $[Bel(\emptyset), Pl(\emptyset)] = [0, 1]$  and  $u(\emptyset) = 0$ ,  $[Bel(\theta_1), Pl(\theta_1)] = [0.5, 0.7]$ ,  $[Bel(\theta_2), Pl(\theta_2)] = [0.3, 0.5]$ , and  $[Bel(\Theta), Pl(\Theta)] = [1, 1]$ . Hence,  $u(\theta_1) = 0.2$ ,  $u(\theta_2) = 0.2$  and  $u(\Theta) = 0$ . Applying (5), one gets  $s(\emptyset) = 0$ ,  $s(\theta_1) \approx 0.377258$ ,  $s(\theta_2) \approx 0.428953$  and  $s(\Theta) \approx 0.321887$ . Using the 2-bits representation with increasing ordering<sup>4</sup>, we encode the elements of the power set as  $\emptyset = 00$ ,  $\theta_1 = 01$ ,  $\theta_2 = 10$  and  $\theta_1 \cup \theta_2 = 11$ . The entropiece vector for this simple example is

$$s(m^{\Theta}) = \begin{bmatrix} s(\emptyset) \\ s(\theta_1) \\ s(\theta_2) \\ s(\theta_1 \cup \theta_2) \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0.377258 \\ 0.428953 \\ 0.321887 \end{bmatrix}. \quad (7)$$

<sup>4</sup>Once the binary values are converted into their digit value with the most significant bit on the left (i.e the least significant bit on the right).

If we use the classical N-bits (here  $N = 2$ ) representation with increasing ordering (as we recommend) the first component of entropiece vector  $s(m^{\Theta})$  will be  $s(\emptyset)$  which is always equal to zero for any BBA  $m$ , hence the first component of  $s(m^{\Theta})$  is always zero. By summing all the components of the entropiece vector  $s(m^{\Theta})$  we obtain the entropy  $U(m^{\Theta}) \approx 1.128098$  nats of the BBA  $m^{\Theta}(\cdot)$ . Note that the components  $s(X)$  (for  $X \neq \emptyset$ ) of the entropieces vector  $s(m^{\Theta})$  are not independent because they are linked to each other through the calculation of  $Bel(X)$  and  $Pl(X)$  values entering in  $u(X)$ .

**Example 2 of entropy calculation:** for the vacuous BBA  $m_v^{\Theta}$ , and when using the binary increasing encoding of elements of  $2^{\Theta}$ , the first component  $s(\emptyset)$  and the last component  $s(\Theta)$  of entropiece vector  $s(m_v^{\Theta})$  will always be equal to zero, and all other components of  $s(m_v^{\Theta})$  will be equal to one. For instance, if we consider  $\Theta = \{\theta_1, \theta_2\}$  and the vacuous BBA  $m_v^{\Theta}(\theta_1) = 0$ ,  $m_v^{\Theta}(\theta_2) = 0$  and  $m_v^{\Theta}(\theta_1 \cup \theta_2) = 1$ , the corresponding entropiece vector  $s(m_v^{\Theta})$  is

$$s(m_v^{\Theta}) = \begin{bmatrix} s(\emptyset) \\ s(\theta_1) \\ s(\theta_2) \\ s(\theta_1 \cup \theta_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}. \quad (8)$$

By summing all the components of the entropiece vector  $s(m_v^{\Theta})$  we obtain the entropy value  $U(m_v^{\Theta}) = 2$  nats for this vacuous BBA  $m_v^{\Theta}(\cdot)$ , which is of course in agreement with the formula (6).

#### IV. INFORMATION CONTENT OF A BBA

We consider a (non-empty) FoD of cardinality  $|\Theta| = N$ , and we model our state of knowledge about the problem under consideration by a BBA defined on  $2^{\Theta}$ . Without more knowledge than the FoD itself (and its cardinality  $N$ ), we are totally ignorant about the solution of the problem we want to solve, and of course we have no clue for making a decision/choice among the elements of the FoD. The BBA reflecting this total ignorant situation is the vacuous BBA  $m_v(\cdot)$ , whose maximal entropy is  $U(m_v) = 2^N - 2$ . In such case, we naturally expect that the information content we have<sup>5</sup> is zero when the uncertainty measure is maximal. In the very opposite case, it is very natural to consider that the information content of a BBA is maximal if the entropy value (the MoU value) of a BBA  $m(\cdot)$  is zero, meaning that we make a choice of one element of the FoD without hesitation. Based on these very simple ideas, we propose to define the information content of any BBA  $m(\cdot)$  as the dual of the effective measure of uncertainty, more precisely by

$$IC(m^{\Theta}) \triangleq U(m_v^{\Theta}) - U(m^{\Theta}) \\ = (2^{|\Theta|} - 2) - \sum_{X \in 2^{\Theta}} s(X), \quad (9)$$

<sup>5</sup>aside of the value of  $N$  of course.

where  $s(X)$  is the *entropiece* of the element  $X \in 2^\Theta$  given by (5), that is

$$s(X) \triangleq -(1 - u(X))m^\Theta(X) \log(m^\Theta(X)) + u(X)(1 - m^\Theta(X)),$$

and where  $u(X)$  is the level of imprecision of the probability  $P(X)$  given by

$$u(X) = Pl^\Theta(X) - Bel^\Theta(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m^\Theta(Y) - \sum_{Y \in 2^\Theta | Y \subseteq X} m^\Theta(Y) \quad (10)$$

From the definition (9), one sees that for  $m^\Theta \neq m_v^\Theta$  one has  $IC(m^\Theta) > 0$  because  $U(m^\Theta) < U(m_v^\Theta)$ , and for  $m^\Theta = m_v^\Theta$  one has  $IC(m_v^\Theta) = 0$ , which is what we naturally expect.

It is worth mentioning that the information content  $IC(m^\Theta)$  of a BBA depends not only of the BBA  $m(\cdot)$  itself but also on the cardinality of the frame of discernment<sup>6</sup>  $\Theta$  because  $IC(m^\Theta)$  requires the knowledge of  $|\Theta|$  to calculate the max entropy value  $U(m_v^\Theta) = 2^{|\Theta|} - 2$  entering in (9). This remark is very important to understand that even if two BBAs (defined on different FoDs) focus entirely on a same focal element, their information contents are necessarily different. For instance, if we consider the Bayesian BBA with  $m^\Theta(\theta_1) = 1$  defined on the FoD  $\Theta = \{\theta_1, \theta_2\}$ , then

$$IC(m^\Theta) = U(m_v^\Theta) - U(m^\Theta) = (2^{|\Theta|} - 2) - 0 = 2 \text{ (nats)},$$

whereas if we consider the Bayesian BBA with  $m^{\Theta'}(\theta_1) = 1$  defined on the larger FoD  $\Theta' = \{\theta_1, \theta_2, \theta_3\}$  (for instance), then

$$IC(m^{\Theta'}) = U(m_v^{\Theta'}) - U(m^{\Theta'}) = (2^{|\Theta'|} - 2) - 0 = 6 \text{ (nats)}.$$

So even if the decision  $\theta_1$  that we would make based either on  $m^\Theta$  or on  $m^{\Theta'}$  is the same, these decisions must not be considered actually with the same strength, and this is what reflects our information content measure.

From this very simple definition of information content, we can also define the Normalized Information Content (NIC) (if needed later in some applications), denoted by  $NIC(m^\Theta)$  by normalizing  $IC(m^\Theta)$  with respect to the maximal value of entropy  $U(m_v^\Theta)$  as

$$NIC(m^\Theta) \triangleq \frac{U(m_v^\Theta) - U(m^\Theta)}{U(m_v^\Theta)} = 1 - \frac{U(m^\Theta)}{U(m_v^\Theta)}. \quad (11)$$

Hence we will have  $NIC(m^\Theta) \in [0, 1]$  and  $NIC(m^\Theta) = 0$  for  $m = m_v$ , and  $NIC(m^\Theta) = 1$  for  $U(m) = 0$  which is obtained when  $m(\cdot)$  is entirely focused on a singleton  $\theta_i \in \Theta$ , that is  $m^\Theta(\theta_i) = 1$  for some  $i \in \{1, 2, \dots, |\Theta|\}$ .

In fact, the (total) information content of a BBA  $IC(m^\Theta)$  is the sum of all the *information granules*  $IG(X|m^\Theta)$  of elements  $X \in 2^\Theta$  carried by a BBA  $m^\Theta$ , that is

$$IC(m^\Theta) = \sum_{X \in 2^\Theta} IG(X|m^\Theta) \quad (12)$$

<sup>6</sup>That is why it is better, we think, to use the notation  $IC(m^\Theta)$  instead of  $IC(m)$ .

where

$$IG(X|m^\Theta) \triangleq \begin{cases} 0, & \text{if } X = \emptyset, \\ -s(X), & \text{if } X = \Theta, \\ 1 - s(X) & \text{otherwise.} \end{cases} \quad (13)$$

We can define the information granules vector<sup>7</sup>  $\mathbf{IG}(m) = [IG(X|m^\Theta), X \in 2^\Theta]^T$  by

$$\mathbf{IG}(m^\Theta) \triangleq \mathbf{s}(m_v^\Theta) - \mathbf{s}(m^\Theta). \quad (14)$$

One sees that the (total) information content  $IC(m^\Theta)$  of a BBA  $m^\Theta$  is just the sum of all components  $IG(X|m^\Theta)$  of the information granules vector  $\mathbf{IG}(m)$ . The information granules vector  $\mathbf{IG}(m)$  is interesting and useful because it helps to see the contribution of each element  $X$  in the whole measure of the information content  $IC(m^\Theta)$  of a BBA  $m^\Theta$ .

**Example 1 (continued):** consider  $\Theta = \{\theta_1, \theta_2\}$  and the BBA  $m^\Theta(\theta_1) = 0.5$ ,  $m^\Theta(\theta_2) = 0.3$  and  $m^\Theta(\theta_1 \cup \theta_2) = 0.2$ . The information granules vector  $\mathbf{IG}(m^\Theta)$  is given by

$$\begin{aligned} \mathbf{IG}(m^\Theta) &= \mathbf{s}(m_v^\Theta) - \mathbf{s}(m^\Theta) \\ &= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0.377258 \\ 0.428953 \\ 0.321887 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.622742 \\ 0.571047 \\ -0.321887 \end{bmatrix}. \end{aligned}$$

By summing all the components of the information granules vector  $\mathbf{IG}(m^\Theta)$  we obtain the (total) information content  $IC(m^\Theta) = 0.871902$  nats of the BBA  $m^\Theta$ , which can of course be calculated directly also as

$$IC(m^\Theta) = U(m_v^\Theta) - U(m^\Theta) = 2 - 1.128098 = 0.871902.$$

However, the information granules vector  $\mathbf{IG}(m^\Theta)$  is interesting to identify the contribution of each element  $X$  in the whole measure of the information content.

## V. INFORMATION GAIN AND INFORMATION LOSS

Once the IC measure is defined for a BBA, it is rather simple to define the information gain and information loss of a BBA with respect to another one, both defined on a same FoD  $\Theta$ . Suppose that we have a first BBA  $m_1^\Theta$  and a second BBA  $m_2^\Theta$ , then we can calculate by formula (9) their respective information contents  $IC(m_1^\Theta)$  and  $IC(m_2^\Theta)$ . The difference of information content measure of  $m_2^\Theta$  with respect to  $m_1^\Theta$  is defined by<sup>8</sup>

$$\Delta_{IC}(m_2|m_1) \triangleq IC(m_2^\Theta) - IC(m_1^\Theta). \quad (15)$$

If we replace  $IC(m_2^\Theta)$  and  $IC(m_1^\Theta)$  by their expressions according to (9), it comes

$$\begin{aligned} \Delta_{IC}(m_2|m_1) &= [U(m_v^\Theta) - U(m_2^\Theta)] - [U(m_v^\Theta) - U(m_1^\Theta)] \\ &= U(m_1^\Theta) - U(m_2^\Theta). \end{aligned} \quad (16)$$

<sup>7</sup>We suppose for convenience that the elements  $X \in 2^\Theta$  are listed in increasing order using the classical  $|\Theta|$ -bits representation with the least significant bit on the right.

<sup>8</sup>Similarly, we can define  $\Delta_{IC}(m_1|m_2) \triangleq IC(m_1^\Theta) - IC(m_2^\Theta) = -\Delta_{IC}(m_2|m_1)$ .

If  $\Delta_{IC}(m_2|m_1) = 0$ , the BBAs  $m_1^\ominus$  and  $m_2^\ominus$  have same measure of information content. So, there is no gain and no loss in information content if one switches from  $m_1^\ominus$  to  $m_2^\ominus$  or vice versa.  $\Delta_{IC}(m_2|m_1) = 0$  does not mean that the decisions based on  $m_1^\ominus$  and on  $m_2^\ominus$  are the same. It does only means that the decision based on  $m_1^\ominus$  must be as easy as the decision made based on  $m_2^\ominus$ . It means that they have the same informational strength. That's it. If  $\Delta_{IC}(m_2|m_1) > 0$ , one has  $IC(m_2^\ominus) > IC(m_1^\ominus)$ , i.e. the BBA  $m_2^\ominus$  is more informative than  $m_1^\ominus$ . In this case we get an information gain if one switches from  $m_1^\ominus$  to  $m_2^\ominus$ , and by duality we get an uncertainty reduction by switching from  $m_1^\ominus$  to  $m_2^\ominus$ . It means that it must be easier to make a decision based on  $m_2^\ominus$  rather on  $m_1^\ominus$ . If  $\Delta_{IC}(m_2|m_1) < 0$ , one has  $IC(m_2^\ominus) < IC(m_1^\ominus)$ , i.e. the BBA  $m_2^\ominus$  is less informative than  $m_1^\ominus$ . In this case we get an information loss if one switches from  $m_1^\ominus$  to  $m_2^\ominus$ , and by duality we get an uncertainty raise by switching from  $m_1^\ominus$  to  $m_2^\ominus$ . It means that it must be easier to make a decision based on  $m_1^\ominus$  rather on  $m_2^\ominus$ .

As simple example, consider  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . For the vacuous BBA one has  $U(m_v^\ominus) = 2^3 - 2 = 6$  nats. Suppose at time  $k = 1$  one has the BBA  $m_1^\ominus(\theta_1 \cup \theta_2) = 0.2$ ,  $m_1^\ominus(\theta_1 \cup \theta_3) = 0.3$ ,  $m_1^\ominus(\theta_1 \cup \theta_2 \cup \theta_3) = 0.5$ , then  $U(m_1^\ominus) \approx 5.1493$  nats, and  $IC(m_1^\ominus) = U(m_v^\ominus) - U(m_1^\ominus) \approx 0.8507$  nats. Suppose that after some information processing (belief revision, or fusion, etc) we come up with the BBA  $m_2^\ominus$  at time  $k = 2$  defined by  $m_2^\ominus(\theta_1) = 0.2$  and  $m_2^\ominus(\theta_1 \cup \theta_3) = 0.8$ , then  $U(m_2^\ominus) \approx 0.5004$  nats and  $IC(m_2^\ominus) = U(m_v^\ominus) - U(m_2^\ominus) \approx 5.4996$  nats. In this case, we get  $\Delta_{IC}(m_2|m_1) = 5.4996 - 0.8507 = 4.6489$  which is positive. Hence we get an information gain by switching from  $m_1^\ominus$  to  $m_2^\ominus$  thanks to the information processing applied.

## VI. CONCLUSIONS

In this paper we have introduced a measure of information content (IC) for any basic belief assignment (BBA). This IC measure based on an effective measure of uncertainty of BBAs is quite simple to calculate, and it reflects somehow the informational strength and easiness ability to make a decision based on any belief mass function. We have also shown how it is possible to identify the contribution of each focal element of the BBA to this information content measure thanks to the information granule vector. This new IC measure is also interesting because it allows to well quantify the information loss or gain between two BBAs, and thus as perspectives we could use it to quantify precisely and compare the performances of information processing using belief functions (like fusion rules, belief conditioning, etc). We hope that this new theoretical IC measure will open interesting tracks for forthcoming research works on reasoning about uncertainty with belief functions.

## REFERENCES

- [1] M. Bovee, R.S. Srivastava, *A Conceptual Framework and Belief-Function Approach to Assessing Overall Information Quality*, Int. J. of Intell. Systems, Vol. 18(1), pp. 51–74, January, 2003.
- [2] Floridi, L., Illari, P. (Editors): *The philosophy of information quality*, Springer International Publishing Switzerland (2014).

- [3] C. Batini, M. Scannapieco, *Data and Information Quality: Dimensions, Principles and Techniques*, Springer International Publishing, Switzerland, 2016.
- [4] R.S. Kenett, G. Shmueli, *Information quality*, John Wiley & Sons, 2017.
- [5] E. Bossé, G.L. Rogova (Editors), *Information Quality in Information Fusion and Decision Making*, Information Fusion and Data Science, Springer Nature, Switzerland, 2019.
- [6] S.A. Bouhamed, I.K. Kalle, R.R. Yager, E. Bossé, B. Solaiman, *An intelligent quality-based approach to fusing multi-source possibilistic information*, Information Fusion, Vol. 55, pp. 68–90, 2020.
- [7] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, in Proc. of Fusion 2022 Int. Conf., Linköping, Sweden, pp. 1–10, July, 2022.
- [8] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, U.S.A., 1976.
- [9] J. Dezert, A. Tchamova, *On Effectiveness of Measures of Uncertainty of Basic Belief Assignments*, Information & Security Journal: An International Journal (ISIJ), Vol. 52, February, 2022.
- [10] M. Tribus, *Rational Descriptions, Decisions and Designs*, Pergamon Press, New York, pp. 26–28, 1969.
- [11] R.E. Bradley, S.J. Petrilli, C.E. Sandifer, *L'Hôpital's analyse des infiniments petits (An annotated translation with source material by Johann Bernoulli)*, Birkhäuser, 311 pages, 2015.
- [12] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379–423 & 623–656, July & October, 1948.
- [13] G.J. Klir, *Principles of uncertainty: What are they? Why do we need them?*, Fuzzy Sets and Systems, Vol. 74, pp. 15–31, 1995.



# Erratum of paper entitled: On the Validity of Dempster’s Fusion Rule and its Interpretation as a Generalization of Bayesian Fusion Rule

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA - DTIS/MIDL, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

Emails: jean.dezert@onera.fr, tchamova@bas.bg

**Abstract**—In this erratum we correct a mathematical mistake included in the paper entitled: *On the Validity of Dempster’s Fusion Rule and its Interpretation as a Generalization of Bayesian Fusion Rule*, published in 2014 in [1]. In taking into account this mathematical correction the Bayesian fusion rule is associative in contrary to what is claimed in the original version of our paper. The comments in our paper remain valid for pages 223 to 238. Corrections in several pages from page 239 to the end of our paper must be done as explained next in this erratum.

**Keywords:** Bayesian fusion, belief functions.

## ERRATUM

In [1] page 239, the general formulas<sup>1</sup> #(34)–#(36) are incorrect. The correct formulas are presented in this erratum.

Based on conditional statistical independence assumption  $P(Z_1, Z_2|X) = P(Z_1|X)P(Z_2|X)$ , we have

$$\begin{aligned} P(X|Z_1 \cap Z_2) &= \frac{P(Z_1 \cap Z_2|X)P(X)}{P(Z_1 \cap Z_2)} \\ &= \frac{P(Z_1|X)P(Z_2|X)P(X)}{P(Z_1 \cap Z_2)} \\ &= \frac{\frac{P(X|Z_1)P(Z_1)}{P(X)} \cdot \frac{P(X|Z_2)P(Z_2)}{P(X)} P(X)}{\sum_{i=1}^N P(X = x_i, Z_1 \cap Z_2)}, \end{aligned} \quad (1)$$

which can be written as

$$P(X|Z_1 \cap Z_2) = \frac{P(X|Z_1)P(X|Z_2)/P(X)}{\sum_{i=1}^N \frac{P(X=x_i|Z_1)P(X=x_i|Z_2)}{P(X=x_i)}}. \quad (2)$$

The formula (2) corresponds to formula #(24) of our original paper [1]. This formula (2) can be rewritten in a symmetrical form as follows

$$\begin{aligned} P(X|Z_1 \cap Z_2) &= \frac{1}{K'(Z_1, Z_2)} \cdot \frac{P(X|Z_1)}{\sqrt{P(X)}} \cdot \frac{P(X|Z_2)}{\sqrt{P(X)}} \\ &= \frac{1}{K'(Z_1, Z_2)} \cdot \frac{P(X|Z_1)}{P^{\frac{1}{2}}(X)} \cdot \frac{P(X|Z_2)}{P^{\frac{1}{2}}(X)}, \end{aligned} \quad (3)$$

<sup>1</sup>For avoiding confusion with formula number in this erratum, we denote the formula number appearing in the original paper [1] by #(xx), where xx is the number under concern.

where the normalization constant  $K'(Z_1, Z_2)$  is given by:

$$\begin{aligned} K'(Z_1, Z_2) &\triangleq \sum_{i=1}^N \frac{P(X = x_i|Z_1)}{\sqrt{P(X = x_i)}} \cdot \frac{P(X = x_i|Z_2)}{\sqrt{P(X = x_i)}} \\ &= \sum_{i=1}^N \frac{P(X = x_i|Z_1)}{P^{\frac{1}{2}}(X = x_i)} \cdot \frac{P(X = x_i|Z_2)}{P^{\frac{1}{2}}(X = x_i)} \end{aligned} \quad (4)$$

The formulas #(24)–#(33) of [1] are correct.

If we generalize the formula (1) for  $s > 2$  conditioning terms, we obtain the following expression

$$\begin{aligned} P(X|Z_1 \cap \dots \cap Z_s) &= \frac{P(Z_1 \cap \dots \cap Z_s|X)P(X)}{P(Z_1 \cap \dots \cap Z_s)} \\ &= \frac{P(Z_1|X) \dots P(Z_s|X)P(X)}{P(Z_1 \cap \dots \cap Z_s)} \\ &= \frac{\frac{P(X|Z_1)P(Z_1)}{P(X)} \dots \frac{P(X|Z_s)P(Z_s)}{P(X)} P(X)}{\sum_{i=1}^N P(X = x_i, Z_1 \cap \dots \cap Z_s)}, \end{aligned} \quad (5)$$

which can be written as

$$P(X|Z_1 \cap \dots \cap Z_s) = \frac{P(X|Z_1) \dots P(X|Z_s)/P^{s-1}(X)}{\sum_{i=1}^N \frac{P(X=x_i|Z_1) \dots P(X=x_i|Z_s)}{P^{s-1}(X=x_i)}}, \quad (6)$$

or equivalently as

$$P(X|Z_1 \cap \dots \cap Z_s) = \frac{\prod_{k=1}^s P(X|Z_k)}{K(X, Z_1, \dots, Z_s)}, \quad (7)$$

where the coefficient  $K(X, Z_1, \dots, Z_s)$  is defined by

$$K(X, Z_1, \dots, Z_s) \triangleq P^{s-1}(X) \sum_{i=1}^N \frac{(\prod_{k=1}^s P(X = x_i|Z_k))}{P^{s-1}(X = x_i)}. \quad (8)$$

The formula #(34) of [1] must be replaced by the formula (8) above.

The symmetrized form of Eq. (6) is:

$$\begin{aligned} P(X|Z_1 \cap \dots \cap Z_s) &= \frac{1}{K'(Z_1, \dots, Z_s)} \cdot \prod_{k=1}^s \frac{P(X|Z_k)}{\sqrt[s]{P^{s-1}(X)}} \\ &= \frac{1}{K'(Z_1, \dots, Z_s)} \cdot \prod_{k=1}^s \frac{P(X|Z_k)}{P^{\frac{s-1}{s}}(X)}, \end{aligned} \quad (9)$$

with the normalization constant  $K'(Z_1, \dots, Z_s)$  given by:

$$\begin{aligned} K'(Z_1, \dots, Z_s) &\triangleq \sum_{i=1}^N \prod_{k=1}^s \frac{P(X = x_i|Z_k)}{\sqrt[s]{P^{s-1}(X = x_i)}} \\ &= \sum_{i=1}^N \prod_{k=1}^s \frac{P(X = x_i|Z_k)}{P^{\frac{s-1}{s}}(X = x_i)} \end{aligned} \quad (10)$$

Hence the incorrect expression #(35) of  $P(X|Z_1 \cap \dots \cap Z_s)$  in [1] must be replaced by the formula (9) above, and the incorrect expression #(36) of  $K'(Z_1, \dots, Z_s)$  must be replaced by the formula (10).

The agreement  $A_s(X)$  of order  $s$ , the global agreement  $GA_s$ , and the global conflict  $GC_s$  for  $s$  sources must be also corrected as follows:

$$A_s(X = x_i) \triangleq \prod_{k=1}^s \frac{P(X = x_i|Z_k)}{\sqrt[s]{P^{s-1}(X = x_i)}},$$

$$\begin{aligned} GA_s &\triangleq \sum_{i_1, \dots, i_s=1}^N \frac{P(X = x_{i_1}|Z_1)}{\sqrt[s]{P^{s-1}(X = x_{i_1})}} \\ &\quad \dots \frac{P(X = x_{i_s}|Z_s)}{\sqrt[s]{P^{s-1}(X = x_{i_s})}}, \end{aligned}$$

$$\begin{aligned} GC_s &\triangleq \sum_{i_1, \dots, i_s=1}^N \frac{P(X = x_{i_1}|Z_1)}{\sqrt[s]{P^{s-1}(X = x_{i_1})}} \\ &\quad \dots \frac{P(X = x_{i_s}|Z_s)}{\sqrt[s]{P^{s-1}(X = x_{i_s})}} - GA_s. \end{aligned}$$

The first consequence of this correction is that the property P1 stated in [1] page 242 must be corrected as (P1): *The PMF  $P(X)$  is a neutral element of Bayes fusion rule.* Remark 2 and formula #(45) on page 242 must be removed.

The remark 3 on page 242 of [1] is incorrect. Indeed, if we take  $P(X|Z_k) = P(X)$  for  $k = 1, \dots, s$  and based on the correct formula (9), we get actually

$$\text{Bayes}(P(X), P(X), \dots, P(X); P(X)) = P(X),$$

and for any type of pmf  $P(X)$  (i.e. uniform, and non-uniform pmf).

The property (P3): *The Bayes fusion rule is in general not associative* stated in [1] on page 242 is incorrect and it must be corrected as (P3): *The Bayes fusion rule is associative.*

**Proof of the property P3** (Associativity of Bayes rule): The expression of  $P(X|Z_1 \cap \dots \cap Z_{s-1})$  is given by formula (9) when using  $s - 1$  conditioning terms. Hence we have

$$P(X|Z_1 \cap \dots \cap Z_{s-1}) = \frac{\prod_{k=1}^{s-1} \frac{P(X|Z_k)}{P^{\frac{s-2}{s-1}}(X)}}{K'(Z_1, \dots, Z_{s-1})}, \quad (11)$$

with the normalization constant  $K'(Z_1, \dots, Z_{s-1})$  given by

$$K'(Z_1, \dots, Z_{s-1}) = \sum_{i=1}^N \prod_{k=1}^{s-1} \frac{P(X = x_i|Z_k)}{P^{\frac{s-2}{s-1}}(X = x_i)}. \quad (12)$$

To calculate  $P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s)$  from  $P(X|Z_1 \cap \dots \cap Z_{s-1})$  and  $P(X|Z_s)$ , we use Bayes formula with the conditional statistical independence assumption, and we get

$$\begin{aligned} P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s) &= \\ &= \frac{P(Z_1 \cap \dots \cap Z_{s-1}|X)P(Z_s|X)P(X)}{\sum_{i=1}^N P(Z_1 \cap \dots \cap Z_{s-1}|X = x_i)P(Z_s|X = x_i)P(X = x_i)}. \end{aligned} \quad (13)$$

Because

$$P(Z_1 \cap \dots \cap Z_{s-1}|X) = \frac{P(X|Z_1 \cap \dots \cap Z_{s-1})P(Z_1 \cap \dots \cap Z_{s-1})}{P(X)},$$

and

$$P(Z_s|X) = \frac{P(X|Z_s)P(Z_s)}{P(X)},$$

the expression of  $P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s)$  given by (13) can be rewritten as

$$\begin{aligned} P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s) &= \\ &= \frac{\frac{P(X|Z_1 \cap \dots \cap Z_{s-1})}{P(X)} \frac{P(X|Z_s)P(Z_s)}{P(X)} P(X)}{\sum_{i=1}^N \frac{P(X = x_i|Z_1 \cap \dots \cap Z_{s-1})}{P(X = x_i)} \frac{P(X = x_i|Z_s)}{P(X = x_i)} P(X = x_i)} \end{aligned} \quad (14)$$

After the simplification by  $P(X)$  in the numerator of (14) and the simplification by  $P(X = x_i)$  in the denominator of (14) it comes

$$\begin{aligned} P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s) &= \\ &= \frac{P(X|Z_1 \cap \dots \cap Z_{s-1}) \frac{P(X|Z_s)}{P(X)}}{\sum_{i=1}^N P(X = x_i|Z_1 \cap \dots \cap Z_{s-1}) \frac{P(X = x_i|Z_s)}{P(X = x_i)}}. \end{aligned} \quad (15)$$

Replacing  $P(X|Z_1 \cap \dots \cap Z_{s-1})$  by its expression given in (11), we have

$$\begin{aligned} P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s) &= \\ &= \frac{\left[ \frac{1}{K'(Z_1, \dots, Z_{s-1})} \cdot \prod_{k=1}^{s-1} \frac{P(X|Z_k)}{P^{\frac{s-2}{s-1}}(X)} \right] \frac{P(X|Z_s)}{P(X)}}{\sum_{i=1}^N \left[ \frac{1}{K'(Z_1, \dots, Z_{s-1})} \cdot \prod_{k=1}^{s-1} \frac{P(X = x_i|Z_k)}{P^{\frac{s-2}{s-1}}(X = x_i)} \right] \frac{P(X = x_i|Z_s)}{P(X = x_i)}}. \end{aligned} \quad (16)$$

After simplification by the constant  $K'(Z_1, \dots, Z_{s-1})$  one gets

$$\begin{aligned}
 P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s) &= \\
 &= \frac{\left[ \prod_{k=1}^{s-1} \frac{P(X|Z_k)}{P^{\frac{s-2}{s-1}}(X)} \right] \frac{P(X|Z_s)}{P(X)}}{\sum_{i=1}^N \left[ \prod_{k=1}^{s-1} \frac{P(X=x_i|Z_k)}{P^{\frac{s-2}{s-1}}(X=x_i)} \right] \frac{P(X=x_i|Z_s)}{P(X=x_i)}} \\
 &= \frac{\frac{1}{P^{s-1}(X)} \prod_{k=1}^s P(X|Z_k)}{\sum_{i=1}^N \frac{1}{P^{s-1}(X=x_i)} \prod_{k=1}^s P(X=x_i|Z_k)}. \quad (17)
 \end{aligned}$$

The formula (17) can be rewritten with an equivalent symmetrical form as

$$\begin{aligned}
 P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s) &= \\
 &= \frac{\prod_{k=1}^s \frac{P(X|Z_k)}{P^{\frac{s-1}{s}}(X)}}{\sum_{i=1}^N \prod_{k=1}^s \frac{P(X=x_i|Z_k)}{P^{\frac{s-1}{s}}(X=x_i)}} = \frac{\prod_{k=1}^s \frac{P(X|Z_k)}{P^{\frac{s-1}{s}}(X)}}{K'(Z_1, \dots, Z_s)}, \quad (18)
 \end{aligned}$$

where  $K'(Z_1, \dots, Z_s) = \sum_{i=1}^N \prod_{k=1}^s \frac{P(X=x_i|Z_k)}{P^{\frac{s-1}{s}}(X=x_i)}$ .

Therefore, we have proved that expression of  $P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s)$  given by (18) is equal to the expression of  $P(X|Z_1 \cap \dots \cap Z_{s-1} \cap Z_s)$  given by (9). This proves the associativity of Bayes fusion rule, i.e. the validity of the property P3. Note that the equality  $P(X|(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s) = P(X|Z_1 \cap \dots \cap Z_{s-1} \cap Z_s)$  does not depend on a particular choice of the intersection of  $s-1$  subsets involved in the conditioning because the intersection operator is associative. Hence the conditioning terms  $(Z_1 \cap \dots \cap Z_{s-1}) \cap Z_s$  and  $Z_1 \cap \dots \cap Z_{s-1} \cap Z_s$  are equal. This implies that the two conditional probabilities must be necessary equal, which is proved by our previous derivations.

With the correct formulas (9)-(10), the numerical application for example 1 on page 243 of [1] gives

$$\begin{aligned}
 P(X = x_1|Z_1 \cap Z_2 \cap Z_3) &= \frac{1}{K_{123}} \frac{0.1}{\sqrt[3]{0.2^2}} \frac{0.5}{\sqrt[3]{0.2^2}} \frac{0.6}{\sqrt[3]{0.2^2}} \\
 &= 0.7273, \\
 P(X = x_2|Z_1 \cap Z_2 \cap Z_3) &= \frac{1}{K_{123}} \frac{0.9}{\sqrt[3]{0.8^2}} \frac{0.5}{\sqrt[3]{0.8^2}} \frac{0.4^2}{\sqrt[3]{0.8^2}} \\
 &= 0.2727.
 \end{aligned}$$

where the normalization constant  $K_{123} = K'(Z_1, Z_2, Z_3)$  is given by (10) for  $s=3$ , i.e.

$$K_{123} = \frac{0.1}{\sqrt[3]{0.2^2}} \frac{0.5}{\sqrt[3]{0.2^2}} \frac{0.6}{\sqrt[3]{0.2^2}} + \frac{0.9}{\sqrt[3]{0.8^2}} \frac{0.5}{\sqrt[3]{0.8^2}} \frac{0.4}{\sqrt[3]{0.8^2}} = 1.0312$$

This corrected result shows that Bayes fusion rule is actually associative because one has

$$\begin{cases} P(X|(Z_1 \cap Z_2) \cap Z_3) = P(X|Z_1 \cap Z_2 \cap Z_3), \\ P(X|Z_1 \cap (Z_2 \cap Z_3)) = P(X|Z_1 \cap Z_2 \cap Z_3), \\ P(X|Z_2 \cap (Z_1 \cap Z_3)) = P(X|Z_1 \cap Z_2 \cap Z_3). \end{cases}$$

As consequence, the property (P4) on page 245 of [1], although being correct, is not necessary.

On page 250 of [1], the sentence:

*Indeed, in Bayes rule one divides each posterior source  $m_i(x_j)$  by  $\sqrt[s]{m_0(x_j)}$ ,  $i=1, 2, \dots, s$ , whereas the prior source  $m_0(\cdot)$  is combined in a pure conjunctive manner by DS rule with the bba's  $m_i(\cdot)$ ,  $i=1, 2, \dots, s$ , as if  $m_0(\cdot)$  was a simple additional source.*

must be corrected as:

*Indeed, in Bayes rule one divides each posterior source  $m_i(x_j)$  by  $\sqrt[s]{m_0^{s-1}(x_j)}$ ,  $i=1, 2, \dots, s$ , whereas the prior source  $m_0(\cdot)$  is combined in a pure conjunctive manner by DS rule with the bba's  $m_i(\cdot)$ ,  $i=1, 2, \dots, s$ , as if  $m_0(\cdot)$  was a simple additional source.*

This erratum concerns also some incorrect formulas appearing in a preliminary version of [1] presented in 2013 in [2].

## REFERENCES

- [1] J. Dezert, A. Tchamova, *On the validity of Dempster's fusion rule and its interpretation as a generalization of Bayesian fusion rule*, International Journal of Intelligent Systems, Special Issue: Advances in Intelligent Systems, Vol. 29, Issue 3, pp. 223–252, March 2014.
- [2] J. Dezert, A. Tchamova, D. Han, J.-M. Tacnet, *Why Dempster's fusion rule is not a generalization of Bayes fusion rule*, in Proc. of Fusion 2013 Int. Conf. on Information Fusion (Fusion 2013), Istanbul, Turkey, July 9–12, 2013.



# On Inequalities Bounding Imprecision and Nonspecificity Measures of Uncertainty

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA - DTIS/MIDL, 91120 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.  
Emails: jean.dezert@onera.fr, tchamova@bas.bg

Originally published as: J. Dezert, A. Tchamova, *On Inequalities Bounding Imprecision and Nonspecificity Measures of Uncertainty*, Information & Security Journal, Vol. 52, pp. 37–51, February 2022, and reprinted with permission.

**Abstract**—In this paper we prove two inequalities about two measures of uncertainty of basic belief assignments, called respectively Imprecision measure and U-uncertainty measure that have been introduced by Dubois and Prade in 1980's. These inequalities have been considered as obvious by these authors, but to prove them rigorously needs some effort, as it will be shown.

**Keywords:** belief functions, measure of uncertainty, convex combination, inequalities.

## I. INTRODUCTION

This paper presents two mathematical proofs of inequalities about two measures of uncertainty of basic belief assignments, called respectively the Imprecision and the U-uncertainty (or nonspecificity) that have been introduced by Dubois and Prade in [1]–[3]. We recall that a Basic Belief Assignment (BBA)  $m$  defined on the power set  $2^\Theta$  of the finite frame of discernment (FoD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  is a mapping  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  such that  $m(\emptyset) = 0$  and  $\sum_{X \subseteq \Theta} m(X) = 1$ . This type of mapping has been introduced by Shafer in [4]. The cardinality of the FoD is  $|\Theta| = n$ . The measures of imprecision  $l(m)$ , and of nonspecificity  $U(m)$  are respectively defined by

$$l(m) = \sum_{X \subseteq \Theta} m(X)|X| = \sum_{X_i \in 2^\Theta} m(X_i)|X_i|, \quad (1)$$

$$U(m) = \sum_{X \subseteq \Theta} m(X) \log(|X|) = \sum_{X_i \in 2^\Theta} m(X_i) \log(|X_i|), \quad (2)$$

where  $X_i$  is the  $i$ -th element of the power set  $2^\Theta$  of the FoD  $\Theta$ , and  $|X_i|$  its cardinality. By convention, and without loss of generality, we will take  $X_1 = \emptyset$  (the empty set), and  $X_{2^n} = \Theta$ . The integer index  $i$  varies from 1 to  $2^n = 2^{|\Theta|}$ .

$m_v$  is the vacuous BBA defined by  $m_v(X) = 1$  if  $X = \Theta$ , and  $m_v(X) = 0$  for all elements  $X \neq \Theta$  of  $2^\Theta$ . This vacuous BBA  $m_v$  characterizes a full ignorant source of evidence.

In the next sections we prove that for any BBA  $m \neq m_v$  defined on  $2^\Theta$  the two following inequalities hold

$$l(m) < l(m_v), \quad (3)$$

and

$$U(m) < U(m_v). \quad (4)$$

We will prove these two inequalities in two ways: 1) by a direct application of the Theorem of convex combination (see Theorem 1, and Theorem 2 in the appendix) by a direct calculation from the mathematical definitions of  $l(m)$  and  $U(m)$  measures of uncertainty.

For proving these inequalities, we first recall that a convex combination, denoted by  $s_n$ , of  $n$  values  $\{z_i, i = 1, 2, \dots, n\}$  is a linear combination of the form

$$s_n = \sum_{i=1}^n w_i z_i, \quad (5)$$

where  $w_i \in [0, 1]$  is the weight of the value  $z_i$  such that  $\sum_{i=1}^n w_i = 1$ .

In the appendix, we prove the following useful theorem that will help us to prove the inequalities (3) and (4) in the next sections.

**Theorem 1:** Let  $s_n = \sum_{i=1}^n w_i z_i$  be a convex combination of  $n$  values  $z_1, \dots, z_n$  with normalized weights  $w_1, \dots, w_n$ , where  $w_i \in [0, 1]$ . Then, we have

$$\min\{z_i \in Z\} \leq s_n \leq \max\{z_i \in Z\}, \quad (6)$$

where  $Z \triangleq \{z_i \in \{z_1, z_2, \dots, z_n\} | w_i > 0\}$ .

**Proof of Theorem 1:** see appendix.

## II. PROOF THAT $l(m) < l(m_v)$ IF $m \neq m_v$

### A. First Proof : using the Theorem of convex combination

The proof of inequality (3) is a direct application of the Theorem 1 when working with  $2^n = 2^{|\Theta|}$  values<sup>1</sup>  $z_i = |X_i|$  and weights  $w_i = m(X_i)$ . We recall that  $w_1 = m(X_1) = m(\emptyset) = 0$  for any BBA  $m$  (by definition of  $m$ ). Therefore one has always at least one weight (i.e.  $w_1$ ) among all  $2^n$  weights equals zero, which justifies the use of Theorem 1, rather than Theorem 2 of appendix.

The imprecision measure  $l(m)$  can also be expressed as  $l(m) = \sum_{i=1}^{2^n} m(X_i)|X_i|$  because

$$\sum_{X_i \in 2^\Theta} m(X_i)|X_i| = \sum_{i=1}^{2^n} m(X_i)|X_i|$$

<sup>1</sup>We recall that integer index  $i$  spans  $\{1, 2, \dots, 2^n\}$ .

Based on Theorem 1, we have

$$\begin{aligned} & \min\{|X_i| \in \{|X_1|, |X_2|, \dots, |X_{2^n}|\} | m(X_i) > 0\} \\ & \leq \sum_{i=1}^{2^n} m(X_i) |X_i| \leq \\ & \max\{|X_i| \in \{|X_1|, |X_2|, \dots, |X_{2^n}|\} | m(X_i) > 0\}. \quad (7) \end{aligned}$$

The upper bound of inequality (7) is always lower than  $|\Theta| = n$  if  $m \neq m_v$ , and it is equal to  $|\Theta| = n$  when  $m = m_v$ . Therefore, one has

$$\sum_{i=1}^{2^n} m(X_i) |X_i| < |\Theta|, \quad (8)$$

and because  $l(m_v) = m_v(\Theta) \cdot |\Theta| = 1 \cdot |\Theta| = |\Theta|$ , one sees that the valid inequality (8) is the same as

$$l(m) < l(m_v), \quad (9)$$

which completes the proof of the inequality (3).

### B. Second proof : using direct calculation

First we note that

$$l(m_v) = m(\Theta) \cdot |\Theta| = 1 \cdot n = n.$$

Because  $m$  is a (normalized) BBA [4] such that  $m(\emptyset) = 0$  and  $\sum_{X \subseteq \Theta} m(X) = 1$ , one has

$$m(\Theta) + \sum_{X \subset \Theta} m(X) = 1,$$

or equivalently

$$m(\Theta) = 1 - \sum_{X \subset \Theta} m(X).$$

Therefore

$$n \cdot m(\Theta) = n \cdot [1 - \sum_{X \subset \Theta} m(X)].$$

The expression of  $l(m)$  can be decomposed as

$$\begin{aligned} l(m) &= \sum_{X \subseteq \Theta} m(X) |X| \\ &= m(\Theta) \cdot |\Theta| + \sum_{X \subset \Theta} m(X) |X| \\ &= n \cdot m(\Theta) + \sum_{X \subset \Theta} m(X) |X| \\ &= n \cdot [1 - \sum_{X \subset \Theta} m(X)] + \sum_{X \subset \Theta} m(X) |X| \\ &= n - [n \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) |X|]. \end{aligned}$$

To prove that  $l(m) < l(m_v)$  is equivalent to prove that

$$n - [n \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) |X|] < n,$$

or to prove

$$n \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) |X| > 0. \quad (10)$$

We can express  $\sum_{X \subset \Theta} m(X) |X|$  as

$$\begin{aligned} \sum_{X \subset \Theta} m(X) |X| &= \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \cdot 1 \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \cdot 2 \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \cdot 3 \\ &+ \dots \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X) \cdot (n-1), \end{aligned}$$

that is

$$\begin{aligned} \sum_{X \subset \Theta} m(X) |X| &= \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \\ &+ 2 \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \\ &+ 3 \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \\ &+ \dots \\ &+ (n-1) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \sum_{X \subset \Theta} m(X) |X| &= \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) + \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) + 2 \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \\ &+ \dots \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X) \\ &+ (n-2) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X), \end{aligned}$$

or equivalently

$$\begin{aligned} \sum_{X \subset \Theta} m(X) |X| &= \sum_{X \subset \Theta} m(X) \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \\ &+ 2 \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \\ &+ \dots \\ &+ (n-2) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X). \end{aligned}$$

Then, for the left hand side of the inequality (10) we obtain the following expression

$$\begin{aligned}
 & n \cdot \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X)|X| = \\
 & (n-1) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \\
 & + (n-2) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \\
 & + (n-3) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \\
 & \quad \quad \quad + \dots \\
 & + (n - (n-1)) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X).
 \end{aligned}$$

The right hand side of the previous expression is clearly strictly positive, that is

$$\begin{aligned}
 & (n-1) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \\
 & + (n-2) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \\
 & + (n-3) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \\
 & \quad \quad \quad + \dots \\
 & + (n - (n-1)) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X) > 0,
 \end{aligned}$$

because  $n > 1$  (the FoD has more than one hypothesis inside), and also because there is at least one element  $X \subset \Theta$  for which  $m(X) > 0$  when  $m \neq m_v$ .

Then we obtain

$$n \cdot \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X)|X| > 0,$$

which completes our second proof of (3) by a direct calculation.

### III. PROOF THAT $U(m) < U(m_v)$ IF $m \neq m_v$

#### A. First Proof : using the Theorem of convex combination

The proof of inequality  $U(m) < U(m_v)$  is similar to the proof of  $l(m) < l(m_v)$  by replacing values  $|X_i|$  by  $\log(|X_i|)$ , and by taking  $m(X_1) \log(|X_1|) = m(\emptyset) \log(|\emptyset|) = 0 \cdot \log(0) = 0$  which is easily justified by continuity since  $x \log(x) \rightarrow 0$  as  $x \rightarrow 0$ . More precisely, we can express  $U(m)$  as

$$\begin{aligned}
 U(m) &= m(\emptyset)(\log(|\emptyset|)) + \sum_{X_i \in 2^\Theta \setminus \{\emptyset\}} m(X_i) \log(|X_i|) \\
 &= \sum_{i=2}^{2^n} m(X_i) \log(|X_i|).
 \end{aligned}$$

Based on Theorem 1, we have

$$\begin{aligned}
 & \min\{\log(|X_i|) \in \{\log(|X_2|), \dots, \log(|X_{2^n}|)\} | m(X_i) > 0\} \\
 & \leq \sum_{i=2}^{2^n} m(X_i) \log(|X_i|) \leq \\
 & \max\{\log(|X_i|) \in \{\log(|X_2|), \dots, \log(|X_{2^n}|)\} | m(X_i) > 0\}.
 \end{aligned}$$

Because  $\log(\cdot)$  is a continuous increasing function, the upper bound of the previous inequality is always lower than  $\log(|\Theta|) = \log(n)$  when  $m \neq m_v$ . Therefore,

$$\sum_{i=2}^{2^n} m(X_i) \log(|X_i|) < \log(|\Theta|), \quad (11)$$

and because  $U(m_v) = m_v(\Theta) \cdot \log(|\Theta|) = 1 \cdot \log(|\Theta|) = \log(|\Theta|)$ , one sees that the valid inequality (11) is the same as

$$U(m) < U(m_v), \quad (12)$$

which completes the proof of the inequality (4).

#### B. Second proof : using direct calculation

We prove the inequality  $U(m) < U(m_v)$  similarly to our second proof for  $l(m) < l(m_v)$  by replacing values  $|X_i|$  by  $\log(|X_i|)$ . We note that

$$U(m_v) = m(\Theta) \cdot \log(|\Theta|) = 1 \cdot \log(n) = \log(n).$$

Because  $m$  is a (normalized) BBA [4] such that  $m(\emptyset) = 0$  and  $\sum_{X \subset \Theta} m(X) = 1$ , one has

$$m(\Theta) + \sum_{X \subset \Theta} m(X) = 1,$$

or equivalently

$$m(\Theta) = 1 - \sum_{X \subset \Theta} m(X).$$

Therefore

$$\log(n) \cdot m(\Theta) = \log(n) \cdot [1 - \sum_{X \subset \Theta} m(X)].$$

The expression of  $U(m)$  can be decomposed as

$$\begin{aligned}
 U(m) &= \sum_{X \subset \Theta} m(X) \log(|X|) \\
 &= m(\Theta) \cdot \log(n) + \sum_{X \subset \Theta} m(X) \log(|X|) \\
 &= \log(n) \cdot [1 - \sum_{X \subset \Theta} m(X)] + \sum_{X \subset \Theta} m(X) \log(|X|) \\
 &= \log(n) - [\log(n) \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) \log(|X|)].
 \end{aligned}$$

To prove that  $U(m) < U(m_v)$  is equivalent to prove that

$$\log(n) - [\log(n) \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) \log(|X|)] < \log(n),$$

or to prove

$$\log(n) \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) \log(|X|) > 0. \quad (13)$$

We can express  $\sum_{X \subset \Theta} m(X) \log(|X|)$  as

$$\begin{aligned} \sum_{X \subset \Theta} m(X) \log(|X|) &= \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \cdot \log(1) \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \cdot \log(2) \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \cdot \log(3) \\ &+ \dots \\ &+ \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X) \cdot \log(n-1). \end{aligned}$$

Then for the left hand side of inequality (13) we obtain:

$$\begin{aligned} \log(n) \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) \log(|X|) &= \\ &[\log(n) - \log(1)] \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \\ &+ [\log(n) - \log(2)] \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \\ &+ [\log(n) - \log(3)] \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \\ &+ \dots \\ &+ [\log(n) - \log(n-1)] \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X). \end{aligned}$$

Because  $\log(1) = 0$ , the equation above can be rewritten as

$$\begin{aligned} \log(n) \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) \log(|X|) &= \\ &\log(n) \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=1} m(X) \\ &+ [\log(n) - \log(2)] \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=2} m(X) \\ &+ [\log(n) - \log(3)] \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=3} m(X) \\ &+ \dots \\ &+ [\log(n) - \log(n-1)] \cdot \sum_{X \subset \Theta \text{ s.t. } |X|=n-1} m(X). \end{aligned}$$

Because  $n > 1$ , and because  $\log(\cdot)$  is an increasing function one always has for  $n > 1$ ,  $\log(n) > 0$ , and  $[\log(n) - \log(n-1)] > 0$ . Because there is at least one element  $X \subset \Theta$  for which  $m(X) > 0$  when  $m \neq m_v$ , we can conclude that

$$\log(n) \sum_{X \subset \Theta} m(X) - \sum_{X \subset \Theta} m(X) \log(|X|) > 0.$$

which completes our second proof of (4) by a direct calculation.

## IV. CONCLUSION

In this paper we have proved that the imprecision measure  $l(m)$  is always lower than  $l(m_v) = |\Theta|$ , and its U-uncertainty (also known as non-specificity) measure  $U(m)$  is always lower than  $U(m_v) = \log(|\Theta|)$  for any non vacuous BBA  $m$ . The proofs presented in this paper have been obtained by two different ways: by the theorem of convex combination, and by direct calculation from the mathematical definitions for  $l(m)$ ,  $l(m_v)$ ,  $U(m)$ , and  $U(m_v)$ . We have shown that the use of the theorem of convex combination provides an elegant and shorter proof of the inequalities. This theorem will be helpful to evaluate the lower and upper bounds of any measures of uncertainty of a BBA that would be based on any convex combination of mass values (chosen as weighting factors) and real values committed to each element of the power set of the frame of discernment.

## APPENDIX

Before proving Theorem 1, we need to establish at first the following theorem.

**Theorem 2:** Let  $s_n = \sum_{i=1}^n w_i z_i$  be a convex combination of  $n$  values  $z_1, \dots, z_n$  with strictly positive normalized weights  $w_1, \dots, w_n$ . Then, we have

$$\min\{z_1, \dots, z_n\} \leq s_n \leq \max\{z_1, \dots, z_n\}. \quad (14)$$

The proof of Theorem 2 is done by induction.

**Proof of the theorem 2:**

- For  $n = 1$ , one has only one value  $z_1$  with weight  $w_1 = 1$ . Hence  $s_1 = w_1 z_1 = z_1$ ,  $\min\{z_1\} = z_1$ , and  $\max\{z_1\} = z_1$ . Therefore,  $\min\{z_1\} = s_1 = \max\{z_1\}$ , which is a special case of the inequality (14). Therefore the inequality (14) is valid for  $n = 1$ .
- For  $n = 2$ , one has two values  $\{z_1, z_2\}$  with (strictly) positive weights  $\{w_1, w_2\}$  and  $s_2 = w_1 z_1 + w_2 z_2$ .
  - 1) if  $z_1 = z_2$ , then  $s_2 = w_1 z_1 + w_2 z_2 = w_1 z_1 + w_2 z_1 = (w_1 + w_2) z_1 = z_1 = z_2$ . Hence one has  $\min\{z_1, z_2\} = z_1 = z_2$  and  $\max\{z_1, z_2\} = z_1 = z_2$ . Therefore, one gets  $\min\{z_1, z_2\} = s_2 = \max\{z_1, z_2\}$ , which is a special case of the inequality (14) for  $n = 2$ .
  - 2) If  $z_1 \neq z_2$ , then two sub-cases are possible:
    - a) Case 1: if  $z_1 < z_2$ , then  $\min\{z_1, z_2\} = z_1$  and  $\max\{z_1, z_2\} = z_2$ . We have

$$\begin{aligned} s_2 &= w_1 z_1 + w_2 z_2 = (z_1 - z_1) + w_1 z_1 + w_2 z_2 \\ &= z_1 - (1 - w_1) z_1 + w_2 z_2 \\ &= z_1 - w_2 z_1 + w_2 z_2 \\ &= z_1 + w_2 (z_2 - z_1) \geq \min\{z_1, z_2\}. \end{aligned}$$

This last inequality comes from the fact that  $w_2 \geq 0$ , and  $z_2 - z_1 \geq 0$  because  $\min\{z_1, z_2\} = z_1$ . So we have proved  $\min\{z_1, z_2\} \leq s_2$ .

Because  $w_2 \in [0, 1]$ , we have  $w_2(z_2 - z_1) \leq z_2 - z_1$ , and therefore

$$\begin{aligned} s_2 &= z_1 + w_2(z_2 - z_1) \\ &\leq z_1 + (z_2 - z_1) = z_2 = \max\{z_1, z_2\}. \end{aligned}$$

This shows that  $s_2 \leq \max\{z_1, z_2\}$ . Therefore, we have proved

$$\min\{z_1, z_2\} \leq s_2 \leq \max\{z_1, z_2\}.$$

We see that the inequality (14) holds for  $n = 2$  for the case 1.

- b) Case 2: if  $z_2 < z_1$ , then  $\min\{z_1, z_2\} = z_2$  and  $\max\{z_1, z_2\} = z_1$ . We have

$$\begin{aligned} s_2 &= w_1 z_1 + w_2 z_2 = (z_2 - z_2) + w_1 z_1 + w_2 z_2 \\ &= z_2 - (1 - w_2)z_2 + w_1 z_1 \\ &= z_2 - w_1 z_2 + w_1 z_1 \\ &= z_2 + w_1(z_1 - z_2) \geq \min\{z_1, z_2\}. \end{aligned}$$

This last inequality comes from the fact that  $w_1 \geq 0$ , and  $z_1 - z_2 \geq 0$  because  $\min\{z_1, z_2\} = z_2$ . So we have proved  $\min\{z_1, z_2\} \leq s_2$ .

Because  $w_1 \in [0, 1]$ , we have  $w_1(z_1 - z_2) \leq z_1 - z_2$ , and therefore

$$\begin{aligned} s_2 &= z_2 + w_1(z_1 - z_2) \\ &\leq z_2 + z_1 - z_2 = z_1 = \max\{z_1, z_2\}. \end{aligned}$$

This shows that  $s_2 \leq \max\{z_1, z_2\}$ . Therefore, we have proved

$$\min\{z_1, z_2\} \leq s_2 \leq \max\{z_1, z_2\}.$$

We see that the inequality (14) holds for  $n = 2$  for the case 2.

Finally, the inequality (14) is always valid for  $n = 2$  in all cases, i.e. when  $z_1 = z_2$ , or  $z_1 < z_2$ , or  $z_2 < z_1$ .

- For  $n > 2$ , we suppose that the inequality (14) holds. That is

$$\min\{z_1, \dots, z_n\} \leq s_n \leq \max\{z_1, \dots, z_n\}. \quad (15)$$

We prove next by induction that this inequality also holds for  $n + 1$ .

- For the induction with  $n + 1$ , we have to consider  $n + 1$  values  $\{z_1, \dots, z_n, z_{n+1}\}$  and  $n + 1$  strictly positive normalized weights  $\{w_1, \dots, w_n, w_{n+1}\}$ , that is  $w_i > 0$  for  $i = 1, 2, \dots, n$  and  $\sum_{i=1}^{n+1} w_i = 1$ . Because all  $w_i > 0$ ,

one has necessarily  $w_{n+1} < 1$ . So, we can always express  $s_{n+1}$  as

$$\begin{aligned} s_{n+1} &= \sum_{i=1}^{n+1} w_i z_i \\ &= w_{n+1} z_{n+1} + \sum_{i=1}^n w_i z_i \\ &= w_{n+1} z_{n+1} + \sum_{i=1}^n \frac{1 - w_{n+1}}{1 - w_{n+1}} w_i z_i \\ &= w_{n+1} z_{n+1} + (1 - w_{n+1}) \sum_{i=1}^n \frac{w_i}{1 - w_{n+1}} z_i \\ &= w_{n+1} z_{n+1} + (1 - w_{n+1}) s_n, \end{aligned}$$

where

$$s_n = \sum_{i=1}^n \frac{w_i}{1 - w_{n+1}} z_i = \sum_{i=1}^n v_i z_i. \quad (16)$$

The new weights involved in  $s_n$  defined by  $v_i \triangleq \frac{w_i}{1 - w_{n+1}}$  are also all strictly positive because  $w_i > 0$  and  $1 - w_{n+1} > 0$ , and they are also normalized because

$$\begin{aligned} \sum_{i=1}^n v_i &= \sum_{i=1}^n \frac{w_i}{1 - w_{n+1}} \\ &= \frac{1}{1 - w_{n+1}} \sum_{i=1}^n w_i \\ &= \frac{1}{1 - w_{n+1}} (1 - w_{n+1}) = 1, \end{aligned}$$

because  $\sum_{i=1}^{n+1} w_i = 1$ , which implies  $\sum_{i=1}^n w_i = 1 - w_{n+1}$ .

Hence, we observe that  $s_n = \sum_{i=1}^n v_i z_i$  is also a convex combination of the  $n$  real values  $\{z_1, \dots, z_n\}$  with normalized and strictly positive weights  $v_i$ , and therefore the inequality (14) holds (by assumption).

One sees that the problem of combination of  $n + 1$  values has been reformulated as a convex combination of two values  $z_{n+1}$  and  $s_n = \sum_{i=1}^n v_i z_i$  with strictly positive and normalized weights  $w_{n+1}$  and  $(1 - w_{n+1})$ . Because the inequality (14) is satisfied for the convex combination of two real values (for  $n = 2$ ), the following inequality holds

$$\min\{s_n, z_{n+1}\} \leq s_{n+1} \leq \max\{s_n, z_{n+1}\}. \quad (17)$$

Because  $\min\{z_1, \dots, z_n\} \leq s_n \leq \max\{z_1, \dots, z_n\}$  is assumed to be true, the inequality (17) can be rewritten as

$$\begin{aligned} \min\{\min\{z_1, z_2, \dots, z_n\}, z_{n+1}\} \\ \leq s_{n+1} \leq \\ \max\{\max\{z_1, z_2, \dots, z_n\}, z_{n+1}\}, \quad (18) \end{aligned}$$

or equivalently

$$\begin{aligned} \min\{z_1, z_2, \dots, z_n, z_{n+1}\} \\ \leq s_{n+1} \leq \\ \max\{z_1, z_2, \dots, z_n, z_{n+1}\}. \end{aligned} \quad (19)$$

Therefore the inequality (14) is also valid for  $n+1$ , which completes the proof of Theorem 2.

We can generalize the Theorem 2 to take into account all the cases where some weights are zero. For this, the set on  $n$  real values denoted by  $\mathcal{Z} = \{z_1, z_2, \dots, z_n\}$  can always be expressed as  $\mathcal{Z} = Z \cup \bar{Z}$ , where  $Z \triangleq \{z_i \in \{z_1, z_2, \dots, z_n\} | w_i > 0\}$  and  $\bar{Z} \triangleq \{z_i \in \{z_1, z_2, \dots, z_n\} | w_i = 0\}$ . The convex combination  $s_n = \sum_{i=1}^n w_i z_i$  can be expressed as

$$s_n = \sum_{z_i \in Z} w_i z_i + \sum_{z_i \in \bar{Z}} w_i z_i. \quad (20)$$

because  $w_i = 0$  for any  $z_i \in \bar{Z}$ , one has  $\sum_{z_i \in \bar{Z}} w_i z_i = 0$ , and therefore  $s_n = \sum_{z_i \in Z} w_i z_i$ , whose bounds are given by Theorem 2. Hence,  $\min\{z_i \in Z\} \leq s_n \leq \max\{z_i \in Z\}$ . This completes the proof of Theorem 1, which is more general than Theorem 2.

#### REFERENCES

- [1] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Comput. Intell., Vol. 4, pp. 244–264, 1988.
- [2] D. Dubois, H. Prade, *A note on measures of specificity for fuzzy sets*, (E)BUSENFAL, Vol. 19, paper # 8, May, 1984.
- [3] D. Dubois, H. Prade, *A note on measures of specificity for fuzzy sets*, Int. J. Gen. Syst., Vol. 10, No. 4, pp. 279–283, 1985.
- [4] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.

# Novel Moderate Transformation of Fuzzy Membership Function into Basic Belief Assignment

Xiaojing Fan<sup>a</sup>, Deqiang Han<sup>a</sup>, Jean Dezert<sup>b</sup>, Yi Yang<sup>c</sup>,

<sup>a</sup>School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

<sup>b</sup>The French Aerospace Lab, ONERA, 91120 Palaiseau, France.

<sup>c</sup>SKLSVMS, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China.

Emails: jjingxiulian@126.com, deqhan@xjtu.edu.cn, jean.dezert@onera.fr, jiafeiyy@mail.xjtu.edu.cn

Originally published as: X. Fan, D. Han, J. Dezert, Y. Yang, *Novel Moderate Transformation of Fuzzy Membership Function into Basic Belief Assignment*, Chinese Journal of Aeronautics, 2022, and reprinted with permission.

**Abstract**—In information fusion, the uncertain information from different sources might be modeled with different theoretical frameworks. When one needs to fuse the uncertain information represented by different uncertainty theories, constructing the transformation between different frameworks is crucial. Various transformations of a Fuzzy Membership Function (FMF) into a Basic Belief Assignment (BBA) have been proposed, where the transformations based on uncertainty maximization and minimization can determine the BBA without preselecting the focal elements. However, these two transformations that based on uncertainty optimization emphasize the extreme cases of uncertainty. To avoid extreme attitudinal bias, a trade-off or moderate BBA with the uncertainty degree between the minimal and maximal ones is more preferred. In this paper, two moderate transformations of an FMF into a trade-off BBA are proposed. One is the weighted average based transformation and the other is the optimization-based transformation with weighting mechanism, where the weighting factor can be user-specified or determined with some prior information. The rationality and effectiveness of our transformations are verified through numerical examples and classification examples.

**Keywords:** Belief functions, basic belief assignment, fuzzy membership function, information fusion, moderate transformation.

## I. INTRODUCTION

In multi-source information fusion, the information obtained from different sources usually have different types of uncertainty. Various kinds of uncertainty theories have been proposed including probability theory, fuzzy set theory [1], possibility theory [2], rough set theory [3] and theory of belief functions [4], etc., for dealing with different types of uncertain information. According to the type of uncertainty, the uncertain information from different sources might be modeled with different theoretical frameworks. Usually, these uncertain information with different representations cannot be directly combined or fused. Therefore, transformations between different frameworks are needed [5], and then, one can fuse them under the same framework.

Random set theory is regarded as a unified framework for various frameworks of uncertainty including probability theory, fuzzy set theory, theory of belief functions, etc [5]. In particular, to fuse the opinion of an expert represented by a Fuzzy Membership Function (FMF) and the output of

a sensor expressed by Basic Belief Assignment (BBA), one can transform the FMF into a BBA and then combine two BBAs. One can also transform the BBA into an FMF and then combine two FMFs. In this paper, we focus on the transformation of an FMF into a BBA. Many transformations have been proposed [5]–[10], which can be categorized into two types.

One type of transformation has to preselect the focal elements. For example, Bi et al. [6] transform an FMF into a BBA by normalizing the given FMF, where the focal elements are preselected as singletons. As a result, the obtained BBA has no compound focal elements. In [5], Florea et al. transform an FMF into a BBA with  $\alpha$ -cut approach, where focal elements are preselected to the “nested in order”. However, a prior selection of focal elements might lead to information loss. The other type of transformation obtains a BBA by solving constrained optimization problems. For example, our previous work [7] proposed two transformations based on uncertainty optimization, which avoid the subjective preselection of focal elements. The difference between the two transformations is the specific optimization criterion, which is the maximization and minimization, respectively. It has been shown in [7] that both transformations based on uncertainty optimization are rational and effective. However, these transformations based on uncertainty optimization seem to be “one-sided” in terms of the uncertainty degrees, since they only focus on the minimal or maximal uncertainty. Either the minimal uncertainty or the maximal uncertainty is an extreme case of uncertainty. If one only pays attention to one of the extreme cases of uncertainty in the process of solving the optimization problem, it would bring the bias of extreme attitudinal on the uncertainty degree, which might bring counter-intuitive results. If we jointly consider two extreme cases of uncertainty, we can obtain the BBA with the degree of uncertainty between the minimal and maximal ones. Such a BBA is more “balanced” and “moderate” than the BBA obtained by pursuing the maximal or minimal uncertainty. In other words, joint consideration of two extremes of uncertainty corresponds to a better moderate attitude (corresponding to a preferred consensual agreement of behavior) for a transformation of FMF into a BBA, and then we can avoid the bias of extreme attitudinal on the uncertainty degree.

In this paper, we aim to obtain such a trade-off BBA to avoid “one-sidedness” on the uncertainty degree, which is based on the two BBAs obtained by the two transformations proposed in our previous work [7]. To transform an FMF into such a trade-off or moderate BBA, a weighting factor is used, which can make the trade-off BBA closer to the BBA obtained with the uncertainty maximization or closer to the BBA obtained with the uncertainty minimization. The weighting factor could be determined by using prior information or user-specified, which reflect the objective situation or meet subjective preferences of users. We propose two “moderate” (i.e., balanced) transformations in this paper. One transformation assigns the weighting factors to the two BBAs obtained by optimization-based transformations. Then, the weighted average of these two BBAs is the trade-off BBA. The other transformation is based on a constrained minimization problem with weighting mechanism. The objective function is constructed by the weighting factor and two degrees of dissimilarity between the trade-off BBA (to determine) and the two BBAs obtained by solving the uncertainty minimization and maximization. According to the given FMF and the legitimate conditions of a BBA, the constraints are constructed. Compared with the transformations based on uncertainty optimization, our transformations can avoid the extreme attitudinal bias and allow users to choose the degree of a trade-off BBA according to their preference.

This paper is an extended version of our previous preliminary work published in [11]. Based on the preliminary work, the main extended work and added value in this paper is as follows. The limitations of available transformations of an FMF into a BBA are analyzed. Examples are given to illustrate the loss of information that might be caused by the preselection of focal elements and the counter-intuitive results that might be caused by the extreme attitudinal of uncertainty. We use another more rational uncertainty measure, which is designed without switching frameworks, to construct the objective function in the transformations proposed in [7]. This is because the uncertainty measure Ambiguity Measure (AM) used in [7] as the objective function has some disputes and limitations, which are mentioned in [7]. Furthermore, to compare our transformations with the others, some numerical examples and a classification example are provided. Compared with the average classification accuracy of available transformations, the moderate transformation has a better classification performance.

The paper is organized as follows. After a brief introduction of the basics for theory of belief functions and some basic concepts of the fuzzy set theory in Section II, some traditional transformations of an FMF into a BBA are reviewed and their limitations are provided in Section III. In Section IV, the transformations of an FMF into a trade-off BBA are presented. In Section V, our transformations are compared with several traditional approaches and related examples are provided. As shown, these new transformations can bring better performances than other transformations in a classification example. Section VI concludes this paper.

## II. PRELIMINARY

### A. Basics for theory of belief functions

The theory of belief functions, also called Dempster-Shafer theory [12], has been applied to information fusion [13]–[15], decision making [16]–[19], pattern recognition [20]–[22], etc. It is a powerful framework for uncertainty modeling and reasoning. The Frame of Discernment (FOD)  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  is mutually exclusive and exhaustive under the closed-world assumption. Based on the power set of  $\Theta$  ( $2^\Theta$ ), a basic belief assignment (BBA, also called a mass function) is defined as

$$\sum_{A \subseteq \Theta} m(A) = 1, \quad m(\emptyset) = 0, \quad (1)$$

where  $A \subseteq \Theta$  is a proposition in the FOD. If  $\forall m(A) > 0$ ,  $A$  is called a focal element of  $m$ .

For all  $A \subseteq \Theta$  the belief function Bel and plausibility function Pl are defined as.

$$\text{Bel}(A) = \sum_{B \subseteq A} m(B), \quad (2)$$

$$\text{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B). \quad (3)$$

Bel( $A$ ) and Pl( $A$ ) are the lower and upper bound of the probability of a focal element  $A$ , respectively. The belief interval of  $A$  is represented by  $[\text{Bel}(A), \text{Pl}(A)]$  whose length is used to describe the imprecision of  $A$ 's probability.

Let's consider two independent BBAs:  $m_1$  and  $m_2$  on the same FOD. Dempster's rule [4] is defined as

$$m_{\text{Demp}}(A) = \begin{cases} 0, & A = \emptyset, \\ \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1-K}, & A \neq \emptyset, \end{cases} \quad (4)$$

where  $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$  denotes the conflict coefficient between two BBAs. Dempster's rule applies only if  $K \neq 1$ . When a high conflict between BBAs exists, Dempster's rule might bring counter-intuitive results [23]–[25]. Alternative rules [26]–[30] have been proposed.

### B. Uncertainty measure of BBA and distance of evidence

The uncertainty measure is used for evaluating the degree of uncertainty in a BBA. There are two types of uncertainty for a BBA in belief functions including the discord and non-specificity, which are collectively known as the ambiguity. Various kinds of uncertainty measures in the theory of belief functions have been proposed [31]–[38].

Let consider the FOD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . One of the total uncertainty measures is the AM [34] defined by

$$\text{AM}(m) = \sum_{\theta \in \Theta} \text{BetP}(\theta) \log_2 \text{BetP}(\theta), \quad (5)$$

where BetP( $\theta$ ) is the pignistic probability of a BBA [39]

$$\text{BetP}(\theta) = \sum_{A \subseteq \Theta} \frac{m(A)}{|A|}. \quad (6)$$

$|A|$  denotes the cardinality of a focal element  $A$ . Actually, in AM, the Shannon entropy of the pignistic probability is used to quantify the uncertainty. The range of AM is  $[0, \log_2(|\Theta|)]$ .

The distance-based total uncertainty measure ( $TU^I$ ) is based on the Wasserstein distance [40] between belief intervals, which is defined as follows:

$$d_W([a_1, b_1], [a_2, b_2]) = \sqrt{\left(\frac{a_1 + b_1}{2} + \frac{a_2 + b_2}{2}\right)^2 + \frac{1}{3}\left(\frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2}\right)^2}. \quad (7)$$

To be more specific, in  $TU^I$ ,  $[a_1, b_1]$  is replaced by  $[\text{Bel}(\{\theta_i\}), \text{Pl}(\{\theta_i\})]$  and  $[a_2, b_2]$  is replaced by  $[0, 1]$ . Then,  $TU^I$  is defined as [36]

$$TU^I(\mathbf{m}) = 1 - \frac{\sqrt{3}}{n} \sum_{i=1}^n d_W([\text{Bel}(\{\theta_i\}), \text{Pl}(\{\theta_i\})], [0, 1]). \quad (8)$$

where  $\sqrt{3}$  is for normalization. The belief interval  $[0, 1]$  is the most uncertain case for a singleton.

Compared with AM,  $TU^I$  is a total uncertainty measure that is defined directly in the theory of belief functions without switching frameworks from the probability theory to the theory of belief functions. The range of  $TU^I$  is  $[0, 1]$ , where  $TU^I = 0$  represents the crispest case, i.e.,  $m(\{\theta_i\}) = 1$ , while  $TU^I = 1$  represents the most uncertain case, i.e.,  $m(\Theta) = 1$ .

The distance of evidence is a metric for the degree of dissimilarity between two BBAs, which can describe how ‘‘far’’ it is from one BBA to the other. The distance of evidence is crucial for many belief functions related applications. Joussemle’s distance between two BBAs  $\mathbf{m}_1$  and  $\mathbf{m}_2$  defined on the same FOD is defined as [41]

$$d_J(\mathbf{m}_1, \mathbf{m}_2) = \sqrt{\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{D}(\mathbf{m}_1 - \mathbf{m}_2)}, \quad (9)$$

where the elements  $D(A, B)$  of Jaccard’s matrix  $\mathbf{D}$  are defined as  $D(A, B) = |A \cap B|/|A \cup B|$  for  $A \subseteq \Theta$  and  $B \subseteq \Theta$ .  $1/2$  is the normalization factor.  $d_J$  is a strict distance satisfying non-negativity, non-degeneracy, symmetry and triangular inequality [41]. There are also various types of distance measures that have been proposed [42]–[47].

### C. Fuzzy set theory

Fuzzy set theory [1] can be used to model the information without a crisp definition or a strict limit (e.g., ‘‘the target is fast’’, ‘‘the target turns quickly’’). A fuzzy set  $A_f$  is defined on a universe of discourse  $\Theta$ , which is equivalent to the FOD in belief functions.  $A_f$  is represented by a Fuzzy Membership Function (FMF)  $\mu_{A_f}(\theta)$ . The value of  $\mu_{A_f}(\theta)$  denotes the degree of membership for  $\theta$  in  $A_f$ .  $\mu_{A_f} : \Theta \rightarrow [0, 1]$ ;  $\theta \mapsto \mu_{A_f}(\theta) \in [0, 1]$ . The sum of  $\mu_{A_f}$  might be equal to, greater than or less than 1. For  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , a fuzzy measure [48] is defined as

$$D(\mu_{A_f}) = \frac{1}{n \ln 2} \sum_{i=1}^n S(\mu_{A_f}(\theta_i)), \quad (10)$$

where  $S(\mu_{A_f}(\theta_i))$  is Shannon’s function defined by

$$S(x) = \begin{cases} -x \ln x - (1-x) \ln(1-x), & 0 < x < 1, \\ 0, & x = 0, 1. \end{cases} \quad (11)$$

In the field of fuzzy set theory, there are many related concepts including  $Z$ -number [49], intuitionistic fuzzy set [50] and Pythagorean fuzzy set [51] which describe fuzzy information from different aspects.  $Z$ -number [49] is defined as an ordered pair of fuzzy numbers containing the reliability of uncertain information and can be denoted as  $Z = (A, B)$ . For the definition of a  $Z$ -number,  $A$  is a possibility restriction and  $B$  denotes the reliability of the possibility measure of  $A$ .

Intuitionistic Fuzzy Set [50] (IFS) is extended based on traditional fuzzy set. For  $\theta \in \Theta$ , an IFS can be represented as  $A = \{(\theta, \mu_A(\theta), \nu_A(\theta)) | \theta \in \Theta\}$ .  $\mu_A(\theta) : \Theta \rightarrow [0, 1]$  is the membership function, which denotes the degree of belonging  $\theta$  to  $A$ .  $\nu_A(\theta) : \Theta \rightarrow [0, 1]$  is the non-membership function, which denotes the degree of non-belonging  $\theta$  to  $A$ . One has  $0 \leq \mu_A(\theta) + \nu_A(\theta) \leq 1$ , and  $1 - \mu_A(\theta) - \nu_A(\theta)$  denotes the degree of hesitation.

Pythagorean Fuzzy Set (PFS) [51] is a non-standard fuzzy subset, which can be represented as  $P = \{(\theta, \mu_P(\theta), \nu_P(\theta)) | \theta \in \Theta\}$ . It satisfies  $0 \leq (\mu_P(\theta))^2 + (\nu_P(\theta))^2 \leq 1$ , where  $\mu_P(\theta) : \Theta \rightarrow [0, 1]$  denotes the membership function, and  $\nu_P(\theta) : \Theta \rightarrow [0, 1]$  denotes the non-membership function.

## III. TRANSFORMATIONS OF FMF INTO BBA

Although fuzzy set theory and the theory of belief functions are two different theoretical frameworks, there are relationships between their basic concepts [52]. The relationships are between the FMF and the singleton plausibility function or singleton belief function, which are as follows.

### A. Relationships between FMF and BBA

Consider the FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The given FMF is denoted by  $\boldsymbol{\mu} = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . The corresponding BBA is denoted by  $\mathbf{m}$ . When  $\sum_{i=1}^n \mu(\theta_i) > 1$ , the FMF is equivalent to a singleton plausibility function:

$$\text{Pl}(\{\theta_i\}) = \sum_{\{\theta_i\} \cap A \neq \emptyset} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta. \quad (12)$$

When the FMF is equivalent to a singleton plausibility function, it is also equivalent to a contour function.

When  $\sum_{i=1}^n \mu(\theta_i) < 1$ , it is equivalent to a singleton belief function:

$$\text{Bel}(\{\theta_i\}) = m(\{\theta_i\}) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta. \quad (13)$$

The detailed proof of two relationships is given in [52].

When  $\sum_{i=1}^n \mu(\theta_i) = 1$ , the FMF can be equivalent to either the singleton plausibility or singleton belief, because when  $\sum_{i=1}^n \mu(\theta_i) = 1$  Eq. (12) and Eq. (13) are equivalent. The proof of their equivalence is in the appendix.

The transformation of an FMF into a BBA is a multi-answer problem [7]. There exists  $2^n - 1$  focal elements of the FOD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , except for  $\emptyset$ . That is to say, at most

$2^n - 1$  unknown variables need to be determined. However, according to Eq. (12) or Eq. (13), one can obtain  $n$  linear equations about the undetermined BBA. In addition, one has  $\sum_{A \subseteq \Theta} m(A) = 1$ , which is one of the legitimate conditions of a BBA. Then, using the  $n + 1$  linear equations to determine  $2^n - 1$  unknown variables is an under-determined problem when  $n \geq 2$ . Thus, the transformation of an FMF into a BBA is a multi-answer (multi-solution) problem.

To deal with the multi-answer problem, various methods have been proposed [5]–[10], which can be categorized into two major types: the transformations with preselection of focal elements, and transformations based on uncertainty optimization.

### B. Available transformations with preselection of focal elements

1) *Normalization based transformation*: Bi et al. [6] transform an FMF into a BBA in the application of the text categorization. They normalize the given FMF to determine a unique BBA. Let the FOD be  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The given FMF is denoted by  $\mu = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . The obtained BBA is denoted by  $m_{\text{norm}}$ . In the work of Bi et al. [6], the BBA is determined as follows:

$$m_{\text{norm}}(\{\theta_i\}) = \mu(\theta_i) / \sum_{j=1}^n \mu(\theta_j), \quad (14)$$

where  $i = 1, 2, \dots, n$ . In the sequel, this transformation is represented by “ $T_{\text{norm}}$ ”. By using  $T_{\text{norm}}$ , one can obtain a Bayesian BBA, i.e., the focal elements are only singletons.

2) *Transformation based on  $\alpha$ -cut approach*: Suppose that an FOD is  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ .  $\mu = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$  denotes the given FMF. Florea et al. [5] transform an FMF into a BBA by using  $\alpha$ -cut approach, where  $\mu$  should be sorted into ascending order. Here, we represent the sort of  $\mu$  as  $0 = \alpha_0 < \alpha_1 < \dots < \alpha_M \leq 1$ , where  $M \leq |\Theta|$ .  $\alpha_j$  ( $j = 1, 2, \dots, M$ ) is the value of the FMF. Then the BBA, denoted by  $m_{\alpha\text{-cut}}$ , is determined as

$$m_{\alpha\text{-cut}}(A_j) = (\alpha_j - \alpha_{j-1}) / \alpha_M, \quad (15)$$

where  $A_j = \{\theta_i \in \Theta \mid \mu(\theta_i) \geq \alpha_j\}$ ,  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, M$ . In the sequel, the above transformation is represented by “ $T_{\alpha\text{-cut}}$ ”. By using  $T_{\alpha\text{-cut}}$ , the supposed structure of focal elements for the obtained BBA is nested in order.

3) *Transformation based on assigning mass to a focal element triplet*: Let the FOD be  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The given FMF is denoted by  $\mu = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . In [10], the authors use a focal element triplet, which is a structure defined as three focal elements  $B_1, B_2$  and  $B_3$ .  $B_1, B_2 \subseteq \Theta$  are singletons, and  $B_3$  is the total set (i.e.,  $\Theta$ ). First, the normalization of the given FMF is calculated using  $m(\{\theta_i\}) = \mu(\theta_i) / \sum_{j=1}^n \mu(\theta_j)$  for  $i = 1, 2, \dots, n$ . Then, the BBA represented by  $m_{\text{tri}}$  can be obtained as follows:

$$\begin{aligned} m_{\text{tri}}(B_1) &= m(\{\theta_s\}), m_{\text{tri}}(B_2) = m(\{\theta_t\}), \\ m_{\text{tri}}(B_3) &= 1 - m(\{\theta_s\}) - m(\{\theta_t\}) \end{aligned} \quad (16)$$

where  $B_1, B_2$  and  $B_3$  are defined by

$$\begin{aligned} B_1 &= \theta_s = \arg \max_{\theta_i \in \Theta} (m(\{\theta_i\})), \\ B_2 &= \theta_t = \arg \max_{\theta_i \in \Theta \setminus \{\theta_s\}} (m(\{\theta_i\})), \\ B_3 &= \Theta. \end{aligned}$$

This transformation is represented by “ $T_{\text{tri}}$ ” in the sequel.

### C. Transformations based on uncertainty optimization

In our previous work [7], the multi-answer problem is formulated as a constrained optimization to obtain a unique BBA without preselecting focal elements. We established two transformations based on uncertainty optimization of an FMF into a BBA, where the uncertainty measure AM (see Eq. (5)) is used as the objective function. When an FMF is given, except for  $m(\emptyset) \neq 0$ , at most  $2^n - 1$  focal elements for the undetermined BBA need to assign the mass. According to the relationships between the given FMF and belief or plausibility function, together with the BBA legitimate conditions, there are  $n + 1$  equations, which are used as the constraints. When the given FMF is equivalent to a singleton plausibility function, the corresponding constraint can also use the contour function.

As analyzed in [7], using AM as the objective function has some disputes and limitations because it actually quantifies the randomness of the pignistic probability measure approximating a BBA, so it does not capture all the aspects of uncertainty (specially the ambiguity) represented by a BBA. AM has also been criticized in [36].  $TU^1$  is a total uncertainty measure without switching frameworks [36], which is based on Wasserstein distance [33] (a strict distance). Therefore, we replace AM with  $TU^1$  (see Eq. (8)) as the objective function of the transformations based on uncertainty optimization in this paper.

1) *Transformation based on uncertainty minimization*: The objective function of the transformation based on uncertainty minimization is  $TU^1$ . The constraints are mainly based on the relationship defined Eq. (12) or Eq. (13). Suppose that the given FMF defined on  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  is  $\mu = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . The BBA, denoted by  $m_{\text{min}}$ , can be obtained as follows:

$$\begin{aligned} &\text{When } \sum_{i=1}^n \mu(\theta_i) \geq 1, \\ m_{\text{min}} &= \arg \min_m TU^1(m) \\ &= \arg \min_m 1 - \frac{\sqrt{3}}{n} \sum_{i=1}^n d_w([\text{Bel}(\{\theta_i\}), \text{Pl}(\{\theta_i\})], [0, 1]) \end{aligned} \quad (17)$$

$$\begin{aligned} &\left\{ \begin{aligned} \sum_{\theta_i \cap A \neq \emptyset} m(A) &= \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta, \\ \sum_{A \subseteq \Theta} m(A) &= 1, \\ 0 \leq m(A) &\leq 1. \end{aligned} \right. \end{aligned}$$

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ ,

$$\begin{aligned} \mathbf{m}_{\min} &= \arg \min_{\mathbf{m}} \text{TU}^I(\mathbf{m}) \\ &= \arg \min_{\mathbf{m}} 1 - \frac{\sqrt{3}}{n} \sum_{i=1}^n d_W([\text{Bel}(\{\theta_i\}), \text{Pl}(\{\theta_i\})], [0, 1]) \end{aligned} \quad (18)$$

$$s.t. \begin{cases} m(\{\theta_i\}) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta, \\ \sum_{A \subseteq \Theta} m(A) = 1, \\ 0 \leq m(A) \leq 1. \end{cases}$$

In the sequel, the transformation based on uncertainty minimization is represented by “ $T_{\min}$ ” for convenience. When  $\sum_{i=1}^n \mu(\theta_i) = 1$ , Eq. (17) and Eq. (18) are equivalent. Therefore when  $\sum_{i=1}^n \mu(\theta_i) = 1$ , one can choose either Eq. (17) or Eq. (18) to do the transformation. The proof of their equivalence is in the Appendix.

#### 2) Transformation based on uncertainty maximization:

The transformation based on the uncertainty maximization has the same objective function as that in  $T_{\min}$ . The corresponding constraints are also mainly based on the relationship between the given FMF and the BBA to determine. For the FOD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ ,  $\boldsymbol{\mu} = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$  is the given FMF. The obtained BBA is represented by  $\mathbf{m}_{\max}$ . The objective function and corresponding constraints are as follows:

When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ ,

$$\begin{aligned} \mathbf{m}_{\max} &= \arg \max_{\mathbf{m}} \text{TU}^I(\mathbf{m}) \\ &= \arg \max_{\mathbf{m}} 1 - \frac{\sqrt{3}}{n} \sum_{i=1}^n d_W([\text{Bel}(\{\theta_i\}), \text{Pl}(\{\theta_i\})], [0, 1]) \end{aligned} \quad (19)$$

$$s.t. \begin{cases} \sum_{\theta_i \cap A \neq \emptyset} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta, \\ \sum_{A \subseteq \Theta} m(A) = 1, \\ 0 \leq m(A) \leq 1. \end{cases}$$

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ ,

$$\begin{aligned} \mathbf{m}_{\max} &= \arg \max_{\mathbf{m}} \text{TU}^I(\mathbf{m}) \\ &= \arg \max_{\mathbf{m}} 1 - \frac{\sqrt{3}}{n} \sum_{i=1}^n d_W([\text{Bel}(\{\theta_i\}), \text{Pl}(\{\theta_i\})], [0, 1]) \end{aligned} \quad (20)$$

$$s.t. \begin{cases} m(\{\theta_i\}) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta, \\ \sum_{A \subseteq \Theta} m(A) = 1, \\ 0 \leq m(A) \leq 1. \end{cases}$$

In the sequel, the transformation based on uncertainty maximization is represented by  $T_{\max}$  for convenience. When  $\sum_{i=1}^n \mu(\theta_i) = 1$ , one can choose either Eq. (19) or Eq. (20) to do the transformation, because when  $\sum_{i=1}^n \mu(\theta_i) = 1$ , Eq. (19) and Eq. (20) are equivalent. The proof of their equivalence is in the Appendix. Although the available transformations can transform a given FMF into a BBA, there are still some limitations and problems described in the next section.

#### D. Limitations of available transformations

1) *Limitations of transformations with preselection of focal elements:* As referred previously, to deal with the under-determined problem for the transformation of an FMF into a BBA, preselecting focal elements or solving the optimization problem are used. Compared with solving the optimization problem ( $T_{\min}$  and  $T_{\max}$ ), preselecting the focal elements without sufficient witness might bring the loss of information. Since the focal elements are preselected, the obtained BBA can only assign the mass to the focal elements specified beforehand (e.g., as with  $T_{\text{norm}}$ ,  $T_{\alpha\text{-cut}}$  or  $T_{\text{tri}}$ ). For the BBA obtained by using  $T_{\text{norm}}$ , there are no compound focal elements. The BBA obtained by using  $T_{\alpha\text{-cut}}$  or  $T_{\text{tri}}$  has a specific structure of focal elements for the given FMF. In addition, for different FMFs (such as in Example 1), the same BBAs might be obtained by using  $T_{\text{norm}}$ ,  $T_{\alpha\text{-cut}}$  or  $T_{\text{tri}}$ .

2) *Example 1:* Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . There are two FMFs  $\boldsymbol{\mu}_1 = [0.3, 0.2, 0.1]$  and  $\boldsymbol{\mu}_2 = [0.9, 0.6, 0.3]$ . The BBAs obtained by  $T_{\text{norm}}$  are as follows:

$$\boldsymbol{\mu}_1 : m_1(\{\theta_1\}) = 1/2, m_1(\{\theta_2\}) = 1/3, m_1(\{\theta_3\}) = 1/6,$$

$$\boldsymbol{\mu}_2 : m_2(\{\theta_1\}) = 1/2, m_2(\{\theta_2\}) = 1/3, m_2(\{\theta_3\}) = 1/6.$$

Using  $T_{\alpha\text{-cut}}$ , the BBAs obtained from  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$  are:

$$\boldsymbol{\mu}_1 : m_3(\{\theta_1\}) = 1/3, m_3(\{\theta_1, \theta_2\}) = 1/3, m_3(\Theta) = 1/3,$$

$$\boldsymbol{\mu}_2 : m_4(\{\theta_1\}) = 1/3, m_4(\{\theta_1, \theta_2\}) = 1/3, m_4(\Theta) = 1/3.$$

Using  $T_{\text{tri}}$ , the BBAs obtained from  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$  are:

$$\boldsymbol{\mu}_1 : m_5(\{\theta_1\}) = 1/2, m_5(\{\theta_2\}) = 1/3, m_5(\Theta) = 1/6,$$

$$\boldsymbol{\mu}_2 : m_6(\{\theta_1\}) = 1/2, m_6(\{\theta_2\}) = 1/3, m_6(\Theta) = 1/6.$$

One can see that  $\mathbf{m}_1 = \mathbf{m}_2$ ,  $\mathbf{m}_3 = \mathbf{m}_4$  and  $\mathbf{m}_5 = \mathbf{m}_6$ . This is not that rational.  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$  are completely different FMFs and have different uncertainty degrees. Using Eq. (10) to calculate the degrees of fuzziness for  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$ , one can verify that  $D(\boldsymbol{\mu}_1) = 0.6907 \neq D(\boldsymbol{\mu}_2) = 0.7737$ . That is to say, given two FMFs with different degrees of fuzziness, the obtained BBAs are respectively identical by using  $T_{\text{norm}}$ ,  $T_{\alpha\text{-cut}}$  or  $T_{\text{tri}}$ .

3) *Limitations of transformations based on uncertainty optimization:* The optimization-based transformations take every possible focal element into consideration to assign the mass, which avoid preselecting the focal elements and deal with multi-answer problem by solving optimization problem. However, in the process of solving optimization problem, the two transformations based on uncertainty optimization consider the minimal and maximal uncertainty degrees respectively, which might lead to extreme attitudinal bias on the uncertainty degree and bring “one-sided” and counter-intuitive results.

4) *Example 2:* Suppose that the given FMF is  $\mu_3 = [0.58, 0.5, 0.42]$ . The given FMF satisfies  $\sum_{i=1}^3 \mu(\theta_i) = 1.5 > 1$ . Using Eq. (19), the BBA obtained by  $T_{\max}$  is

$$\begin{aligned} \mu_3 : m_7(\{\theta_1\}) &= 0.09, m_7(\{\theta_2\}) = 0.17, \\ m_7(\{\theta_1, \theta_2\}) &= 0.32, m_7(\{\theta_3\}) = 0.25, \\ m_7(\{\theta_1, \theta_3\}) &= 0.16, m_7(\Theta) = 0.01. \end{aligned}$$

One can calculate pignistic probability of  $m_7$  with Eq. (6).

$$\text{BetP}_7(\theta_1) = 1/3, \text{BetP}_7(\theta_2) = 1/3, \text{BetP}_7(\theta_3) = 1/3.$$

As can be seen,

$$\text{BetP}_7(\theta_1) = \text{BetP}_7(\theta_2) = \text{BetP}_7(\theta_3).$$

However, the values of  $\mu_3$  have obvious difference and thus this equation is not that rational. When only emphasizing the maximal uncertainty, a certain degree of difference might be ignored.

Suppose that the given FMF is  $\mu_4 = [0.5001, 0.5, 0.4999]$ , the BBA obtained by  $T_{\min}$  is as follows:

$$\begin{aligned} \mu_4 : m_8(\{\theta_1\}) &= 0.5, m_8(\{\theta_1, \theta_2\}) = 0.0001, \\ m_8(\{\theta_2, \theta_3\}) &= 0.4999. \end{aligned}$$

The corresponding pignistic probability is  $\text{BetP}_8(\theta_1) \approx 0.5$ ,  $\text{BetP}_8(\theta_2) \approx 0.25$  and  $\text{BetP}_8(\theta_3) \approx 0.25$ . When only emphasizing the minimal uncertainty,  $\text{BetP}_8(\theta_1)$  is overemphasized.

The optimization-based transformations emphasize the extreme cases of uncertainty and might bring counter-intuitive results. In fact, a trade-off, i.e., a more “moderate” (or balanced) BBA, is more natural than the obtention of BBA based on extreme (min, or max) strategies. To avoid being “one-sided”, we propose “moderate” transformations to obtain the trade-off BBA with an uncertainty between the minimal and maximal uncertainty as presented in the next section.

#### IV. MODERATE TRANSFORMATIONS WITH WEIGHTING FACTOR

To obtain a BBA with a trade-off or moderate uncertainty degree, one can use weighting factors, which can be user-specified or determined by some prior information. Considering different preference or requirements of users, the value of the weighting factor can be determined according to the prior information or directly determined by the user. We use the weighting factor to transform an FMF into a trade-off BBA based on the two BBAs  $m_{\min}$  and  $m_{\max}$ , where  $m_{\min}$  is obtained by  $T_{\min}$ , and  $m_{\max}$  is obtained by  $T_{\max}$ . Let  $\beta$  ( $0 \leq \beta \leq 1$ ) be the weighting factor, the trade-off BBA satisfies the following conditions:

- (1) When  $\beta \rightarrow 0$ , the trade-off BBA becomes closer to  $m_{\min}$ ;
- (2) When  $\beta \rightarrow 1$ , the trade-off BBA becomes closer to  $m_{\max}$ .

Meanwhile, the trade-off BBA's corresponding singleton belief or singleton plausibility should be equivalent to the given FMF. Here, we propose two transformations of an FMF into such a trade-off BBA.

#### A. Weighted average based transformation

Consider the frame of discernment  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The given FMF is  $\mu = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . First, we calculate  $m_{\min}$  and  $m_{\max}$  by  $T_{\min}$  and  $T_{\max}$ , respectively. A trade-off BBA, denoted by  $m_{\text{wa}}$ , is a weighted average of  $m_{\min}$  and  $m_{\max}$  as shown in Eq. (21):

$$m_{\text{wa}} = (1 - \beta)m_{\min} + \beta m_{\max}, \quad (21)$$

where  $A \subseteq \Theta$ . In the sequel, the transformation based on the Weighted Average (WA) is represented by “ $T_{\text{wa}}$ ”.

The trade-off BBA obtained by Eq. (21) satisfies the following legitimate conditions of a BBA:

$$\sum_{A \subseteq \Theta} m_{\text{wa}}(A) = 1, \text{ and } 0 \leq m_{\text{wa}}(A) \leq 1, \forall A \subseteq \Theta.$$

In addition, the singleton belief or singleton plausibility for the trade-off BBA obtained by  $T_{\text{wa}}$  is equivalent to the given FMF. When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ , both  $m_{\min}$  and  $m_{\max}$  satisfy Eq. (12). Then, we have  $\text{Pl}_{\min}(\{\theta_i\}) = \text{Pl}_{\max}(\{\theta_i\}) = \mu(\theta_i)$ ,  $i = 1, 2, \dots, n$ , where  $\text{Pl}_{\min}$  and  $\text{Pl}_{\max}$  are the plausibility functions for  $m_{\min}$  and  $m_{\max}$ , respectively. According to Eq. (21), one can deduce that

$$\begin{aligned} \text{Pl}_{\text{wa}}(\{\theta_i\}) &= \sum_{\{\theta_i\} \cap A \neq \emptyset} m_{\text{wa}}(A) \\ &= \sum_{\{\theta_i\} \cap A \neq \emptyset} (1 - \beta)m_{\min}(A) + \beta m_{\max}(A) \\ &= (1 - \beta) \sum_{\{\theta_i\} \cap A \neq \emptyset} m_{\min}(A) \\ &\quad + \beta \sum_{\{\theta_i\} \cap A \neq \emptyset} m_{\max}(A) \quad (22) \\ &= (1 - \beta)\text{Pl}_{\min}(\{\theta_i\}) + \beta \text{Pl}_{\max}(\{\theta_i\}) \\ &= (1 - \beta)\mu(\theta_i) + \beta \mu(\theta_i) = \mu(\theta_i) \end{aligned}$$

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ , both  $m_{\min}$  and  $m_{\max}$  satisfy Eq. (13). Then, we have  $\text{Bel}_{\min}(\{\theta_i\}) = \text{Bel}_{\max}(\{\theta_i\}) = \mu(\theta_i)$ ,  $i = 1, 2, \dots, n$ , where  $\text{Bel}_{\min}$  and  $\text{Bel}_{\max}$  are the plausibility functions for  $m_{\min}$  and  $m_{\max}$ , respectively. According to Eq. (21), one can deduce that

$$\begin{aligned} \text{Bel}_{\text{wa}}(\{\theta_i\}) &= \sum_{A=\{\theta_i\}} m_{\text{wa}}(A) \\ &= \sum_{A=\{\theta_i\}} (1 - \beta) \cdot m_{\min}(A) + \beta \cdot m_{\max}(A) \\ &= (1 - \beta) \cdot \sum_{A=\{\theta_i\}} m_{\min}(A) \\ &\quad + \beta \cdot \sum_{A=\{\theta_i\}} m_{\max}(A) \quad (23) \\ &= (1 - \beta) \cdot \text{Bel}_{\min}(\{\theta_i\}) + \beta \cdot \text{Bel}_{\max}(\{\theta_i\}) \\ &= (1 - \beta) \cdot \mu(\theta_i) + \beta \cdot \mu(\theta_i) = \mu(\theta_i) \end{aligned}$$

According to Eq. (21), the trade-off BBA obtained by using  $T_{\text{wa}}$  conforms to the conditions aforementioned at the

beginning of Section IV. When  $\beta \rightarrow 0$ ,  $m_{wa}$  approximates to  $m_{\min}$ . When  $\beta \rightarrow 1$ ,  $m_{wa}$  approximates to  $m_{\max}$ .

Beside the directly weighted approach, we can also use degree of dissimilarity between the trade-off BBA and the BBA obtained by using  $m_{\min}$  or  $m_{\max}$ . Then, the weighting factor can influence the relationship between two degrees of dissimilarity to obtain the trade-off BBA.

### B. User-specified optimization based transformation

We use the distance of evidence between two BBAs in this transformation. A trade-off BBA to determine is represented by  $m_{uo}$ , where the subscript “uo” denotes the “user-specified optimization”.  $m_{\min}$  is obtained by  $T_{\min}$ .  $m_{\max}$  is obtained by  $T_{\max}$ .  $d(m_{uo}, m_{\min})$  denotes the distance of evidence between  $m_{uo}$  and  $m_{\min}$ , while  $d(m_{uo}, m_{\max})$  denotes the distance of evidence between  $m_{uo}$  and  $m_{\max}$ . When  $d(m_{uo}, m_{\max})$  is larger,  $m_{uo}$  is closer to  $m_{\min}$ . When  $d(m_{uo}, m_{\min})$  is larger,  $m_{uo}$  is closer to  $m_{\max}$ .

Then, we can use the weighting factor  $\beta$  ( $0 \leq \beta \leq 1$ ) to influence how close  $m_{uo}$  is to  $m_{\min}$  or  $m_{\max}$ . The relationship between two distances of evidence (i.e.,  $d(m_{uo}, m_{\min})$  and  $d(m_{uo}, m_{\max})$ ) is defined based on  $\beta$  as follows:

$$\frac{d(m_{uo}, m_{\min})}{d(m_{uo}, m_{\max})} = \frac{\beta}{1 - \beta} \quad (24)$$

Here,  $\beta$  can be regarded as the weight of  $d(m_{uo}, m_{\min})$ , and  $1 - \beta$  can be regarded as the weight of  $d(m_{uo}, m_{\max})$ . The visualized illustration of Eq. (24) can be shown in Fig. 1.

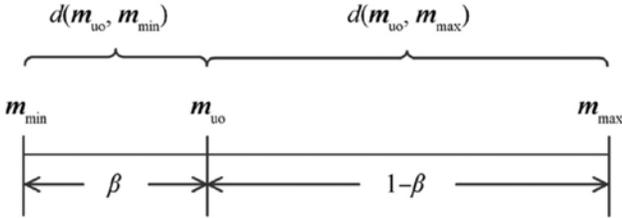


Figure 1. Illustration of Eq. (24).

As we see in Fig. 1, when the value of  $\beta$  gradually decreases,  $d(m_{uo}, m_{\min})$  gradually decreases. Meanwhile, the value of  $d(m_{uo}, m_{\max})$  gradually increases. Then  $m_{uo}$  becomes similar to  $m_{\min}$ . Conversely, when the value of  $\beta$  increases gradually, the value of  $d(m_{uo}, m_{\min})$  increases and the value of  $d(m_{uo}, m_{\max})$  decreases gradually. Then  $m_{uo}$  becomes similar to  $m_{\max}$ . However, a BBA strictly satisfied Eq. (24) might not always exist. Therefore, we rewrite Eq. (24) as

$$\text{obj}(m_{uo}) = [(1 - \beta) \cdot d(m_{uo}, m_{\min}) - \beta \cdot d(m_{uo}, m_{\max})]^2 \quad (25)$$

where the value range of  $\beta$  can be  $[0, 1]$ .

Although Eq. (24) cannot strictly hold sometime, so long as the value of  $\text{obj}(m_{uo})$  is small enough, Eq. (24) holds approximately. Then, we establish a minimization problem, whose

objective function is Eq. (25), to obtain the trade-off BBA. The constraints of this minimization problem are mainly based on the given FMF. Consider an FOD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ , the given FMF is represented by  $\mu = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . Here, we use Jousselme’s distance (see Eq. (9)) to construct the objective function.

When  $\sum_{i=1}^n \mu(\theta_i) \geq 1$ ,

$$\begin{aligned} m_{uo} = \arg \min_m & [(1 - \beta) \cdot d_J(m, m_{\min}) - \beta \cdot d_J(m, m_{\max})]^2 \\ \text{s.t.} & \begin{cases} \sum_{\theta_i \cap A \neq \emptyset} m(A) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta, \\ \sum_{A \subseteq \Theta} m(A) = 1, \\ 0 \leq m(A) \leq 1. \end{cases} \end{aligned} \quad (26)$$

When  $\sum_{i=1}^n \mu(\theta_i) \leq 1$ ,

$$\begin{aligned} m_{uo} = \arg \min_m & [(1 - \beta) \cdot d_J(m, m_{\min}) - \beta \cdot d_J(m, m_{\max})]^2 \\ \text{s.t.} & \begin{cases} m(\{\theta_i\}) = \mu(\theta_i), \forall \{\theta_i\} \subseteq \Theta, \\ \sum_{A \subseteq \Theta} m(A) = 1, \\ 0 \leq m(A) \leq 1. \end{cases} \end{aligned} \quad (27)$$

In the sequel, the user-specified optimization-based transformation is represented by  $T_{uo}$  for convenience. When  $\sum_{i=1}^n \mu(\theta_i) = 1$ , one can choose either Eq. (26) or Eq. (27) to do the transformation, because when  $\sum_{i=1}^n \mu(\theta_i) = 1$ , Eq. (26) and Eq. (27) are equivalent. The proof of their equivalence is in the Appendix.

We find that using different evidence distances to construct the objective function might transform a given FMF into different BBAs. But the difference between these BBAs is relatively small. Therefore, we use Jousselme’s distance, one of the representative evidence distances, as the objective function of  $T_{uo}$  in this paper.

For convenience, we list all the transformations aforementioned and their abbreviations in Table 1 together with the symbols of corresponding BBAs obtained.

Table I  
TRANSFORMATIONS AND THEIR ABBREVIATIONS.

Approach	Description of Transformation	BBA
$T_{\text{norm}}$	Normalization based transformation cf Eq. (14)	$m_{\text{norm}}$
$T_{\alpha\text{-cut}}$	Transformation based on $\alpha$ -cut cf Eq. (15)	$m_{\alpha\text{-cut}}$
$T_{\text{tri}}$	Transformation based on assigning mass to a focal element triplet, cf Eq. (16)	$m_{\text{tri}}$
$T_{\text{min}}$	Transformation based on uncertainty minimization cf Eq. (17) or Eq. (18)	$m_{\text{min}}$
$T_{\text{max}}$	Transformation based on uncertainty maximization cf Eq. (19) or Eq. (20)	$m_{\text{max}}$
$T_{\text{wa}}$	Weighted average based transformation cf Eq. (21)	$m_{\text{wa}}$
$T_{\text{uo}}$	User-specified optimization based transformation cf Eq. (26) or Eq. (27)	$m_{\text{uo}}$

### C. Example 3 for illustration

Here, an illustrative example to show the computation procedure of our transformations is provided. Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The given FMF is  $\mu(\theta_1) = 0.9, \mu(\theta_2) = 0.7, \mu(\theta_3) = 0.3$ . The weighting factor is  $\beta, (0 \leq \beta \leq 1)$ . The possible focal elements for this example are  $A_1 = \{\theta_1\}, A_2 = \{\theta_2\}, A_3 = \{\theta_1, \theta_2\}, A_4 = \{\theta_3\}, A_5 = \{\theta_1, \theta_3\}, A_6 = \{\theta_2, \theta_3\},$  and  $A_7 = \Theta$ .

First,  $\mathbf{m}_{\min}$  and  $\mathbf{m}_{\max}$  are calculated. Using  $T_{\min}$  and  $T_{\max}$  we have

$$\begin{aligned} m_{\min(A_1)} &= 0.3 \\ m_{\min(A_3)} &= 0.4 \\ m_{\min(A_6)} &= 0.1 \\ m_{\min(A_7)} &= 0.2 \end{aligned}$$

$$\begin{aligned} m_{\max(A_1)} &= 0.05 \\ m_{\max(A_2)} &= 0.03 \\ m_{\max(A_3)} &= 0.62 \\ m_{\max(A_4)} &= 0.07 \\ m_{\max(A_5)} &= 0.18 \\ m_{\max(A_7)} &= 0.05 \end{aligned}$$

Second, using  $T_{\text{wa}}$ ,  $\mathbf{m}_{\text{wa}}$  can be obtained. For instance, if we choose  $\beta = 0.8$ , then we have

$$\begin{aligned} m_{\text{wa}}(A_1) &= (1 - 0.8) \times 0.3 + 0.8 \times 0.05 = 0.1 \\ m_{\text{wa}}(A_2) &= 0.8 \times 0.03 = 0.024 \\ m_{\text{wa}}(A_3) &= (1 - 0.8) \times 0.4 + 0.8 \times 0.62 = 0.576 \\ m_{\text{wa}}(A_4) &= 0.8 \times 0.07 = 0.056 \\ m_{\text{wa}}(A_5) &= 0.8 \times 0.18 = 0.144 \\ m_{\text{wa}}(A_6) &= (1 - 0.8) \times 0.1 = 0.02 \\ m_{\text{wa}}(A_7) &= (1 - 0.8) \times 0.2 + 0.8 \times 0.05 = 0.08 \end{aligned}$$

According to  $\sum_{i=1}^3 \mu(\theta_i) = 1.9 > 1$ , the given FMF is equivalent to the singleton plausibility. For the user-specified optimization-based transformations, Eq. (26) is used to transform the given FMF into a trade-off BBA.

$$\begin{aligned} \mathbf{m}_{\text{uo}} &= \arg \min_{\mathbf{m}} [(1 - \beta) \cdot d_J(\mathbf{m}, \mathbf{m}_{\min}) - \beta \cdot d_J(\mathbf{m}, \mathbf{m}_{\max})]^2 \\ \text{s.t. } &\begin{cases} m(A_1) + m(A_3) + m(A_5) + m(A_7) = \mu(\theta_1) = 0.9 \\ m(A_2) + m(A_3) + m(A_6) + m(A_7) = \mu(\theta_2) = 0.6 \\ m(A_4) + m(A_5) + m(A_6) + m(A_7) = \mu(\theta_3) = 0.3 \\ \sum_{i=1}^7 m(A_i) = 1, \\ 0 \leq m(A_i) \leq 1, i = 1, 2, \dots, 7 \end{cases} \end{aligned}$$

The trade-off BBAs can be obtained by using  $T_{\text{uo}}$ . Here,  $\beta = 0.8$  and

$$\begin{aligned} m_{\text{uo}}(A_1) &= 0.1013 \\ m_{\text{uo}}(A_2) &= 0.0390 \\ m_{\text{uo}}(A_3) &= 0.5597 \\ m_{\text{uo}}(A_4) &= 0.0610 \\ m_{\text{uo}}(A_5) &= 0.1377 \\ m_{\text{uo}}(A_7) &= 0.1013 \end{aligned}$$

As expected, when  $\beta = 0.8$ ,  $\mathbf{m}_{\text{wa}}$  and  $\mathbf{m}_{\text{uo}}$  are closer to  $\mathbf{m}_{\max}$ , respectively.

## V. EXAMPLES

We use different approaches to transform an FMF into a BBA in this section. Some numerical examples are provided to illustrate the difference between different transformations (including all the transformations mentioned in Table I). In addition, a classification example is provided to compare our transformations with other transformations.

### A. Example 4

This example is a revisiting of the two examples in III-D2 and III-D4 with  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . As mentioned above, using transformations with preselection of focal elements might transform different FMFs (e.g.,  $\mu_1 = [0.3, 0.2, 0.1]$  and  $\mu_2 = [0.9, 0.6, 0.3]$  in Example 1) into the same BBA. Then we need to know the results of using transformations based on uncertainty optimization and moderate transformations. Here, we use  $T_{\min}, T_{\max}, T_{\text{wa}}$  and  $T_{\text{uo}}$  to transform  $\mu_1$  and  $\mu_2$  into BBAs, respectively. The corresponding obtained BBAs are listed in Table II and Table III, respectively. The results of moderate transformations (i.e.,  $T_{\text{wa}}$  and  $T_{\text{uo}}$ ) are obtained with  $\beta = 0.7$ .

According to the results of each transformation listed in Tables II and III, it can be seen that the BBAs transformed from  $\mu_1$  and  $\mu_2$  are different. Their degrees of uncertainty are different. That is, using optimization-based transformations can avoid transforming different FMFs into the same BBAs due to the preselection of focal elements. Although only the results of moderate transformations with  $\beta = 0.7$  are listed, the trade-off BBAs transformed from  $\mu_1$  and  $\mu_2$  would be different at any value of  $\beta$ .

For  $\mu_3 = [0.58, 0.5, 0.42]$  and  $\mu_4 = [0.5001, 0.5, 0.4999]$  in Example 2, the aforementioned results in Section III-D4 show that using transformations based on uncertainty optimization might lead to extreme attitudinal bias on the uncertainty degree during the process of solving optimization problem and bring counter-intuitive results. Can moderate transformations avoid such a bias?

In this example, we use the moderate transformations to obtain BBAs from  $\mu_3$  and  $\mu_4$ , respectively. Then, we calculate the corresponding pignistic probability. The trade-off BBAs transformed from  $\mu_3$  and  $\mu_4$  are listed in Table IV and Table V, respectively. Here,  $\beta = 0.7$ .

Table II  
BBAS TRANSFORMED FROM  $\mu_1$ .

Approach	BBA	TU <sup>1</sup>
$T_{\min}$	$m_{\min}(A_1) = 0.3, m_{\min}(A_2) = 0.2, m_{\min}(A_3) = 0.4, m_{\min}(A_4) = 0.1$	0.4997
$T_{\max}$	$m_{\max}(A_1) = 0.3, m_{\max}(A_2) = 0.2, m_{\max}(A_4) = 0.1, m_{\max}(A_7) = 0.4$	0.6318
$T_{wa}$	$m_{wa}(A_1) = 0.3, m_{wa}(A_2) = 0.2, m_{wa}(A_3) = 0.12, m_{wa}(A_4) = 0.1, m_{wa}(A_7) = 0.28$	0.5923
$T_{uo}$	$m_{uo}(A_1) = 0.3, m_{uo}(A_2) = 0.2, m_{uo}(A_3) = 0.12, m_{uo}(A_4) = 0.1,$ $m_{uo}(A_5) = 0.0335, m_{uo}(A_6) = 0.1849, m_{uo}(A_7) = 0.1779$	0.5795

Table III  
BBAS TRANSFORMED FROM  $\mu_2$ .

Approach	BBA	TU <sup>1</sup>
$T_{\min}$	$m_{\min}(A_1) = 0.4, m_{\min}(A_3) = 0.3, m_{\min}(A_6) = 0.1, m_{\min}(A_7) = 0.2$	0.5131
$T_{\max}$	$m_{\max}(A_1) = 0.1, m_{\max}(A_2) = 0.0364, m_{\max}(A_3) = 0.5636, m_{\max}(A_4) = 0.0636, m_{\max}(A_5) = 0.2364$	0.6155
$T_{wa}$	$m_{wa}(A_1) = 0.19, m_{wa}(A_2) = 0.0255, m_{wa}(A_3) = 0.4845, m_{wa}(A_4) = 0.0445,$ $m_{wa}(A_5) = 0.1655, m_{wa}(A_6) = 0.03, m_{wa}(A_7) = 0.06$	0.5896
$T_{uo}$	$m_{uo}(A_1) = 0.1922, m_{uo}(A_2) = 0.0490, m_{uo}(A_3) = 0.4588, m_{uo}(A_4) = 0.0506,$ $m_{uo}(A_5) = 0.1572, m_{uo}(A_6) = 0.0005, m_{uo}(A_7) = 0.0917$	0.5931

[htp]

Table IV  
BBAS TRANSFORMED FROM  $\mu_3$ .

Approach	BBA	[BetP( $\theta_1$ ), BetP( $\theta_2$ ), BetP( $\theta_3$ )]
$T_{wa}$	$m_{wa}(A_1) = 0.297, m_{wa}(A_2) = 0.1361, m_{wa}(A_3) = 0.1469,$ $m_{wa}(A_4) = 0.1579, m_{wa}(A_5) = 0.0451, m_{wa}(A_6) = 0.126, m_{wa}(A_7) = 0.091$	[0.4233, 0.3029, 0.2738]
$T_{uo}$	$m_{uo}(A_1) = 0.2283, m_{uo}(A_2) = 0.2355, m_{uo}(A_3) = 0.1162, m_{uo}(A_4) = 0.0747,$ $m_{uo}(A_5) = 0.1969, m_{uo}(A_6) = 0.1098, m_{uo}(A_7) = 0.0386$	[0.3977, 0.3613, 0.2410]

Table V  
BBAS TRANSFORMED FROM  $\mu_4$ .

Approach	BBA	[BetP( $\theta_1$ ), BetP( $\theta_2$ ), BetP( $\theta_3$ )]
$T_{wa}$	$m_{wa}(A_1) = 0.3385, m_{wa}(A_2) = 0.0885, m_{wa}(A_3) = 0.0731,$ $m_{wa}(A_4) = 0.0886, m_{wa}(A_5) = 0.0729, m_{wa}(A_6) = 0.3228, m_{wa}(A_7) = 0.0156$	[0.4167, 0.2917, 0.2916]
$T_{uo}$	$m_{uo}(A_1) = 0.3584, m_{uo}(A_2) = 0.1393, m_{uo}(A_3) = 0.0024, m_{uo}(A_4) = 0.0741,$ $m_{uo}(A_5) = 0.0675, m_{uo}(A_6) = 0.2864, m_{uo}(A_7) = 0.0719$	[0.4173, 0.3077, 0.2750]

As can be seen in Table IV, the three values of the pignistic probability of each trade-off BBA are different instead of  $\text{BetP}_7(\theta_1) = \text{BetP}_7(\theta_2) = \text{BetP}_7(\theta_3)$  in Example 2. Using the moderate transformation can obtain a trade-off or balanced BBA to express even a small difference of the given FMF (e.g.,  $\mu_3$ ) and bring a rational result.

According to the results listed in Table V, there is no overemphasis on  $\text{BetP}(\theta_1)$  (the pignistic probability in Example 2 is  $\text{BetP}_8(\theta_1) = 0.5, \text{BetP}_8(\theta_2) = 0.25$  and  $\text{BetP}_8(\theta_3) = 0.25$ ). One can transform such an FMF like  $\mu_4$  into a trade-off BBA by using the moderate transformations to avoid the overemphasis caused by “one-sidedness” on the uncertainty degree. Even if the difference of values of  $\mu_4$  is tiny, they are different and there exists  $\mu_4(\theta_1) > \mu_4(\theta_2) > \mu_4(\theta_3)$ . As can be seen in Table V, the trade-off BBAs obtained by using  $T_{wa}$  and  $T_{uo}$  can represent the tiny difference between the values of  $\mu_4$ , respectively.

**B. Example 5**

This example is a revisiting of the illustrative example (Example 3) in Section IV-C. In addition to using the moderate

transformations, we also use  $T_{\text{norm}}, T_{\alpha\text{-cut}}$  and  $T_{\text{tri}}$  to transform the given FMF  $\mu = [0.9, 0.7, 0.3]$  into a BBA. The obtained BBAs are listed in Table VI. Furthermore, the trade-off BBAs obtained by using moderate transformations with  $\beta = 0.4$  (we did use  $\beta = 0.8$  in Section IV-C) are listed in the same table.

The corresponding degrees of uncertainty are also listed. Furthermore, we provide the degrees of uncertainty for the trade-off BBAs obtained with  $\beta = 0.8$  as follows:

$$\begin{aligned} \text{TU}^1(m_{\min}) &= 0.5785 \\ \text{TU}^1(m_{\max}) &= 0.6532 \\ \text{TU}^1(m_{wa}) &= 0.6458 \\ \text{TU}^1(m_{uo}) &= 0.6484 \end{aligned}$$

As shown in Table VI,  $m_{\text{norm}}$  only has singleton focal elements. The structure of focal elements for  $m_{\alpha\text{-cut}}$  is nested and that for  $m_{\text{tri}}$  depends on the ordering of values of the given FMF. According to the results in Example 3 and Table VI, using the uncertainty optimization based transformations and moderate transformations can consider more focal elements to assign mass. Moreover, the focal element structure of the

obtained BBA is not fixed. When the FMF is given, using the transformations with preselection of focal elements can only obtain a BBA with a certain structure of focal elements.

For the moderate transformations, the degrees of uncertainty of the trade-off BBAs are both between the minimal and maximal degrees of uncertainty (i.e., the degrees of uncertainty of  $m_{\min}$  and  $m_{\max}$ ). When  $\beta = 0.4$ , the trade-off BBAs are closer to  $m_{\min}$ . Meanwhile, for the degree of uncertainty of the trade-off BBA, it is closer to the minimal degree of uncertainty. When  $\beta = 0.8$ , the trade-off BBAs are closer to  $m_{\max}$ . Contrary to the situation when  $\beta = 0.4$ , the degrees of uncertainty of the trade-off BBA are closer to the maximal degree of uncertainty when  $\beta = 0.8$ . It can be indicated that the obtained trade-off BBAs satisfy the conditions for the trade-off BBA mentioned at the beginning of Section IV. In addition, although there exist  $TU^I(\mathbf{m}_{\text{norm}}) < TU^I(\mathbf{m}_{\min})$ ,  $TU^I(\mathbf{m}_{\alpha\text{-cut}}) < TU^I(\mathbf{m}_{\min})$  and  $TU^I(\mathbf{m}_{\text{tri}}) < TU^I(\mathbf{m}_{\min})$ , the three BBAs  $\mathbf{m}_{\text{norm}}$ ,  $\mathbf{m}_{\alpha\text{-cut}}$  and  $\mathbf{m}_{\text{tri}}$  do not satisfy the relationship as Eq. (12).

### C. Example 6

Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ . The given FMF is  $\mu(\theta_1) = 0.4$ ,  $\mu(\theta_2) = 0.3$ ,  $\mu(\theta_3) = 0.2$ , and  $\mu(\theta_4) = 0.1$ . The weighting factor is  $\beta$ , ( $0 \leq \beta \leq 1$ ). The possible focal elements of the obtained BBAs for this example are denoted as follows:  $A_1 = \{\theta_1\}$ ,  $A_2 = \{\theta_2\}$ ,  $A_3 = \{\theta_1, \theta_2\}$ ,  $A_4 = \{\theta_3\}$ ,  $A_5 = \{\theta_1, \theta_3\}$ ,  $A_6 = \{\theta_2, \theta_3\}$ ,  $A_7 = \{\theta_1, \theta_2, \theta_3\}$ ,  $A_8 = \{\theta_4\}$ ,  $A_9 = \{\theta_1, \theta_4\}$ ,  $A_{10} = \{\theta_2, \theta_4\}$ ,  $A_{11} = \{\theta_1, \theta_2, \theta_4\}$ ,  $A_{12} = \{\theta_3, \theta_4\}$ ,  $A_{13} = \{\theta_1, \theta_3, \theta_4\}$ ,  $A_{14} = \{\theta_2, \theta_3, \theta_4\}$ , and  $A_{15} = \Theta$ .

Using all the transformations in Table I, one can transform the given FMF into a BBA. According to  $\sum_{i=1}^4 \mu(\theta_i) = 1$ , the FMF is equivalent to the singleton plausibility or singleton belief. The proof of this equivalent is in the appendix. The obtained BBAs and corresponding values of uncertainty degrees are listed in Table VII. For the moderate transformations, we just list the corresponding BBAs for  $\beta = 0.1$  and  $\beta = 0.9$ .

In Table VII,  $\mathbf{m}_{\text{norm}}$  only has singleton focal elements.  $\mathbf{m}_{\alpha\text{-cut}}$  has four nested focal elements. No matter how many elements the FOD contains,  $\mathbf{m}_{\text{tri}}$  always has only three focal elements. According the criterion of preselection of each transformation with preselection of focal elements and the values of the given FMF, the corresponding BBA has a specific structure of focal elements. As can be seen,  $\mathbf{m}_{\min}$ ,  $\mathbf{m}_{\max}$ ,  $\mathbf{m}_{\text{wa}}$  and  $\mathbf{m}_{\text{uo}}$  in Table VII are identical. No matter what the value of  $\beta$  is, no matter which transformation is used, the trade-off BBAs are identical. This is normal in this particular example because  $m_{\min} = m_{\max}$ . Although there exist  $TU^I(\mathbf{m}_{\alpha\text{-cut}}) > TU^I(\mathbf{m}_{\max})$  and  $TU^I(\mathbf{m}_{\text{tri}}) > TU^I(\mathbf{m}_{\max})$ , the two BBAs  $\mathbf{m}_{\alpha\text{-cut}}$  and  $\mathbf{m}_{\text{tri}}$  do not satisfy the relationship as Eq. (12) or Eq. (13).

### D. Example 7

Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The given FMF is  $\mu(\theta_1) = 0.3$ ,  $\mu(\theta_2) = 0.3$ , and  $\mu(\theta_3) = 0.3$ . The weighting factor is  $\beta$ , ( $0 \leq \beta \leq 1$ ). The possible focal elements of the

obtained BBAs for this example are denoted as follows:  $A_1 = \{\theta_1\}$ ,  $A_2 = \{\theta_2\}$ ,  $A_3 = \{\theta_1, \theta_2\}$ ,  $A_4 = \{\theta_3\}$ ,  $A_5 = \{\theta_1, \theta_3\}$ ,  $A_6 = \{\theta_2, \theta_3\}$ , and  $A_7 = \Theta$ .

We use all the transformations in Table I to transform the given FMF into a BBA. The obtained BBAs and corresponding values of uncertainty are listed in Table VIII. According to  $\sum_{i=1}^3 \mu(\theta_i) = 0.9 < 1$ , the given FMF is equivalent to the singleton belief. For  $T_{\text{wa}}$  and  $T_{\text{uo}}$ , we just list the corresponding results for  $\beta = 0.2$  and  $\beta = 0.6$ .

As shown in Table VIII,  $\mathbf{m}_{\text{norm}}$  only has singleton focal elements and the mass values are the same for all singleton focal elements. By using  $T_{\alpha\text{-cut}}$ ,  $\Theta$  is the only focal element of the obtained BBA. This is because  $\mu(\theta_1) = \mu(\theta_2) = \mu(\theta_3) = 0.3$ . All mass values assigned to the total set means the most uncertain case. For  $\mathbf{m}_{\text{tri}}$ ,  $m_{\text{tri}}(A_1) = 0$ , but  $m_{\text{tri}}(A_2) = m_{\text{tri}}(A_4) = 1/3$ . According to  $A_2 = \{\theta_2\}$  and  $A_4 = \{\theta_3\}$ , there exists overemphasis on  $\{\theta_2\}$  and  $\{\theta_3\}$ . If the given FMF has more than two identical values and uses  $T_{\text{tri}}$  to obtain a BBA, the BBA would overemphasize two singletons when assigning mass, i.e., suppose there are more than two same values in the given FMF (e.g.,  $\mu(\theta_1) = \mu(\theta_2) = \mu(\theta_3) = \mu(\theta_4) = \mu(\theta_5)$ ), two of them (e.g.,  $\mu(\theta_1)$  and  $\mu(\theta_2)$ ) are overemphasized. This is not that rational.

More focal elements are considered to assign mass by using transformations based on uncertainty optimization and moderate transformations. However,  $\mathbf{m}_{\min}$  overemphasize  $A_3 = \{\theta_1, \theta_2\}$  and do not assign mass to other compound focal elements. Using moderate transformations can avoid overemphasizing a certain focal element. In this example,  $\mathbf{m}_{\text{wa}} = \mathbf{m}_{\text{uo}}$ . When  $\beta = 0.2$ ,  $\mathbf{m}_{\text{wa}}$  and  $\mathbf{m}_{\text{uo}}$  are closer to  $\mathbf{m}_{\min}$ , respectively. The values of  $TU^I$  for the corresponding trade-off BBAs are between the minimal and maximal values of  $TU^I$  and closer to  $TU^I(\mathbf{m}_{\min})$ . When  $\beta = 0.6$ , the trade-off BBAs are closer to  $\mathbf{m}_{\max}$ . The values of uncertainty degrees are between the minimal and maximal degrees of uncertainty and closer to  $TU^I(\mathbf{m}_{\max})$ .

In addition, although some of the degrees of uncertainty of the BBAs obtained by using transformations with preselection of focal elements are greater (or less) than the maximal degree of uncertainty (or the minimal degree of uncertainty), the three BBAs  $\mathbf{m}_{\text{norm}}$ ,  $\mathbf{m}_{\alpha\text{-cut}}$  and  $\mathbf{m}_{\text{tri}}$  do not satisfy the relationship as Eq. (13).

### E. Example 8

To verify the effectiveness for the moderate transformations of an FMF into a trade-off BBA, we compare the average accuracy of 300-run experiments of all the mentioned transformations in a classification problem. Note that we only aim to show the impact of different transformations on the classification results, rather than improve the classification accuracy.

We use three datasets of open UCI database [53] including iris dataset, wheat seeds dataset and wine dataset. Iris dataset has 150 samples including 3 classes, each of which has 50 samples. Every sample has 4 features and all data of 4 features for 150 samples are complete. Wheat seeds dataset has 210

Table VI  
OBTAINED BBAS IN EXAMPLE 5.

Approach	BBA	TU <sup>l</sup>
$T_{\text{norm}}$	$m_{\text{norm}}(A_1) = 0.4737, m_{\text{norm}}(A_2) = 0.3684, m_{\text{norm}}(A_4) = 0.1579$	0.3910
$T_{\alpha\text{-cut}}$	$m_{\alpha\text{-cut}}(A_1) = 2/9, m_{\alpha\text{-cut}}(A_3) = 4/9, m_{\alpha\text{-cut}}(A_7) = 1/3$	0.5667
$T_{\text{tri}}$	$m_{\text{tri}}(A_1) = 0.4737, m_{\text{tri}}(A_2) = 0.3684, m_{\text{tri}}(A_7) = 0.1579$	0.4321
$T_{\text{wa}}$ with $\beta = 0.4$	$m_{\text{wa}}(A_1) = 0.2, m_{\text{wa}}(A_2) = 0.012, m_{\text{wa}}(A_3) = 0.488, m_{\text{wa}}(A_4) = 0.028$ $m_{\text{wa}}(A_5) = 0.072, m_{\text{wa}}(A_6) = 0.06, m_{\text{wa}}(A_7) = 0.14$	0.6154
$T_{\text{uo}}$ with $\beta = 0.4$	$m_{\text{uo}}(A_1) = 0.2061, m_{\text{uo}}(A_2) = 0.0578, m_{\text{uo}}(A_3) = 0.4361, m_{\text{uo}}(A_4) = 0.0168,$ $m_{\text{uo}}(A_5) = 0.0771, m_{\text{uo}}(A_6) = 0.0255, m_{\text{uo}}(A_7) = 0.1806$	0.6180

Table VII  
OBTAINED BBAS IN EXAMPLE 6.

Approach	BBA	TU <sup>l</sup>
$T_{\text{norm}}$	$m_{\text{norm}}(A_1) = 0.4, m_{\text{norm}}(A_2) = 0.3, m_{\text{norm}}(A_4) = 0.2, m_{\text{norm}}(A_8) = 0.1$	0.3218
$T_{\alpha\text{-cut}}$	$m_{\alpha\text{-cut}}(A_1) = 0.25, m_{\alpha\text{-cut}}(A_3) = 0.25, m_{\alpha\text{-cut}}(A_7) = 0.25, m_{\alpha\text{-cut}}(A_{15}) = 0.25$	0.5625
$T_{\text{tri}}$	$m_{\text{tri}}(A_1) = 0.4, m_{\text{tri}}(A_2) = 0.3, m_{\text{tri}}(A_{15}) = 0.2$	0.4697
$T_{\text{min}}$	$m_{\text{min}}(A_1) = 0.4, m_{\text{min}}(A_2) = 0.3, m_{\text{min}}(A_4) = 0.2, m_{\text{min}}(A_8) = 0.1$	0.3218
$T_{\text{max}}$	$m_{\text{max}}(A_1) = 0.4, m_{\text{max}}(A_2) = 0.3, m_{\text{max}}(A_4) = 0.2, m_{\text{max}}(A_8) = 0.1$	0.3218
$T_{\text{wa}}$ with $\beta = 0.1$	$m_{\text{wa}}(A_1) = 0.4, m_{\text{wa}}(A_2) = 0.3, m_{\text{wa}}(A_4) = 0.2, m_{\text{wa}}(A_8) = 0.1$	0.3218
$T_{\text{wa}}$ with $\beta = 0.9$	$m_{\text{wa}}(A_1) = 0.4, m_{\text{wa}}(A_2) = 0.3, m_{\text{wa}}(A_4) = 0.2, m_{\text{wa}}(A_8) = 0.1$	0.3218
$T_{\text{uo}}$ with $\beta = 0.1$	$m_{\text{uo}}(A_1) = 0.4, m_{\text{uo}}(A_2) = 0.3, m_{\text{uo}}(A_4) = 0.2, m_{\text{uo}}(A_8) = 0.1,$	0.3218
$T_{\text{uo}}$ with $\beta = 0.9$	$m_{\text{uo}}(A_1) = 0.4, m_{\text{uo}}(A_2) = 0.3, m_{\text{uo}}(A_4) = 0.2, m_{\text{uo}}(A_8) = 0.1,$	0.3218

Table VIII  
OBTAINED BBAS IN EXAMPLE 7.

Approach	BBA	TU <sup>l</sup>
$T_{\text{norm}}$	$m_{\text{norm}}(A_1) = 1/3, m_{\text{norm}}(A_2) = 1/3, m_{\text{norm}}(A_4) = 1/3, m_{\text{norm}}(A_8) = 0.1$	0.4226
$T_{\alpha\text{-cut}}$	$m_{\alpha\text{-cut}}(A_7) = 1$	1
$T_{\text{tri}}$	$m_{\text{tri}}(A_2) = 1/3, m_{\text{tri}}(A_4) = 1/3, m_{\text{tri}}(A_7) = 1/3$	0.5556
$T_{\text{min}}$	$m_{\text{min}}(A_1) = 0.3, m_{\text{min}}(A_2) = 0.3, m_{\text{min}}(A_3) = 0.1, m_{\text{min}}(A_4) = 0.3$	0.4508
$T_{\text{max}}$	$m_{\text{max}}(A_1) = 0.3, m_{\text{max}}(A_2) = 0.3, m_{\text{max}}(A_4) = 0.3, m_{\text{max}}(A_7) = 0.1$	0.4804
$T_{\text{wa}}$ with $\beta = 0.2$	$m_{\text{wa}}(A_1) = 0.3, m_{\text{wa}}(A_2) = 0.3, m_{\text{wa}}(A_3) = 0.08, m_{\text{wa}}(A_4) = 0.3, m_{\text{wa}}(A_7) = 0.02$	0.4568
$T_{\text{wa}}$ with $\beta = 0.6$	$m_{\text{wa}}(A_1) = 0.3, m_{\text{wa}}(A_2) = 0.3, m_{\text{wa}}(A_3) = 0.04, m_{\text{wa}}(A_4) = 0.3, m_{\text{wa}}(A_7) = 0.06$	0.4687
$T_{\text{uo}}$ with $\beta = 0.2$	$m_{\text{uo}}(A_1) = 0.3, m_{\text{uo}}(A_2) = 0.3, m_{\text{uo}}(A_3) = 0.08, m_{\text{uo}}(A_4) = 0.3, m_{\text{uo}}(A_7) = 0.02$	0.4568
$T_{\text{uo}}$ with $\beta = 0.6$	$m_{\text{uo}}(A_1) = 0.3, m_{\text{uo}}(A_2) = 0.3, m_{\text{uo}}(A_3) = 0.04, m_{\text{uo}}(A_4) = 0.3, m_{\text{uo}}(A_7) = 0.06$	0.4687

samples including 3 classes and each of which has 70 samples. Every sample has 7 features and all data of 7 features for 210 samples are complete. Wine dataset has 178 samples including 3 classes, each of which has 59, 71 and 48 samples. Every sample has 13 features and all data of 13 features for 178 samples are complete. The transformation process is given.

In this example, only iris dataset is used to illustrate the detailed process of classification and the four features are denoted by  $f_1, f_2, f_3$  and  $f_4$ . We randomly select 70% of samples from each class as samples of the training set and the test set consist of the rest samples. In the training set, the minimal values, average values and maximal values of each feature for each class are used as parameters. Then, the FMF can be defined as follows:

$$\mu(\theta_j) = \begin{cases} \frac{x_i - \min_i}{\text{ave}_i - \min_i}, & \text{for } \min_i \leq x_i \leq \text{ave}_i, \\ \frac{x_i - \max_i}{\text{ave}_i - \max_i}, & \text{for } \text{ave}_i \leq x_i \leq \max_i, \\ 0, & \text{otherwise.} \end{cases} \quad (28)$$

where  $i = 1, 2, 3, 4$ .  $\min_i$  is the minimal value of Class  $j$  ( $\theta_j, j = 1, 2, 3$ ).  $\text{ave}_i$  is the average value of the Class  $j$ .  $\max_i$  is the maximal value of the Class  $j$ .

In Table IX, the corresponding parameters of training set are listed. The test sample, randomly selected from the test set, is  $x = [5.6, 3, 4.5, 1.5]$ , which belongs to Class 2 (i.e.  $\theta_2$ ). According to the test sample and the corresponding parameters listed in Table IX, the triangular fuzzy membership functions of each feature for each class are shown in Fig. 2.

According to Eq. (28), the four corresponding FMFs are determined as follows:

$$\text{FMF}_1 : \mu_1(\theta_1) = 0.1446, \mu_1(\theta_2) = 0.6171, \mu_1(\theta_3) = 0.4344$$

$$\text{FMF}_2 : \mu_2(\theta_1) = 0.1955, \mu_2(\theta_2) = 0.6422, \mu_2(\theta_3) = 0.9777$$

$$\text{FMF}_3 : \mu_3(\theta_1) = 0, \mu_3(\theta_2) = 0.7692, \mu_3(\theta_3) = 0$$

$$\text{FMF}_4 : \mu_4(\theta_1) = 0, \mu_4(\theta_2) = 0.6481, \mu_4(\theta_3) = 0.1502$$

Using all the transformations listed in Table 1, one can transform the above four FMFs into BBAs, respectively. Let the FOD be  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The possible focal elements are represented as follows:  $A_1 = \{\theta_1\}, A_2 = \{\theta_2\}, A_3 = \{\theta_1, \theta_2\}, A_4 = \{\theta_3\}, A_5 = \{\theta_1, \theta_3\}, A_6 = \{\theta_2, \theta_3\},$  and  $A_7 = \Theta$ .

The corresponding BBAs obtained by using transformations with preselection of focal elements, transformations based

Table IX  
PARAMETERS OF TRAINING SET.

Feature	Class 1 ( $\theta_1$ )			Class 2 ( $\theta_2$ )			Class 3 ( $\theta_3$ )		
	min <sub>1</sub>	ave <sub>1</sub>	max <sub>1</sub>	min <sub>2</sub>	ave <sub>2</sub>	max <sub>2</sub>	min <sub>3</sub>	ave <sub>3</sub>	max <sub>3</sub>
$f_1$	4.3	5.0086	5.7	4.9	6.0343	7	4.9	6.5114	7.9
$f_2$	2.9	3.4114	4.4	2.2	2.771	3.4	2.5	3.0117	3.8
$f_3$	1	1.4571	1.9	3	4	5.1	4.5	5.5314	6.9
$f_4$	0.1	0.2571	0.6	1	1.3371	1.8	1.4	2.0657	2.5

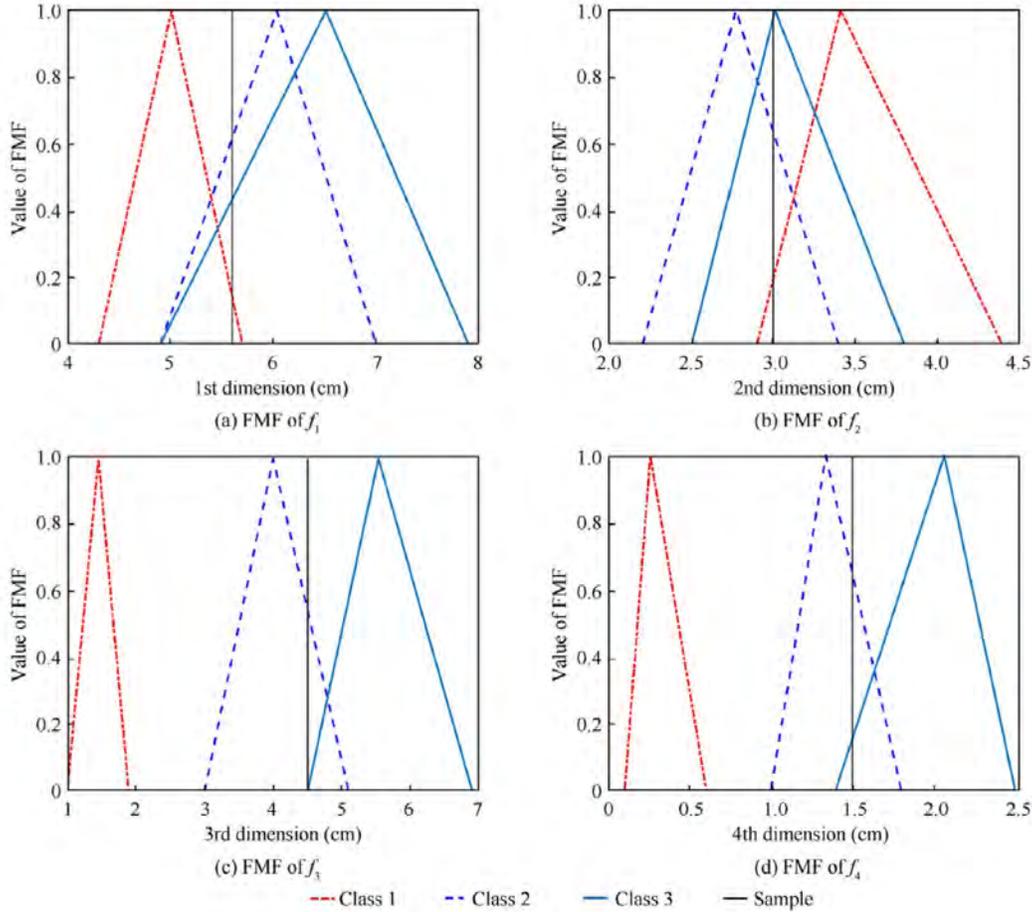


Figure 2. FMFs of four features.

on uncertainty optimization and moderate transformations are listed in Table X. We use moderate transformations to determine trade-off BBAs with  $\beta = 0.2$ .

When the BBAs of each feature are determined, one can combine the corresponding four BBAs of each transformation. For convenience and simplicity, Dempster’s rule of combination is used. The combined BBAs of each transformation are listed in Table XI.

According to the results of Table XI, the corresponding pignistic probabilities can be calculated by using Eq. (6). As we can see in Table XII, the values of  $\text{BetP}(\theta_2)$  for all the transformations are the largest, which means that the

classifications are correct.

In order to compare the moderate transformations with the traditional transformations, we provide 300-run experiments on three datasets (including iris dataset, wheat seeds dataset and wine dataset) to obtain the average accuracy, respectively. All the transformations in Table I are used and the transformation process is the same as above. We use all the features in iris dataset. In wheat seeds dataset and wine dataset, we randomly select 4 features and 7 features to classify, respectively. Reducing the feature dimensions used for classification is to simplify the experiment process.

Table X  
OBTAINED BBAS IN EXAMPLE 8.

Approach	Feature	BBA
$T_{\text{norm}}$	$f_1$	$m_{\text{norm}}(A_1) = 0.1209, m_{\text{norm}}(A_2) = 0.5159, m_{\text{norm}}(A_4) = 0.3632$
	$f_2$	$m_{\text{norm}}(A_1) = 0.1077, m_{\text{norm}}(A_2) = 0.3538, m_{\text{norm}}(A_4) = 0.5385$
	$f_3$	$m_{\text{norm}}(A_2) = 1$
	$f_4$	$m_{\text{norm}}(A_2) = 0.8118, m_{\text{norm}}(A_4) = 0.1882$
$T_{\alpha\text{-cut}}$	$f_1$	$m_{\alpha\text{-cut}}(A_2) = 0.2961, m_{\alpha\text{-cut}}(A_6) = 0.4696, m_{\alpha\text{-cut}}(A_7) = 0.2343$
	$f_2$	$m_{\alpha\text{-cut}}(A_4) = 0.3431, m_{\alpha\text{-cut}}(A_6) = 0.4569, m_{\alpha\text{-cut}}(A_7) = 0.2$
	$f_3$	$m_{\alpha\text{-cut}}(A_2) = 1$
	$f_4$	$m_{\alpha\text{-cut}}(A_2) = 0.7682, m_{\alpha\text{-cut}}(A_6) = 0.2318$
$T_{\text{tri}}$	$f_1$	$m_{\text{tri}}(A_2) = 0.5159, m_{\text{tri}}(A_4) = 0.3632, m_{\text{tri}}(A_7) = 0.1209$
	$f_2$	$m_{\text{tri}}(A_2) = 0.3538, m_{\text{tri}}(A_4) = 0.5385, m_{\text{tri}}(A_7) = 0.1077$
	$f_3$	$m_{\text{tri}}(A_2) = 1$
	$f_4$	$m_{\text{tri}}(A_2) = 0.8118, m_{\text{tri}}(A_4) = 0.1882$
$T_{\text{min}}$	$f_1$	$m_{\text{min}}(A_2) = 0.5656, m_{\text{min}}(A_4) = 0.2383, m_{\text{min}}(A_5) = 0.1446, m_{\text{min}}(A_6) = 0.0515$
	$f_2$	$m_{\text{min}}(A_3) = 0.0223, m_{\text{min}}(A_4) = 0.3578, m_{\text{min}}(A_6) = 0.4467, m_{\text{min}}(A_7) = 0.1732$
	$f_3$	$m_{\text{min}}(A_2) = 0.7692, m_{\text{min}}(A_3) = 0.1887, m_{\text{min}}(A_6) = 0.0421$
	$f_4$	$m_{\text{min}}(A_2) = 0.6482, m_{\text{min}}(A_4) = 0.1502, m_{\text{min}}(A_6) = 0.2016$
$T_{\text{max}}$	$f_1$	$m_{\text{max}}(A_1) = 0.1446, m_{\text{max}}(A_2) = 0.4210, m_{\text{max}}(A_4) = 0.2382, m_{\text{max}}(A_6) = 0.1962$
	$f_2$	$m_{\text{max}}(A_1) = 0.0115, m_{\text{max}}(A_2) = 0.0069, m_{\text{max}}(A_4) = 0.1623, m_{\text{max}}(A_5) = 0.18, m_{\text{max}}(A_6) = 0.6353$
	$f_3$	$m_{\text{max}}(A_2) = 0.7692, m_{\text{max}}(A_5) = 0.2308$
	$f_4$	$m_{\text{max}}(A_2) = 0.6481, m_{\text{max}}(A_4) = 0.1502, m_{\text{max}}(A_5) = 0.1739, m_{\text{max}}(A_7) = 0.0278$
$T_{\text{wa}},$ $\beta = 0.2$	$f_1$	$m_{\text{wa}}(A_1) = 0.0289, m_{\text{wa}}(A_2) = 0.5367, m_{\text{wa}}(A_4) = 0.2382, m_{\text{wa}}(A_5) = 0.1157, m_{\text{wa}}(A_6) = 0.0805$
	$f_2$	$m_{\text{wa}}(A_1) = 0.0031, m_{\text{wa}}(A_2) = 0.0014, m_{\text{wa}}(A_3) = 0.0179, m_{\text{wa}}(A_4) = 0.3187,$ $m_{\text{wa}}(A_5) = 0.036, m_{\text{wa}}(A_6) = 0.4844, m_{\text{wa}}(A_7) = 0.1385$
	$f_3$	$m_{\text{wa}}(A_2) = 0.7692, m_{\text{wa}}(A_3) = 0.1509, m_{\text{wa}}(A_5) = 0.0348, m_{\text{wa}}(A_6) = 0.0337$
	$f_4$	$m_{\text{wa}}(A_2) = 0.6481, m_{\text{wa}}(A_4) = 0.1502, m_{\text{wa}}(A_5) = 0.0348, m_{\text{wa}}(A_6) = 0.1613, m_{\text{wa}}(A_7) = 0.0056$
$T_{\text{uo}},$ $\beta = 0.2$	$f_1$	$m_{\text{uo}}(A_1) = 0.0289, m_{\text{uo}}(A_2) = 0.5637, m_{\text{uo}}(A_4) = 0.2382, m_{\text{uo}}(A_5) = 0.1157, m_{\text{uo}}(A_6) = 0.0805$
	$f_2$	$m_{\text{uo}}(A_3) = 0.0223, m_{\text{uo}}(A_4) = 0.3185, m_{\text{uo}}(A_5) = 0.0393, m_{\text{uo}}(A_6) = 0.486, m_{\text{uo}}(A_7) = 0.1339$
	$f_3$	$m_{\text{uo}}(A_2) = 0.7692, m_{\text{uo}}(A_3) = 0.1383, m_{\text{uo}}(A_5) = 0.0002, m_{\text{uo}}(A_6) = 0.0923$
	$f_4$	$m_{\text{uo}}(A_2) = 0.6481, m_{\text{uo}}(A_3) = 0.0421, m_{\text{uo}}(A_4) = 0.1506, m_{\text{uo}}(A_6) = 0.1592$

Table XI  
COMBINED BBAS IN TABLE X.

Approach	BBA
$T_{\text{norm}}$	$m_{\text{norm}}(A_2) = 1$
$T_{\alpha\text{-cut}}$	$m_{\alpha\text{-cut}}(A_2) = 1$
$T_{\text{tri}}$	$m_{\text{tri}}(A_2) = 1$
$T_{\text{min}}$	$m_{\text{min}}(A_2) = 0.9816, m_{\text{min}}(A_4) = 0.0176, m_{\text{min}}(A_6) = 0.0008$
$T_{\text{max}}$	$m_{\text{max}}(A_1) = 0.0054, m_{\text{max}}(A_2) = 0.8520, m_{\text{max}}(A_4) = 0.1426$
$T_{\text{wa}}$	$m_{\text{wa}}(A_1) = 0.0006, m_{\text{wa}}(A_2) = 0.9621, m_{\text{wa}}(A_4) = 0.0363, m_{\text{wa}}(A_6) = 0.0001, m_{\text{wa}}(A_7) = 0.0009$
$T_{\text{uo}}$	$m_{\text{uo}}(A_1) = 0.0005, m_{\text{uo}}(A_2) = 0.9646, m_{\text{uo}}(A_4) = 0.0328, m_{\text{uo}}(A_6) = 0.0021$

Table XII  
PIGNISTIC PROBABILITIES IN EXAMPLE 8.

Approach	BetP( $\theta_1$ )	BetP( $\theta_2$ )	BetP( $\theta_3$ )
$T_{\text{norm}}$	0	1	0
$T_{\alpha\text{-cut}}$	0	1	0
$T_{\text{tri}}$	0	1	0
$T_{\text{min}}$	0	0.9820	0.0180
$T_{\text{max}}$	0.0054	0.8520	0.1426
$T_{\text{wa}}$	0.0006	0.9626	0.0368
$T_{\text{uo}}$	0.0005	0.9656	0.0339

Table XIII  
AVERAGE CLASSIFICATION ACCURACY (%).

Approach	Iris	Wheat	Wine
$T_{\text{norm}}$	87.92 ± 4.79	84.14 ± 4.80	70.88 ± 5.08
$T_{\alpha\text{-cut}}$	87.92 ± 4.79	84.14 ± 4.80	70.88 ± 5.08
$T_{\text{iri}}$	88.19 ± 4.65	83.86 ± 4.73	70.47 ± 5.14
$T_{\text{min}}$	92.81 ± 4.31	87.60 ± 3.73	78.93 ± 4.04
$T_{\text{max}}$	92.16 ± 4.65	87.01 ± 3.70	77.81 ± 3.93
$T_{\text{wa}}, \beta = 0.2$	93.76 ± 3.49	89.29 ± 3.69	80.99 ± 3.96
$T_{\text{uo}}, \beta = 0.2$	93.83 ± 3.52	89.16 ± 3.73	81.10 ± 3.86
$T_{\text{wa}}, \beta = 0.5$	93.53 ± 3.75	88.73 ± 3.61	80.07 ± 3.94
$T_{\text{uo}}, \beta = 0.5$	93.54 ± 3.60	88.14 ± 3.78	80.04 ± 4.05
$T_{\text{wa}}, \beta = 0.8$	93.73 ± 4.31	88.70 ± 3.51	79.02 ± 4.01
$T_{\text{uo}}, \beta = 0.8$	93.31 ± 4.57	88.24 ± 3.88	79.69 ± 4.04

On each run, 70% of each class samples of iris dataset and wheat seeds dataset are for training samples, and the rest samples are for testing. For wine dataset, 34 samples (the class with the smallest samples contains 48 samples, 70% of which is about 34) of each class samples are for training samples and the rest samples are for testing. The training samples are select randomly. For moderate transformations, we specify  $\beta = 0.2, 0.5, 0.8$  to obtain the trade-off BBA, respectively. The results for average classification accuracy of three datasets are listed in Table XIII.

According to the results in Table XIII, there is a gap between the classification accuracy we obtained and the best possible classification accuracy for each dataset (e.g., for iris, the best possible accuracy with other classification approach can be beyond 95%). Here, we only aim to compare the impact of different BBA transformations on the classification performance.

All the results based on optimization-based transformations are better than those based on transformations with preselection of focal elements, i.e., considering more possible focal elements might reduce the loss of information due to the preselection of focal elements, thereby improving the classification accuracy. Meanwhile, the moderate transformations achieve higher classification accuracy than other transformations. The moderate transformations do not pursue the minimal or maximal degree of uncertainty on the basis of considering all possible focal elements, since the extreme attitudinal bias on the uncertainty degree might bring counter-intuitive results and a moderate (or balanced) BBA without the minimal or maximal degree of uncertainty is more natural.

Besides, we note two cases of samples of 300-run experiments:

- **Case 1:** the classification results of transformations based on uncertainty optimization are wrong and that of moderate transformations is correct.
- **Case 2:** the classification results of moderate transformations are wrong and that of transformations based on uncertainty optimization is correct.

In this example, the test sets of three datasets have 45, 63 and 76 samples, respectively. Here, we count the number of samples for Case 1 in each run experiment and calculate the average. The average numbers of samples of Case 1 are

2.2167 (4.93%), 1.8833 (3.00%) and 4.08 (5.37%) for three datasets, respectively, i.e., the moderate transformations can bring better results. We find that the samples belonging to Case 1 in each dataset contain at least one dimension feature with small difference in values of different classes. Compared with the transformations based on uncertainty optimization, the moderate transformations can better represent the uncertainty contained in the FMFs obtained according to the samples, e.g., the samples marked in Fig. 3 are the samples of Case 1 in iris dataset after 300-run experiments (repetitive samples are marked only once). In Fig. 3, samples of Class 1 are marked in red points; samples of Class 2 are marked in blue solid diamonds; samples of Class 3 are marked in cyan solid triangles. We use red circles to mark the samples of Case 1 of Class 1; we use blue diamonds to mark the samples of Case 1 of Class 2; we use cyan triangles to mark the samples of Case 1 of Class 3.

As we can see in Fig. 3(b), the samples of Class 1 can be clearly distinguished from the samples of the other two classes. However, the values of  $f_1$  and  $f_2$  of Class 1 have small difference with the other two classes. Taking a sample  $x = [7, 3.2, 4.7, 1.4]$  of Case 1 of iris dataset as an example, this test sample belongs to Class 2. The corresponding FMFs are:

$$\text{FMF}_1 : \mu_1(\theta_1) = 0, \mu_1(\theta_2) = 0, \mu_1(\theta_3) = 0.6618$$

$$\text{FMF}_2 : \mu_2(\theta_1) = 0.8119, \mu_2(\theta_2) = 0.3017, \mu_2(\theta_3) = 0.7095$$

$$\text{FMF}_3 : \mu_3(\theta_1) = 0, \mu_3(\theta_2) = 0.4389, \mu_3(\theta_3) = 0$$

$$\text{FMF}_4 : \mu_4(\theta_1) = 0, \mu_4(\theta_2) = 0.7349, \mu_4(\theta_3) = 0.0001$$

According to these four FMFs, we can transform them into BBAs, respectively. The results are listed in Table XIV.

One can combine the corresponding four BBAs of each transformation by using Dempster's rule of combination and then obtain the following results by using Eq. (6), yielding

$$\text{BetP}_{\text{min}}(\theta_1) = 0.2427$$

$$\text{BetP}_{\text{min}}(\theta_2) = 0.3426$$

$$\text{BetP}_{\text{min}}(\theta_3) = 0.4147$$

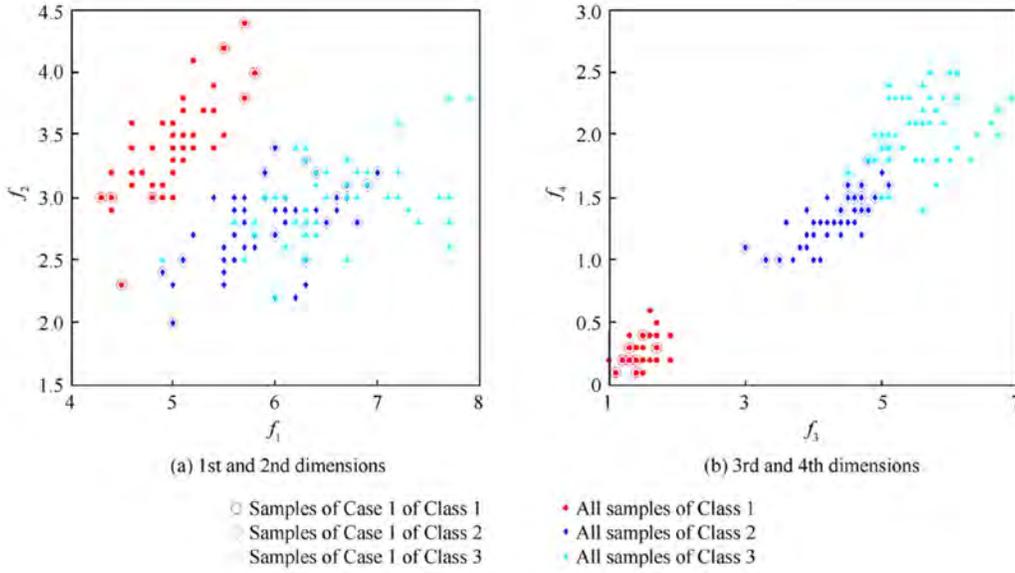


Figure 3. Samples of Case 1 of iris dataset.

 Table XIV  
 OBTAINED BBAs OF A SAMPLE OF CASE 1 IN EXAMPLE 8.

Approach	Feature	BBA
$T_{\min}$	$f_1$	$m_{\min}(A_4) = 0.6618, m_{\min}(A_5) = 0.2859, m_{\min}(A_6) = 0.0523$
	$f_2$	$m_{\min}(A_1) = 0.2906, m_{\min}(A_5) = 0.4077, m_{\min}(A_6) = 0.1881, m_{\min}(A_7) = 0.1136$
	$f_3$	$m_{\min}(A_2) = 0.4389, m_{\min}(A_3) = 0.3351, m_{\min}(A_6) = 0.2260$
	$f_4$	$m_{\min}(A_2) = 0.7394, m_{\min}(A_4) = 0.1417, m_{\min}(A_6) = 0.1189$
$T_{\max}$	$f_1$	$m_{\max}(A_3) = 0.3309, m_{\max}(A_4) = 0.6617, m_{\max}(A_7) = 0.0074$
	$f_2$	$m_{\max}(A_1) = 0.0941, m_{\max}(A_2) = 0.1329, m_{\max}(A_3) = 0.0636, m_{\max}(A_4) = 0.0553, m_{\max}(A_5) = 0.5489, m_{\max}(A_7) = 0.1053$
	$f_3$	$m_{\max}(A_2) = 0.4389, m_{\max}(A_5) = 0.2194, m_{\max}(A_6) = 0.3417$
	$f_4$	$m_{\max}(A_2) = 0.7395, m_{\max}(A_5) = 0.2605$
$T_{\text{wa}}, \beta = 0.5$	$f_1$	$m_{\text{wa}}(A_3) = 0.1654, m_{\text{wa}}(A_4) = 0.6618, m_{\text{wa}}(A_5) = 0.1430, m_{\text{wa}}(A_6) = 0.0261, m_{\text{wa}}(A_7) = 0.0037$
	$f_2$	$m_{\text{wa}}(A_1) = 0.1923, m_{\text{wa}}(A_2) = 0.0664, m_{\text{wa}}(A_3) = 0.0318, m_{\text{wa}}(A_4) = 0.0277, m_{\text{wa}}(A_5) = 0.4783, m_{\text{wa}}(A_6) = 0.0941, m_{\text{wa}}(A_7) = 0.1094$
	$f_3$	$m_{\text{wa}}(A_2) = 0.4389, m_{\text{wa}}(A_3) = 0.1676, m_{\text{wa}}(A_5) = 0.1097, m_{\text{wa}}(A_6) = 0.1130, m_{\text{wa}}(A_7) = 0.1708$
	$f_4$	$m_{\text{wa}}(A_2) = 0.7395, m_{\text{wa}}(A_3) = 0.0708, m_{\text{wa}}(A_5) = 0.1303, m_{\text{wa}}(A_6) = 0.0594$
$T_{\text{uo}}, \beta = 0.5$	$f_1$	$m_{\text{uo}}(A_3) = 0.1249, m_{\text{uo}}(A_4) = 0.6618, m_{\text{uo}}(A_5) = 0.0816, m_{\text{uo}}(A_6) = 0.1054, m_{\text{uo}}(A_7) = 0.0263$
	$f_2$	$m_{\text{uo}}(A_1) = 0.1852, m_{\text{uo}}(A_2) = 0.0002, m_{\text{uo}}(A_3) = 0.1051, m_{\text{uo}}(A_4) = 0.1615, m_{\text{uo}}(A_5) = 0.3516, m_{\text{uo}}(A_6) = 0.0264, m_{\text{uo}}(A_7) = 0.1700$
	$f_3$	$m_{\text{uo}}(A_2) = 0.4389, m_{\text{uo}}(A_3) = 0.1440, m_{\text{uo}}(A_5) = 0.1537, m_{\text{uo}}(A_6) = 0.1907, m_{\text{uo}}(A_7) = 0.0727$
	$f_4$	$m_{\text{uo}}(A_2) = 0.7394, m_{\text{uo}}(A_3) = 0.0574, m_{\text{uo}}(A_5) = 0.0973, m_{\text{uo}}(A_6) = 0.0032, m_{\text{uo}}(A_7) = 0.1027$

$$\text{BetP}_{\max}(\theta_1) = 0.2372$$

$$\text{BetP}_{\max}(\theta_2) = 0.3512$$

$$\text{BetP}_{\max}(\theta_3) = 0.4116$$

$$\text{BetP}_{\text{wa}}(\theta_1) = 0.1927$$

$$\text{BetP}_{\text{wa}}(\theta_2) = 0.4161$$

$$\text{BetP}_{\text{wa}}(\theta_3) = 0.3912$$

$$\text{BetP}_{\text{uo}}(\theta_1) = 0.1282$$

$$\text{BetP}_{\text{uo}}(\theta_2) = 0.4643$$

$$\text{BetP}_{\text{uo}}(\theta_3) = 0.4075$$

We can see that the classification results are correct except for the results obtained using  $T_{\min}$  and  $T_{\max}$ .

On the other hand, the average numbers of samples for Case 2 (the moderate transformations bring out incorrect results, while  $T_{\min}$  and  $T_{\max}$  bring out correct results) in 300-run experiment are 0.07 (0.16%), 0.2967 (0.47%) and 0.1433 (0.19%) for three datasets, respectively. Compared with the average numbers of samples for Case 1, moderate transformations are better than transformations based on uncertainty optimization. Using moderate transformations can avoid “one-sidedness” in terms of the uncertainty degrees. Meanwhile, the trade-off BBA can represent the small differences between the values of a given FMF. In summary, the moderate transformations can bring better classification results compared with that of

the transformations based on uncertainty optimization in a statistical sense.

## VI. CONCLUSIONS

In this paper, we propose two transformations with a weighting factor to transform a given FMF into a trade-off or moderate BBA. The weighting factor could be determined by using prior information or user-specified to reflect the objective situation or meet subjective preferences of users. Numerical examples and classification results validate the effectiveness of the two moderate transformations. Comparing these two transformations, the computational complexity of  $T_{wa}$  is lower, and  $T_{uo}$  can bring a better classification performance. In practical applications, users can choose  $T_{wa}$  or  $T_{uo}$  according to the demands of applications. Note that our transformations have been evaluated through some numerical and classification examples, within which, the design of numerical examples is usually subjective. In fact, the related fields of belief functions, including the generation of a BBA, lack objective and reasonable evaluation criteria. The conclusions obtained by numerical examples are incomplete. An objective evaluation criterion can help to obtain better related tools or approaches to deal with belief functions. Therefore, we will focus on the objective evaluation criteria of the belief functions in our future work. With the increase of FOD's cardinality, the possible focal elements in a BBA will grow exponentially, i.e., the unknown variables that need to be determined in our formulated optimization problem will grow exponentially. The exponential growth of computational complexity is a limitation of our transformations. In the future work, we will attempt to use more simple and effective approaches beside the optimization-based transformations to transform an FMF into a trade-off BBA.

## ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (No. 61671370), Postdoctoral Science Foundation of China (No. 2016M592790) and Postdoctoral Science Research Foundation of Shaanxi Province, China (No. 2016BSHEDZZ46).

## APPENDIX

Proof of the equivalence between Eq. (12) and Eq. (13) when  $\sum_{i=1}^n \mu(\theta_i) = 1$ .

Consider the FOD  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ . The given FMF is denoted by  $\boldsymbol{\mu} = [\mu(\theta_1), \mu(\theta_2), \dots, \mu(\theta_n)]$ . According to Eq. (12), the FMF is equivalent to a singleton plausibility and then we have the following  $n$  equations.

$$\begin{cases} m(\{\theta_1\}) + m(\{\theta_1, \theta_2\}) + \dots + m(\{\theta_1, \theta_n\}) + m(\{\theta_1, \theta_2, \theta_3\}) \\ + \dots + m(\{\theta_1, \theta_{n-1}, \theta_n\}) + \dots + m(\Theta) = \mu(\theta_1) \\ m(\{\theta_2\}) + m(\{\theta_1, \theta_2\}) + \dots + m(\{\theta_2, \theta_n\}) + m(\{\theta_1, \theta_2, \theta_3\}) \\ + \dots + m(\{\theta_2, \theta_{n-1}, \theta_n\}) + \dots + m(\Theta) = \mu(\theta_2) \\ \vdots \\ m(\{\theta_n\}) + m(\{\theta_1, \theta_2\}) + \dots + m(\{\theta_{n-1}, \theta_n\}) + m(\{\theta_1, \theta_2, \theta_n\}) \\ + \dots + m(\{\theta_{n-2}, \theta_{n-1}, \theta_n\}) + \dots + m(\Theta) = \mu(\theta_n) \end{cases} \quad (\text{A.1})$$

By adding the left and right sides of these  $n$  equations, then

$$\begin{aligned} & m(\{\theta_1\}) + m(\{\theta_2\}) + \dots + m(\{\theta_n\}) \\ & + 2[m(\{\theta_1, \theta_2\}) + \dots + m(\{\theta_{n-1}, \theta_n\})] \\ & \quad + 3[m(\{\theta_1, \theta_2, \theta_3\}) + \dots \\ & \quad + m(\{\theta_1, \theta_2, \theta_n\}) + \dots + m(\{\theta_{n-2}, \theta_{n-1}, \theta_n\})] \\ & + \dots + nm(\Theta) = \mu(\theta_1) + \mu(\theta_2) + \dots + \mu(\theta_n) = 1 \end{aligned} \quad (\text{A.2})$$

According to  $\sum_{A \subseteq \Theta} m(A) = 1$ , Eq. (A.2) can be rewritten as follows:

$$\begin{aligned} & m(\{\theta_1, \theta_2\}) + \dots + m(\{\theta_{n-1}, \theta_n\}) + 2[m(\{\theta_1, \theta_2, \theta_3\}) \\ & \quad + \dots + m(\{\theta_1, \theta_2, \theta_n\})] \\ & \quad + \dots + m(\{\theta_{n-2}, \theta_{n-1}, \theta_n\}) \\ & \quad + \dots + (n-1)m(\Theta) = 0 \end{aligned} \quad (\text{A.3})$$

Because  $0 \leq m(A) \leq 1$ , ( $A \subseteq \Theta$ ). This means that the focal elements at the left side of Eq. (A.3) are 0. Then we have

$$\begin{aligned} m(\{\theta_1, \theta_2\}) &= \dots = m(\{\theta_{n-1}, \theta_n\}) = m(\{\theta_1, \theta_2, \theta_3\}) \\ &= \dots = m(\{\theta_{n-2}, \theta_{n-1}, \theta_n\}) = \dots = m(\Theta) \\ &= 0 \end{aligned} \quad (\text{A.4})$$

Then, Eq. (A.1) can be rewritten as.

$$\begin{cases} m(\{\theta_1\}) = \mu(\theta_1) \\ m(\{\theta_2\}) = \mu(\theta_2) \\ \vdots \\ m(\{\theta_n\}) = \mu(\theta_n) \end{cases} \quad (\text{A.5})$$

which means that the FMF is also equivalent to both the singleton belief function and singleton plausibility function when  $\sum_{i=1}^n \mu(\theta_i) = 1$ .

## REFERENCES

- [1] L.A. Zadeh, *Fuzzy sets*, Inf. Control, Vol. 8(3), pp. 338–53, 1965.
- [2] L.A. Zadeh, *Fuzzy sets as a basis for a theory of possibility*, Fuzzy Sets Syst., Vol. 1(1), pp. 3–28, 1978.
- [3] Z. Pawlak, *Rough sets*, Int. J. Comput. Inf. Sci., Vol. 11(5), pp. 341–356, 1982.
- [4] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [5] M.C. Florea, A.-L. Jousselme, D. Grenier, É. Bossé, *Approximation techniques for the transformation of fuzzy sets into random sets*, Fuzzy Sets Syst., Vol. 159(3), pp. 270–288, 2008.

- [6] Y. Bi, D. Bell, H. Wang, G. Guo, K. Greer, *Combining multiple classifiers using Dempster's rule of combination for text categorization*, in Proc. of Int. Conf. on Modeling Decisions for Artificial Intelligence, pp. 127–38, Barcelona, Spain, Aug 2–4, 2004.
- [7] D. Han, C. Han, Y. Deng, *Novel approaches for the transformation of fuzzy membership function into basic probability assignment based on uncertainty optimization*, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 21(2), pp. 289–322, 2013.
- [8] Y. Yang, X.-R. Li, D. Han, *An improved  $\alpha$ -cut approach to transforming fuzzy membership function into basic belief assignment*, Chinese Journal of Aeronautics, Vol. 29(4), pp. 1042–1051, 2016.
- [9] Z. Li, F. Xiao, L. Fei, S. Mahadevan, Y. Deng, *An evidential failure mode and effects analysis using linguistic terms*, Quality and Reliability Engineering International, Vol. 33 (5), pp. 993–1010, 2017.
- [10] Y. Bi, D. Bell, H. Wang, G. Guo, J. Guan, *Combining multiple classifiers using Dempster's rule for text categorization*, Applied Artificial Intelligence, Vol. 21(3), pp. 211–239, 2007
- [11] X.J. Fan, D. Han, J. Dezert, Y. Yang, *User-specified optimization based transformation of fuzzy membership into basic belief assignment*, in Proc. of the 22th Int. Conf. on Information Fusion (Fusion 2019), Ottawa, Canada, July 2–5, 2019.
- [12] P. Smets, *Analyzing the combination of conflicting belief functions*, Information Fusion, Vol. 8(4), pp. 387–412, 2007
- [13] T.L. Yao, R. Miao, W. Wang, Z. Li, J. Dong, Y. Gu, X. Yan, *Synthetic damage effect assessment through evidential reasoning approach and neural fuzzy inference: application in ship target*, Chinese Journal of Aeronautics, Vol. 35 (4), pp. 362–375, 2022.
- [14] H.P. Zhang, Y. Deng, *Weighted belief function of sensor data fusion in engine fault diagnosis*, Soft computing, Vol. 24(3), pp. 2329–2339, 2020.
- [15] K. Guo, L. Zhang, *Multi-source information fusion for safety risk assessment in underground tunnels*, Knowledge-Based Systems, Vol.227, Sept. 2021.
- [16] L. Chen, Z. Zhou, C. Hu, R. Yue, Z. Feng, *Performance evaluation of complex systems using evidential reasoning approach with uncertain parameters*, Chinese Journal of Aeronautics, Vol. 34(1), pp. 194–208, 2021.
- [17] P. Liu, X. Zhang, *A new hesitant fuzzy linguistic approach for multiple attribute decision making based on Dempster-Shafer evidence theory*, Applied Soft Computing, Vol. 86, Jan. 2020.
- [18] F. Xiao, *EFMCDM: Evidential fuzzy multicriteria decision making based on belief entropy*, IEEE Trans. on Fuzzy Systems, Vol. 28 (7), pp. 1477–1491, 2020.
- [19] I. Ullah, J. Youn, Y.H. Han, *Multisensor data fusion based on modified belief entropy in Dempster-Shafer theory for smart environment*, IEEE Access, Vol. 9, pp. 37813–37822, March 2021.
- [20] Z. Liu, Q. Pan, J. Dezert, J.-W. Han, Y. He, *Classifier fusion with contextual reliability evaluation*, IEEE Trans. on Cybernetics, Vol. 48(5), pp. 1605–1618, May 2018.
- [21] Z. Liu, Y. Liu, J. Dezert, F. Cuzzolin, *Evidence combination based on credal belief redistribution for pattern classification*, IEEE Trans. on Fuzzy Systems, Vol. 28(4), pp. 618–631, 2020.
- [22] Z. Tong, P. Xu, T. Deneux, *An evidential classifier based on Dempster-Shafer theory and deep learning*, Neurocomputing, Vol. 450, pp. 275–293, August 2021.
- [23] J. Li, W. Wang, X. Gao, L. Zhou, *Strong conflicting evidences fusion based on belief interval distance measurement*, Cluster Computing, Vol. 22(S3), pp. 6589–6598, 2019.
- [24] X. Deng, D. Han, J. Dezert, Y. Deng, Y. Shyr, *Evidence combination from an evolutionary game theory perspective*, IEEE Trans. on Cybernetics, Vol.46(9), pp. 2070–2082, 2016.
- [25] D. Han, Y. Deng, C. Han, *Sequential weighted combination for unreliable evidence based on evidence variance*, Decision Support Systems, Vol. 56, pp. 387–393, 2013.
- [26] D. Dubois, H. Prade, *Representation and combination of uncertainty with belief functions and possibility measures*, Computational intelligence, Vol. 4(3), pp. 244–264, 1988.
- [27] R. Ilin, E. Blasch, *Information fusion with belief functions: a comparison of proportional conflict redistribution PCR5 and PCR6 rules for networked sensors*, in Proc. of the 18th Int. Conf. on Information Fusion (Fusion 2015), pp. 2084–2091, Washington, USA, July 6–9, 2015.
- [28] F. Smarandache, J. Dezert, *On the consistency of PCR6 with the averaging rule and its application to probability estimation*, in Proc. of the 16th Int. Conf. on Information Fusion (Fusion 2013), pp. 1119–1126, Istanbul, Turkey, July 9–12, 2013.
- [29] H. Xu, Y. Deng, *Dependent evidence combination based on decision-making trial and evaluation laboratory method*, Int. Journal of Intelligent Systems, Vol. 34(7), pp. 1555–1571, July 2019.
- [30] J. Wang, K. Qiao, Z. Zhang, *An improvement for combination rule in evidence theory*, Future Generation Computer Systems, Vol. 91, pp. 1–9, Feb; 2019.
- [31] A. Bronevich, G.J. Klir, *Measures of uncertainty for imprecise probabilities: an axiomatic approach* Int. J. of Approximate Reasoning, Vol. 51(4), pp. 365–390, 2010.
- [32] R.R. Yager, *Entropy and specificity in a mathematical theory of evidence*, International Journal of General System, Vol. 9(4), pp. 249–260, 1983.
- [33] D. Dubois, H. Prade, *A note on measures of specificity for fuzzy sets*, International Journal of General System, Vol. 10(4), pp. 279–283, 1985.
- [34] A.-L. Jousselme, C. Liu, D. Grenier, E. Bossé, *Measuring ambiguity in the evidence theory*, IEEE Trans. on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 36(5), pp. 890–903, 2006.
- [35] D. Harmanec, G.J. Klir, *Measuring total uncertainty in Dempster-Shafer theory: A novel approach*, International Journal of General System, Vol. 22(4), pp. 405–419, 1994.
- [36] Y. Yang, D. Han, *A new distance-based total uncertainty measure in the theory of belief functions*, Knowledge-Based Systems, Vol. 94, pp. 114–123, Feb. 2016.
- [37] X. Gao, F. Liu, L. Pan, Y. Deng, S.-B. Tsai, *Uncertainty measure based on Tsallis entropy in evidence theory*, Int. J. of Intelligent Systems, Vol. 34(11), pp. 3105–3120, 2019.
- [38] F. Xiao, *A new divergence measure for belief functions in D-S evidence theory for multisensor data fusion*, Information Sciences, Vol. 514, pp. 462–483, April 2020.
- [39] P. Smets, R. Kennes, *The transferable belief model*, Artificial Intelligence, Vol. 66(2), pp. 191–234, 1994.
- [40] A. Irpino, R. Verde, *Dynamic clustering of interval data using a Wasserstein-based distance*, Pattern Recognition Letters, Vol. 29(11), pp. 1648–1658, 2008.
- [41] A.-L. Jousselme, D. Grenier, É. Bossé, *A new distance between two bodies of evidence*, Information Fusion, Vol. 2(2), pp. 91–101, 2001.
- [42] A.-L. Jousselme, P. Maupin, *Distances in evidence theory: comprehensive survey and generalizations*, Int. J. of Approximate Reasoning, Vol. 53(2), pp. 118–145, 2012.
- [43] J. Zhu, X. Wang, Y. Song, *A new distance between BPAs based on the power-set-distribution pignistic probability function*, Applied Intelligence, Vol. 48(6), pp. 1506–1518, 2018.
- [44] C. Cheng, F. Xiao, *A new distance measure of belief function in evidence theory*, IEEE Access, Vol. 7, pp. 68607–68617, 2019
- [45] B. Tessem, *Approximations for efficient computation in the theory of evidence*, Artificial Intelligence, Vol. 61(2), pp. 315–329, 1993.
- [46] D. Han, J. Dezert, C. Han, Y. Yang, *New dissimilarity measures in evidence theory*, in Proc. of the 14th Int. Conf. on Information Fusion (Fusion 2011), Chicago, USA, July 5–8, 2011.
- [47] D. Han, J. Dezert, Y. Yang, *Belief interval-based distance measures in the theory of belief functions*, IEEE Trans. on Systems, Man, and Cybernetics: Systems, Vol. 48(6), pp. 833–850, 2018.
- [48] X. Liu, *Entropy, distance measure and similarity measure of fuzzy sets and their relations*, Fuzzy Sets and Systems, Vol. 52(3), pp. 305–318, 1992
- [49] L.A. Zadeh, *A note on Z-numbers*, Information sciences, Vol. 181(14), pp. 2923–2932, 2011.
- [50] K. Atanassov, *Intuitionistic fuzzy sets*, Fuzzy Sets and Systems, Vol. 20(1), pp. 87–96, 1986.
- [51] R.R. Yager, A.M. Abbasov, *Pythagorean membership grades, complex numbers, and decision making*, Int. J. of Intelligent Systems, Vol. 28(5), pp. 436–452, 2013.
- [52] J. Kampé de Fériet, *Interpretation of fuzzy membership functions of fuzzy sets in terms of plausibility and belief*, Fuzzy Information and Decision Processes, pp. 93–98, edited by M.M. Gupta and E. Sanchez, North-Holland, Amsterdam, 1982.
- [53] UCI repository of machine learning databases [Internet]. Available from: <http://archive.ics.uci.edu/ml/index.php>.



# Cross-Entropy and Relative Entropy of Basic Belief Assignments

Jean Dezert, Frédéric Dambreville  
Department of Information Processing and Systems  
The French Aerospace Lab - ONERA  
Palaiseau, France.

Emails: jean.dezert@onera.fr, frederic.dambreville@onera.fr

Originally published as: J. Dezert, F. Dambreville, *Cross-Entropy and Relative Entropy of Basic Belief Assignments*, in Proc. of Int. Conf. on Information Fusion (Fusion 2023), Charleston, SC, USA, June 2023, and reprinted with permission.

**Abstract**—This paper introduces the concept of cross-entropy and relative entropy of two basic belief assignments. It is based on the new entropy measure presented recently. We prove that the cross-entropy satisfies a generalized Gibbs-alike inequality from which a generalized Kullback-Leibler divergence measure can be established in the framework of belief functions. We show on a simple illustrating example how these concepts can be used for decision-making under uncertainty.

**Keywords:** generalized entropy, cross-entropy, relative entropy, Kullback-Leibler divergence, belief functions

## I. INTRODUCTION

In Shannon’s theory of communication developed in 1948 [1], [2], the measure of uncertainty (MoU), also called entropy, for characterizing a source of information (from signal transmission standpoint) is defined by Shannon entropy. This entropy measures the randomness of a probability distribution  $P$  and is usually noted by  $H(P)$ . Shannon entropy does not concern the semantic aspects of the content of a message but only its transmission [3]–[5].  $H(P)$  has played a very important role in the development of modern communication systems and cryptography [6] until today. According to Cover and Thomas [7], the cross-entropy denoted by  $H(P, Q)$  is the average number of bits needed to encode data coming from a source with a probability distribution  $P$  when we use a distribution model  $Q$  to define our codebook. Cross-entropy is commonly used in machine learning as a loss function [8], and the cross-entropy method is often used in practice to estimate an unknown true pmf (probability mass function) based on a test set where  $Q$  is the assumed (or eventually empirical) pmf model. The minimization of the cross-entropy is related with the principle of the maximization of the likelihood. That is why cross-entropy plays a major role in many statistical applications. The relative entropy, often referred as Kullback-Leibler divergence [9], is the difference between the cross-entropy and Shannon entropy, and so it is  $H(P, Q) - H(P)$ . All these aforementioned basic concepts have been well established (and strongly justified) from the mid of 20th century, and all use the theory of probability as the fundamental underlying mathematical framework.

In this paper we go beyond the classical probabilistic framework because we want to work possibly with epistemic uncertainty represented by non-probabilistic models thanks to the mathematical framework of belief functions introduced by

Shafer [10], and in this context the legitimate and important question is to know if it is possible, or not, to extend the concepts of entropy, cross-entropy and relative entropy for the belief functions. Concerning the concept of entropy, the answer is affirmative and very recently a new generalized entropy measure has been proposed in [11] in the framework of the theory of belief functions. Concerning the second and third theoretical questions about cross-entropy and relative entropy concepts, we give new comprehensive and better answers to these questions in this paper. This is our new theoretical contribution in the field. The concrete meaning of relative entropy and cross-entropy measures in the belief functions framework is a challenging question because the entropy of belief function is merely related to the uncertainty of epistemic knowledge rather than of statistical knowledge. No concrete meaning of these notions has been firmly established so far. This interesting open question is left for future research works.

To make the material of this paper quite self-contained, we recall the basic classical concepts related to entropy (Shannon entropy, cross-entropy, and relative entropy) in the section II, and we present the basics of belief functions [10] in Section III with the new concept of entropy measure of basic belief mass assignment (BBA) [11] in the section IV. After recalling a very recent definition of cross-entropy of BBAs [12] based on the non effective Deng’s entropy definition [13], we present in the section V a new cross-entropy definition based on our new effective entropy definition. The section VI presents the concept of relative entropy of BBAs which can be interpreted as a generalization of the Kullback-Leibler divergence measure for belief functions. An example of the use of these concepts for decision-making under uncertainty is given in the section VII. Concluding remarks and perspectives are given in the section VIII.

## II. CLASSICAL NOTIONS RELATED TO ENTROPY

### A. Shannon entropy

Consider a discrete random variable  $\theta$  represented by a probability mass function (pmf)  $P_N = (p_1, p_2, \dots, p_N)$ , where  $p_i = P(\theta_i)$  is the probability of the  $i$ -th state  $\theta_i$  (i.e. outcome) of  $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ . Shannon was interested in communication systems where the various events were the carriers of coded messages, and he did justify his entropy measure as appropriate measure of average uncertainty (or

measure of randomness) of a random variable [1], [2], [6], [7]. The entropy of a random variable is the average level of *surprisal*, or *uncertainty* inherent in the variable's possible outcomes [14]. Shannon entropy is defined by<sup>1</sup>

$$H(P_N) \triangleq - \sum_{i=1}^{|\Theta|} P(\theta_i) \log(P(\theta_i)). \quad (1)$$

By convention,  $P(\theta_i) \log(P(\theta_i)) = 0$  if  $P(\theta_i) = 0$ . This is easily justified by continuity because  $\lim_{x \rightarrow 0^+} x \log x = 0$ , which can be proved using L'Hôpital rule [15]. Adding terms of zero probability does not change the entropy value. In (1) we use the natural logarithm (i.e. base  $e$  logarithm) and in this case the Shannon entropy value is expressed in *nats* unity. We can also use the base 2 logarithm ( $\log_2$ ) function instead of the natural logarithm, and if so the Shannon entropy value will be expressed in *bits*. Shannon entropy can be interpreted as a generalization of Hartley entropy (1928) [16] when presuming the pmf of equally probable states (i.e. uniform<sup>2</sup> pmf  $P_N^{\text{unif}}$ ), hence getting  $H(P_N^{\text{unif}}) = \log(|\Theta|) = \log(N)$ . Note that if we have a uniform pmf  $P_N^{\text{unif}}$  defined on  $\Theta$  with  $|\Theta| = N$  and another uniform pmf  $P_{N'}^{\text{unif}}$  defined on  $\Theta'$  with  $|\Theta'| = N'$ , and if  $|\Theta| < |\Theta'|$  then  $H(P_N^{\text{unif}}) < H(P_{N'}^{\text{unif}})$  because  $\log(|\Theta|) < \log(|\Theta'|)$  since  $\log(x)$  is an increasing function. The minimum value of Shannon entropy is zero, which characterizes a *non-random* (or *sure*) event  $\theta_j$  for which  $P(\theta_j) = 1$ .

The main algebraic properties of Shannon entropy are, see [17] (p. 30) for details: the symmetry, the normality<sup>3</sup>, expansibility, decisivity, sub-additivity and recursivity. We recall that Shannon entropy value  $H(P_N)$  is always smaller than  $H(P_N^{\text{unif}})$  if  $P_N \neq P_N^{\text{unif}}$ , expressing the fact that the uniform pmf is the only pmf giving the maximal Shannon entropy value, and characterizing the maximum of uncertainty (or randomness), which is called the maximality property.

### B. The cross-entropy

Consider a finite set of exhaustive events  $\Theta = \{\theta_1, \dots, \theta_n\}$  where  $\theta_i$  are mutually exclusive (i.e.  $\theta_i \cap \theta_j = \emptyset$  if  $i \neq j$ ). Suppose that  $P = \{P(\theta_1) = p_1, \dots, P(\theta_n) = p_n\}$  is a probability distribution over the set  $\Theta$ . Then for any other probability distribution  $Q = \{Q(\theta_1) = q_1, \dots, Q(\theta_n) = q_n\}$  the Gibbs inequality holds [18]

$$- \sum_{i=1}^n p_i \log(q_i) \geq - \sum_{i=1}^n p_i \log(p_i). \quad (2)$$

The cross-entropy between probability distributions  $P$  and  $Q$  over the same underlying set of events  $\Theta$  is defined by

$$H(P, Q) = - \sum_{i=1}^n p_i \log(q_i) = - \sum_{X \in \Theta} P(X) \log(Q(X)). \quad (3)$$

One can easily verify that  $H(P, Q) = H(P)$  when  $Q = P$ , i.e. when the probability distribution  $Q$  coincides with the

true probability distribution  $P$  the cross-entropy value equals Shannon entropy of  $P$ . Gibbs inequality is  $H(P, Q) \geq H(P)$ .

### C. The relative entropy

The difference between the cross-entropy  $H(P, Q)$  and Shannon entropy  $H(P)$  is named the relative entropy or the Kullback-Leibler (KL) divergence [9]), and is often denoted by<sup>4</sup>

$$D_{\text{KL}}(P \parallel Q) \triangleq H(P, Q) - H(P) = \sum_{i=1}^n p_i \log(p_i/q_i). \quad (4)$$

$D_{\text{KL}}(P \parallel Q)$  measures how the probability distribution  $P$  is different from a second, reference probability distribution  $Q$ . It corresponds to the expectation of the logarithmic difference between the probability distributions  $P$  and  $Q$ , where the expectation is taken using the distribution  $P$ . In general the relative entropy  $D_{\text{KL}}(P \parallel Q)$  is not symmetric under interchange of the distributions  $P$  and  $Q$  and we have  $D_{\text{KL}}(P \parallel Q) \neq D_{\text{KL}}(Q \parallel P)$ . Therefore,  $D_{\text{KL}}$  is not strictly a distance even if it is often abusively called a *distance* in the literature, even by Cover in [7]. This relative entropy (i.e. divergence measure) is important in pattern recognition and neural networks for making classification, as well as in information theory. Kullback and Leibler also proposed a symmetrized measure in [9] defined as  $D_{\text{KL}}(P \parallel Q) + D_{\text{KL}}(Q \parallel P)$ . Another renown symmetric version of the KL divergence is the Jensen-Shannon (JS) divergence defined by Lin in [19]

$$D_{\text{JS}}(P \parallel Q) \triangleq \frac{1}{2} D_{\text{KL}} \left( P \parallel \frac{P+Q}{2} \right) + \frac{1}{2} D_{\text{KL}} \left( Q \parallel \frac{P+Q}{2} \right). \quad (5)$$

The Jensen-Shannon divergence can be interpreted as the total Kullback-Leibler divergence to the average probability distribution  $(P + Q)/2$ . This JS divergence is often used in practice because its square root is a metric often referred to as Jensen-Shannon distance [20], that is

$$d_{\text{JS}}(P, Q) \triangleq \sqrt{D_{\text{JS}}(P \parallel Q)}. \quad (6)$$

Jensen-Shannon divergence has been applied in different fields of applications (e.g. bioinformatics, social sciences, fire experiments, machine learning, in deep learning for studying generative adversarial networks, etc), see [21].

## III. BELIEF FUNCTIONS

The belief functions (BF) were introduced by Shafer [10] for modeling epistemic uncertainty, reasoning about uncertainty and combining distinct sources of evidence. The answer of the problem under concern is assumed to belong to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \dots, \theta_N\}$  where all elements (i.e. members) of  $\Theta$  are exhaustive and exclusive. The set of all subsets of  $\Theta$  (including empty set

<sup>1</sup>The symbol  $\triangleq$  means *equal by definition*.

<sup>2</sup>for which  $P(\theta_i) = 1/N$  for  $i = 1, 2, \dots, N$ .

<sup>3</sup>This stipulates that  $H(P_2^{\text{unif}}) = 1$  using base 2 logarithm function in (1).

<sup>4</sup>As in [7] (p. 19), in the formula (4) we use the conventions that  $0 \log(0/0) = 0$ ,  $0 \log(0/q) = 0$ , and  $p \log(p/0) = \infty$ . So, if there is any  $X \in \Theta$  such that  $P(X) > 0$  and  $Q(X) = 0$ ,  $D_{\text{KL}}(P \parallel Q) = \infty$ .

$\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of the power-set is  $2^{|\Theta|}$ . A (normalized) basic belief assignment (BBA) associated with a given source of evidence is a mapping  $m^\Theta(\cdot) : 2^\Theta \rightarrow [0, 1]$  such that<sup>5</sup>  $m^\Theta(\emptyset) = 0$  and  $\sum_{X \in 2^\Theta} m^\Theta(X) = 1$ . A BBA  $m^\Theta(\cdot)$  characterizes a source of evidence related with a FoD  $\Theta$ . For notation shorthand, we can omit the superscript  $\Theta$  in  $m^\Theta(\cdot)$  notation if there is no ambiguity on the FoD we work with. The quantity  $m(X)$  is called the mass of belief of  $X$ .  $X \in 2^\Theta$  is called a focal element (FE) of  $m(\cdot)$  if  $m(X) > 0$ . The set of all focal elements of  $m(\cdot)$  is denoted by  $\mathcal{F}_\Theta(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$ . The belief and the plausibility of  $X$  are respectively defined for any  $X \in 2^\Theta$  by [10]

$$Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y), \quad (7)$$

$$Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) = 1 - Bel(\bar{X}), \quad (8)$$

where  $\bar{X} \triangleq \Theta \setminus \{X\}$  is the complement of  $X$  in  $\Theta$ .

One has always  $0 \leq Bel(X) \leq Pl(X) \leq 1$ , see [10]. For  $X = \emptyset$ ,  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and for  $X = \Theta$  one has  $Bel(\Theta) = Pl(\Theta) = 1$ .  $Bel(X)$  and  $Pl(X)$  are often interpreted as the lower and upper bounds of unknown probability  $P(X)$  of  $X$ , that is  $Bel(X) \leq P(X) \leq Pl(X)$ . To quantify the uncertainty (i.e. the imprecision) of  $P(X) \in [Bel(X), Pl(X)]$ , we use  $u(X) \in [0, 1]$  defined by

$$u(X) \triangleq Pl(X) - Bel(X). \quad (9)$$

The quantity  $u(X) = 0$  if  $Bel(X) = Pl(X)$  which means that  $P(X)$  is known precisely, and one has  $P(X) = Bel(X) = Pl(X)$ . One has  $u(\emptyset) = 0$  because  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and one has  $u(\Theta) = 0$  because  $Bel(\Theta) = Pl(\Theta) = 1$ . If all focal elements of  $m(\cdot)$  are singletons of  $2^\Theta$  the BBA  $m(\cdot)$  is a Bayesian BBA because  $\forall X \in 2^\Theta$  one has  $Bel(X) = Pl(X) = P(X)$  and  $u(X) = 0$ . Hence the belief and plausibility of  $X$  coincide with a probability measure  $P(X)$  defined on the FoD  $\Theta$ . The vacuous BBA characterizing a totally ignorant source of evidence is defined by  $m_v(X) = 1$  for  $X = \Theta$ , and  $m_v(X) = 0$  for all  $X \in 2^\Theta$  different of  $\Theta$ . This very particular BBA plays a major role in the establishment of a new effective measure of uncertainty for BBA.

#### IV. ENTROPY OF BASIC BELIEF ASSIGNMENTS

In [22] we did analyze in details forty-five measures of uncertainty (MoU) of BBAs by covering 40 years of research works on this topic. Some of these MoUs capture only a particular aspect of the uncertainty inherent to a BBA (typically, the non-specificity and the conflict). Other MoUs propose a total uncertainty measure to capture jointly several aspects of the uncertainty. Unfortunately, most of these MoUs fail to satisfy four very simple reasonable and essential desiderata, and so they cannot be considered as really effective and

useful. Actually only five MoUs can be considered as effective from the mathematical sense presented next, but unfortunately they appear as conceptually defective and disputable, see discussions in [22]. That is why, a better effective measure of uncertainty (MoU), i.e. generalized entropy of BBAs has been developed and presented in [11]. The mathematical definition of this new effective entropy is given by

$$U(m) = \sum_{X \in 2^\Theta} s(X) \quad (10)$$

with

$$s(X) \triangleq -m(X)(1 - u(X)) \log(m(X)) + u(X)(1 - m(X)). \quad (11)$$

$s(X)$  is the uncertainty contribution related to  $X$  named the *entropiece* of  $X$ . This entropiece  $s(X)$  involves  $m(X)$  and the imprecision  $u(X) = Pl(X) - Bel(X)$  about the unknown probability of  $X$  in a subtle interwoven manner. Because  $u(X) \in [0, 1]$  and  $m(X) \in [0, 1]$  one has  $s(X) \geq 0$ , and  $U(m) \geq 0$ . The quantity  $U(m)$  is expressed in *nats* because we use the natural logarithm.  $U(m)$  can be expressed in *bits* by dividing the  $U(m)$  value in *nats* by  $\log(2) = 0.69314718\dots$ . This measure of uncertainty  $U(m)$  is a continuous function in its basic belief mass arguments because it is a summation of continuous functions. In formula (11), we always take  $m(X) \log(m(X)) = 0$  when  $m(X) = 0$  because  $\lim_{m(X) \rightarrow 0^+} m(X) \log(m(X)) = 0$ . Note that for any BBA  $m$ , one has  $s(\emptyset) = 0$  because  $m(\emptyset) = 0$  and  $u(\emptyset) = 0$ . For the vacuous BBA, one has  $s(\Theta) = 0$  because  $m_v(\Theta) = 1$  and  $u(\Theta) = 0$ .

This measure of uncertainty  $U(m)$  is effective because it can be proved (see proofs in [11]) that it satisfies the following four essential properties:

- 1)  $U(m) = 0$  for any BBA  $m(\cdot)$  focused on a singleton  $X$  of  $2^\Theta$ .
- 2)  $U(m_v^\Theta) < U(m_v^{\Theta'})$  if  $|\Theta| < |\Theta'|$ .
- 3)  $U(m) = -\sum_{X \in \Theta} m(X) \log(m(X))$  if the BBA  $m(\cdot)$  is a Bayesian BBA. Hence,  $U(m)$  reduces to Shannon entropy [1] in this case.
- 4)  $U(m) < U(m_v)$  for any non-vacuous BBA  $m(\cdot)$  and for the vacuous BBA  $m_v(\cdot)$  defined with respect to the same FoD.

The maximum of entropy value is obtained for the vacuous BBA  $m_v$  over a FoD  $\Theta$ , because  $m_v$  characterizes a source of evidence with a full lack of information. This maximum entropy value is  $U(m_v^\Theta) = 2^{|\Theta|} - 2$  (see derivation in [11]) and it represents the sum of all imprecisions of  $P(X)$  for all  $X \in 2^\Theta$ . Because for all  $X \in 2^\Theta \setminus \{\emptyset, \Theta\}$  one has  $u(X) = 1$  because  $[Bel(X), Pl(X)] = [0, 1]$ , and one has  $u(\emptyset) = 0$  and  $u(\Theta) = 0$  when considering the vacuous BBA then the sum of all imprecisions  $u(X)$  about  $P(X)$  is equal to  $2^{|\Theta|} - 2$ . It is worth mentioning that one has always  $U(m_v^\Theta) > \log(|\Theta|)$  which means that the vacuous BBA has always an entropy greater than the maximum of Shannon entropy  $\log(|\Theta|)$  obtained with the uniform pmf on  $\Theta$ .

<sup>5</sup>In Shafer's theory of BFs we work with a closed FoD and the mass of the empty set must always be equal to zero.

V. CROSS-ENTROPY OF TWO BBAS

A. Cross-entropy derived from Deng’s entropy

Very recently in [12], Gao et al. proposed a definition of the cross-entropy of two BBAs inspired by the non-effective Deng’s entropy  $E_d(m)$  proposed earlier by Deng in [13] and defined as follows:

$$E_d(m) = - \sum_{X \subseteq \Theta} m(X) \log\left(\frac{m(X)}{2^{|X|} - 1}\right) \quad (12)$$

where  $m(X)$  is the mass of belief of any subset  $X$  of the frame of discernment  $\Theta$ , and where  $|X|$  is the cardinality of  $X$ . If  $m(X) = 0$ , the term  $m(X) \log\left(\frac{m(X)}{2^{|X|} - 1}\right)$  is set to zero.

Deng’s entropy definition is unfortunately not recommended because it is non-effective. Indeed, we can have  $E_d(m) > E_d(m_v)$  indicating that a non-vacuous BBA  $m(\cdot)$  can be more uncertain than the vacuous BBA  $m_v(\cdot)$ , which obviously is not appropriate because the vacuous BBA characterizes the state of total ignorance. As a simple counterexample of Deng’s entropy consider  $\Theta = \{A, B, C\}$  the vacuous BBA  $m_v(\cdot)$  with  $m_v(A \cup B \cup C) = 1$ , and the non-vacuous BBA  $m(\cdot)$  with  $m(A \cup B) = m(A \cup C) = m(B \cup C) = 1/3$ . Clearly, one gets  $E_d(m) > E_d(m_v)$ . See the paper [22] for more discussions about other non-effective entropy proposals. For this counterexample, the values of Deng’s entropies are

$$\begin{aligned} E_d(m_v) &= -m_v(A \cup B \cup C) \log\left(\frac{m_v(A \cup B \cup C)}{2^{|A \cup B \cup C|} - 1}\right) \\ &= -1 \cdot \log\left(\frac{1}{2^3 - 1}\right) = -\log\left(\frac{1}{7}\right) \approx 1.9459, \end{aligned}$$

$$\begin{aligned} E_d(m) &= -m(A \cup B) \log\left(\frac{m(A \cup B)}{2^{|A \cup B|} - 1}\right) \\ &\quad - m(A \cup C) \log\left(\frac{m(A \cup C)}{2^{|A \cup C|} - 1}\right) \\ &\quad - m(B \cup C) \log\left(\frac{m(B \cup C)}{2^{|B \cup C|} - 1}\right) \\ &= -3 \cdot \frac{1}{3} \cdot \log\left(\frac{1}{3} \cdot \frac{1}{2^2 - 1}\right) = -\log(1/9) \approx 2.1972. \end{aligned}$$

Based on this non-effective entropy measure, the cross-entropy defined by Gao et al. [12] between BBAs  $m_1$  and  $m_2$  is based on a mimicry of the classical cross-entropy definition using Deng’s entropy, that is

$$C(m_1, m_2) = - \sum_{X \subseteq \Theta} m_1(X) \log\left(\frac{m_2(X)}{2^{|X|} - 1}\right).$$

Similarly, the cross-entropy between  $m_2$  and  $m_1$  is

$$C(m_2, m_1) = - \sum_{X \subseteq \Theta} m_2(X) \log\left(\frac{m_1(X)}{2^{|X|} - 1}\right).$$

Because Deng’s entropy is non effective, we have serious doubt on the validity of the cross-entropy concept defined by  $C(m_1, m_2)$  and  $C(m_2, m_1)$  formulas. This matter of fact justifies the necessity of using a better entropy measure [11] defined by (10)–(11), and the development of a better cross-entropy measure. This is what we present in the next section.

B. A new definition of cross-entropy

Based on the definition (3) of cross-entropy in the probabilistic framework, and the definition of the effective generalized entropy  $U(m)$  given in (10), it seems quite natural to try to extend directly the concept of cross-entropy of two pdfs  $p$  and  $q$  to the cross-entropy of two BBAs  $m_1$  and  $m_2$  defined over the same FoD  $\Theta$ . The extension of the classical cross-entropy formula (3) applied with generalized entropy  $U(m)$  given in (10) suggests directly the following generic formula of the cross-entropy between two BBAs

$$U(m_1, m_2) = \sum_{X \in 2^\Theta} s_{1,2}(X) \quad (13)$$

with

$$\begin{aligned} s_{1,2}(X) \triangleq & -m_1(X)(1 - u_i(X)) \log(m_2(X)) \\ & + u_j(X)(1 - m_k(X)) \end{aligned} \quad (14)$$

where indexes  $i, j$  and  $k$  have to belong to the set  $\{1, 2\}$ .

From this generic formulation, one sees that we could a priori define eight different cross-entropies between two BBAs depending on the choice of indexes  $(i, j, k)$  listed in Table I.

Table I  
POSSIBLE TRIPLETS  $(i, j, k)$ .

Triplet $T = (i, j, k)$	Value
$T_1$	(1,1,1)
$T_2$	(1,1,2)
$T_3$	(1,2,1)
$T_4$	(1,2,2)
$T_5$	(2,1,1)
$T_6$	(2,1,2)
$T_7$	(2,2,1)
$T_8$	(2,2,2)

It is worth mentioning that if  $m_2 = m_1$  the cross-entropy measure coincides with the entropy measure, that is  $U(m_1, m_2) = U(m_1, m_1) = U(m_1)$ .

What is the best definition of the cross-entropy of two BBAs among the eight possible definitions? Or equivalently, what is the most suitable triplet of indexes  $(i, j, k)$  to plug in the generic cross-entropy formula (14)? To answer to this important question, we propose to consider as the effective choice of triplet  $(i, j, k)$  the one which allows the information entropy of a BBA  $m_1$  to be less than or equal to its cross-entropy with any other BBA  $m_2$ . More precisely, select the triplet  $(i, j, k)$  such that for any BBAs  $m_1$  and  $m_2$  defined on the same FoD, the following inequality holds

$$U(m_1, m_2) \geq U(m_1). \quad (15)$$

Actually for the eight a priori possible definitions of cross-entropy drawn from (13)–(14), one can easily find by Monte-Carlo simulations of random pairs  $(m_1, m_2)$  of BBAs that the choices of triplets  $(1, 1, 1)$ ,  $(1, 2, 1)$ ,  $(1, 2, 2)$ ,  $(2, 1, 1)$ ,  $(2, 1, 2)$ ,  $(2, 2, 1)$ , and  $(2, 2, 2)$  are not judicious because the inequality (15) can be violated, see some examples in the appendix. Because our Monte-Carlo analysis based on 100000

random pairs  $(m_1, m_2)$  of BBAs revealed that the inequality (15) was satisfied only for the triplet  $(i, j, k) = (1, 1, 2)$  for different cardinalities of frames of discernment tested up to  $|\Theta| = 10$ , we did conjecture that the satisfactory definition of a cross-entropy of two BBAs satisfying inequality (15) is mathematically defined by (13) with

$$s_{1,2}(X) \triangleq -m_1(X)(1 - u_1(X)) \log(m_2(X)) + u_1(X)(1 - m_2(X)). \quad (16)$$

The term  $s_{1,2}(X)$  defined in (16) is called the *cross-entropiece* of  $X$ .

**Theorem 1:** Let  $m_1$  and  $m_2$  be BBAs defined on the same frame of discernment. The cross-entropy  $U(m_1, m_2)$  defined by (13) and (16) always satisfies the inequality  $U(m_1, m_2) \geq U(m_1)$ , with equality only if  $m_1 = m_2$ .

**Proof:** see appendix.

**Proposition:** If the BBAs  $m_1$  and  $m_2$  are Bayesian the cross-entropy defined by (13) and (16) coincides with the classical cross-entropy given by (3).

**Proof:** Since  $u_1(X) = 0$  for all  $X \in 2^\Theta$  for any Bayesian BBA  $m_1(\cdot)$ , the proposition is immediate.

## VI. RELATIVE ENTROPY OF TWO BBAS

It is worth mentioning that the inequality (15) is a generalization of the well-known Gibbs inequality (2), and it coincides with Gibbs inequality when the BBAs  $m_1$  and  $m_2$  are Bayesian BBAs. The generalized relative entropy (GRE) of two BBAs  $m_1$  and  $m_2$  that are defined over the same frame of discernment  $\Theta$  is naturally defined by

$$U(m_1 \parallel m_2) \triangleq U(m_1, m_2) - U(m_1). \quad (17)$$

Because Theorem 1 holds, one has always  $U(m_1 \parallel m_2) \geq 0$ , with equality if  $m_1$  equals  $m_2$ . As for the classical relative entropy defined by (4), the GRE is not symmetric under the interchange of the BBAs  $m_1$  and  $m_2$ , so that in general  $U(m_1 \parallel m_2) \neq U(m_2 \parallel m_1)$ . Therefore GRE must also not be considered as a distance. This GRE is a direct generalization of Kullback-Leibler (KL) divergence measure in the framework of belief functions. Using expressions (16) and<sup>6</sup> (11) the mathematical definition of  $U(m_1 \parallel m_2)$  is

$$U(m_1 \parallel m_2) = \sum_{X \subseteq \Theta} [m_1(X)(1 - u_1(X)) \cdot (\log(m_1(X)) - \log(m_2(X))) + u_1(X)(m_1(X) - m_2(X))]. \quad (18)$$

GRE coincides with *KL*-divergence formula (4) when the BBAs  $m_1$  and  $m_2$  are Bayesian because if focal elements of  $m_1$  and  $m_2$  are singletons of  $2^\Theta$  then  $u_1(X) = 0$  and

<sup>6</sup>with  $m$  replaced by  $m_1$ .

$$\begin{aligned} U(m_1 \parallel m_2) &= \sum_{X \subseteq \Theta} m_1(X)(\log(m_1(X)) - \log(m_2(X))) \\ &= \sum_{X \subseteq \Theta} m_1(X) \log \left( \frac{m_1(X)}{m_2(X)} \right) \end{aligned} \quad (19)$$

which is equivalent to formula (4) when interpreting the Bayesian BBA  $m_1$  as a probability measure  $p$ , and the Bayesian BBA  $m_2$  as a probability measure  $q$  over the set  $\Theta$ .

## VII. EXAMPLE OF APPLICATION

In this section we present an example of the use of the entropy, cross-entropy and relative entropy concepts defined in this paper for the purpose of decision-making under uncertainty. More precisely, given a BBA  $m(\cdot)$  defined over a FoD  $\Theta$ , how to make a decision based on  $m(\cdot)$  and how to select the most pertinent element  $\theta_i$  of  $\Theta$ ?

### A. Decision using relative entropy

Classically the decision-making from a BBA is based on the max of  $Pl(\cdot)$ , on the max of  $Bel(\cdot)$ , or on the max of pignistic probability depending on the attitude chosen by the decision-maker (resp. optimistic, pessimistic or in-between attitudes). Here we propose to make the decision based on the relative entropy measure. More precisely, from any BBA  $m$  defined over a FoD  $\Theta = \{\theta_i, i = 1, \dots, n\}$ , we calculate the divergences  $U(m_i \parallel m)$  for  $i = 1, 2, \dots, n$ , where  $m_i$  is the BBA focused on the element  $\theta_i \in \Theta$  such that  $m_i(\theta_i) = 1$ . We will take as decision  $\hat{\theta}$  the element  $\theta_i$  for which the divergence between  $m$  and  $m_i$  is minimal, that is  $\hat{\theta} = \theta_{i^*}$  with  $i^* = \arg \min_{i \in \{1, 2, \dots, n\}} U(m_i \parallel m)$ .

**Example:** Consider the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ , and after some fusion processing suppose we obtain the following BBA  $m(\cdot)$  defined by  $m(\theta_1) = 0.1$ ,  $m(\theta_2) = 0.2$ ,  $m(\theta_3) = 0.3$ ,  $m(\theta_1 \cup \theta_2) = 0.01$ ,  $m(\theta_1 \cup \theta_3) = 0.02$ ,  $m(\theta_2 \cup \theta_3) = 0.07$  and  $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.3$ . Then we get  $U(m_1 \parallel m) \approx 2.30$ ,  $U(m_2 \parallel m) \approx 1.60$  and  $U(m_3 \parallel m) \approx 1.20$ . Based on this result the decision will be  $\hat{\theta} = \theta_3$  because the divergence  $U(m_3 \parallel m) = 1.20$  is the least value among the values 2.30, 1.60 and 1.20. This decision is consistent with what we intuitively expect because  $[Bel(\theta_1), Pl(\theta_1)] = [0.10, 0.43]$ ,  $[Bel(\theta_2), Pl(\theta_2)] = [0.20, 0.58]$  and  $[Bel(\theta_3), Pl(\theta_3)] = [0.30, 0.69]$  showing that  $\theta_3$  is the element of  $\Theta$  that has the maximum of belief and also the maximum of plausibility.

**Remark 1:** We could not use  $U(m \parallel m_i)$  instead of  $U(m_i \parallel m)$ . Indeed, we get  $U(m \parallel m_i) = +\infty$ , and thus cannot decide. But the use of  $U(m_i \parallel m)$  is however not completely satisfactory because for  $m_i$  we have  $u_i(X) = 0$  for all  $X \in 2^\Theta$  and  $U(m_i) = 0$ , so that  $U(m_i \parallel m) = U(m_i, m) - U(m_i) = -\log(m(\theta_i))$ . Thus, the decision is made with only part of the information of  $m$  about  $\theta_i$  and not with the other mass values of non-singleton focal elements of  $m$  (if any). Subsequently, a pseudo-distance inspired by Jensen-Shannon is proposed which uses the whole BBA information.

### B. Decision using Jensen-Shannon pseudo-distance

In [23] we did propose a decision-making method based on the minimum of belief-interval distance that used Wasserstein distance. We take for decision  $\hat{\theta}$  the element  $\theta_i$  for which the distance  $d(m, m_i)$  between  $m$  and  $m_i$  is minimal, that is  $\hat{\theta} = \theta_{i^*}$  with  $i^* = \arg \min_{i \in \{1, 2, \dots, n\}} d(m, m_i)$ . This method implicitly assumes the uniform distribution of the probability  $P(X)$  in  $[Bel(X), Pl(X)]$  which is disputable because we cannot check in practice if this assumption is true, or not. To circumvent this problem, we propose to replace the belief-interval distance between BBAs by the Jensen-Shannon-like pseudo-distance derived from our relative entropy concept, which would be defined by

$$d(m, m') \triangleq \sqrt{\frac{1}{2} [U(m \parallel \frac{m+m'}{2}) + U(m' \parallel \frac{m+m'}{2})]}. \quad (20)$$

Note that  $d(m, m')$  coincides with Jensen-Shannon distance (6) when the BBAs  $m$  and  $m'$  are bayesian BBAs. One has also  $d(m, m') = d(m', m)$ ,  $d(m, m') \geq 0$  and  $d(m, m') = 0$  when  $m = m'$  because  $U(m \parallel \frac{m+m}{2}) = U(m \parallel m)$ , and  $U(m \parallel m) = U(m, m) - U(m) = 0$ .

In our example, we obtain the following pseudo-distances:  $d(m, m_1) \approx 0.67$ ,  $d(m, m_2) \approx 0.60$ , and  $d(m, m_3) \approx 0.55$ . Based on these values we will take the decision  $\hat{\theta} = \theta_3$ .

Note also that if  $m$  is the vacuous BBA (i.e.  $m = m_v$ ), then in this particular case we will obtain  $d(m_v, m_1) = d(m_v, m_2) = d(m_v, m_3) = 0.6656$  so that no clear decision can be drawn from the vacuous BBA since it does not contain useful information, which makes perfect sense. Note that the inequality  $(d(m, m_1) = 0.6763) > (d(m_v, m_1) = 0.6656)$  is not surprising because the BBA  $m$  is more unfavorable to  $\theta_1$  than the vacuous BBA  $m_v$  is.

**Remark 2:** We tested (20) against the triangular inequality  $d(m, m') + d(m', m'') \geq d(m, m'')$ . A crude Monte Carlo analysis based on millions of random BBAs generated uniformly over different frames of discernment up to cardinality  $|\Theta| = 13$  revealed no counterexample. This indicates that such counterexamples are rare events. However, we tried a refined Monte Carlo analysis, where the set of focal elements were generated prior to the BBA. On the basis of 10000 different generated BBAs and near  $500 \cdot 10^9$  combination cases, we have found a rate of  $2 \cdot 10^{-5}$  counterexamples to the triangular inequality. This is quite small. More interestingly, the degree of violation of the triangular inequality was small, since we found 1.17 as the maximum value for  $\frac{d(m, m'')}{d(m, m') + d(m', m'')}$ . That is why we consider  $d(m, m')$  only as a pseudo-distance, i.e. a semimetric. But our simulations suggest that this semimetric satisfies a sharp  $\rho$ -relaxed triangle inequality:

$$d(m, m'') \leq \rho(d(m, m') + d(m', m'')) \quad \text{with } \rho \geq 1.2.$$

In conclusion, the topology induced by this semimetric is certainly very close to a true metric topology.

### Counterexample of triangular inequality:

Consider  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the three BBAs  $m$ ,  $m'$  and  $m''$  as follows:

$$\begin{aligned} m(\theta_3) &= 0.25, m(\theta_1 \cup \theta_3) = 0.19, m(\theta_2 \cup \theta_3) = 0.21, m(\Theta) = 0.35, \\ m'(\theta_1 \cup \theta_3) &= 0.25, m'(\theta_2 \cup \theta_3) = 0.26, m'(\Theta) = 0.49, \\ m''(\theta_1 \cup \theta_2) &= 0.44, m''(\Theta) = 0.56. \end{aligned}$$

We get  $d(m, m') \approx 0.1144$ ,  $d(m', m'') \approx 0.1800$  and  $d(m, m'') \approx 0.3306$ . Hence  $d(m, m') + d(m', m'') = 0.2945$  which is smaller than  $d(m, m'') = 0.3306$ . So there, the triangular inequality  $d(m, m') + d(m', m'') \geq d(m, m'')$  is violated.

## VIII. CONCLUSION

In this paper we have proposed new measures of cross-entropy and relative entropy of two basic belief assignments based on the new effective measure of entropy of belief function presented in 2022. These new concepts are mathematically well-defined and are direct generalizations of their classical formulations drawn of the probabilistic framework. It is expected that these new theoretical concepts will become useful in some applications for decision-making under uncertainty. As research perspectives, we hope to improve them a bit more in order to provide a true Jensen-Shannon metric for belief functions in a near future. Also, applications of these new concepts are under development and they will be reported in future publications.

## APPENDIX

### A. Counterexamples of inequality (15)

We consider the FoD  $\Theta = \{A, B, C\}$  and we give BBAs<sup>7</sup>  $m_1(\cdot)$  and  $m_2(\cdot)$  such that inequality (15) is violated for the different choices of triplet  $(i, j, k)$  used in the formula (14).

- Consider  $(i, j, k) = (1, 1, 1)$  and the BBAs of Table II. We get  $U(m_1) = 3.9742$  and  $U(m_1, m_2) = 3.9432$ . The inequality (15) is violated because  $U(m_1) > U(m_1, m_2)$ .
- Consider  $(i, j, k) = (1, 2, 1)$  and the BBAs of Table III. We get  $U(m_1) = 3.7447$  and  $U(m_1, m_2) = 2.4995$ . The inequality (15) is violated because  $U(m_1) > U(m_1, m_2)$ .
- Consider  $(i, j, k) = (1, 2, 2)$  and the BBAs of Table IV. We get  $U(m_1) = 3.9568$  and  $U(m_1, m_2) = 2.5086$ . The inequality (15) is violated because  $U(m_1) > U(m_1, m_2)$ .
- Consider  $(i, j, k) = (2, 1, 1)$  and the BBAs of Table V. We get  $U(m_1) = 3.2115$  and  $U(m_1, m_2) = 2.8616$ . The inequality (15) is violated because  $U(m_1) > U(m_1, m_2)$ .
- Consider  $(i, j, k) = (2, 1, 2)$  and the BBAs of Table VI. We get  $U(m_1) = 2.5542$  and  $U(m_1, m_2) = 2.2147$ . The inequality (15) is violated because  $U(m_1) > U(m_1, m_2)$ .
- Consider  $(i, j, k) = (2, 2, 1)$  and the BBAs of Table VII. We get  $U(m_1) = 4.5243$  and  $U(m_1, m_2) = 3.8714$ . The inequality (15) is violated because  $U(m_1) > U(m_1, m_2)$ .

<sup>7</sup>The numerical values entering in the tables have been approximated to their fourth decimal for convenience.

- Consider  $(i, j, k) = (2, 2, 2)$  and the BBAs of Table VIII. We get  $U(m_1) = 3.8858$  and  $U(m_1, m_2) = 3.0406$ . The inequality (15) is violated because  $U(m_1) > U(m_1, m_2)$ .

Table II  
BBAS  $m_1(\cdot)$  AND  $m_2(\cdot)$ .

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.2094	0.1199
B	0.0537	0.0885
$A \cup B$	0.3016	0.2833
C	0.0054	0.0112
$A \cup C$	0.0713	0.0539
$B \cup C$	0.0712	0.0646
$A \cup B \cup C$	0.2874	0.3786

Table III  
BBAS  $m_1(\cdot)$  AND  $m_2(\cdot)$ .

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.1443	0.4612
B	0.0695	0.0657
$A \cup B$	0.0128	0.0291
C	0.0903	0.2119
$A \cup C$	0.2922	0.1056
$B \cup C$	0.2324	0.0305
$A \cup B \cup C$	0.1585	0.0960

Table IV  
BBAS  $m_1(\cdot)$  AND  $m_2(\cdot)$ .

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.1585	0.2677
B	0.0180	0.2017
$A \cup B$	0.2202	0.0566
C	0.0758	0.2432
$A \cup C$	0.1396	0.0120
$B \cup C$	0.1681	0.1051
$A \cup B \cup C$	0.2198	0.1137

Table V  
BBAS  $m_1(\cdot)$  AND  $m_2(\cdot)$ .

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.1718	0.0852
B	0.1045	0.0246
$A \cup B$	0.1721	0.0551
C	0.1329	0.0598
$A \cup C$	0.1721	0.2664
$B \cup C$	0.2078	0.2391
$A \cup B \cup C$	0.0388	0.2698

**B. Proof of the Theorem 1**

Subsequently, log is the natural logarithm function to the base of the mathematical Euler constant  $e$ . To prove the Theorem 1, we first prove the theorem 2 below.

**Theorem 2:** Let  $\mathcal{M}(\Theta)$  be the set of basic belief assignments over  $\Theta$ . Then:

$$\arg \min_{m \in \mathcal{M}(\Theta)} U(m_1, m) = \{m_1\}.$$

Table VI  
BBAS  $m_1(\cdot)$  AND  $m_2(\cdot)$ .

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.2228	0.0284
B	0.2112	0.0617
$A \cup B$	0.0767	0.3397
C	0.2523	0.0428
$A \cup C$	0.0726	0.1748
$B \cup C$	0.1196	0.2164
$A \cup B \cup C$	0.0448	0.1362

Table VII  
BBAS  $m_1(\cdot)$  AND  $m_2(\cdot)$ .

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.0443	0.1234
B	0.0892	0.2140
$A \cup B$	0.2486	0.1321
C	0.0473	0.0962
$A \cup C$	0.0562	0.2928
$B \cup C$	0.0803	0.0245
$A \cup B \cup C$	0.4341	0.1170

Table VIII  
BBAS  $m_1(\cdot)$  AND  $m_2(\cdot)$ .

Focal Elem.	$m_1(\cdot)$	$m_2(\cdot)$
A	0.0164	0.2007
B	0.0196	0.0967
$A \cup B$	0.2219	0.1484
C	0.0746	0.2518
$A \cup C$	0.1052	0.0996
$B \cup C$	0.4906	0.1538
$A \cup B \cup C$	0.0717	0.0490

**Proof:** Let  $F_1 \subset 2^\Theta \setminus \{\emptyset\}$  be the set of focal elements of  $m_1$ . First at all, it is noticed that  $u_1(X) < 1$  for all  $X \in F_1$ . Moreover, if there are  $X \in F_1$  and  $m \in \mathcal{M}(\Theta)$  such that  $m(X) = 0$ , then  $U(m_1, m) = +\infty$ . As a consequence, if  $m$  minimizes  $U(m_1, m)$ , then its set of focal elements contains the set of focal elements of  $m_1$ .

*Optimizations.* Let  $F$  be such that  $F_1 \subset F \subset 2^\Theta \setminus \{\emptyset\}$ . The proof is done by solving:

$$\min_{m: F \rightarrow \mathbb{R}_+^*} f(m) \tag{21}$$

under constraint

$$\sum_{X \in F} m(X) = 1 \tag{22}$$

where

$$f(m) = U(m_1, m) = \sum_{\emptyset \neq X \neq \Theta} u_1(X)(1 - m(X)) + \sum_{X \in F_1} -m_1(X)(1 - u_1(X)) \log(m(X)). \tag{23}$$

It is worth noting that  $m(X)$  is nothing but the component of index  $X$  of the unknown map vector  $m : F \rightarrow \mathbb{R}_+^*$ . The optimization (21) with equality constraint (22) could typically be solved by means of Lagrangian multiplier method.

Because  $\log(m(X))$  is a concave function of  $m(X)$ , the term  $-m_1(X)(1 - u_1(X)) \log(m(X))$  is proportional to  $-\log(m(X))$  and is a convex function of  $m(X)$ . And because  $u_1(X)(1 - m(X))$  is a linear function of  $m(X)$ , the term  $-m_1(X)(1 - u_1(X)) \log(m(X)) + u_1(X)(1 - m(X))$  is a convex function of  $m(X)$ . Therefore, the function  $f(m)$  is a convex function. We are then ensured that Lagrangian multiplier condition will point, if it is fulfilled, to the minima of the function.

Lagrangian multiplier is defined for this problem by:

$$L(m, \lambda) \triangleq \sum_{X \in F_1} -m_1(X)(1 - u_1(X)) \log(m(X)) + \sum_{\emptyset \neq X \neq \Theta} u_1(X)(1 - m(X)) + \lambda[1 - \sum_{X \in F} m(X)]. \quad (24)$$

The optimality conditions are:

$$D_{m(X)}L(m, \lambda) = 0 \quad \text{for all } X \in F.$$

Where  $D_{m(X)}L(m, \lambda)$  is the differential of  $L(m, \lambda)$  with respect to  $m(X)$  given by

$$D_{m(X)}L(m, \lambda) = \frac{-m_1(X)(1 - u_1(X))}{m(X)} - u_1(X) - \lambda,$$

for  $X \in F_1$ , and:

$$D_{m(X)}L(m, \lambda) = -u_1(X) - \lambda, \quad \text{for } X \in F \setminus F_1.$$

Then, the optimal solution for (21) is  $m^{\text{opt}}$  such that:

$$m^{\text{opt}}(X) = \frac{m_1(X)(1 - u_1(X))}{(-\lambda - u_1(X))}, \quad \text{for all } X \in F_1, \quad (25)$$

with  $\lambda$  chosen such that:

$$-\lambda = u_1(X) \quad \text{for all } X \in F \setminus F_1, \quad (26)$$

$$\sum_{X \in F \setminus F_1} m^{\text{opt}}(X) + \sum_{X \in F_1} \frac{m_1(X)(1 - u_1(X))}{(-\lambda - u_1(X))} = 1. \quad (27)$$

Noticed that (25) implies  $-\lambda - u_1(X) > 0$  for all  $X \in F_1$ .

**Case  $F \neq F_1$ :** Condition (26) implies  $-\lambda \leq 1$  and then:

$$\sum_{X \in F_1} \frac{m_1(X)(1 - u_1(X))}{(-\lambda - u_1(X))} \geq \sum_{X \in F_1} m_1(X) = 1.$$

Then by (27), it comes  $m^{\text{opt}}(X) = 0$  for  $X \in F \setminus F_1$ , which contradicts hypothesis that  $F$  is the set of focal elements of  $m$ . *There is no solution with more focal elements than  $m_1$ .*

**Case  $F = F_1$ :** Choice  $\lambda = -1$  is obvious. Therefore, *the*

*unique minimizer  $m^{\text{opt}} = m_1$  is obtained.*

**Conclusion.** It has been shown that minimizer of  $U(m_1, m)$  only exists if it has the same set of focal elements than  $m_1$ . Moreover, it is shown in that case that the only minimizer is  $m_1$ . As a consequence:

$$\arg \min_{m \in \mathcal{M}(\Theta)} U(m_1, m) = \{m_1\} \quad \text{for all } m \in \mathcal{M}(\Theta).$$

Because Theorem 2 holds, we have  $U(m_1, m) > U(m_1)$  when  $m \neq m_1$ , and  $U(m_1, m_1) = U(m_1)$  when  $m = m_1$ . Therefore,  $U(m_1, m) \geq U(m_1)$  for any BBA  $m \in \mathcal{M}(\Theta)$ . Thus the inequality (15) holds, with equality only if  $m_1 = m_2$ , which completes the proof of the Theorem 1.

## REFERENCES

- [1] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379-423 & 623-656, July & October 1948 (reprinted in [6]).
- [2] C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: The University of Illinois Press, 1-117, 1949.
- [3] M. Higashi, G.J. Klir, *Measures of uncertainty and information based on possibility distributions*, Int. J. Gen. Systems, Vol. 9, pp. 43-58, 1983.
- [4] O. Lombardi, F. Holik, L. Vanni, *What is Shannon information?*, Synthese 193, pp. 1983-2012, 2016.
- [5] O. Rioul, *This is IT: A Primer on Shannon's Entropy and Information*, L'Information, Séminaire Poincaré XXIII, 43-77, 2018.
- [6] N.J.A. Sloane, A.D. Wyner (Editors), *Claude Elwood Shannon - Collected Papers*, IEEE Press, 924 pages, 1993.
- [7] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Hoboken, New Jersey, Second Edition, 2006.
- [8] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
- [9] S. Kullback, R.A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics, Vol. 22, No. 1, pp. 79-86, 1951.
- [10] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [11] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, Fusion 2022 Conf., Linköping, Sweden, pp. 1-10, July, 2022.
- [12] X. Gao, L. Pan, Y. Deng, *Cross entropy of mass function and its application in similarity measure*, Applied Intelligence, Vol. 52, pp. 8337-8350, 2022.
- [13] Y. Deng, *Deng entropy*, Chaos, Solitons & Fractals, Vol. 91, pp. 549-553, Oct. 2016.
- [14] [https://en.wikipedia.org/wiki/Entropy\\_\(information\\_theory\)](https://en.wikipedia.org/wiki/Entropy_(information_theory))
- [15] R.E. Bradley, S.J. Petrilli, C.E. Sandifer, *L'Hôpital's analyse des infimiments petits (An annotated translation with source material by Johann Bernoulli)*, Birkhäuser, 311 pages, 2015.
- [16] R. Hartley, *Transmission of information*, Bell System Tech. J., Vol. 7, pp. 535-563, 1928.
- [17] J. Aczél, Z. Daróczy, *On Measures of Information and Their Characterizations*, Academic Press, 1975.
- [18] D.J.C. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.
- [19] J. Lin, *Divergence measures based on the Shannon entropy*, IEEE Trans. on Information Theory, Vol. 37, No. 1, pp. 145-151, 1991.
- [20] D.M. Endres, J.E. Schindelin, *A new metric for probability distributions*, IEEE Trans. Inf. Theory, Vol. 49, No. 7, pp. 1858-1860, July 2003.
- [21] [https://en.wikipedia.org/wiki/Jensen-Shannon\\_divergence](https://en.wikipedia.org/wiki/Jensen-Shannon_divergence)
- [22] J. Dezert, A. Tchamova, *On Effectiveness of Measures of Uncertainty of Basic Belief Assignments*, Information & Security Journal: An International Journal (ISIJ), Vol. 52, Feb. 2022.
- [23] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-Making with Belief Interval Distance*, in Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21-23, 2016.

# On Monotonicity Desideratum for an Efficient Entropy Measure of Basic Belief Assignments

Jean Dezert, Frédéric Dambreville  
 Department of Information Processing and Systems  
 The French Aerospace Lab - ONERA  
 Palaiseau, France.

Emails: jean.dezert@onera.fr, frederic.dambreville@onera.fr

**Abstract**—In this short paper we discuss the monotonicity desideratum for defining an efficient entropy measure of basic belief assignments. We browse some alternatives of the effective entropy measure developed recently, and we show that all these new alternatives for an entropy measure are not efficient. Only one appears to be quasi-efficient for the frame of discernment of dimension 2.

**Keywords:** entropy, cross-entropy, belief functions.

## I. INTRODUCTION

We assume that the formula of the entropy for basic belief assignment (BBA) is of the general form, see [1] for details

$$U(m) = \sum_{X \in 2^\Theta} s(X) \quad (1)$$

with

$$s(X) \triangleq -\alpha(u(X))m(X)\log(m(X)) + \beta(u(X), m(X)) \quad (2)$$

where  $u(X) = Pl(X) - Bel(X)$ ,  $\alpha(u(X)) \in [0, 1]$  is a weighting function of the surprisal  $-\log(m(X))$ , and  $\beta(u(X), m(X))$  is a function that must increase the entropy  $U(m)$  as soon as there is some imprecision on unknown probability  $P(X)$  of  $X$ .  $s(X)$  has been called the *entropiece* of subset  $X$ .

In [1], the function  $U(m)$  has been defined as *effective* if it satisfies the four natural following desiderata:

**D1:** For any non-empty frame of discernment  $\Theta$  and for any BBA  $m(\cdot)$  focused on a singleton  $X$  of  $2^\Theta$  one must have

$$U(m) = 0 \quad (3)$$

**D2:** The measure of uncertainty of a total ignorant source of evidence must increase with the cardinality of the frame of discernment. That is

$$U(m_v^\Theta) < U(m_v^{\Theta'}), \quad \text{if } |\Theta| < |\Theta'|. \quad (4)$$

**D3:** The measure of uncertainty  $U(m)$  must coincide with Shannon entropy [2]–[4] if the BBA  $m(\cdot)$  is a Bayesian

BBA. This desideratum is mathematically expressed for any Bayesian BBA  $m(\cdot)$  defined on the FoD  $\Theta$  by the condition<sup>1</sup>

$$U(m) = - \sum_{X \in \Theta} m(X) \log(m(X)) \quad (5)$$

**D4:** For any non-vacuous BBA  $m(\cdot)$  and for the vacuous BBA  $m_v(\cdot)$  defined with respect to the same FoD one must have

$$U(m) < U(m_v) \quad (6)$$

It has been proved in [1] that  $U(m)$  is effective in particular if one takes

$$\alpha(u(X)) = 1 - u(X) \quad (7)$$

$$\beta(u(X), m(X)) = u(X)(1 - m(X)) \quad (8)$$

There is unfortunately no unicity for the choice of  $\alpha(u(X))$  and  $\beta(u(X), m(X))$  functions, even if this particular choice has a quite simple interpretation. The interest of this choice is that it allows to define easily the cross-entropy  $U(m_1, m_2)$  of BBAs in a simple way satisfying the Gibbs-alike inequality [5]  $U(m_1, m_2) > U(m_1)$ . However this effective entropy formula is not entirely satisfying because the triangular inequality for Jensen-Shannon-alike distance can be violated in rare situations for some distributions of BBAs. This matter of fact motivates us to search for improved effective entropy formulas.

For this, we would like that the entropy satisfies a 5th desideratum D5 about the monotonicity of  $U(m)$ . More precisely, we want that a reduction of mass of  $X \subset Y$  transferred to its superset  $Y$  increases the entropy value, and we want that any reduction of mass of  $m(Y)$  transferred to one of its subset  $X$  decreases the entropy value. As a simple example, for  $\Theta = \{A, B\}$  if we consider the BBAs  $m(\cdot)$  and  $m_\epsilon(\cdot)$  defined by

$$\begin{cases} m(\emptyset) = 0 \\ m(A) = a \\ m(B) = b \\ m(A \cup B) = 1 - a - b \end{cases} \quad \text{and} \quad \begin{cases} m_\epsilon(\emptyset) = 0 \\ m_\epsilon(A) = \epsilon \cdot a \\ m_\epsilon(B) = b \\ m_\epsilon(A \cup B) = 1 - (\epsilon \cdot a) - b \end{cases}$$

with  $0 \leq \epsilon < 1$ . We would like to have  $U(m) < U(m_\epsilon)$  because the BBA  $m_\epsilon$  is less specific than  $U(m)$  because the

<sup>1</sup>Shannon entropy [2] is given here in *nats*, and we take  $0 \log(0) = 0$  because  $\lim_{x \rightarrow 0^+} x \log(x) = 0$  which is proved using L'Hôpital's rule.

mass of ambiguity (or disjunction)  $A \cup B$  for  $m_\epsilon$  is bigger than for the BBA  $m$ .

So we would like that the entropy of a BBA satisfies the extra desiderata D5 (i.e. the monotonicity desideratum) stated as follows:

**D5** (monotonicity):  $U(m)$  must increase when a mass of a proposition  $Y$  increases while the mass of one of its subset  $X \subset Y$  decreases by the same amount, and vice versa.

Note that D5 is equivalent to the mathematical condition

$$\frac{\partial U(m)}{\partial m(X)} < \frac{\partial U(m)}{\partial m(Y)}$$

for all  $X$  and  $Y$  such that  $m(X) > 0$  and  $X \subset Y$ .

It is worth mentioning that if  $U(m)$  satisfies D5 then it will satisfy D4, but the converse is unfortunately false for the entropy proposed in [1] because it can be easily shown based on a simple counter-example that the  $U(m)$  based on (7) and (8) (which satisfies D4) can occasionally violate D5.

**Example:** Consider  $\Theta = \{A, B\}$ ,  $\epsilon = 0.5$  and the BBAs

$$\begin{cases} m(\emptyset) = 0 \\ m(A) = 0.0067 \\ m(B) = 0.8645 \\ m(A \cup B) = 0.1288 \end{cases} \quad \text{and} \quad \begin{cases} m_{\epsilon=0.5}(\emptyset) = 0 \\ m_{\epsilon=0.5}(A) = 0.00335 \\ m_{\epsilon=0.5}(B) = 0.8645 \\ m_{\epsilon=0.5}(A \cup B) = 0.13215 \end{cases}$$

We get  $U(m) = 0.548244475651207$  and  $U(m_\epsilon) = 0.542869517947531$ , and we observe that  $U(m) > U(m_\epsilon)$ . This proves that D5 is not satisfied in this counter-example with the effective entropy measure defined in [1] by the formulas (1) and (2).

A measure of uncertainty  $U(m)$  that will satisfy desiderata D1, D2, D3 and D5 (and thus D4 too) will be named an *efficient entropy*.

II. ATTEMPTS FOR IMPROVING ENTROPY FORMULATION

In the spirit of original effective entropy formula defined in [1] by the formulas (1) and (2), we did explore other possible formulas by slightly changing the  $\alpha(u(X))$  and  $\beta(u(X), m(X))$  functions involved in the derivation of the entropieces  $s(X)$ . The Monte-Carlo evaluation of these modifications with respect to the satisfaction of desiderata D4 and D5 are reported in the next section. For D5, we did only make the evaluation based on the example presented before when considering  $m$  and  $m_\epsilon$  defined only over  $2^\Theta = \{A, B\}$ .

To keep the spirit of principle of entropy we choose positive functions  $\alpha(u(X))$  such that  $\alpha(u(X)) = 0$  for  $u(X) = 1$ , and  $\alpha(u(X)) = 1$  for  $u(X) = 0$ , so that the first part of entropy formula (1) remains compatible with Shannon entropy for Bayesian BBA. We also choose positive functions  $\beta(u(X), m(X))$  having same behavior as  $u(X)(1 - m(X))$  at the limits when  $u(X)$  equals zero or one, and when  $m(X)$  equals zero or one.

We make a behavior analysis of  $U(m)$  with the change of the functions  $\alpha(u(X))$  and  $\beta(u(X), m(X))$ . We refer a particular choice of couple of functions  $(\alpha(u(X)), \beta(u(X), m(X)))$  by a version number  $v_i$ . The original version  $v_0$  corresponds to the choice used in effective definition presented in [1], i.e.  $\alpha(u(X)) = 1 - u(X)$  and  $\beta(u(X), m(X)) = u(X)(1 - m(X))$ .

First, we analyze the behavior of  $U(m)$  with respect to the change of  $\alpha(u(X))$  function as those tested in Table I.

Version #	$\alpha(u(X))$	$\beta(u(X), m(X))$
$v_0$	$1 - u(X)$	$u(X)(1 - m(X))$
$v_1$	$\frac{1-u(X)}{1+(u(X)/ X )}$	$u(X)(1 - m(X))$
$v_2$	$\frac{1-u(X)}{1+u(X)}$	$u(X)(1 - m(X))$
$v_3$	$\frac{1-u(X)}{1+ X  \cdot u(X)}$	$u(X)(1 - m(X))$
$v_4$	$\frac{1-u(X)}{1-(u(X)/ X )}$	$u(X)(1 - m(X))$
$v_5$	$\frac{1-u(X)}{1-(u(X)/2)}$	$u(X)(1 - m(X))$
$v_6$	$\frac{1-u(X)}{1-(1-\frac{1}{ X })u(X)}$	$u(X)(1 - m(X))$

Table I  
FUNCTIONS ANALYZED.

The functions  $\alpha(u(X))$  used in versions  $v_1, v_2$  and  $v_3$  under-amplify the discounting because these functions are under the line  $(1 - u(X))$ . The functions  $\alpha(u(X))$  used in versions  $v_4, v_5$  and  $v_6$  over-amplify the discounting because these functions are above the line  $(1 - u(X))$ .

Secondly, we analyze the behavior of  $U(m)$  with respect to the change of the  $u(X)$  function appearing in  $\beta(u(X), m(X))$  of version  $v_0$  by those tested in Table II.

Version #	$\alpha(u(X))$	$\beta(u(X), m(X))$
$v_0$	$1 - u(X)$	$u(X)(1 - m(X))$
$v_7$	$1 - u(X)$	$(1 - \frac{1-u(X)}{1-(u(X)/ X )})(1 - m(X))$
$v_8$	$1 - u(X)$	$(1 - \frac{1-u(X)}{1-(u(X)/2)})(1 - m(X))$
$v_9$	$1 - u(X)$	$(1 - \frac{1-u(X)}{1-(1-\frac{1}{ X })u(X)})(1 - m(X))$
$v_{10}$	$1 - u(X)$	$(1 - \frac{1-u(X)}{1+u(X)/ X })(1 - m(X))$
$v_{11}$	$1 - u(X)$	$(1 - \frac{1-u(X)}{1+u(X)})(1 - m(X))$
$v_{12}$	$1 - u(X)$	$(1 - \frac{1-u(X)}{1+ X  \cdot u(X)})(1 - m(X))$

Table II  
FUNCTIONS ANALYZED.

The functions in  $v_7, v_8$  and  $v_9$  under-amplify  $u(X)(1 - m(X))$  because they are below the line  $u(X)(1 - m(X))$ , whereas the functions in  $v_{10}, v_{11}$  and  $v_{12}$  over-amplify  $u(X)(1 - m(X))$  because they are above the line  $u(X)(1 - m(X))$ .

Thirdly, we analyze the behavior of  $U(m)$  with respect to the change of the linear term  $1 - m(X)$  entering in  $\beta(u(X), m(X))$  formula of the original entropy version  $v_0$ , and we consider the functions listed in the Table III.

Version #	$\alpha(u(X))$	$\beta(u(X), m(X))$
v <sub>0</sub>	$1 - u(X)$	$u(X)(1 - m(X))$
v <sub>13</sub>	$1 - u(X)$	$u(X) \frac{1-m(X)}{1+(m(X)/ X )}$
v <sub>14</sub>	$1 - u(X)$	$u(X) \frac{1-m(X)}{1+m(X)}$
v <sub>15</sub>	$1 - u(X)$	$u(X) \frac{1-m(X)}{1+ X  \cdot m(X)}$
v <sub>16</sub>	$1 - u(X)$	$u(X) \frac{1-m(X)}{1-(m(X)/ X )}$
v <sub>17</sub>	$1 - u(X)$	$u(X) \frac{1-m(X)}{1-(m(X)/2)}$
v <sub>18</sub>	$1 - u(X)$	$u(X) \frac{1-m(X)}{1-(1-\frac{1}{ X })m(X)}$

Table III  
FUNCTIONS ANALYZED.

### III. SIMULATION RESULTS

The following tables presents the results of the tests of satisfaction of D4 and D5 based on  $N_{MC} = 500000$  random generations of BBAs used in our Monte-Carlo simulation.  $N(D4)$  is the number of cases where  $U(m) > U(m_v)$  has occurred, i.e. the number of violations of D4 for the case where  $|\Theta| = 5$ , and where  $m_v$  is the vacuous BBA for which  $m_v(\Theta) = 1$ .  $N_e(D5)$  is the number of cases where  $U(m) > U(m_e)$  has occurred, i.e. the number of violations of D5 when working with the frame of discernment  $\Theta = \{A, B\}$ .

Table IV gives the results corresponding to functions listed in Table I.

Version #	$(N(D4), U(m_v))$	$N_{0.9}(D5)$	$N_{0.5}(D5)$	$N_{0.2}(D5)$
v <sub>0</sub>	(0, 30)	15380	20248	26798
v <sub>1</sub>	(0, 30)	6048	7775	10117
v <sub>2</sub>	(0, 30)	6040	7699	10365
v <sub>3</sub>	(0, 30)	5851	7716	10474
v <sub>4</sub>	(0, 30)	34488	45352	63153
v <sub>5</sub>	(0, 30)	25667	33673	45652
v <sub>6</sub>	(0, 30)	15294	19931	26900

Table IV  
NUMBERS OF FAILURES AMONG  $N_{MC} = 500000$  RANDOM RUNS.

We observe from the results of Table IV that all expressions of  $U(m)$  satisfy D4, and more we discount  $m(A)$  (i.e. closer to zero is the  $\epsilon$ -factor) more failures we get for the D5 test. We also observe that using an **under-amplifying**  $\alpha(u(X))$  function as in versions v<sub>1</sub>, v<sub>2</sub> and v<sub>3</sub> reduces notably (by almost a factor 3) the number of failures for the D5 test with respect to the original linear weighting function  $1 - u(X)$  used in original entropy formula of v<sub>0</sub>. We observe conversely that using **over-amplifying**  $\alpha(u(X))$  function as in versions v<sub>4</sub>, v<sub>5</sub> and v<sub>6</sub> increase drastically the number of failures of the D5 test with respect to the original linear weighting function  $1 - u(X)$  used in original entropy formula of v<sub>0</sub>. Therefore, the choice for an **under-amplifying**  $\alpha(u(X))$  function is recommended.

Table V gives the results corresponding to functions listed in Table II.

As we observe in Table V, replacing  $u(X)$  by an **under-amplifying function** as in versions v<sub>7</sub>, v<sub>8</sub>, and v<sub>9</sub> has an impact on number of failures of D5 test but it changes

Version #	$(N(D4), U(m_v))$	$N_{0.9}(D5)$	$N_{0.5}(D5)$	$N_{0.2}(D5)$
v <sub>0</sub>	(0, 30)	15380	20248	26798
v <sub>7</sub>	(0, $\infty$ )	1452	1899	2585
v <sub>8</sub>	(0, 60)	3818	4954	6528
v <sub>9</sub>	(0, 75)	15542	20067	27014
v <sub>10</sub>	(0, 30)	12653	16280	22376
v <sub>11</sub>	(0, 30)	12617	16396	22230
v <sub>12</sub>	(0, 30)	12603	15999	22090

Table V  
NUMBERS OF FAILURES AMONG  $N_{MC} = 500000$  RANDOM RUNS.

substantially the entropy of vacuous BBA. If we replace  $u(X)$  by an **over-amplifying function** as in versions v<sub>10</sub>, v<sub>11</sub>, and v<sub>12</sub> we do not change the entropy of vacuous BBA which remains equal to  $U(m_v) = 2^{|\Theta|} - 2$ , and we reduce moderately the number of failures of D5 test and all results with versions v<sub>10</sub>, v<sub>11</sub>, and v<sub>12</sub> are very similar.

Table VI gives the results corresponding to functions listed in Table III.

Version #	$(N(D4), U(m_v))$	$N_{0.9}(D5)$	$N_{0.5}(D5)$	$N_{0.2}(D5)$
v <sub>0</sub>	(0, 30)	15380	20248	26798
v <sub>13</sub>	(0, 30)	10574	13614	18423
v <sub>14</sub>	(0, 30)	10645	13857	18521
v <sub>15</sub>	(0, 30)	10506	13645	18578
v <sub>16</sub>	(0, 30)	12496	16401	22062
v <sub>17</sub>	(0, 30)	16226	20251	28595
v <sub>18</sub>	(0, 30)	15513	19848	26816

Table VI  
NUMBERS OF FAILURES AMONG  $N_{MC} = 500000$  RANDOM RUNS.

We observe that using an **under-amplifying function** of  $1 - m(X)$  as with versions v<sub>13</sub>, v<sub>14</sub>, and v<sub>15</sub> reduces by about 1/3 the number of D5 failures, and there is no so much differences between results of these three versions. One sees that using **over-amplifying functions** as those in versions v<sub>16</sub>, v<sub>17</sub>, and v<sub>18</sub> does nor reduce substantially the number of failures of D5 test, and it can even be slightly worse than result of v<sub>0</sub> as we can see when using functions of v<sub>17</sub>.

Our analysis of the behavior of  $U(m)$  done with all functions tested in the Tables I, II and III reveals that the most important effect on the reduction of the number of failures of D5 desideratum is obtained when using an under-amplifying function  $\alpha(u(X))$ .

In order to reduce a bit more the number of failures of D5, we combine the expressions of  $\alpha(u(X))$  and  $\beta(u(X), m(X))$  that provide the minimum of failures to obtain the most failure reduction of D5 test. Hence if we consider the expression of  $\alpha(u(X))$  in v<sub>2</sub>, and the expression of  $\beta(u(X), m(X))$  in v<sub>13</sub>,

we get the following entropiece expression (denoted  $v_{19}$ )

$$s(X) = -\frac{1-u(X)}{1+u(X)}m(X)\log(m(X)) + u(X)\frac{1-m(X)}{1+(m(X)/|X|)} \quad (9)$$

Using this entropiece formulation, named  $v_{19}$ , we get the results of Table VII.

Version #	$(N(D4), U(m_v))$	$N_{0.9}(D5)$	$N_{0.5}(D5)$	$N_{0.2}(D5)$
$v_0$	(0, 30)	15380	20248	26798
$v_{19}$	(0, 30)	4180	5424	7250

Table VII

NUMBERS OF FAILURES AMONG  $N_{MC} = 500000$  RANDOM RUNS.

We see that using this new expression  $v_{19}$  for entropiece, we drastically reduce the number of failures of D5 to approximately  $(5000/500000) \cdot 100 = 1\%$ . Even if this failure percentage is quite small, it is not equal to zero. Therefore, the entropy measure based on this version  $v_{19}$  for entropiece definition is not totally efficient, but almost efficient for the case with  $|\Theta| = 2$ . Of course a more deeper and general analysis of test D5 with random BBA generated with frames of discernment  $\Theta$  having more than two elements should be tested to see if a similar small percentage of failure are obtained with the entropiece defined in the version  $v_{19}$ , and in this case we could conjecture that this new entropy measure is *quasi-efficient*.

#### IV. CONCLUSION

The existence of an entropy formulation that would warrant the efficiency (i.e. D1, D2, D3 and D5 desiderata) of the entropy measure of any BBA remains an open very challenging problem. We hope that a theoretical solution of this interesting problem exists, and if so the proof of its unicity (if any) would also be very welcome.

#### REFERENCES

- [1] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, Fusion 2022 Conf., Linköping, Sweden, pp. 1–10, July, 2022.
- [2] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379-423 & 623-656, July & October 1948 (reprinted in [4]).
- [3] C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, IL: The University of Illinois Press, 1–117, 1949.
- [4] N.J.A. Sloane, A.D. Wyner (Editors), *Claude Elwood Shannon - Collected Papers*, IEEE Press, 924 pages, 1993.
- [5] J. Dezert, F. Dambreville, *Cross-Entropy and Relative Entropy of Basic Belief Assignments*, Submitted to Fusion 2023 Int. Conf, Charleston, USA, June 2023.

# Involutory Negator of Basic Belief Assignments Applied to Information Fusion

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, F-91761 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.  
Emails: jean.dezert@onera.fr, tchamova@bas.bg

Originally published as: J. Dezert, A. Tchamova, *Involutory Negator of Basic Belief Assignments Applied to Information Fusion*, Cybernetics and Information Technologies Journal, Sofia, Vol. 23(3), September 2023, and reprinted with permission.

**Dedication:** This paper is dedicated to the memory of our colleague and friend Ludmil Bojilov (1944–2023).

**Abstract**—This paper analyzes the different definitions of a negator of a probability mass function (pmf) and a basic belief assignment (BBA) available in the literature. To overcome their limitations we propose an involutory negator of BBA, and we present a new indirect information fusion method based on this negator which can simplify the conflict management problem. The direct and indirect information fusion strategies are analyzed for three interesting examples of fusion of two BBAs. We also propose two methods for using the whole available information (the original BBAs and their negators) for decision-making support. The first method is based on the combination of the direct and indirect fusion strategies, and the second method selects the most reasonable fusion strategy to apply (direct, or indirect) based on the maximum entropy principle.

**Keywords:** belief functions, BBA negator, information fusion, measure of uncertainty, entropy.

## I. INTRODUCTION

This paper is an extended version of our paper published in Cybernetics and Information Technologies (CIT) journal [1]. Due to page limit restrictions of the CIT journal, we were not able to provide all technical details and the examples, and that is why we propose this extended version for the readers and researchers interested in this topic.

This paper deals with basic belief assignments (BBAs) introduced by Shafer in his mathematical theory of evidence [2] known also as Dempster-Shafer Theory (DST) in the literature. We focus on the construction of an involutory negator of a BBA, and its application for information fusion. The concept of the complement of a body of evidence (i.e. a negator) has been introduced by Dubois and Prade [3] in 1986, and re-examined by Yager in [4] who has attracted a new interest of the research community working with the belief functions. The main disadvantage with these negators (and of the most recent proposals) is that they are not involutory<sup>1</sup> in general so that the information content of the negator of a negator of a BBA is not equal to the information content of the original BBA. This is problematic from the informational standpoint because we naturally expect that working with negator of negator

<sup>1</sup>An involutory function (or involution) is a function  $f$  that is its own inverse, that is  $f(f(x)) = x$  for all  $x$  in the domain of  $f$ . This means that applying  $f$  twice produces the original value.

of evidence should be equivalent to working with original evidence. The problem we address in this paper can be stated as follows: let's consider a frame of discernment (FoD)  $\Theta$  of a problem under concern. Knowing a first expert providing a BBA  $m(\cdot)$  defined on the power set  $2^\Theta$ , is it possible to find a second expert with a BBA  $\bar{m}(\cdot)$  defined on the power set  $2^\Theta$  that expresses the opposite (or negation) assessment of the first expert? How can this be done effectively? Based on which principle and justifications? The second problem we address is the use of negator of BBAs for the information fusion and their possible advantages for decision-making support. It is worth mentioning that the negation of a BBA (i.e a BBA negator) must not be confused with negative values for masses of belief which are not allowed in Shafer's mathematical theory of evidence. This work focuses on the search for an involutory negator of BBA which can be interpreted as a dual approach of the characterization of any source of evidence.

This paper is organized as follows. Section II recalls the basic notions of Belief Functions (BF) and the entropy of BBAs. In Section III a detailed review and examples of several negators of probability mass function (pmf) and BBA proposed in the literature up to now is made. Section IV introduces a new involutory negator for BBAs. Section V recalls the principle of the classical direct fusion approach and describes the principle of a new indirect fusion approach based on the new involutory negator for BBA. In Section VI three interesting examples related to conflicting sources of evidences are described. The results obtained on the base of direct and indirect fusion approaches using Dempster's rule and Proportional Conflict Redistribution rule No.6 of combination are analyzed. Section VII discusses two important remarks about the indirect fusion based on conjunctive rule and about the entropy change due to the use of BBA's negator. Section VIII is devoted to the management of direct and indirect fusions for decision-making support. Conclusion is done in Section IX.

<sup>2</sup>By definition,  $2^\Theta = \{X | X \subseteq \Theta\}$ , which is the set of all subsets of  $\Theta$  (empty set  $\emptyset$  and  $\Theta$  included). We usually omit  $\{$  and  $\}$  characters for denoting the elements of the power-set because it makes the notation simpler and shorter. For instance, if the frame of discernment is  $\Theta = \{A, B\}$ , the power-set  $2^\Theta$  will be denoted by  $\{\emptyset, A, B, A \cup B\}$  with our notation which uses only 11 characters, instead of using the classical notation  $\{\emptyset, \{A\}, \{B\}, \{A, B\}\}$  which would require 17 characters. We can make the notation even more shorter by writing  $2^\Theta$  as  $\{\emptyset, A, B, \Theta\}$  using only 9 characters.

## II. BELIEF FUNCTIONS AND ENTROPY

The belief functions were introduced by Shafer [2] for modeling epistemic uncertainty, for reasoning about uncertainty, and for combining distinct sources of evidence (SoEs). The answer of the problem under concern is assumed to belong to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \dots, \theta_N\}$  where all elements (i.e. members) of  $\Theta$  are exhaustive and exclusive. The set of all subsets of  $\Theta$  (including empty set  $\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of the power-set is  $2^{|\Theta|}$ .

## A. Basic definitions

A normalized basic belief assignment<sup>3</sup> (BBA), associated with a given source of evidence is a mapping  $m^\Theta(\cdot) : 2^\Theta \rightarrow [0, 1]$  such that  $m^\Theta(\emptyset) = 0$  and  $\sum_{X \in 2^\Theta} m^\Theta(X) = 1$ . A BBA  $m^\Theta(\cdot)$  characterizes a source of evidence related with a FoD  $\Theta$ . For notation shorthand, we can omit the superscript  $\Theta$  in  $m^\Theta(\cdot)$  notation if there is no ambiguity on the FoD we work with. The quantity  $m(X)$  is called the mass of belief of  $X$ . The element  $X \in 2^\Theta$  is called a focal element (FE) of  $m(\cdot)$  if  $m(X) > 0$ . The set of all focal elements of  $m(\cdot)$  is denoted<sup>4</sup> by  $\mathcal{F}_\Theta(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$ . The belief and the plausibility of  $X$  are respectively defined for any  $X \in 2^\Theta$  by [2]

$$Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y), \quad (1)$$

$$Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) = 1 - Bel(\bar{X}). \quad (2)$$

where  $\bar{X} \triangleq \Theta \setminus \{X\}$  is the complement of  $X$  in  $\Theta$ .

One has always  $0 \leq Bel(X) \leq Pl(X) \leq 1$ , see [2]. For  $X = \emptyset$ ,  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and for  $X = \Theta$  one has  $Bel(\Theta) = Pl(\Theta) = 1$ .  $Bel(X)$  and  $Pl(X)$  are often interpreted as the lower and upper bounds of unknown probability  $P(X)$  of  $X$ , that is  $Bel(X) \leq P(X) \leq Pl(X)$ . To quantify the uncertainty (i.e. the imprecision) of  $P(X) \in [Bel(X), Pl(X)]$ , we use  $u(X) \in [0, 1]$  defined by

$$u(X) \triangleq Pl(X) - Bel(X). \quad (3)$$

The quantity  $u(X) = 0$  if  $Bel(X) = Pl(X)$ , which means that  $P(X)$  is known precisely, and one has  $P(X) = Bel(X) = Pl(X)$ . One has  $u(\emptyset) = 0$  because  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and one has  $u(\Theta) = 0$  because  $Bel(\Theta) = Pl(\Theta) = 1$ . If all focal elements of  $m(\cdot)$  are singletons of  $2^\Theta$  the BBA  $m(\cdot)$  is a Bayesian BBA because  $\forall X \in 2^\Theta$  one has  $Bel(X) = Pl(X) = P(X)$  and  $u(X) = 0$ . Hence the belief and plausibility of  $X$  coincide with a probability measure  $P(X)$  defined on the FoD  $\Theta$ . The vacuous BBA characterizing a totally ignorant source of evidence is defined by  $m_v(X) = 1$  for  $X = \Theta$ , and  $m_v(X) = 0$  for all  $X \in 2^\Theta$  different of  $\Theta$ . This very particular BBA has played a major role in the establishment of a new effective measure of uncertainty (i.e. entropy) for BBA in [5].

<sup>3</sup>also referred as a normal BBA, or a proper BBA in the literature.

<sup>4</sup>The symbol  $\triangleq$  means *equals by definition*.

## B. Entropy of a BBA

In [6] we did analyze in details forty-eight measures of uncertainty (MoU) of BBAs by covering 40 years of research works on this topic. Some of these MoUs capture only a particular aspect of the uncertainty inherent to a BBA (typically, the non-specificity and the conflict). Other MoUs propose a total uncertainty measure to capture jointly several aspects of the uncertainty. Unfortunately, most of these MoUs fail to satisfy four very simple reasonable and essential desiderata, and so they cannot be considered as really effective and useful. Actually only six MoUs can be considered as effective from the mathematical sense presented next, but unfortunately they appear as conceptually defective and disputable, see discussions in [6]. That is why, a better effective measure of uncertainty (MoU), i.e. generalized entropy of BBAs has been developed and presented in [5]. The mathematical definition of this new effective entropy is given by

$$U(m) = \sum_{X \in 2^\Theta} s(X), \quad (4)$$

with

$$s(X) \triangleq -m(X)(1 - u(X)) \log(m(X)) + u(X)(1 - m(X)). \quad (5)$$

$s(X)$  is the uncertainty contribution related to  $X$  named the *entropiece* of  $X$ . This entropiece  $s(X)$  involves  $m(X)$  and the imprecision  $u(X) = Pl(X) - Bel(X)$  about the unknown probability of  $X$  in a subtle interwoven manner. Because  $u(X) \in [0, 1]$  and  $m(X) \in [0, 1]$  one has  $s(X) \geq 0$ , and  $U(m) \geq 0$ . The quantity  $U(m)$  is expressed in *nats* because we use the natural logarithm.  $U(m)$  can be expressed in *bits* by dividing the  $U(m)$  value in *nats* by  $\log(2) \approx 0.69314718$ . This measure of uncertainty  $U(m)$  is a continuous function in its basic belief mass arguments because it is a summation of continuous functions. In formula (5), we always take  $m(X) \log(m(X)) = 0$  when  $m(X) = 0$  because  $\lim_{m(X) \rightarrow 0^+} m(X) \log(m(X)) = 0$ . Note that for any BBA  $m$ , one has  $s(\emptyset) = 0$  because  $m(\emptyset) = 0$  and  $u(\emptyset) = 0$ . For the vacuous BBA, one has  $s(\Theta) = 0$  because  $m_v(\Theta) = 1$  and  $u(\Theta) = 0$ . This measure of uncertainty  $U(m)$  is effective because it can be proved [5] that it satisfies the following four essential desiderata:

- 1)  $U(m) = 0$  for any BBA  $m(\cdot)$  focused on a singleton  $X$  of  $2^\Theta$ .
- 2)  $U(m_v^\Theta) < U(m_v^{\Theta'})$  if  $|\Theta| < |\Theta'|$ .
- 3)  $U(m) = -\sum_{X \in \Theta} m(X) \log(m(X))$  if the BBA  $m(\cdot)$  is a Bayesian BBA. Hence,  $U(m)$  reduces to Shannon entropy [7] in this case.
- 4)  $U(m) < U(m_v)$  for any non-vacuous BBA  $m(\cdot)$  and for the vacuous BBA  $m_v(\cdot)$  defined with respect to the same FoD  $\Theta$ .

The maximum of entropy value is obtained for the vacuous BBA  $m_v$  over a FoD  $\Theta$ , because  $m_v$  characterizes a source of evidence with a full lack of information. This maximum entropy value is  $U(m_v^\Theta) = 2^{|\Theta|} - 2$  (see derivation in [5]),

and it represents the sum of all imprecisions of  $P(X)$  for all  $X \in 2^\Theta$ . Because for all  $X \in 2^\Theta \setminus \{\emptyset, \Theta\}$  one has<sup>5</sup>  $u(X) = 1$ , and one has  $u(\emptyset) = 0$  and  $u(\Theta) = 0$  when considering the vacuous BBA, the sum of all imprecisions  $u(X)$  about  $P(X)$  is equal to  $2^{|\Theta|} - 2$ . It is worth mentioning that one has always  $U(m_v^\Theta) > \log(|\Theta|)$ , which means that the vacuous BBA has always an entropy greater than the maximum of Shannon entropy  $\log(|\Theta|)$  obtained with the uniform probability mass function (pmf) on the frame of discernment  $\Theta$ .

### III. NEGATORS OF PMF AND BBA IN THE LITERATURE

In this section we present several negators proposed in the literature with some examples, and we comment them.

#### A. Dubois and Prade non-involutory negator of a BBA (1986)

In 1986, Dubois and Prade introduced in [3] (pp. 202–203) for the first time the concept of negation of a BBA which has been adopted later by Smets [9] in his transferable belief model framework. Dubois and Prade (DP) negator is defined for any  $X \subseteq \Theta$  by

$$\bar{m}(X) = m(\bar{X}), \quad (6)$$

where  $\bar{X} = \Theta \setminus \{X\}$  is the complement of  $X$  in the FoD  $\Theta$ .

This simple definition is quite natural except that it does not satisfy the involutory property because  $\bar{\bar{m}} \neq m$  in general. Because we consider that this must be a very a natural and desired property to satisfy by an effective negator, we do not consider DP negator as effective. Moreover, it is clear that the DP negator of the vacuous BBA given by  $\bar{m}_v(\emptyset) = m_v(\Theta) = 1$  is not a proper BBA.

#### B. Yager's non-involutory negator of a pmf (2015)

Yager introduced the concept of the negation of a probability distribution  $P$  in [4] which was raised by Zadeh in his Berkeley Initiative in Soft Computing (BISC) blog. By the term *negation* Yager means the representation of the knowledge we use if we have the statement *not*  $P$ . The negation of a probability mass function (pmf)  $P(\cdot)$  over a reference set  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  has been defined by Yager as follows

$$\bar{P}(\theta_i) = \frac{1}{\lambda} P(\bar{\theta}_i), \quad (7)$$

where  $\bar{\theta}_i = \Theta \setminus \{\theta_i\}$  is the complement of  $\theta_i$  in the set  $\Theta$ ,  $P(\bar{\theta}_i) = 1 - P(\theta_i)$ , and where  $\lambda$  is a normalization factor given by

$$\lambda = \sum_{i=1}^n P(\bar{\theta}_i) = \sum_{i=1}^n (1 - P(\theta_i)) = n - 1. \quad (8)$$

The definition (7) is called Yager's negator in the literature [10]. Yager's justification for his definition is based on the maximal entropy principle of the weights associated with each focal element. As Yager pointed out in [4] the definition (7) does not satisfy the double negation (i.e. involutory or involutory) property in general (when  $|\Theta| > 2$  and  $P(\cdot)$  is

not the uniform pmf), that is  $\bar{\bar{P}}(\cdot) \neq P(\cdot)$ . Yager's negator of a probability distribution is the one that provides the maximum entropy among all possible negation definitions. The iterative application of Yager's negator converges towards the uniform pmf for which the entropy is maximal in the framework of the probability theory.

**Example 1:** Consider the set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the pmf  $P(\cdot)$  with  $P(\theta_1) = 1$  (i.e.  $\theta_1$  is a sure event). Based on (7), we get  $\bar{P}(\theta_1) = 0$ ,  $\bar{P}(\theta_2) = 1/2$  and  $\bar{P}(\theta_3) = 1/2$ . In this example the whole probability mass  $P(\theta_1) = 1$  is equally distributed back to singletons  $\theta_2$  and  $\theta_3$  of  $\bar{\theta}_1 = \theta_2 \cup \theta_3$ . The double Yager's negator of  $P(\cdot)$  is

$$\begin{aligned} \bar{\bar{P}}(\theta_1) &= \frac{1-0}{3-1} = 0.5 \neq P(\theta_1), \\ \bar{\bar{P}}(\theta_2) &= \frac{1-0.5}{3-1} = 0.25 \neq P(\theta_2), \\ \bar{\bar{P}}(\theta_3) &= \frac{1-0.5}{3-1} = 0.25 \neq P(\theta_3). \end{aligned}$$

**Example 2:** Consider the set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the uniform pmf  $P(\cdot)$  with  $P(\theta_1) = P(\theta_2) = P(\theta_3) = 1/3$ . Based on (7) we get also the uniform pmf because

$$\bar{P}(\theta_1) = \bar{P}(\theta_2) = \bar{P}(\theta_3) = \frac{1 - (1/3)}{3 - 1} = \frac{2/3}{2} = 1/3.$$

As a general result, the negation of any uniform pmf defined over  $\Theta$  of cardinality  $n > 1$  is always equal to the uniform pmf, i.e. the negation operator has no impact on the uniform pmf, and if  $P(\cdot)$  is uniform we always have  $\bar{\bar{P}}(\cdot) = P(\cdot)$ . The uniform pmf is the fixed point of Yager's negator.

**Example 3:** Consider the set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the non-uniform pmf  $P(\cdot)$  with  $P(\theta_1) = 0.6$ ,  $P(\theta_2) = 0.3$  and  $P(\theta_3) = 0.1$ . Based on (7) we get the negator

$$\begin{aligned} \bar{P}(\theta_1) &= \frac{1-0.6}{3-1} = 0.20, \\ \bar{P}(\theta_2) &= \frac{1-0.3}{3-1} = 0.35, \\ \bar{P}(\theta_3) &= \frac{1-0.1}{3-1} = 0.45. \end{aligned}$$

Note that in the very particular case where  $\Theta = \{\theta_1\}$  (i.e. there is only one element in the reference set), we have  $n = 1$  and necessarily  $P(\theta_1) = 1$ . Therefore by applying (7) we obtain  $\bar{P}(\theta_1) = \frac{1-P(\theta_1)}{n-1} = 0/0$  which is indeterminate.

A generalization of Yager's negator has been proposed in [10] by considering

$$\bar{P}(\theta_i) = \frac{1}{\lambda} (1 - d \cdot P(\theta_i)), \quad (9)$$

where  $d \in [0, 1]$  is a tuning parameter, and

$$\lambda = \sum_{i=1}^n (1 - d \cdot P(\theta_i)) = n - d. \quad (10)$$

The analysis of the new properties of Yager's negator has been done by Srivastava et al. in [11], [12]. An extension of Yager's negator based on Tsallis entropy has been proposed by

<sup>5</sup>Because  $[Bel(X), Pl(X)] = [0, 1]$ .

Zhang et al. in [13]. More recently a non-involutory exponential negator for pmf has been proposed by Wu et al. in [14] which unfortunately does not bring useful advantages w.r.t. Yager's negator for applications up to now. Wu's exponential negator has a better convergence rate towards uniform pmf by the repetitive application of it, but the real interest of this behaviour for practical applications is questionable and remains to be demonstrated.

### C. Yin's non-involutory negator of a BBA (2019)

In 2019, Yin et al. [15] proposed a definition of the negation of a BBA as a three steps procedure which can be expressed more concisely as follows for any focal element  $X$  of a BBA  $m(\cdot)$  defined over a frame of discernment  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ :

$$\bar{m}(X) = \frac{1}{\lambda} \cdot (1 - m(X)), \quad (11)$$

where  $\lambda$  is the normalization constant defined by

$$\lambda = \sum_{X \in 2^\Theta | m(X) > 0} (1 - m(X)) = N - 1, \quad (12)$$

and where  $N$  is the number of focal elements of  $m(\cdot)$ .

Clearly, Yin's negator is directly inspired by Yager's negator, but it works with BBA instead of pmf. Yin's negator is disputable because it is easy to check that it is non-involutory as shown in the example 4 next.

**Example 4:** Consider the set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the BBA  $m(\cdot)$  with  $m(\theta_1) = 0.1$ ,  $m(\theta_1 \cup \theta_2) = 0.2$ ,  $m(\theta_2 \cup \theta_3) = 0.3$ ,  $m(\Theta) = 0.4$ . Here  $N = 4$  because we the BBA  $m(\cdot)$  has four focal elements. Based on (11), we get

$$\begin{aligned} \bar{m}(\theta_1) &= \frac{1 - 0.1}{4 - 1} = 0.30, \\ \bar{m}(\theta_1 \cup \theta_2) &= \frac{1 - 0.2}{4 - 1} \approx 0.27, \\ \bar{m}(\theta_2 \cup \theta_3) &= \frac{1 - 0.3}{4 - 1} \approx 0.23, \\ \bar{m}(\Theta) &= \frac{1 - 0.4}{4 - 1} = 0.20. \end{aligned}$$

The double Yin's negator of  $m(\cdot)$  is

$$\begin{aligned} \bar{\bar{m}}(\theta_1) &= \frac{1 - 0.3}{4 - 1} \approx 0.23 \neq m(\theta_1), \\ \bar{\bar{m}}(\theta_1 \cup \theta_2) &\approx \frac{1 - 0.27}{4 - 1} \approx 0.24 \neq m(\theta_1 \cup \theta_2), \\ \bar{\bar{m}}(\theta_2 \cup \theta_3) &\approx \frac{1 - 0.23}{4 - 1} \approx 0.26 \neq m(\theta_2 \cup \theta_3), \\ \bar{\bar{m}}(\Theta) &= \frac{1 - 0.20}{4 - 1} \approx 0.27 \neq m(\Theta). \end{aligned}$$

So, we see that  $\bar{\bar{m}}(\cdot)$  obtained with Yin's negator is not equal to the original BBA  $m(\cdot)$ . This simple counter-example suffices to prove that Yin's negator is non-involutory.

More problematic, Yin's negator is indeterminate for the vacuous BBA  $m_v(\cdot)$  for which  $m_v(\Theta) = 1$ , because in this case  $\Theta$  is the only focal element so that  $N = 1$ , and from

(11) we get  $\bar{m}(\Theta) = \frac{1 - m(\Theta)}{N - 1} = 0/0$  which is indeterminate. Actually, this serious problem occurs not only for the vacuous BBA, but for any BBA focused on only one focal element. Therefore, Yin's negator is not appropriate for the negation of a BBA. It is worth mentioning that the iterative application of Yin's negator converges towards the uniform distribution of masses on all focal elements of  $m(\cdot)$  (assuming  $N > 1$ ) [15], which does not coincide to the vacuous BBA that must give the maximum of entropy [5].

We mention that Yin's negator has been presented also by Gao and Deng in [16]. Unfortunately this reference contains several mistakes, and the authors do not apply correctly Yin's definition in some of their examples. For instance, in their first specific example considering  $\Theta = \{a\}$  with  $m(a) = 1$  (i.e. we have only  $N = 1$  focal element for  $m(\cdot)$ ) the author claim that  $\bar{m}(a) = 0$ . This is obviously wrong because one has  $\bar{m}(a) = (1 - m(a))/(N - 1) = 0/0$  which is indeterminate. Gao and Deng write also: *Assumed a BPA that contains only one focal element (e.g. the  $m(a) = 0$ ), the negation of BPA can be defined by  $\bar{m}(a) = 1$ .* This claim is incorrect because any focal element  $X$  must be such that  $m(X) > 0$  by definition [2]. So, there is no proper BBA that contains only one focal element satisfying  $m(a) = 0$ . In their example 1 (see of Section IV-A of [16]) the same authors consider  $\Theta = \{a, b\}$  with  $m(a) = m(b) = 0.5$  (i.e. a Bayesian BBA with  $N = 2$  focal elements). Applying (12) we must obtain  $\bar{m}(a) = (1 - 0.5)/(2 - 1) = 0.5$  and  $\bar{m}(b) = (1 - 0.5)/(2 - 1) = 0.5$ , and not  $\bar{m}(a) = 0.25$ ,  $\bar{m}(b) = 0.25$  and  $\bar{m}(a \cup b) = 0.5$  as the authors claim. This casts doubts on the correctness of the technical content of Gao and Deng paper [16].

### D. Xie and Xiao non-involutory negator of a BBA (2019)

An other non-involutory negator of a BBA has been proposed by Xie and Xiao in [17]. This negator is defined by

$$\bar{\mathbf{m}} = \mathbf{E} \cdot \mathbf{m}, \quad (13)$$

where  $\mathbf{m}$  is the BBA  $m(\cdot)$  expressed as a vertical vector of size  $2^{|\Theta|}$ , that is  $\mathbf{m} = [m(\emptyset), m(\theta_1), \dots, m(\Theta)]^T$ , and  $\bar{\mathbf{m}}$  is the negation vector of the BBA vector  $\mathbf{m}$  which characterizes the negation of  $m(\cdot)$ . The matrix  $\mathbf{E}$  is a *negation symmetrical matrix*  $\mathbf{E} = [e_{ij}]$  of size  $2^{|\Theta|} \times 2^{|\Theta|}$  whose elements  $e_{ij}$  are defined as follows

$$e_{ij} = \begin{cases} 0, & \text{for } i = 1, \dots, 2^{|\Theta|} \text{ and } j = 1, \\ 0, & \text{for } j = 1, \dots, 2^{|\Theta|} \text{ and } i = 1, \\ 0, & \text{for } i = j \text{ and } i \neq 2^{|\Theta|}, \\ 1, & \text{for } i = j = 2^{|\Theta|}, \\ \frac{|X_i \cap \bar{X}_j|}{\sum_{X_k \in 2^\Theta \setminus \{\emptyset, \Theta\}} |X_k \cap \bar{X}_j|}, & \text{otherwise.} \end{cases} \quad (14)$$

where  $X_k$  is an element of the power-set  $2^\Theta$  of the FoD  $\Theta$ . By convention  $X_1 = \emptyset$  and  $X_{2^{|\Theta|}} = \Theta$ , that is  $k \in \{1, 2, \dots, 2^{|\Theta|}\}$ .  $\bar{X}_j = \Theta - \{X_j\}$  is the complement of  $X_j$  in the FoD  $\Theta$ .

Xie and Xiao proposal for a BBA negator is based on redistribution factors defined by (14) which appear rather ad-hoc and counter-intuitive as the following example 5 demonstrates.

**Example 5:** Consider the set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the BBA entirely focused on  $\theta_1$ . Hence we have the BBA vector<sup>6</sup>

$$\mathbf{m} = \begin{bmatrix} m(\emptyset) \\ m(\theta_1) \\ m(\theta_2) \\ m(\theta_3) \\ m(\theta_1 \cup \theta_2) \\ m(\theta_1 \cup \theta_3) \\ m(\theta_2 \cup \theta_3) \\ m(\Theta) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The negation matrix  $\mathbf{E} = [e_{ij}]$  is given by (see example 3 in [17])

$$\mathbf{E} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/6 & 1/6 & 0 & 0 & 1/3 & 0 \\ 0 & 1/6 & 0 & 1/6 & 0 & 1/3 & 0 & 0 \\ 0 & 1/6 & 1/6 & 0 & 1/3 & 0 & 0 & 0 \\ 0 & 1/6 & 1/6 & 1/3 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/6 & 1/3 & 1/6 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/6 & 1/6 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Therefore, the negation of the BBA  $m(\cdot)$  is given by

$$\bar{\mathbf{m}} = \begin{bmatrix} \bar{m}(\emptyset) \\ \bar{m}(\theta_1) \\ \bar{m}(\theta_2) \\ \bar{m}(\theta_3) \\ \bar{m}(\theta_1 \cup \theta_2) \\ \bar{m}(\theta_1 \cup \theta_3) \\ \bar{m}(\theta_2 \cup \theta_3) \\ \bar{m}(\Theta) \end{bmatrix} = \mathbf{E} \cdot \mathbf{m} = \begin{bmatrix} 0 \\ 0 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/6 \\ 1/3 \\ 0 \end{bmatrix}.$$

This result is very counter-intuitive because this negator commits some mass of belief to elements that have a non-empty intersection with  $\theta_1$ . This behavior does not make sense because the complement of  $\theta_1$  must have an empty intersection with  $\theta_1$  so the mass of  $\theta_1$  must be redistributed only to elements of the power set that have an empty intersection with  $\theta_1$  or eventually their disjunction. Moreover, Xie and Xiao present their analysis of their negator using Deng's entropy concept which is known to be non effective [6]. It is worth mentioning that Xie and Xiao negator is of course not involutory because in this very simple example we get

$$\bar{\bar{\mathbf{m}}} = \begin{bmatrix} \bar{\bar{m}}(\emptyset) \\ \bar{\bar{m}}(\theta_1) \\ \bar{\bar{m}}(\theta_2) \\ \bar{\bar{m}}(\theta_3) \\ \bar{\bar{m}}(\theta_1 \cup \theta_2) \\ \bar{\bar{m}}(\theta_1 \cup \theta_3) \\ \bar{\bar{m}}(\theta_2 \cup \theta_3) \\ \bar{\bar{m}}(\Theta) \end{bmatrix} = \mathbf{E} \cdot \bar{\mathbf{m}} = \begin{bmatrix} 0 \\ 1/6 \\ 1/12 \\ 1/12 \\ 1/4 \\ 1/4 \\ 1/6 \\ 0 \end{bmatrix} \neq \mathbf{m}.$$

A variant of Xie and Xiao approach has been published by Luo and Deng in [18]. Unfortunately, Luo and Deng negator is not involutory and it suffers of the same problems as Xie and Xiao negator.

<sup>6</sup>The elements of  $2^\Theta$  are listed as done by Xie and Xiao in [17].

*E. Deng and Jiang non-involutory negator of a BBA (2020)*

In 2020, Deng and Jiang proposed a new negator for any BBA defined as follows [19] over a frame of discernment  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$

$$\bar{m}(X) = \sum_{Y \in 2^\Theta | \bigcup_{\theta_i \in Y} (\Theta \setminus \{\theta_i\}) = X} m(Y). \quad (15)$$

As explained in [19] (p. 348) the authors consider that the negation of a singleton focal element  $X = \theta_i$  is  $\bar{X} = \bar{\theta}_i = \Theta \setminus \{\theta_i\}$ , and if a focal element  $X$  is not a singleton its negation is equal to  $\bar{X} = \bigcup_{\theta_i \in X} (\Theta \setminus \{\theta_i\}) = \Theta$ . This is what we call here the Deng-Jiang complementation principle. Unfortunately, there is no strong justification for adopting this complementation principle which is ad-hoc and very counter-intuitive because the negation of all non-singleton focal elements will correspond to the same complement element  $\Theta$  which is the whole FoD. This principle is actually inappropriate. The application of formula (15) is illustrated in the Example 6 drawn from [19].

**Example 6:** Consider the FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and  $m(\cdot)$  with  $m(\theta_1) = 0.7$ ,  $m(\theta_2 \cup \theta_3) = 0.1$ , and  $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.2$ .

Because the focal element  $\theta_1$  is a singleton, its Deng-Jiang complement is the classical complement  $\bar{\theta}_1 = \Theta \setminus \{\theta_1\} = \theta_2 \cup \theta_3$ , and we have

$$\bar{m}(\bar{\theta}_1) = \bar{m}(\theta_2 \cup \theta_3) = m(\theta_1) = 0.7.$$

Because the focal element  $\theta_2 \cup \theta_3$  is not a singleton, its Deng-Jiang complement is (by definition) taken as  $\bar{\theta}_2 \cup \bar{\theta}_3 = \Theta$ , and we have

$$\bar{m}(\overline{\theta_2 \cup \theta_3}) = \bar{m}(\Theta) = m(\theta_2 \cup \theta_3) = 0.1.$$

Because the focal element  $\Theta = \theta_1 \cup \theta_2 \cup \theta_3$  is not a singleton, its Deng-Jiang complement is (by definition) taken as  $\overline{\theta_1 \cup \theta_2 \cup \theta_3} = \Theta$ , and we have

$$\bar{m}(\overline{\theta_1 \cup \theta_2 \cup \theta_3}) = \bar{m}(\Theta) = m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.2.$$

Finally the two negator masses  $\bar{m}(\Theta) = 0.1$  and  $\bar{\bar{m}}(\Theta) = 0.2$  are added together to give the final result  $\bar{\bar{m}}(\theta_1 \cup \theta_2 \cup \theta_3) = 0.3$ . This is how formula (15) works.

Besides its weird complementation principle, Deng and Jiang's negator is not involutory in general. Indeed, if we apply this negator on the negator  $\bar{m}(\cdot)$  of example 5 we obtain  $\bar{\bar{m}}(\theta_1 \cup \theta_2 \cup \theta_3) = 1$  which is the vacuous BBA and not the original BBA  $m(\cdot)$  of example 5. This non-involutory property also appears in Table 4 of [19]. We point out also a flaw of Deng and Jiang negator which has a problem in the very special case where  $\Theta = \{\theta_1\}$  and  $m(\theta_1) = 1$ . In this case because the focal  $\theta_1$  is a singleton and based on Deng and Jiang's complementation we have  $\bar{\theta}_1 = \Theta \setminus \{\theta_1\} = \emptyset$  and applying Deng and Jiang's formula (15) we will get  $\bar{m}(\theta_1) = \bar{m}(\emptyset) = m(\theta_1) = 1$  which is not a proper BBA according to Shafer's definition [2]. As for Yin's negator, we consider that Deng and Jiang negator is not appropriate, and not effective from the theoretical standpoint, and we do not recommend its use for applications.

### F. Batyrshin's involutory negator of a pmf (2021)

In [20], [21] Batyrshin proposed an involutory negator of a pmf  $P$  defined over a reference set  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  by

$$\bar{P}(\theta_i) = \frac{MP - P(\theta_i)}{n \cdot MP - 1}, \quad (16)$$

where  $MP \triangleq \max P + \min P$ .

**Example 1** (continued): Consider the set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the pmf  $P(\cdot)$  with  $P(\theta_1) = 1$  (i.e.  $\theta_1$  is a sure event). Based on (16), we get  $MP = 1$  and Batyrshin's negator  $\bar{P}(\theta_1) = \frac{1-1}{3-1} = 0$ ,  $\bar{P}(\theta_2) = \frac{1-0}{3-1} = 1/2$  and  $\bar{P}(\theta_3) = \frac{1-0}{3-1} = 1/2$ . This result coincides with Yager's result based on (7). The double Batyrshin's negator of  $P(\cdot)$  is now given by

$$\begin{aligned} \bar{\bar{P}}(\theta_1) &= \frac{0.5 - 0}{3 \cdot 0.5 - 1} = 0.5/0.5 = 1 = P(\theta_1), \\ \bar{\bar{P}}(\theta_2) &= \frac{0.5 - 0.5}{3 \cdot 0.5 - 1} = 0/0.5 = 0 = P(\theta_2), \\ \bar{\bar{P}}(\theta_3) &= \frac{0.5 - 0.5}{3 \cdot 0.5 - 1} = 0/0.5 = 0 = P(\theta_3). \end{aligned}$$

**Example 2** (continued): Consider the set  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the uniform pmf  $P(\cdot)$  with  $P(\theta_1) = P(\theta_2) = P(\theta_3) = 1/3$ . Based on (16), we get  $MP = \frac{1}{3} + \frac{1}{3} = 2/3$  and Batyrshin's negator

$$\bar{P}(\theta_1) = \bar{P}(\theta_2) = \bar{P}(\theta_3) = \frac{(2/3) - (1/3)}{3 \cdot (2/3) - 1} = 1/3,$$

which corresponds also to the uniform pmf, and thus the double Batyrshin's negator of uniform pmf  $P(\cdot)$  is equal to itself. Actually, the uniform pmf is a fixed point for Batyrshin's negator (as it is for Yager's negator too), see [20] for details.

**Example 3** (continued): We consider  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and the non-uniform pmf  $P(\cdot)$  with  $P(\theta_1) = 0.6$ ,  $P(\theta_2) = 0.3$  and  $P(\theta_3) = 0.1$ . Batyrshin's negator of  $P$  is given by<sup>7</sup>

$$\begin{aligned} \bar{P}(\theta_1) &= \frac{0.7 - 0.6}{3 \cdot 0.7 - 1} = \frac{0.1}{1.1} \approx 0.09, \\ \bar{P}(\theta_2) &= \frac{0.7 - 0.3}{3 \cdot 0.7 - 1} = \frac{0.4}{1.1} \approx 0.36, \\ \bar{P}(\theta_3) &= \frac{0.7 - 0.1}{3 \cdot 0.7 - 1} = \frac{0.6}{1.1} \approx 0.55. \end{aligned}$$

It is worth mentioning that Batyrshin's negator of  $P(\cdot)$  equals  $P(\cdot)$  in the very special case where  $\Theta = \{\theta_1\}$ , because in this case one has  $n = 1$  and necessarily  $P(\theta_1) = 1$ . Therefore,  $\min P = \max P = 1$  and  $MP = 2$ , and from (16) we obtain Batyrshin's negator  $\bar{P}(\theta_1) = \frac{MP - P(\theta_1)}{n \cdot MP - 1} = \frac{2-1}{1 \cdot 2 - 1} = 1 = P(\theta_1)$ .

Even if this negator is appealing from the theoretical standpoint when working with probabilities its real usefulness has to be shown in real applications. Batyrshin's negator has not yet been extended for the framework of the theory of belief functions, and it may be interesting to extend it (if possible) for the theory of evidence

<sup>7</sup>because  $MP = 0.7$ .

### G. Liu's non-involutory negator of a BBA (2023)

In 2023 Liu et al [22] did propose a new negator of a BBA  $m(\cdot)$  defined over a frame of discernment  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  by

$$\bar{m}(X) = \frac{1}{\lambda} \cdot \frac{2^{|X|} - 1}{\sum_{Y \in 2^\Theta | Y \neq X} 2^{|Y|} - 1} (1 - m(X)), \quad (17)$$

where  $X \in 2^\Theta$ , and  $\lambda$  is the normalization constant defined by

$$\lambda = \sum_{X \in 2^\Theta} \frac{2^{|X|} - 1}{\sum_{Y \in 2^\Theta | Y \neq X} 2^{|Y|} - 1} (1 - m(X)). \quad (18)$$

This new negator is unfortunately not involutory as proved by the authors in [22], and they justify this new negator based on Deng's entropy concept which is non effective [5], [6]. So, their justification is flawed. It is also clear that the concept of complementation used by Liu et al. is inappropriate as shown in the very simple following example.

**Example 7:** Consider the FoD  $\Theta = \{A, B\}$  and the vacuous BBA  $m_v(\cdot)$  defined on this FoD by  $m_v(\emptyset) = 0$ ,  $m_v(A) = 0$ ,  $m_v(B) = 0$ , and  $m_v(A \cup B) = 1$ . By applying (17) we will obtain<sup>8</sup> the following Liu's negator

$$\begin{aligned} \bar{m}_v(\emptyset) &= \frac{1}{\lambda} \cdot \frac{(2^{|\emptyset|} - 1)(1 - m_v(\emptyset))}{\sum_{Y \in 2^\Theta | Y \neq \emptyset} 2^{|Y|} - 1} = \frac{0 \cdot 1}{1 + 1 + 3} = 0, \\ \bar{m}_v(A) &= \frac{1}{\lambda} \cdot \frac{(2^{|A|} - 1)(1 - m_v(A))}{\sum_{Y \in 2^\Theta | Y \neq A} 2^{|Y|} - 1} = \frac{1 \cdot 1}{0 + 1 + 3} = 0.5, \\ \bar{m}_v(B) &= \frac{1}{\lambda} \cdot \frac{(2^{|B|} - 1)(1 - m_v(B))}{\sum_{Y \in 2^\Theta | Y \neq B} 2^{|Y|} - 1} = \frac{1 \cdot 1}{0 + 1 + 3} = 0.5, \\ \bar{m}_v(A \cup B) &= \frac{1}{\lambda} \cdot \frac{(2^{|A \cup B|} - 1)(1 - m_v(A \cup B))}{\sum_{Y \in 2^\Theta | Y \neq A \cup B} 2^{|Y|} - 1} = \frac{3 \cdot 0}{0 + 1 + 1} = 0. \end{aligned}$$

because  $\lambda = (0/4) + (2/4) + (2/4) + (0/2) = 1$ .

One sees clearly that Liu's negator is inappropriate because  $A$  and  $B$  cannot be considered as valid complements of  $A \cup B$  because  $A \cap (A \cup B) \neq \emptyset$  and  $B \cap (A \cup B) \neq \emptyset$ .

## IV. A NEW INVOLUTORY NEGATOR FOR BBAS

We have shown in the previous section that most of negators developed previously are not involutory functions except Batyrshin's negator which applies only to probabilities, and not to non-Bayesian BBAs. Consequently, these negators increase in general the entropy when negator applies iteratively and this iterative application of negator makes the result to converge towards uniform pmf or BBA, which is not very useful in practice. In this section we present a new simple definition for an involutory negator of any BBA  $m(\cdot) : 2^\Theta \rightarrow [0, 1]$  which expresses the opposite evidence of any source of evidence characterized by  $m(\cdot)$ . The opposite (i.e. the negator) of the BBA  $m(\cdot)$  is denoted by  $\bar{m}(\cdot)$ , and it is simply defined by

$$\bar{m}(X) = \begin{cases} 0, & \text{if } X = \emptyset, \\ m(\bar{X}), & \forall X \neq \emptyset \subset \Theta, \\ m(\Theta), & \text{if } X = \Theta. \end{cases} \quad (19)$$

<sup>8</sup>Because  $|\emptyset| = 0$  and  $2^0 = 1$  we have  $2^{|\emptyset|} - 1 = 0$ .

$\bar{X}$  is the complement of the subset  $X$  in the FoD  $\Theta$ , that is  $\bar{X} = \Theta \setminus \{X\}$ .

This new negator defined by (19) is actually a revised definition of Dubois and Prade negator (6). This definition may appear strange at the first glance for some readers because of the conditions  $\bar{m}(\emptyset) = 0$  and  $\bar{m}(\Theta) = m(\Theta)$ . Some readers may dispute why the mass of belief committed to the whole ignorance proposition  $\Theta$  is kept unchanged in the expression (19) of the negator of the source of evidence. This is a legitimate question because the (classical) complement  $\bar{\Theta}$  of  $\Theta$  in  $\Theta$  is equal to the empty set, and because the (classical) complement  $\bar{\emptyset}$  of the empty set in  $\Theta$  is equal to  $\Theta$ . As Dubois and Prade did, we could consider a priori taking  $\bar{m}(\emptyset) = m(\Theta)$  and  $\bar{m}(\Theta) = m(\emptyset)$ . We think however that this option is actually not very reasonable because it would mean that the negation of a BBA will not be a proper BBA (as defined by Shafer in [2]). In fact, we would have  $\bar{m}(\emptyset) > 0$  when  $m(\Theta) > 0$ , and we would always have  $\bar{m}(\Theta) = 0$  because  $m(\emptyset) = 0$  which is very restrictive. We consider that the most reasonable solution is to consider that the negation of the BBA  $m(\cdot)$  is better defined by (19). This new very simple definition for the negator of a BBA presents the great advantage to preserve the involutory property of the negator concept of a BBA so that  $\bar{\bar{m}}(\cdot) = m(\cdot)$ . Note that the negator of any BBA  $m(\cdot)$  defined by  $\bar{m}(\cdot)$  in (19) is a proper BBA because  $\bar{m}(X) \in [0, 1]$ ,  $\bar{m}(\emptyset) = 0$  and  $\sum_{X \in 2^\Theta} \bar{m}(X) = 1$  because the focal elements of  $\bar{m}(\cdot)$  belong to  $2^\Theta$  and they correspond to the complement of the focal elements of  $m(\cdot)$  which is a proper BBA.

We mention that the negator of a Bayesian BBA is not a Bayesian BBA in general as soon as the FoD  $\Theta$  has more than two elements. This is normal because if a focal element, say  $\theta_i \in \Theta$ , of  $m(\cdot)$  is a singleton with  $m(\theta_i) > 0$ , then its complement  $\bar{\theta}_i = \Theta \setminus \{\theta_i\}$  is not singleton of  $\Theta$  if  $|\Theta| > 2$ , and we have  $\bar{m}(\bar{\theta}_i) = m(\theta_i) > 0$  with  $|\bar{\theta}_i| > 1$  which indicates that the BBA  $\bar{m}(\cdot)$  is not Bayesian. We also mention that the negator of the vacuous BBA  $m_v(\cdot)$  is equal to itself, which indicates that the vacuous source of evidence plays a neutral role with respect to this new negator concept. This is not very surprising because from no useful information (characterized by a fully ignorant source whose BBA is the vacuous BBA) we cannot draw any conclusion for making a decision in favor of one hypothesis or its opposite. This makes the definition (19) coherent with the intuition when working with vacuous BBA and the negator concept.

Of course, it is always possible to approximate any non-Bayesian BBA (or any non-Bayesian negator of a BBA) by a pmf (if we want or we need for any reason) thanks to different techniques of approximation, for instance using BetP, or DSMP transformations [23], [24]. As a simple example consider the

FoD  $\Theta = \{A, B, C\}$  and the Bayesian BBAs  $m(\cdot)$  defined by<sup>9</sup>

$$m(A) = 0.9, m(B) = 0, m(C) = 0.1.$$

Its negator is the non-Bayesian BBA defined as

$$\bar{m}(B \cup C) = 0.9, \bar{m}(A \cup B) = 0.1.$$

If one approximates  $\bar{m}(\cdot)$  by a probability measure thanks to the BetP transformation for instance, we will obtain the Bayesian negation of  $m(\cdot)$  denoted either as  $\bar{m}_{Bayesian}(\cdot)$  or more concisely as  $Bet\bar{P}(\cdot)$  which is given by

$$\begin{aligned} Bet\bar{P}(A) &= \frac{1}{2}\bar{m}(A \cup B) = 0.05, \\ Bet\bar{P}(B) &= \frac{1}{2}\bar{m}(B \cup C) + \frac{1}{2}\bar{m}(A \cup B) = 0.50, \\ Bet\bar{P}(C) &= \frac{1}{2}\bar{m}(B \cup C) = 0.45. \end{aligned}$$

This result is quite reasonable because based on the fact that  $\bar{m}(B \cup C) = 0.9$  and  $\bar{m}(A \cup B) = 0.1$  (when considering the negator of  $m(\cdot)$  as valid input information) it makes sense that  $B$  has the most chance to occur among  $A$ ,  $B$  and  $C$ , and  $A$  has the second best chance to occur. This is what reflects the  $Bet\bar{P}(\cdot)$  distribution for this example.

## V. DIRECT AND INDIRECT FUSION APPROACHES

In this section we recall the principle of the classical direct fusion approach, and we describe also the principle of the indirect fusion approach based on the involutory negator of the BBAs described in the previous section.

### A. Direct fusion approach

Before presenting the application of the BBA negator for information fusion, we recall several well-known fusion rules used to combine directly distinct bodies of evidence represented by the BBAs  $m_1, m_2, \dots, m_S$  defined over the same FoD  $\Theta$ . This is what we call the direct fusion approach.

To make this presentation simple, we present the formulas for the combination of two BBAs only (i.e.  $S = 2$ ). General formulas for more than two BBAs can be encountered in the literature, see [2], [25], [26] for instance. A survey of more fusion rules can be found in [27].

- Conjunctive rule of combination:  $\forall X \in 2^\Theta$ ,

$$m_{1,2}^\cap(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2). \quad (20)$$

- Disjunctive rule of combination:  $\forall X \in 2^\Theta$ ,

$$m_{1,2}^\cup(X) = \sum_{\substack{X_1, X_2 \in 2^\Theta \\ X_1 \cup X_2 = X}} m_1(X_1)m_2(X_2). \quad (21)$$

<sup>9</sup>Here we voluntarily indicate that  $m(B) = 0$  for convenience to point out that we have three elements in the frame of discernment. The notation  $m(B) = 0$  could be omitted of course because only  $A$  and  $C$  are the focal elements of the BBA  $m(\cdot)$ . In fact all elements of the power set  $2^\Theta$  have masses equal to zero except  $A$  and  $C$  in this example. In general, only masses of focal elements need to be listed because all other masses equal zero.

- Dempster-Shafer (DS) rule of combination [2]:  
 $m_{1,2}^{DS}(\emptyset) = 0$  and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{1,2}^{DS}(X) = \frac{\sum_{X_1, X_2 \in 2^\Theta, X_1 \cap X_2 = X} m_1(X_1)m_2(X_2)}{1 - m_{1,2}^{\cap}(\emptyset)}. \quad (22)$$

- Proportional Conflict Redistribution Rule no. 6 (PCR6) [26]:  $m_{1,2}^{PCR6}(\emptyset) = 0$  and  $\forall X \in 2^\Theta \setminus \{\emptyset\}$

$$m_{1,2}^{PCR6}(X) = m_{1,2}^{\cap}(X) + \sum_{\substack{Y \in 2^\Theta \\ X \cap Y = \emptyset}} \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)}. \quad (23)$$

In this paper we consider only PCR6 rule because we use examples for the fusion of only two BBAs to keep the presentation as simple as possible. If one needs to combine three BBAs (or more) altogether, we recommend to use the improved PCR6 rule (denoted by PCR6<sup>+</sup>) which is presented in details<sup>10</sup> in [26] with Matlab codes.

If the sources of evidence are considered fully reliable the conjunctive fusion rule applies, but it happens that the sources to combine are conflicting if  $m_{1,2}^{\cap}(\emptyset) > 0$ . In this case  $m_{1,2}^{\cap}(\cdot)$  is not strictly a proper BBA. To overcome this problem, Dempster-Shafer (DS) rule of combination or PCR6 fusion rule can be used to obtain a normalized and combined BBA. DS rule offers the main advantage of being associative making its use quite easy for the applications, and DS preserves the neutrality<sup>11</sup> of the vacuous BBA  $m_v$  which is generally considered as a good property for a fusion rule. DS rule being associative, the sequential DS fusion of many sources of evidence is independent of the sequence order which is appealing. Unfortunately, DS rule exhibits counter-intuitive dictatorial behavior in high and low conflict situations as well [28], [29]. This is one of the main reasons<sup>12</sup> why DS rule has been abandoned by many researchers and engineers working with belief functions. If the two sources are in total conflict (i.e.  $m_{1,2}^{\cap}(\emptyset) = 1$ ), DS rule does not work because of the division by zero in (22). PCR6 rule provides more reasonable fusion results, and it works in low and high conflicting situations as well. PCR6 does not behave dictatorially. The main disadvantage of PCR6 rule is its high complexity because it is not associative that is why all the sources of evidence must be combined altogether (not sequentially) with this rule. PCR6 does not preserve the neutrality of the vacuous BBA  $m_v$  when combining more than two BBAs altogether, but an improved version denoted by PCR6<sup>+</sup> preserves the neutrality of  $m_v$ , see [26] for details. If one of the sources of evidence is not reliable and we do not know which one, the disjunctive fusion rule applies.

<sup>10</sup>Note that PCR6<sup>+</sup> and PCR6 rules coincide for the fusion of two BBAs.

<sup>11</sup>The vacuous BBA  $m_v$  has no impact on the fusion result when combined to a BBA  $m \neq m_v$  with DS rule.

<sup>12</sup>The second main reason is that Shafer's BBA conditioning based on DS rule is not consistent with lower and upper bounds of conditional probability [30].

Finally the direct fusion approach of  $S$  BBAs  $m_1, m_2, \dots, m_S$  defined over the same FoD  $\Theta$  is denoted symbolically by

$$m_{1,2,\dots,S}^{DF} = F(m_1, m_2, \dots, m_S), \quad (24)$$

where DF means the chosen Direct Fusion (DF) rule used for the combination of the  $S$  sources of evidence. Typically DF = DS if we use Dempster-Shafer rule for making the direct fusion of the  $S$  BBAs, or DF = PCR6 if we use the PCR6 fusion rule for making the direct fusion, etc.

### B. Indirect fusion using the involutory negator of BBAs

In some information fusion situations the combination of BBAs  $m_1(\cdot), m_2(\cdot), \dots, m_S(\cdot)$  (with  $S > 1$ ) is problematic if there exist some conflicts between the sources of evidence. This means that  $m_1(X_1)m_2(X_2)\dots m_S(X_S) > 0$  when  $X_1 \cap X_2 \cap \dots \cap X_S = \emptyset$  for some focal elements  $X_1, X_2, \dots$ , and  $X_S$ . When conflicts occur the simple conjunctive rule of combination (20) is not able to provide an acceptable fusion result because it commits a strictly positive mass of belief to the impossible event (i.e. to the empty set), that is  $m_{1,2,\dots,S}^{\cap}(\emptyset) > 0$ . Therefore, it is generally necessary to manage the existing conflicts between the sources efficiently to obtain what we consider as reasonable fusion result for decision-making support. That is the reason why many fusion rules of combination have been developed and proposed in the literature during the last decades [25], [27].

In this work, we propose a new generic approach to combine the sources of evidence thanks to their involutory negator of the BBAs, which is what we call the indirect fusion approach. The idea behind the indirect fusion approach is rather simple. Instead of combining directly the original BBAs by some fusion rules (typically Dempster-Shafer (DS) rule [2], PCR6 rule [25], [26], Dubois-Prade rule [3], etc) to directly obtain the fusion result, we propose to compute the fusion result indirectly using the negators of BBAs. This new indirect fusion approach is based on the following three simple steps:

- **Step-1** (Calculation of the BBAs negators):

Calculate the involutory negator of BBAs  $m_1(\cdot), m_2(\cdot), \dots, m_S(\cdot)$  using (19) to get  $\bar{m}_1(\cdot), \bar{m}_2(\cdot), \dots, \bar{m}_S(\cdot)$ .

- **Step-2:** (Combination of the negators)

Combine (i.e. fuse) the  $S > 1$  BBAs  $\bar{m}_1(\cdot), \bar{m}_2(\cdot), \dots, \bar{m}_S(\cdot)$  by a chosen fusion rule denoted symbolically by DF to get the direct fusion of negators, that is

$$\bar{m}_{1,2,\dots,S}^{DF}(\cdot) = F(\bar{m}_1, \bar{m}_2, \dots, \bar{m}_S). \quad (25)$$

The choice of the direct fusion rule DF for combining the negators is left to the fusion system designer. Proponents of DST will prefer Dempster-Shafer's rule of combination (22), while opponents of DS rule will use other fusion rules (typically PCR6 rule (23), etc).

- **Step-3** (Negation of the fused negators):

Once the fusion result  $\bar{m}_{1,2,\dots,S}^{DF}(\cdot)$  is obtained, one calculates its negator to get the final Indirect Fusion (IF)

result of the original BBAs thanks to the definition (19) where  $m(\cdot)$  is replaced by  $\bar{m}_{1,2,\dots,S}^{\text{DF}}(\cdot)$ , that is

$$m_{1,2,\dots,S}^{\text{IF}}(X) = \begin{cases} 0, & \text{if } X = \emptyset, \\ \bar{m}_{1,2,\dots,S}^{\text{DF}}(\bar{X}), & \forall X \neq \emptyset \subset \Theta, \\ \bar{m}_{1,2,\dots,S}^{\text{DF}}(\Theta), & \text{if } X = \Theta. \end{cases} \quad (26)$$

More concisely, we will write steps 1, 2 and 3 by the symbolic expression

$$m_{1,2,\dots,S}^{\text{IF}} = \bar{m}_{1,2,\dots,S}^{\text{DF}}, \quad (27)$$

where the negator operator used in (27) (represented by a bar symbol) is the involutory negator defined in (19).

As it will be discussed in Section VIII, in general we have  $m_{1,2,\dots,S}^{\text{IF}} \neq m_{1,2,\dots,S}^{\text{DF}}$ . This means that the direct and indirect fusion methods provide different results depending on the fusion rule chosen, and on the distribution of masses of belief to focal elements. This is because the fusion rules do not satisfy De Morgan's law when a conflict exists between the sources of evidence. Only in the case where  $S = 2$  and  $m_1(\cdot) = m_v$ , or  $m_2(\cdot) = m_v$ , one has  $m_{1,2}^{\text{IF}} = m_{1,2}^{\text{DF}}$  because there is no conflict between the two sources of evidence to deal with in this very particular case.

## VI. SOME INTERESTING EXAMPLES

In this section we examine three interesting examples where a conflict exists between two sources of evidence, and we compare the result based on indirect fusion method with the result obtained with the classical direct fusion approach using DS and PCR6 rules of combination.

### A. Zadeh's example (two Bayesian BBAs)

Consider the famous Zadeh's example [28] where  $\Theta = \{A, B, C\}$  represents three hypotheses about the origin of a diseases of a patient, and two Bayesian BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  expressed by two doctors after the examination of the same patient. These BBAs are given as follows

$$\begin{aligned} m_1(A) &= 0.9, m_1(C) = 0.1, \\ m_2(B) &= 0.9, m_2(C) = 0.1. \end{aligned}$$

- Direct fusion with DS rule:

By applying DS rule (22), we obtain the Bayesian BBA  $m_{1,2}^{\text{DS}}(C) = 1$ , which is considered as a counter-intuitive result by Zadeh and by many authors because this result means that the hypothesis  $C$  is diagnosed for sure for the origin of the disease by DS rule even if both doctors agree in committing a low belief to the origin  $C$ . This example is important because it has served as a starting point to question the validity of DS rule in Shafer's theory of belief functions by Zadeh. This result has however been justified by a first school of proponents of DS theory by the fact that the two sources of evidence are highly conflicting in this example because

$$\begin{aligned} m_{1,2}^{\cap}(\emptyset) &= m_1(A)m_2(B) + m_1(A)m_2(C) \\ &+ m_1(C)m_2(B) = 0.81 + 0.09 + 0.09 = 0.99 \end{aligned}$$

Therefore, the proponents of this rule argue that DS rule should not be applied without preprocessing (i.e. discounting) the sources of evidence in this situation. Other proponents of DS rule belonging to the second school of proponents of DS rule argue that DS result makes perfectly sense in Zadeh's example. Both schools of proponents defend DS rule but their conclusions are based on very different contradictory arguments, which amplify the suspicion about the validity of DS rule as pointed out by Zadeh in [28]. Actually the two types of arguments used to defend DS rule are flawed because DS rule behaves dictatorially even in low conflict situation as well as reported by the authors in [29]. This will be shown in the problematic example of Section VI-B.

- Direct fusion with PCR6 rule

By applying PCR6 fusion rule (23) to combine  $m_1(\cdot)$  and  $m_2(\cdot)$  we get

$$\begin{aligned} m_{1,2}^{\text{PCR6}}(A) &= 0.486, \\ m_{1,2}^{\text{PCR6}}(B) &= 0.486, \\ m_{1,2}^{\text{PCR6}}(C) &= 0.028. \end{aligned}$$

This Bayesian PCR6 result is more reasonable than DS result because it clearly points out the difficulty to make a choice between hypotheses  $A$  and  $B$  because of the disagreement of two doctors while rejecting both the third hypothesis  $C$ .

- Indirect fusion approach:

By applying the indirect fusion approach, after step 1 we get the following BBAs negators

$$\begin{aligned} \bar{m}_1(B \cup C) &= 0.9, \bar{m}_1(A \cup B) = 0.1, \\ \bar{m}_2(A \cup C) &= 0.9, \bar{m}_2(A \cup B) = 0.1. \end{aligned}$$

We observe that there is no conflict between these two negators so that the conjunctive fusion rule can be used, and there is no need to adopt a specific management of conflicting masses either by DS rule, or by PCR6 rule because results from both rules are equal to the result computed with the conjunctive rule, when no conflict occurs.

At step-2, we use the conjunctive fusion of  $\bar{m}_1$  and  $\bar{m}_2$  because there is no conflict between these negators, and we get

$$\begin{aligned} \bar{m}_{1,2}^{\cap}(A) &= \bar{m}_1(A \cup B)\bar{m}_2(A \cup C) = 0.09, \\ \bar{m}_{1,2}^{\cap}(B) &= \bar{m}_1(B \cup C)\bar{m}_2(A \cup B) = 0.09, \\ \bar{m}_{1,2}^{\cap}(C) &= \bar{m}_1(B \cup C)\bar{m}_2(A \cup C) = 0.81, \\ \bar{m}_{1,2}^{\cap}(A \cup B) &= \bar{m}_1(A \cup B)\bar{m}_2(A \cup B) = 0.01. \end{aligned}$$

At step 3, we take the negator of  $\bar{m}_{1,2}^\cap(\cdot)$  as the final indirect fusion result. We obtain<sup>13</sup>

$$\begin{aligned} m_{1,2}^{\text{IF}\cap}(A \cup B) &= \bar{m}_{1,2}^\cap(C) = 0.81, \\ m_{1,2}^{\text{IF}\cap}(A \cup C) &= \bar{m}_{1,2}^\cap(B) = 0.09, \\ m_{1,2}^{\text{IF}\cap}(B \cup C) &= \bar{m}_{1,2}^\cap(A) = 0.09, \\ m_{1,2}^{\text{IF}\cap}(C) &= \bar{m}_{1,2}^\cap(A \cup B) = 0.01. \end{aligned}$$

This non-Bayesian indirect fusion result is more acceptable than DS result because it reveals clearly the uncertainty between hypotheses  $A$  and  $B$ , while reinforcing the disbelief of hypothesis  $C$  as we intuitively expect. We observe that  $m_{1,2}^{\text{IF}\cap} \neq m_{1,2}^{\text{DS}}$  and  $m_{1,2}^{\text{IF}\cap} \neq m_{1,2}^{\text{PCR6}}$ . However, if we approximate  $m_{1,2}^{\text{IF}\cap}$  by a probability measure thanks to BetP transform we obtain

$$\begin{aligned} \text{BetP}^{\text{IF}\cap}(A) &= \frac{m_{1,2}^{\text{IF}\cap}(A \cup B)}{2} + \frac{m_{1,2}^{\text{IF}\cap}(A \cup C)}{2} = 0.45, \\ \text{BetP}^{\text{IF}\cap}(B) &= \frac{m_{1,2}^{\text{IF}\cap}(A \cup B)}{2} + \frac{m_{1,2}^{\text{IF}\cap}(B \cup C)}{2} = 0.45, \\ \text{BetP}^{\text{IF}\cap}(C) &= m_{1,2}^{\text{IF}\cap}(C) + \frac{m_{1,2}^{\text{IF}\cap}(A \cup C)}{2} \\ &\quad + \frac{m_{1,2}^{\text{IF}\cap}(B \cup C)}{2} = 0.1. \end{aligned}$$

This  $\text{BetP}^{\text{IF}\cap}(\cdot)$  result is without doubt closer to the direct PCR6 fusion result than the DS result although not strictly equal. One observes that the  $\text{BetP}^{\text{IF}\cap}(\cdot)$  distribution obtained from this indirect fusion result coincides with the simple averaging fusion rule which is a common fusion rule adopted by users not familiar with belief functions. This behavior is, we think, another argument against the direct fusion result provided by DS rule.

### B. Dezert-Tchamova example (two non-Bayesian BBAs)

Here we consider another problematic example presented by Dezert and Tchamova in [29] to show the dictatorial behavior of DS rule of combination in high and low conflicting situations as well. An infinity of problematic examples like this one can be defined, see [32] for more examples. We consider the FoD  $\Theta = \{A, B, C\}$  with the following two (generic) non-Bayesian BBAs

$$m_1(A) = a, m_1(A \cup B) = 1 - a,$$

$$m_2(A \cup B) = b_1, m_2(C) = 1 - b_1 - b_2, m_2(A \cup B \cup C) = b_2,$$

with  $0 < a, b_1, b_2 < 1$  and  $b_1 + b_2 < 1$ .

The conflict of these two BBAs is actually independent of the BBA  $m_1(\cdot)$  because

$$\begin{aligned} m_{1,2}^\cap(\emptyset) &= m_1(A)m_2(C) + m_1(A \cup B)m_2(C) \\ &= m_2(C) = 1 - b_1 - b_2. \end{aligned}$$

One can easily verify that the direct Dempster-Shafer's fusion of these two BBAs gives  $m_{1,2}^{\text{DS}}(A) = m_1(A) = a$  and

<sup>13</sup>We use the notation  $m_{1,2}^{\text{IF}\cap}(\cdot)$  to explicitly specify that the indirect fusion (IF) has been done with the conjunctive rule (symbolized by the  $\cap$  symbol).

$m_{1,2}^{\text{DS}}(A \cup B) = m_1(A \cup B) = 1 - a$  which indicates that the fusion result is actually independent of the BBA  $m_2(\cdot)$  even if  $m_2(\cdot)$  is not the vacuous BBA and the conflict degree can be taken as high or as low as we want. This behavior of DS rule is of course counter-intuitive and dictatorial, and that is why we do not recommend its use in applications.

**Example 8:** We consider here Dezert-Tchamova example with parameters  $a = 0.3$ ,  $b_1 = 0.2$  and  $b_2 = 0.3$ . Hence,

$$m_1(A) = 0.3, m_1(A \cup B) = 0.7,$$

$$m_2(A \cup B) = 0.2, m_2(C) = 0.5, m_2(A \cup B \cup C) = 0.3.$$

For this numerical example, using the conjunctive fusion rule, we obtain

$$\begin{aligned} m_{1,2}^\cap(\emptyset) &= 0.3 \cdot 0.5 + 0.7 \cdot 0.5 = 0.50, \\ m_{1,2}^\cap(A) &= 0.3 \cdot 0.2 + 0.3 \cdot 0.3 = 0.15, \\ m_{1,2}^\cap(A \cup B) &= 0.7 \cdot 0.2 + 0.7 \cdot 0.3 = 0.35. \end{aligned}$$

One sees that there exists a positive conflict  $m_{1,2}^\cap(\emptyset)$  between these two sources of evidence that needs to be redistributed in order to obtain a proper resulting BBA.

- Direct fusion with DS rule:

By applying DS rule (22), we obtain

$$\begin{aligned} m_{1,2}^{\text{DS}}(\emptyset) &= 0, \\ m_{1,2}^{\text{DS}}(A) &= \frac{0.3 \cdot 0.2 + 0.3 \cdot 0.3}{0.5} = 0.3 = m_1(A), \\ m_{1,2}^{\text{DS}}(A \cup B) &= \frac{0.7 \cdot 0.2 + 0.7 \cdot 0.3}{0.5} = 0.7 = m_1(A \cup B). \end{aligned}$$

Same dictatorial DS fusion result would be obtained for other numerical values of positive parameters  $a$ ,  $b_1$  and  $b_2$  with  $b_1 + b_2 < 1$ .

- Direct fusion with PCR6 rule:

By applying PCR6 rule<sup>14</sup> (23), we obtain

$$\begin{aligned} m_{1,2}^{\text{PCR6}}(\emptyset) &= 0, \\ m_{1,2}^{\text{PCR6}}(A) &= 0.2062, \\ m_{1,2}^{\text{PCR6}}(A \cup B) &= 0.5542, \\ m_{1,2}^{\text{PCR6}}(C) &= 0.2396. \end{aligned}$$

We see that the PCR6 fusion rule does not behave dictatorially. It can be easily verified that the PCR6 fusion result changes with different values of  $b_1$  and  $b_2$ .

- Indirect fusion with DS rule:

If we apply the indirect fusion approach, the negators of  $m_1(\cdot)$  and  $m_2(\cdot)$  are given by

$$\bar{m}_1(B \cup C) = 0.3, \bar{m}_1(C) = 0.7,$$

$$\bar{m}_2(C) = 0.2, \bar{m}_2(A \cup B) = 0.5, \bar{m}_2(A \cup B \cup C) = 0.3.$$

<sup>14</sup>see formulas (12)–(14) in [31] for details. Note that PCR5 and PCR6 formulas for the fusion of two BBAs provide the same result.

Hence the conjunctive fusion of negators gives

$$\begin{aligned}\bar{m}_{1,2}^{\cap}(\emptyset) &= 0.7 \cdot 0.5 = 0.35, \\ \bar{m}_{1,2}^{\cap}(B) &= 0.3 \cdot 0.5 = 0.15, \\ \bar{m}_{1,2}^{\cap}(C) &= 0.3 \cdot 0.2 + 0.7 \cdot 0.2 + 0.7 \cdot 0.3 = 0.41, \\ \bar{m}_{1,2}^{\cap}(B \cup C) &= 0.3 \cdot 0.3 = 0.09.\end{aligned}$$

Note that  $\bar{m}_{1,2}^{\cap}(\cdot) \neq m_{1,2}^{\cap}(\cdot)$ . Applying DS rule of combination of these negators we obtain

$$\begin{aligned}\bar{m}_{1,2}^{DS}(\emptyset) &= 0, \\ \bar{m}_{1,2}^{DS}(B) &= 0.15/0.65 \approx 0.23, \\ \bar{m}_{1,2}^{DS}(C) &= 0.41/0.65 \approx 0.63, \\ \bar{m}_{1,2}^{DS}(B \cup C) &= 0.09/0.65 \approx 0.14.\end{aligned}$$

After taking the negator of  $\bar{m}_{1,2}^{DS}(\cdot)$  we obtain using the indirect DS (i.e. IF-DS) fusion approach the following final result

$$\begin{aligned}m_{1,2}^{IF-DS}(\emptyset) &= \bar{m}_{1,2}^{DS}(\emptyset) = 0, \\ m_{1,2}^{IF-DS}(A \cup B) &= \bar{m}_{1,2}^{DS}(C) \approx 0.63, \\ m_{1,2}^{IF-DS}(A \cup C) &= \bar{m}_{1,2}^{DS}(B) \approx 0.23, \\ m_{1,2}^{IF-DS}(A) &= \bar{m}_{1,2}^{DS}(B \cup C) \approx 0.14.\end{aligned}$$

This result appears a bit more acceptable than the direct DS fusion result without being dictatorial because  $m_{1,2}^{IF-DS}(\cdot) \neq m_1(\cdot)$  and  $m_{1,2}^{IF-DS}(\cdot) \neq m_2(\cdot)$ . It is worth mentioning that this new indirect DS fusion approach does not always circumvent the bad dictatorial behavior of DS rule in general thanks to the negators and their DS fusion. To validate this remark, it is easy to built another (dual) Dezert-Tchamova example where the fusion of negators of BBAs really provides a dictatorial behavior instead of the direct DS fusion. For instance, consider  $\Theta = \{A, B, C\}$  and the following BBAs

$$m_1(B \cup C) = a, m_1(C) = 1 - a,$$

$$m_2(C) = b_1, m_2(A \cup B) = 1 - b_1 - b_2, m_2(A \cup B \cup C) = b_2.$$

Then we have

$$\bar{m}_1(A) = a, \bar{m}_1(A \cup B) = 1 - a,$$

$$\bar{m}_2(A \cup B) = b_1, \bar{m}_2(C) = 1 - b_1 - b_2, \bar{m}_2(A \cup B \cup C) = b_2.$$

The fusion of  $\bar{m}_1(\cdot)$  and  $\bar{m}_2(\cdot)$  with Dempster-Shafer's rule exhibits a dictatorial behavior because one gets  $\bar{m}_{12}^{DS}(A) = a$  and  $\bar{m}_{12}^{DS}(A \cup B) = 1 - a$ , and the final result based on these negators and indirect DS fusion is dictatorial and given by

$$\begin{aligned}m_{12}^{IF-DS}(B \cup C) &= \bar{m}_{12}^{DS}(A) = a, \\ m_{12}^{IF-DS}(C) &= \bar{m}_{12}^{DS}(A \cup B) = 1 - a.\end{aligned}$$

Hence, we get  $m_{12}^{IF-DS}(B \cup C) = a = m_1(B \cup C)$  and  $m_{12}^{IF-DS}(C) = 1 - a = m_1(C)$ . So the use of Dempster-Shafer's rule in the information fusion method based on the negators of BBA remains also disputable in this case.

That is why in any strategy of fusion chosen (direct and indirect) we cannot recommend seriously Dempster-Shafer rule of combination because of its potential dictatorial behavior.

- Indirect fusion with PCR6 rule:

If we apply the indirect PCR6 fusion approach of the negators of  $m_1(\cdot)$  and  $m_2(\cdot)$  we obtain

$$\begin{aligned}\bar{m}_{1,2}^{PCR6}(\emptyset) &= 0, \\ \bar{m}_{1,2}^{PCR6}(B) &= \bar{m}_{1,2}^{\cap}(B) = 0.1500, \\ \bar{m}_{1,2}^{PCR6}(C) &= \bar{m}_{1,2}^{\cap}(C) + \frac{\bar{m}_1(C)^2 \bar{m}_2(A \cup B)}{\bar{m}_1(C) + \bar{m}_2(A \cup B)} = 0.6142, \\ \bar{m}_{1,2}^{PCR6}(B \cup C) &= \bar{m}_{1,2}^{\cap}(B \cup C) = 0.0900, \\ \bar{m}_{1,2}^{PCR6}(A \cup B) &= \frac{\bar{m}_1(C) \bar{m}_2(A \cup B)^2}{\bar{m}_1(C) + \bar{m}_2(A \cup B)} = 0.1458.\end{aligned}$$

After taking the negator of  $\bar{m}_{1,2}^{PCR6}(\cdot)$  using Indirect PCR6 (IF-PCR6) fusion approach we obtain the following final result

$$\begin{aligned}m_{1,2}^{IF-PCR6}(\emptyset) &= \bar{m}_{1,2}^{PCR6}(\emptyset) = 0, \\ m_{1,2}^{IF-PCR6}(A \cup B) &= \bar{m}_{1,2}^{PCR6}(C) = 0.6142, \\ m_{1,2}^{IF-PCR6}(A \cup C) &= \bar{m}_{1,2}^{PCR6}(B) = 0.1500, \\ m_{1,2}^{IF-PCR6}(A) &= \bar{m}_{1,2}^{PCR6}(B \cup C) = 0.0900, \\ m_{1,2}^{IF-PCR6}(C) &= \bar{m}_{1,2}^{PCR6}(A \cup B) = 0.1458.\end{aligned}$$

We observe that direct and indirect PCR6-based fusion methods provide distinct results because  $m_{1,2}^{IF-PCR6}(\cdot) \neq m_{1,2}^{PCR6}(\cdot)$ . It is worth noting also that the indirect fusion results based on DS rule (IF-DS) and PCR6 rule (IF-PCR6) provide quite similar maximal mass value for the same focal element  $A \cup B$  because  $m_{1,2}^{IF-DS}(A \cup B) \approx 0.63$  and  $m_{1,2}^{IF-PCR6}(A \cup B) = 0.6142$ . However, we observe that the set of focal elements of  $m_{1,2}^{IF-DS}(\cdot)$  and  $m_{1,2}^{IF-PCR6}(\cdot)$  are not the same because IF-PCR6 commits a mass specifically to the element  $C$  which is not a focal element of  $m_{1,2}^{IF-DS}(\cdot)$ . Therefore the structures (i.e. the set of focal elements) of BBAs  $m_{1,2}^{IF-DS}(\cdot)$  and  $m_{1,2}^{IF-PCR6}(\cdot)$  are different.

### C. Blackman's example (Bayesian and non-Bayesian BBAs)

This simple example has been introduced by Blackman in [33] and analyzed by the authors in [34]. We consider the FoD  $\Theta = \{A, B\}$  and the following two BBAs

$$m_1(A) = 0.5, m_1(B) = 0.5, m_1(A \cup B) = 0,$$

$$m_2(A) = 0.1, m_2(B) = 0.1, m_2(A \cup B) = 0.8.$$

We see that there is no way to decide either  $A$  or  $B$  in this particular example because each source of evidence does not bring useful information to help for decision-making. Each BBA  $m_1(\cdot)$  and  $m_2(\cdot)$  is completely symmetrical to  $A$  and  $B$ . So intuitively, there is no reason to expect an improvement in the decision-making based on the fusion of these two BBAs.

We mention that  $m_1(\cdot)$  is a Bayesian BBA, and  $m_2(\cdot)$  is a non-Bayesian BBA in Blackman's example.

The conjunctive fusion of  $m_1(\cdot)$  and  $m_2(\cdot)$  yields

$$\begin{aligned} m_{1,2}^\cap(\emptyset) &= m_1(A)m_2(B) + m_1(B)m_2(A) = 0.10, \\ m_{1,2}^\cap(A) &= m_1(A)m_2(A) + m_1(A)m_2(A \cup B) = 0.45, \\ m_{1,2}^\cap(B) &= m_1(B)m_2(B) + m_1(B)m_2(A \cup B) = 0.45, \\ m_{1,2}^\cap(A \cup B) &= 0. \end{aligned}$$

We see that the conflicting mass  $m_{1,2}^\cap(\emptyset) = 0.10$  must be redistributed to some elements of  $2^\Theta \setminus \{\emptyset\}$  in order to get a proper fused BBA.

- Direct fusion with DS rule:

By applying DS rule (22), we obtain

$$\begin{aligned} m_{1,2}^{\text{DS}}(\emptyset) &= 0 \\ m_{1,2}^{\text{DS}}(A) &= \frac{m_{1,2}^\cap(A)}{1 - m_{1,2}^\cap(\emptyset)} = 0.45/0.9 = 0.5, \\ m_{1,2}^{\text{DS}}(B) &= \frac{m_{1,2}^\cap(B)}{1 - m_{1,2}^\cap(\emptyset)} = 0.45/0.9 = 0.5, \\ m_{1,2}^{\text{DS}}(A \cup B) &= \frac{m_{1,2}^\cap(A \cup B)}{1 - m_{1,2}^\cap(\emptyset)} = 0.00/0.9 = 0. \end{aligned}$$

- Direct fusion with PCR6 rule:

By applying PCR6 rule (23), we obtain

$$\begin{aligned} m_{1,2}^{\text{PCR6}}(\emptyset) &= 0, \\ m_{1,2}^{\text{PCR6}}(A) &= m_{1,2}^\cap(A) + \frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} \\ &\quad + \frac{m_1(B) m_2(A)^2}{m_1(B) + m_2(A)} = 0.5, \\ m_{1,2}^{\text{PCR6}}(B) &= m_{1,2}^\cap(B) + \frac{m_1(B)^2 m_2(A)}{m_1(B) + m_2(A)} \\ &\quad + \frac{m_1(A) m_2(B)^2}{m_1(A) + m_2(B)} = 0.5, \\ m_{1,2}^{\text{PCR6}}(A \cup B) &= m_{1,2}^\cap(A \cup B) = 0. \end{aligned}$$

As intuitively expected, the direct fusion results based on DS rule and on PCR6 rule do not help to make a rational decision in favor of  $A$  or  $B$ .

- Indirect fusion with DS and PCR6 rules:

Applying BBA negator defined by (19), we obtain

$$\begin{aligned} \bar{m}_1(B) &= 0.5, \quad \bar{m}_1(A) = 0.5, \quad \bar{m}_1(A \cup B) = 0, \\ \bar{m}_2(B) &= 0.1, \quad \bar{m}_2(A) = 0.1, \quad \bar{m}_2(A \cup B) = 0.8. \end{aligned}$$

Because  $|\Theta| = 2$ , we observe that we have in this example  $\bar{m}_1(\cdot) = m_1(\cdot)$  and  $\bar{m}_2(\cdot) = m_2(\cdot)$ . Therefore, we will get the same result with the conjunctive fusion of  $\bar{m}_1(\cdot)$  and  $\bar{m}_2(\cdot)$  as for the direct conjunctive fusion of  $m_1(\cdot)$  and  $m_2(\cdot)$ . The direct or indirect fusion methods

based on DS and PCR6 rules will yield actually to the same fusion result, that is

$$\begin{aligned} m_{1,2}^{\text{DS}}(A) &= m_{1,2}^{\text{IF-DS}}(A) = m_{1,2}^{\text{PCR6}}(A) = m_{1,2}^{\text{IF-PCR6}}(A) = 0.5, \\ m_{1,2}^{\text{DS}}(B) &= m_{1,2}^{\text{IF-DS}}(B) = m_{1,2}^{\text{PCR6}}(B) = m_{1,2}^{\text{IF-PCR6}}(B) = 0.5. \end{aligned}$$

This example is interesting because it clearly shows that there exist some situations where there is no advantage of using direct fusion w.r.t. indirect fusion, and vice-versa.

#### Extension of Blackman's example

We extend Blackman's example using a bigger FoD as follows. We consider  $\Theta = \{A, B, C\}$  with the two following BBAs

$$m_1(A) = m_1(B) = m_1(C) = 1/3,$$

$$m_2(A) = m_2(B) = m_2(C) = 0.1, \quad m_2(A \cup B \cup C) = 0.7.$$

As in the previous Blackman's example we see that there is no way to decide either  $A$ ,  $B$  or  $C$  because each BBA  $m_1(\cdot)$  and  $m_2(\cdot)$  is completely symmetrical to  $A$ ,  $B$  and  $C$ . So there is no rational reason to expect an improvement in the decision-making based on the fusion of these two BBAs. Here also  $m_1(\cdot)$  is a Bayesian BBA, and  $m_2(\cdot)$  is a non-Bayesian BBA in this example.

The conjunctive fusion of  $m_1(\cdot)$  and  $m_2(\cdot)$  yields

$$m_{1,2}^\cap(\emptyset) = 0.6/3, \quad m_{1,2}^\cap(A) = m_{1,2}^\cap(B) = m_{1,2}^\cap(C) = 0.8/3.$$

The direct DS fusion and the direct PCR6 fusion give the same result which is

$$\begin{aligned} m_{1,2}^{\text{DS}}(\emptyset) &= m_{1,2}^{\text{PCR6}}(\emptyset) = 0, \\ m_{1,2}^{\text{DS}}(A) &= m_{1,2}^{\text{PCR6}}(A) = 1/3, \\ m_{1,2}^{\text{DS}}(B) &= m_{1,2}^{\text{PCR6}}(B) = 1/3, \\ m_{1,2}^{\text{DS}}(C) &= m_{1,2}^{\text{PCR6}}(C) = 1/3. \end{aligned}$$

If we want to apply the indirect fusion methods, the negators of  $m_1(\cdot)$  and  $m_2(\cdot)$  are the non-Bayesian BBAs  $\bar{m}_1(\cdot)$  and  $\bar{m}_2(\cdot)$  given by

$$\begin{aligned} \bar{m}_1(B \cup C) &= \bar{m}_1(A \cup C) = \bar{m}_1(A \cup B) = 1/3, \\ \bar{m}_2(B \cup C) &= \bar{m}_2(A \cup C) = \bar{m}_2(A \cup B) = 0.1, \\ \bar{m}_2(A \cup B \cup C) &= 0.7. \end{aligned}$$

The conjunctive fusion of  $\bar{m}_1(\cdot)$  and  $\bar{m}_2(\cdot)$  gives

$$\begin{aligned} \bar{m}_{1,2}^\cap(\emptyset) &= 0, \\ \bar{m}_{1,2}^\cap(A) &= \bar{m}_{1,2}^\cap(B) = \bar{m}_{1,2}^\cap(C) = 0.2/3, \\ \bar{m}_{1,2}^\cap(A \cup B) &= \bar{m}_{1,2}^\cap(A \cup C) = \bar{m}_{1,2}^\cap(B \cup C) = 0.8/3. \end{aligned}$$

Because there is no conflicting mass to redistribute, there is no need to apply indirect DS fusion of  $\bar{m}_1(\cdot)$  and  $\bar{m}_2(\cdot)$ , nor indirect PCR6 fusion of  $\bar{m}_1(\cdot)$  and  $\bar{m}_2(\cdot)$ . More precisely if we apply DS rule or PCR6 rule we will obtain

$$\bar{m}_{1,2}^{\text{DS}}(\cdot) = \bar{m}_{1,2}^{\text{PCR6}}(\cdot) = \bar{m}_{1,2}^\cap(\cdot).$$

After taking the negator of  $\bar{m}_{12}^{\cap}(\cdot)$  we obtain using indirect fusion approaches the following final result

$$m_{1,2}^{\text{IF-DS}}(\cdot) = m_{1,2}^{\text{IF-PCR6}}(\cdot) = m_{12}^{\text{IF-}\cap}(\cdot) = \bar{\bar{m}}_{12}^{\cap}(\cdot),$$

where

$$\begin{aligned} m_{12}^{\text{IF-}\cap}(\emptyset) &= \bar{\bar{m}}_{12}^{\cap}(\emptyset) = 0, \\ m_{12}^{\text{IF-}\cap}(B \cup C) &= \bar{\bar{m}}_{12}^{\cap}(A) = 0.2/3, \\ m_{12}^{\text{IF-}\cap}(A \cup C) &= \bar{\bar{m}}_{12}^{\cap}(B) = 0.2/3, \\ m_{12}^{\text{IF-}\cap}(A \cup B) &= \bar{\bar{m}}_{12}^{\cap}(C) = 0.2/3, \\ m_{12}^{\text{IF-}\cap}(C) &= \bar{\bar{m}}^{\cap}(A \cup B) = 0.8/3, \\ m_{12}^{\text{IF-}\cap}(B) &= \bar{\bar{m}}^{\cap}(A \cup C) = 0.8/3, \\ m_{12}^{\text{IF-}\cap}(A) &= \bar{\bar{m}}^{\cap}(B \cup C) = 0.8/3. \end{aligned}$$

As we can see in this extended Blackman's example, the indirect fusion approach eliminates the problem of conflict management because there is no conflict to deal with when working with the negators. Of course the fusion result is less specific than the one we obtain using the direct fusion approaches (based on DS and on PCR6 rules), and  $m_{12}^{\text{IF-}\cap}(\cdot)$  result is no more helpful for decision-making standpoint which is normal in such situation, but this indirect fusion approach could be interesting to use if other sources of evidence may become available in the fusion system.

## VII. TWO IMPORTANT REMARKS

**Remark 1:** As shown in Zadeh's example of Section VI-A the indirect fusion method gives

$$\begin{aligned} m_{1,2}^{\text{IF-}\cap}(A \cup B) &= \bar{\bar{m}}_{1,2}^{\cap}(C) = 0.81, \\ m_{1,2}^{\text{IF-}\cap}(A \cup C) &= \bar{\bar{m}}_{1,2}^{\cap}(B) = 0.09, \\ m_{1,2}^{\text{IF-}\cap}(B \cup C) &= \bar{\bar{m}}_{1,2}^{\cap}(A) = 0.09, \\ m_{1,2}^{\text{IF-}\cap}(C) &= \bar{\bar{m}}_{1,2}^{\cap}(A \cup B) = 0.01. \end{aligned}$$

It is interesting to observe that this result coincides with the fusion result obtained with the disjunctive rule of combination (21). Indeed, we have

$$\begin{aligned} m_{1,2}^{\cup}(A \cup B) &= m_1(A)m_2(B) = 0.81, \\ m_{1,2}^{\cup}(A \cup C) &= m_1(A)m_2(C) = 0.09, \\ m_{1,2}^{\cup}(B \cup C) &= m_1(C)m_2(B) = 0.09, \\ m_{1,2}^{\cup}(C) &= m_1(C)m_2(C) = 0.01. \end{aligned}$$

We may question if the equality  $m_{1,2}^{\text{IF-}\cap}(\cdot) = m_{1,2}^{\cup}(\cdot)$  is a general property satisfied, or only just a coincidence. In fact this equality does not hold in general but it is due to the very particular structure of focal elements of the BBAs involved in Zadeh's example. This equality does not hold even if there is no conflict between the negators (as it appears in Zadeh's example). As a simple counter-example, consider again the extended Blackman's example of Section VI-C where no

conflict exists between the negators  $\bar{m}_1(\cdot)$  and  $\bar{m}_2(\cdot)$ . The indirect fusion approach gives the final result

$$\begin{aligned} m_{12}^{\text{IF-}\cap}(\emptyset) &= 0, \\ m_{12}^{\text{IF-}\cap}(A) &= 0.8/3, \\ m_{12}^{\text{IF-}\cap}(B) &= 0.8/3, \\ m_{12}^{\text{IF-}\cap}(C) &= 0.8/3, \\ m_{12}^{\text{IF-}\cap}(A \cup B) &= 0.2/3, \\ m_{12}^{\text{IF-}\cap}(A \cup C) &= 0.2/3, \\ m_{12}^{\text{IF-}\cap}(B \cup C) &= 0.2/3, \\ m_{12}^{\text{IF-}\cap}(A \cup B \cup C) &= 0. \end{aligned}$$

The fusion result obtained with the disjunctive rule of combination (21) for this extended Blackman's example is

$$\begin{aligned} m_{1,2}^{\cup}(\emptyset) &= 0 \\ m_{1,2}^{\cup}(A) &= m_1(A)m_2(A) = 0.1/3, \\ m_{1,2}^{\cup}(B) &= m_1(B)m_2(B) = 0.1/3, \\ m_{1,2}^{\cup}(C) &= m_1(C)m_2(C) = 0.1/3, \\ m_{1,2}^{\cup}(A \cup B) &= m_1(A)m_2(B) + m_1(B)m_2(A) = 0.2/3, \\ m_{1,2}^{\cup}(A \cup C) &= m_1(A)m_2(C) + m_1(C)m_2(A) = 0.2/3, \\ m_{1,2}^{\cup}(B \cup C) &= m_1(B)m_2(C) + m_1(C)m_2(B) = 0.2/3, \\ m_{1,2}^{\cup}(A \cup B \cup C) &= m_1(A)m_2(A \cup B \cup C) \\ &\quad + m_1(B)m_2(A \cup B \cup C) \\ &\quad + m_1(C)m_2(A \cup B \cup C) = 0.7. \end{aligned}$$

We see clearly that  $m_{1,2}^{\text{IF-}\cap}(\cdot) \neq m_{1,2}^{\cup}(\cdot)$  in this example, so the property  $m_{1,2}^{\text{IF-}\cap}(\cdot) = m_{1,2}^{\cup}(\cdot)$  is not always satisfied. This means that De Morgan's law does not hold in general in information fusion. More precisely, the direct disjunctive fusion of BBAs is not necessarily equivalent to the negator of the conjunctive fusion of negators. Similarly, the direct conjunctive fusion of BBAs is not necessarily equivalent to the negator of the disjunctive fusion of the negators.

**Remark 2:** The negation of a BBA does not necessarily increase the entropy contrary to what is claimed in the literature in some papers cited in Section III. To prove this, just consider the FoD  $\Theta = \{A, B, C\}$  and the BBA  $m(\cdot)$  given by

$$m(A \cup B) = 0.7, m(A \cup C) = 0.2, m(A \cup B \cup C) = 0.1.$$

It can be easily verified that the entropy of the BBA  $m(\cdot)$  obtained by the formula (4) of effective entropy definition is (expressed in nats)

$$U(m) \approx 4.299.$$

The negator of  $m(\cdot)$  based on the definition (19) is

$$\bar{m}(C) = 0.7, \bar{m}(B) = 0.2, \bar{m}(A \cup B \cup C) = 0.1,$$

whose entropy is (expressed in nats)

$$U(\bar{m}) \approx 1.254.$$

One sees that one has  $U(\bar{m}) < U(m)$  in this simple example. Therefore, the negation of a BBA  $m(\cdot)$  does not necessarily increase the entropy. This really depends on the distribution of the mass of belief committed to focal elements of the original BBA  $m(\cdot)$ .

### VIII. MANAGEMENT OF DIRECT AND INDIRECT FUSIONS

As shown in the examples of the Section VI the results obtained with direct fusion approach and indirect fusion approach do not coincide but in very particular cases. In general, we have  $m_{1,2,\dots,S}^{\text{IF}} \neq m_{1,2,\dots,S}^{\text{DF}}$ . Therefore, at this stage of our research work we are facing to a new problem: what to do with these two fusion results  $m_{1,2,\dots,S}^{\text{DF}}(\cdot)$  and  $m_{1,2,\dots,S}^{\text{IF}}(\cdot)$  for decision-making support? This section provides two possible answers to this important question.

#### A. Answer 1: Fuse $m_{1,2,\dots,S}^{\text{DF}}$ with $m_{1,2,\dots,S}^{\text{IF}}$

The first intuitive answer to the aforementioned question would consist in fusing (i.e. combining) the two fusion results  $m_{1,2,\dots,S}^{\text{DF}}$  with  $m_{1,2,\dots,S}^{\text{IF}}$  by some chosen appropriate rule of combination, typically the PCR6 rule (or the PCR6<sup>+</sup> rule if  $S > 2$ , see [26]). This first answer is unfortunately not very satisfactory and not recommended from the theoretical standpoint because the fusion results  $m_{1,2,\dots,S}^{\text{DF}}$  and  $m_{1,2,\dots,S}^{\text{IF}}$  are actually based on exactly the same original inputs corresponding to BBAs  $m_1(\cdot)$ ,  $m_2(\cdot)$ , ...,  $m_S(\cdot)$ . Therefore, the inputs  $m_{1,2,\dots,S}^{\text{DF}}$  and  $m_{1,2,\dots,S}^{\text{IF}}$  cannot be considered as (cognitively) independent and their fusion is not recommended because of redundant information which may generate some biases in the final result, and induce decision-making mistakes.

If this approach is however used in applications by some users, we suggest at least to take into account the quality of each source  $m_{1,2,\dots,S}^{\text{DF}}$  and  $m_{1,2,\dots,S}^{\text{IF}}$  characterized somehow by their entropies  $U(m_{1,2,\dots,S}^{\text{DF}})$  and  $U(m_{1,2,\dots,S}^{\text{IF}})$ .

A very simple fusion method would consist for instance to apply the weighted averaging fusion of  $m_{1,2,\dots,S}^{\text{DF}}$  with  $m_{1,2,\dots,S}^{\text{IF}}$  defined for any  $X \in 2^\Theta$  by

$$m(X) = w^{\text{DF}} m_{1,2,\dots,S}^{\text{DF}}(X) + w^{\text{IF}} m_{1,2,\dots,S}^{\text{IF}}(X), \quad (28)$$

where the importance weighting factors  $w^{\text{DF}}$  and  $w^{\text{IF}}$  belong to  $[0, 1]$  and satisfy  $w^{\text{DF}} + w^{\text{IF}} = 1$ . These factors should depend on the quality of the BBAs  $m_{1,2,\dots,S}^{\text{DF}}$  and  $m_{1,2,\dots,S}^{\text{IF}}$  which is related to their entropy. Hence we could take

$$w^{\text{DF}} = \frac{U_v - U(m_{1,2,\dots,S}^{\text{DF}})}{U_v - U(m_{1,2,\dots,S}^{\text{DF}}) + U_v - U(m_{1,2,\dots,S}^{\text{IF}})},$$

$$w^{\text{IF}} = \frac{U_v - U(m_{1,2,\dots,S}^{\text{IF}})}{U_v - U(m_{1,2,\dots,S}^{\text{DF}}) + U_v - U(m_{1,2,\dots,S}^{\text{IF}})},$$

where  $U_v$  is the maximum value of the entropy corresponding to the vacuous BBA  $m_v(\cdot)$ . This max value is given by (see [5] for details)

$$U_v = 2^{|\Theta|} - 2$$

Other fusion methods based on discounting techniques and entropies could be eventually developed also, but fundamentally we do not recommend to combine  $m_{1,2,\dots,S}^{\text{DF}}$  with

$m_{1,2,\dots,S}^{\text{IF}}$  for the aforementioned reason of underlying dependency of original BBAs that have been used to generate direct and indirect fusion results  $m_{1,2,\dots,S}^{\text{DF}}$  and  $m_{1,2,\dots,S}^{\text{IF}}$ .

#### B. Answer 2: Select either $m_{1,2,\dots,S}^{\text{DF}}$ or $m_{1,2,\dots,S}^{\text{IF}}$

Because we consider that the intuitive previous answer is not satisfactory, we need to seriously consider a second option of management of direct and indirect fusion results  $m_{1,2,\dots,S}^{\text{DF}}$  and  $m_{1,2,\dots,S}^{\text{IF}}$ . This second option consists in selecting only one BBA  $m_{1,2,\dots,S}^{\text{DF}}$  or  $m_{1,2,\dots,S}^{\text{IF}}$  for decision-making support. But which one to select? How?

For selecting the BBA  $m_{1,2,\dots,S}^{\text{DF}}$  or  $m_{1,2,\dots,S}^{\text{IF}}$  we propose to adopt the maximum entropy principle which states we should select the BBA which leaves us the largest remaining uncertainty. More precisely, we will select  $m_{1,2,\dots,S}^{\text{DF}}$  if  $U(m_{1,2,\dots,S}^{\text{DF}}) > U(m_{1,2,\dots,S}^{\text{IF}})$ , and we will select  $m_{1,2,\dots,S}^{\text{IF}}$  if  $U(m_{1,2,\dots,S}^{\text{IF}}) > U(m_{1,2,\dots,S}^{\text{DF}})$ . In the very rare cases where  $m_{1,2,\dots,S}^{\text{DF}} = m_{1,2,\dots,S}^{\text{IF}}$ , no selection is needed because the two BBAs  $m_{1,2,\dots,S}^{\text{DF}}$  and  $m_{1,2,\dots,S}^{\text{IF}}$  coincide. This maximum entropy principle is rather simple to use in practice because we need only to calculate the entropies  $U(m_{1,2,\dots,S}^{\text{DF}})$  and  $U(m_{1,2,\dots,S}^{\text{IF}})$ .

We now provide more details on how to proceed in the interesting examples considered in Section VI.

#### For Zadeh's example (two Bayesian BBAs)

- With direct fusion using DS rule:

We obtain the Bayesian BBA  $m_{1,2}^{\text{DS}}(C) = 1$ . The entropy of  $m_{1,2}^{\text{DS}}$  is  $U(m_{1,2}^{\text{DS}}) = 0$  nats. This stipulates that there is no uncertainty carried by this very specific BBA which is a counter-intuitive result as explained in [28].

- With direct fusion using PCR6 rule

We obtain

$$m_{1,2}^{\text{PCR6}}(A) = m_{1,2}^{\text{PCR6}}(B) = 0.486, m_{1,2}^{\text{PCR6}}(C) = 0.028.$$

The entropy of this Bayesian BBA  $m_{1,2}^{\text{PCR6}}$  based on the formula (4) is  $U(m_{1,2}^{\text{PCR6}}) \approx 0.8014$  nats.

- With indirect fusion approach:

We obtain (see section VI)

$$m_{1,2}^{\text{IF} \cap}(A \cup B) = 0.81,$$

$$m_{1,2}^{\text{IF} \cap}(A \cup C) = 0.09,$$

$$m_{1,2}^{\text{IF} \cap}(B \cup C) = 0.09,$$

$$m_{1,2}^{\text{IF} \cap}(C) = 0.01.$$

The entropy of this non-Bayesian BBA  $m_{1,2}^{\text{IF} \cap}$  based on the formula (4) is  $U(m_{1,2}^{\text{IF} \cap}) \approx 3.8714$  nats.

Clearly, the BBA to use for decision-making support corresponds to the indirect fusion result  $m_{1,2}^{\text{IF} \cap}$  because  $U(m_{1,2}^{\text{IF} \cap}) > U(m_{1,2}^{\text{PCR6}})$ . From the selected BBA  $m_{1,2}^{\text{IF} \cap}$  the final decision can be done thanks to different techniques that are detailed in [35]. In Zadeh's example the hypothesis  $C$  will be rejected, even

if there is a tie between  $A$  and  $B$ . This tie can be eliminated arbitrarily (if we want), or randomly by a uniform random draw (i.e. perfect coin tossing) between  $A$  and  $B$ .

*For Dezert-Tchamova example (two non-Bayesian BBAs)*

We consider the example 7 given in section VI-B.

- For direct fusion with DS rule:

We have

$$\begin{aligned} m_{1,2}^{\text{DS}}(A) &= 0.3 = m_1(A), \\ m_{1,2}^{\text{DS}}(A \cup B) &= 0.7 = m_1(A \cup B). \end{aligned}$$

The entropy of  $m_{1,2}^{\text{DS}}$  is  $U(m_{1,2}^{\text{DS}}) \approx 0.6108$  nats.

- For direct fusion with PCR6 rule:

We have

$$\begin{aligned} m_{1,2}^{\text{PCR6}}(A) &= 0.2062, \\ m_{1,2}^{\text{PCR6}}(A \cup B) &= 0.5542, \\ m_{1,2}^{\text{PCR6}}(C) &= 0.2396. \end{aligned}$$

The entropy of  $m_{1,2}^{\text{PCR6}}$  is  $U(m_{1,2}^{\text{PCR6}}) \approx 2.917$  nats.

- For indirect fusion with DS rule:

We have

$$\begin{aligned} m_{1,2}^{\text{IF-DS}}(A \cup B) &\approx 0.63, \\ m_{1,2}^{\text{IF-DS}}(A \cup C) &\approx 0.23, \\ m_{1,2}^{\text{IF-DS}}(A) &\approx 0.14. \end{aligned}$$

The entropy of  $m_{1,2}^{\text{IF-DS}}$  is  $U(m_{1,2}^{\text{IF-DS}}) \approx 3.4175$  nats.

- For indirect fusion with PCR6 rule:

We have

$$\begin{aligned} m_{1,2}^{\text{IF-PCR6}}(A) &= 0.0900, \\ m_{1,2}^{\text{IF-PCR6}}(C) &= 0.1458, \\ m_{1,2}^{\text{IF-PCR6}}(A \cup B) &= 0.6142, \\ m_{1,2}^{\text{IF-PCR6}}(A \cup C) &= 0.1500. \end{aligned}$$

The entropy of  $m_{1,2}^{\text{IF-PCR6}}$  is  $U(m_{1,2}^{\text{IF-PCR6}}) \approx 3.4358$  nats.

One sees that if DS rule is used by the user (for his own reason) and because  $U(m_{1,2}^{\text{IF-DS}}) > U(m_{1,2}^{\text{DS}})$ , it will be more reasonable for the user to select  $m_{1,2}^{\text{IF-DS}}$  rather than  $m_{1,2}^{\text{DS}}$  to draw the final decision. Because we do not recommend DS fusion rule in general due to its bad dictatorial behavior, we will actually select  $m_{1,2}^{\text{IF-PCR6}}$  for decision-making because  $U(m_{1,2}^{\text{IF-PCR6}}) > U(m_{1,2}^{\text{PCR6}})$ . For this example and based on  $m_{1,2}^{\text{IF-PCR6}}$  we will finally decide  $A$  because  $m_{1,2}^{\text{IF-PCR6}}$  is closest to the sure BBA defined by  $m_A(A) = 1$  than to the sure BBAs defined by  $m_B(B) = 1$  and by  $m_C(C) = 1$ . More precisely, for this numerical example we get  $d_{\text{BI}}(m_{1,2}^{\text{IF-PCR6}}, m_A) = 0.5019$ ,  $d_{\text{BI}}(m_{1,2}^{\text{IF-PCR6}}, m_B) = 0.6456$  and  $d_{\text{BI}}(m_{1,2}^{\text{IF-PCR6}}, m_C) = 0.7093$ , where  $d_{\text{BI}}(\cdot, \cdot)$  is the Euclidean belief interval distance between two BBAs, see [35]

for details. Note that the same decision  $A$  will be drawn incidentally from  $m_{1,2}^{\text{IF-DS}}$ .

*For Blackman's example (Bayesian and non-Bayesian BBAs)*

For the simple Blackman's example of Section VI-C we have

$$\begin{aligned} m_{1,2}^{\text{DS}}(A) &= m_{1,2}^{\text{IF-DS}}(A) = m_{1,2}^{\text{PCR6}}(A) = m_{1,2}^{\text{IF-PCR6}}(A) = 0.5, \\ m_{1,2}^{\text{DS}}(B) &= m_{1,2}^{\text{IF-DS}}(B) = m_{1,2}^{\text{PCR6}}(B) = m_{1,2}^{\text{IF-PCR6}}(B) = 0.5. \end{aligned}$$

Therefore there is no BBA selection to do because all coincide and we have

$$U(m_{1,2}^{\text{DS}}) = U(m_{1,2}^{\text{IF-DS}}) \approx 0.6931 \text{ nats},$$

and

$$U(m_{1,2}^{\text{PCR6}}) = U(m_{1,2}^{\text{IF-PCR6}}) \approx 0.6931 \text{ nats}.$$

Because all the masses of belief of  $A$  and  $B$  are equal there is no way to make a rational decision towards  $A$ , or towards  $B$ . The final decision-making in this situation (where there is a tie) can be done based either on an arbitrary choice between  $A$  and  $B$ , or by a (uniform) random choice between  $A$  and  $B$  based on a perfect coin tossing experiment. Eventually in a given practical fusion problem (for instance in a tracking application) where a tie occurs we would estimate the main consequences generated by the arbitrary (or random) decision chosen (in term of costs and benefits for instance) to select the best one. This tie elimination method needs of course extra knowledge about the problem under concern. This goes beyond the scope of this paper.

For the extended Blackman's example of Section VI-C the direct DS fusion and the direct PCR6 fusion give the same following result

$$\begin{aligned} m_{12}^{\text{DS}}(A) &= m_{12}^{\text{PCR6}}(A) = 1/3, \\ m_{12}^{\text{DS}}(B) &= m_{12}^{\text{PCR6}}(B) = 1/3, \\ m_{12}^{\text{DS}}(C) &= m_{12}^{\text{PCR6}}(C) = 1/3. \end{aligned}$$

Therefore,

$$U(m_{1,2}^{\text{DS}}) = U(m_{1,2}^{\text{PCR6}}) \approx 1.0986 \text{ nats}.$$

If we apply the indirect fusion approach, we obtain for this extended Blackman's example

$$m_{1,2}^{\text{IF-DS}}(\cdot) = m_{1,2}^{\text{IF-PCR6}}(\cdot) = m_{12}^{\text{IF-}\cap}(\cdot),$$

where

$$\begin{aligned} m_{12}^{\text{IF-}\cap}(A) &= 0.8/3, \\ m_{12}^{\text{IF-}\cap}(B) &= 0.8/3, \\ m_{12}^{\text{IF-}\cap}(C) &= 0.8/3, \\ m_{12}^{\text{IF-}\cap}(A \cup B) &= 0.2/3, \\ m_{12}^{\text{IF-}\cap}(A \cup C) &= 0.2/3, \\ m_{12}^{\text{IF-}\cap}(B \cup C) &= 0.2/3. \end{aligned}$$

Therefore,

$$U(m_{1,2}^{\text{IF-DS}}) = U(m_{1,2}^{\text{IF-PCR6}}) \approx 2.0524 \text{ nats}.$$

We observe that  $U(m_{1,2}^{\text{IF-PCR6}}) > U(m_{1,2}^{\text{PCR6}})$ . This inequality indicates that the BBA  $m_{1,2}^{\text{IF-PCR6}}(\cdot)$  (which coincides also with  $m_{1,2}^{\text{IF-DS}}(\cdot)$  and  $m_{1,2}^{\text{IF-}\cap}(\cdot)$ ) is selected for the decision-making support. Because of the same repartition of mass of belief committed to  $A$ ,  $B$  and  $C$  and their disjunctions there is no way to make a rational decision towards  $A$ ,  $B$ , or towards  $C$  in this very particular tied situation unless an arbitrary or random decision strategy is adopted. So, there is no real advantage of using indirect fusion w.r.t. direct fusion in Blackman's example. However, things could obviously become different if a third source of evidence enters in the fusion problem.

## IX. CONCLUSION

In this paper we have analyzed different definitions of a negator of a probability mass function, and of a basic belief assignment (BBA) existing so far in the literature. In order to overcome their limitations we have introduced a new involutory negator of BBA. Based on it, a new indirect information fusion method was proposed which can circumvent the conflict management problem in difficult fusion situations. The classical direct and the new indirect information fusion strategies were analyzed for three interesting examples of fusion of two BBAs. In order to manage properly these two types of fusion strategies, two methods for using the whole available information (the original BBAs and their negators) for decision-making support were presented. The first method is based on the combination of the direct and indirect fusion strategies. The second one selects the most reasonable fusion strategy (direct, or indirect) to apply based on the maximum entropy principle. A deep analysis of the advantages and drawbacks of these two methods was made. We will evaluate these new fusion approaches in different fields of applications (multi-sensor data association for tracking, multi-criteria decision-making under uncertainty, perception in robotics, risk assessment, etc) in our future research works. We also invite the users of belief functions and the fusion system designers to share and report their evaluation of this new approach on their own applications in future publications.

## REFERENCES

- [1] J. Dezert, A. Tchamova, *Involutory Negator of Basic Belief Assignments*, Cybernetics and Information Technologies (CIT) journal, Sofia, Vol. 23(3), Sept. 2023.
- [2] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [3] D. Dubois, H. Prade, *A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets*, International Journal of General Systems, Vol. 12, pp. 193–226, 1986.
- [4] R. Yager, *On the Maximum Entropy Negation of a Probability Distribution*, IEEE Trans. on Fuzzy Systems, Vol. 23, No. 5, pp. 1899–1902, Oct. 2015.
- [5] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, Fusion 2022 Conf., Linköping, Sweden, pp. 1–10, July, 2022.
- [6] J. Dezert, A. Tchamova, *On Effectiveness of Measures of Uncertainty of Basic Belief Assignments*, Information & Security Journal: An International Journal (ISIJ), Vol. 52, Feb. 2022.
- [7] C.E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal, Vol. 27, pp. 379–423 & 623–656, July & October 1948 (reprinted in [8]).
- [8] N.J.A. Sloane, A.D. Wyner (Editors), *Claude Elwood Shannon - Collected Papers*, IEEE Press, 924 pages, 1993.
- [9] P. Smets, *The application of the matrix calculus to belief functions*, Int. J. Approx. Reason., Vol. 31, pp. 1–30, 2002.
- [10] I.Z. Batyrshin, N.I. Kubysheva, V.R. Bayrasheva, O. Kosheleva, V. Kreinovich, *Negations of Probability Distributions: A Survey*, Computación y Sistemas, Vol. 25, No. 4, pp. 775–781, 2021.
- [11] A. Srivastava, S. Maheshwari, *Some New Properties of Negation of a Probability Distribution*, Int. J. of Intelligent Systems, Vol. 33(6), pp. 1133–1145, June 2018.
- [12] A. Srivastava, L. Kaur, *Uncertainty and negation – Information theoretic applications*, Int. J. of Intelligent Systems, Vol. 34(6), pp. 1248–1260, June 2019.
- [13] J. Zhang, R. Liu, J. Zhang, B. Kang, *Extension of Yager's negation of a probability distribution based on Tsallis entropy*, International Journal of Intelligent Systems, Vol. 35(1), pp. 72–84, 2020.
- [14] Q. Wu, Y. Deng, N. Xiong, *Exponential negation of a probability distribution*, Soft Computing, Vol. 26, pp. 2147–2156, 2022.
- [15] L. Yin, X. Deng, Y. Deng, *The negation of a basic probability assignment*, IEEE Trans. Fuzzy Syst., Vol. 27, No. 1, pp. 135–143, January 2019.
- [16] X. Gao, Y. Deng, *The negation of basic probability assignment*, IEEE Access 7, 107006–107014, 2019.
- [17] K. Xie, F. Xiao, *Negation of belief function based on the total uncertainty measure*, Entropy, Vol. 21(1), pp. 73, 2019.
- [18] Z. Luo, Y. Deng, *A matrix method of basic belief assignment's negation in Dempster-Shafer theory*, IEEE Trans. on Fuzzy Systems, Vol. 28(9), pp. 2270–2276, 2020.
- [19] X. Deng, W. Jiang, *On the negation of a Dempster-Shafer belief structure based on maximum uncertainty allocation*, Information Sciences, Vol. 516, pp. 346–352, 2020.
- [20] I.Z. Batyrshin, *Contracting and involutory negations of probability distributions*, Mathematics, 2021, 9, 2389. arXiv preprint arXiv:2103.16176.
- [21] I.Z. Batyrshin et al., *Generating negations of probability distributions*, Soft Computing, Vol. 25, pp. 7929–7935, 2021.
- [22] R. Liu, Y. Deng, Z. Li, *The maximum entropy negation of basic probability assignment*, Soft Computing, 2023 (published online 12 April 2023).
- [23] P. Smets, R. Kennes, *The transferable belief model*, Artificial Intelligence, Vol. 66, No. 2, pp. 191–234, 1994.
- [24] J. Dezert, F. Smarandache, *A new probabilistic transformation of belief mass assignment*, Proc. of Fusion 2008 Int. Conf., K'oln, Germany pp. 1–8, June 30th–July 3rd, 2008.
- [25] F. Smarandache, J. Dezert (Editors), *Advances and applications of DSMT for information Fusion (Collected works)*, Vol. 2–4, American Research Press, 2006, 2009 & 2015.
- [26] T. Dezert, J. Dezert, F. Smarandache, *Improvement of Proportional Conflict Redistribution Rules of Combination of Basic Belief Assignments*, in Journal of Advances in Information Fusion (JAIF), Vol. 16, No. 1, June 2021.
- [27] F. Smarandache, *A in-depth look at quantitative information fusion rules*, in [25], Chap. 8, Vol. 2, pp. 205–236, June 2009.
- [28] L.A. Zadeh, *On the validity of Dempster's rule of combination*, Memo M79/24, Univ. of California, Berkeley, U.S.A., 1979.
- [29] J. Dezert, P. Wang, A. Tchamova, *On the validity of Dempster-Shafer theory*, Proc. of 15th Int. Conf. on Information Fusion (Fusion 2012), pp. 655–660, Singapore, July 9–12, 2012.
- [30] J. Dezert, A. Tchamova, D. Han, *Total Belief Theorem and Generalized Bayes' Theorem*, Proc. of Fusion 2018 Conf., Cambridge, UK, July 10–13, 2018.
- [31] A. Tchamova, J. Dezert, *On the Behavior of Dempster's Rule of Combination and the Foundations of Dempster-Shafer Theory*, IEEE IS'2012, Sofia, Bulgaria, Sept. 6–8, 2012.
- [32] F. Smarandache, J. Dezert, A. Tchamova A., *Examples where Dempster's rule is insensitive to the conflict level between the sources of evidence*, Octagon Mathematical Magazine, Vol. 25, No. 2, pp. 284–290, October 2017.
- [33] S. Blackman, R. Popoli, *Design and analysis of modern tracking systems*, Artech House, 1986.
- [34] A. Tchamova, J. Dezert, F. Smarandache, *New fusion rules for solving Blackman's association problem*, Chap 15 of [25], Vol.3, pp. 425–436, 2009.
- [35] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-Making with Belief Interval Distance*, in Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21–23, 2016.

# On Kenn's Rule of Combination Applied to Breast Cancer Precision Therapy

Jean Dezert<sup>a</sup>, Albena Tchamova<sup>b</sup>

<sup>a</sup>The French Aerospace Lab, ONERA, F-91761 Palaiseau, France.

<sup>b</sup>Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria.

Emails: jean.dezert@onera.fr, tchamova@bas.bg

**Abstract**—This short paper points out an erroneous claim about a new rule of combination of basic belief assignments presented recently by Kenn et al. in [1], referred as Kenn's rule of combination (or just as KRC for short). We prove thanks a very simple counter-example that Kenn's rule is not associative. Consequently, the results of the method proposed by Kenn et al. highly depends on the ad-hoc sequential order chosen for the fusion process as proposed by the authors. This serious problem casts in doubt the interest of this method and its real ability to provide trustful results and to make good decisions to help for precise breast cancer therapy.

**Keywords:** belief functions, rule of combination, Kenn's rule.

## I. INTRODUCTION

Recently a paper devoted to the Breast Cancer Precision Therapy by Kenn et al. [1] attracted our attention for two main reasons: 1) this application of information fusion is very interesting and important; 2) Kenn's et al. method is based on a new rule of combination of basic belief assignments (BBAs). Because we did some theoretical contributions in this field [2] we wanted to examine this new combination rule in detail. So, we have read with interest Kenn's et al. paper. Unfortunately we quickly discovered a serious erroneous claim about Kenn's rule of combination (KRC) and this has strong consequences on the methodology presented by Kenn. In this short technical note we warn the readers of the risk of potential therapy errors if such a method is used in practice. We clearly explain the problem of the method presented by Kenn et al. To make the paper self-containing, we recall briefly the basics of belief functions in the next section, and the KRC in the section III. In section IV we prove thanks a very simple numerical counter-example that KRC is not associative. Conclusion and recommendations are given in the last section of this note.

## II. BELIEF FUNCTIONS

The belief functions (BF) were introduced by Shafer [3] for modeling epistemic uncertainty, reasoning about uncertainty and combining distinct sources of evidence. The answer of the problem under concern is assumed to belong to a known finite discrete frame of discernment (FoD)  $\Theta = \{\theta_1, \dots, \theta_N\}$  where all elements (i.e. members) of  $\Theta$  are exhaustive and exclusive. The set of all subsets of  $\Theta$  (including empty set

$\emptyset$ , and  $\Theta$ ) is the power-set of  $\Theta$  denoted by  $2^\Theta$ . The number of elements (i.e. the cardinality) of the power-set is  $2^{|\Theta|}$ . A (normalized) basic belief assignment (BBA) associated with a given source of evidence is a mapping  $m^\Theta(\cdot) : 2^\Theta \rightarrow [0, 1]$  such that  $m^\Theta(\emptyset) = 0$  and  $\sum_{X \in 2^\Theta} m^\Theta(X) = 1$ . A BBA  $m^\Theta(\cdot)$  characterizes a source of evidence related with a FoD  $\Theta$ . For notation shorthand, we can omit the superscript  $\Theta$  in  $m^\Theta(\cdot)$  notation if there is no ambiguity on the FoD we work with. The quantity  $m(X)$  is called the mass of belief of  $X$ . The element  $X \in 2^\Theta$  is called a focal element (FE) of  $m(\cdot)$  if  $m(X) > 0$ . The set of all focal elements of  $m(\cdot)$  is denoted<sup>1</sup> by  $\mathcal{F}_\Theta(m) \triangleq \{X \in 2^\Theta | m(X) > 0\}$ . The belief and the plausibility of  $X$  are respectively defined for any  $X \in 2^\Theta$  by [3]

$$Bel(X) = \sum_{Y \in 2^\Theta | Y \subseteq X} m(Y) \quad (1)$$

$$Pl(X) = \sum_{Y \in 2^\Theta | X \cap Y \neq \emptyset} m(Y) = 1 - Bel(\bar{X}). \quad (2)$$

where  $\bar{X} \triangleq \Theta \setminus \{X\}$  is the complement of  $X$  in  $\Theta$ .

One has always  $0 \leq Bel(X) \leq Pl(X) \leq 1$ , see [3]. For  $X = \emptyset$ ,  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and for  $X = \Theta$  one has  $Bel(\Theta) = Pl(\Theta) = 1$ .  $Bel(X)$  and  $Pl(X)$  are often interpreted as the lower and upper bounds of unknown probability  $P(X)$  of  $X$ , that is  $Bel(X) \leq P(X) \leq Pl(X)$ . To quantify the uncertainty (i.e. the imprecision) of  $P(X) \in [Bel(X), Pl(X)]$ , we use  $u(X) \in [0, 1]$  defined by

$$u(X) \triangleq Pl(X) - Bel(X) \quad (3)$$

If  $u(X) = 0$ ,  $Bel(X) = Pl(X)$  and therefore  $P(X)$  is known precisely because  $P(X) = Bel(X) = Pl(X)$ . One has  $u(\emptyset) = 0$  because  $Bel(\emptyset) = Pl(\emptyset) = 0$ , and one has  $u(\Theta) = 0$  because  $Bel(\Theta) = Pl(\Theta) = 1$ . If all focal elements of  $m(\cdot)$  are singletons of  $2^\Theta$  the BBA  $m(\cdot)$  is a Bayesian BBA because  $\forall X \in 2^\Theta$  one has  $Bel(X) = Pl(X) = P(X)$  and  $u(X) = 0$ . Hence the belief and plausibility of  $X$  coincide with a probability measure  $P(X)$  defined on the FoD  $\Theta$ . The vacuous BBA characterizing a totally ignorant source of evidence is defined by  $m_v(X) = 1$  for  $X = \Theta$ , and  $m_v(X) = 0$  for all  $X \in 2^\Theta$  different of  $\Theta$ .

<sup>1</sup> $\triangleq$  means equal by definition.

In the Mathematical Theory of Evidence of Glenn Shafer, the combination of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined over the same FoD  $\Theta$  is obtained with Dempster's rule. More precisely<sup>2</sup> by  $m_{12}^{DS}(\emptyset) = 0$ , and for any  $X \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{12}^{DS}(X) = \frac{\sum_{\substack{X_1, X_2 \subseteq \Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2)}{1 - \sum_{\substack{X_1, X_2 \subseteq \Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2)} \quad (4)$$

The value  $K_{12} = \sum_{\substack{X_1, X_2 \subseteq \Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2)$  is classically interpreted as the degree of conflict between the BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$ . When the degree of conflict is maximum one has  $K_{12} = 1$ , and in this particular case Dempster-Shafer rule cannot be applied because of division by zero in the formula (4). This rule can be easily directly generalized for the combination of more than two BBAs.

The DS rule has had a great success during the past decades in expert systems and artificial intelligence mainly because it is a commutative and associative rule of combination able to deal with (possibly epistemic) uncertainty and incomplete information based on an appealing mathematical framework. This makes its use very attractive from the implementation standpoint in decision-making support systems. Indeed, because of its associativity we can apply DS rule sequentially when we have more than two sources of evidence to fuse, and the sequence order will not impact the DS fusion result. Unfortunately, DS rule of combination generates counter-intuitive results due to the normalization step in DS formula (not only in high conflicting situations but also in low conflicting situations as well), and it generates very controversial dictatorial behaviors, see [4], [5] for discussions with examples. That is why many alternatives of DS rule have been proposed in the literature to circumvent these serious problems of DS rule. Unfortunately, there is no general consensus in the scientific community about the choice of the *best* rule of combination of belief functions to make for the applications.

### III. KENN'S RULE OF COMBINATION

Kenn's rule of combination (KRC) proposed in [1] is a slight modification of DS rule introducing a tuning parameter  $\lambda \in [0, 1]$ . The KRC of two BBAs  $m_1(\cdot)$  and  $m_2(\cdot)$  defined over the same FoD  $\Theta$  is denoted symbolically  $m_1 \oplus_\lambda m_2$  in [1]. Its mathematical expression is given by<sup>3</sup>  $m_{12}^{KRC}(\emptyset) = 0$ , for  $X \in 2^\Theta \setminus \{\emptyset\}$  by

$$m_{12}^{KRC}(X) = [m_1 \oplus_\lambda m_2](X) = \frac{\sum_{\substack{X_1, X_2 \subseteq \Theta \\ X_1 \cap X_2 = X}} m_1(X_1)m_2(X_2)}{1 - \lambda \sum_{\substack{X_1, X_2 \subseteq \Theta \\ X_1 \cap X_2 = \emptyset}} m_1(X_1)m_2(X_2)} \quad (5)$$

<sup>2</sup>DS upper index in formula (4) stands for Dempster-Shafer because this rule is often referred also as Dempster-Shafer rule of combination in the literature.

<sup>3</sup>see formula (4) in [1]

and for  $X = \Theta$ , by

$$m_{12}^{KRC}(\Theta) = [m_1 \oplus_\lambda m_2](\Theta) = 1 - \sum_{X \subset \Theta} m_{12}^{KRC}(X) \quad (6)$$

For  $\lambda = 1$ , KRC coincides with Dempster-Shafer rule and consequently it will suffer of same problems as DS rule in this particular case. According to Kenn et al., the parameter  $\lambda$  in the formula (5) provides flexibility to adapt to circumstances and the restriction to  $\lambda \leq 1$  is motivated by restricting the authors to an interpolation type evidential combination rule. KRC is claimed associative and commutative by Kenn et al. (see page 5 of [1]). We prove in the next section that KRC is in fact not associative. Because of non-associativity of KRC, the methodology proposed in [1] becomes very disputable and doubtful, and potentially very dangerous for breast cancer therapy application addressed by Kenn et al., and for any other applications using sequential fusion of sources of evidences based on KRC.

### IV. COUNTER-EXAMPLE OF ASSOCIATIVITY OF KRC

To prove that KRC is not associative it suffices to verify that

$$(m_1 \oplus_\lambda m_2) \oplus_\lambda m_3 \neq m_1 \oplus_\lambda (m_2 \oplus_\lambda m_3) \quad (7)$$

To prove (7) when  $\lambda < 1$ , just consider for instance  $\lambda = 0.2$ , the FoD  $\Theta = \{A, B\}$  and the three BBAs given in Table I

Table I  
THREE BASIC BELIEF ASSIGNMENTS.

Elements	$m_1$	$m_2$	$m_3$
$\emptyset$	0	0	0
A	0.2	0.8	0.4
B	0.7	0.1	0.3
$A \cup B$	0.1	0.1	0.3

#### A. Derivation of $(m_1 \oplus_{0.2} m_2) \oplus_{0.2} m_3$

For the combination of  $m_1$  with  $m_2$  we have the degree of conflict

$$K_{12} = m_1(A)m_2(B) + m_1(B)m_2(A) = 0.2 \cdot 0.1 + 0.7 \cdot 0.8 = 0.58$$

The results of the conjunctive fusion of  $m_1$  with  $m_2$  for A and B are

$$\begin{aligned} m_{12}(A) &= m_1(A)m_2(A) + m_1(A)m_2(A \cup B) \\ &\quad + m_2(A)m_1(A \cup B) \\ &= 0.2 \cdot 0.8 + 0.2 \cdot 0.1 + 0.8 \cdot 0.1 = 0.26 \\ m_{12}(B) &= m_1(B)m_2(B) + m_1(B)m_2(A \cup B) \\ &\quad + m_2(B)m_1(A \cup B) \\ &= 0.7 \cdot 0.1 + 0.7 \cdot 0.1 + 0.1 \cdot 0.1 = 0.15 \end{aligned}$$

For KRC of  $m_1$  with  $m_2$  we get (taking  $\lambda = 0.2$ )  $m_{12}^{KRC}(\emptyset) = 0$  and

$$\begin{aligned} m_{12}^{KRC}(A) &= [m_1 \oplus_\lambda m_2](A) \\ &= \frac{m_{12}(A)}{1 - \lambda K_{12}} = \frac{0.26}{1 - 0.2 \cdot 0.58} \approx 0.2941 \end{aligned}$$

$$\begin{aligned} m_{12}^{KRC}(B) &= [m_1 \oplus_\lambda m_2](B) = \frac{m_{12}(B)}{1 - \lambda K_{12}} \\ &= \frac{0.15}{1 - 0.2 \cdot 0.58} \approx 0.1697 \end{aligned}$$

$$\begin{aligned} m_{12}^{KRC}(A \cup B) &= [m_1 \oplus_\lambda m_2](A \cup B) \\ &= 1 - \frac{0.26}{1 - 0.2 \cdot 0.58} - \frac{0.15}{1 - 0.2 \cdot 0.58} \\ &\approx 0.5362 \end{aligned}$$

For the combination of  $m_{12}^{KRC} = m_1 \oplus_\lambda m_2$  with  $m_3$  we have the degree of conflict

$$\begin{aligned} K_{(12)3} &= m_{12}^{KRC}(A)m_3(B) + m_{12}^{KRC}(B)m_3(A) \\ &\approx 0.2941 \cdot 0.3 + 0.1697 \cdot 0.4 \approx 0.1561 \end{aligned}$$

The results of the conjunctive fusion of  $m_{12}^{KRC}$  with  $m_3$  for  $A$  and  $B$  are

$$\begin{aligned} m_{(12)3}(A) &= m_{12}^{KRC}(A)m_3(A) + m_{12}^{KRC}(A)m_3(A \cup B) \\ &\quad + m_3(A)m_{12}^{KRC}(A \cup B) \\ &\approx 0.2941 \cdot 0.4 + 0.2941 \cdot 0.3 + 0.4 \cdot 0.5362 \\ &= 0.4204 \end{aligned}$$

$$\begin{aligned} m_{(12)3}(B) &= m_{12}^{KRC}(B)m_3(B) + m_{12}^{KRC}(B)m_3(A \cup B) \\ &\quad + m_3(B)m_{12}^{KRC}(A \cup B) \\ &\approx 0.1697 \cdot 0.3 + 0.1697 \cdot 0.3 + 0.3 \cdot 0.5362 \\ &= 0.2627 \end{aligned}$$

Therefore, the KRC of  $m_{12}^{KRC}$  with  $m_3$  yields  $m_{(12)3}^{KRC}(\emptyset) = 0$  and

$$\begin{aligned} m_{(12)3}^{KRC}(A) &= [m_{12}^{KRC} \oplus_\lambda m_3](A) \\ &= \frac{m_{(12)3}(A)}{1 - \lambda K_{(12)3}} \\ &= \frac{0.4204}{1 - 0.2 \cdot 0.1561} \approx 0.4339 \end{aligned}$$

$$\begin{aligned} m_{(12)3}^{KRC}(B) &= [m_{12}^{KRC} \oplus_\lambda m_3](B) \\ &= \frac{m_{(12)3}(B)}{1 - \lambda K_{(12)3}} \\ &= \frac{0.2627}{1 - 0.2 \cdot 0.1561} \approx 0.2711 \end{aligned}$$

$$\begin{aligned} m_{(12)3}^{KRC}(A \cup B) &= [m_{12}^{KRC} \oplus_\lambda m_3](A \cup B) \\ &\approx 1 - 0.4339 - 0.2711 \approx 0.2950 \end{aligned}$$

Hence for the fusion  $(m_1 \oplus_{0.2} m_2) \oplus_{0.2} m_3$  we get finally

$$\begin{aligned} m_{(12)3}^{KRC}(A) &= [(m_1 \oplus_{0.2} m_2) \oplus_{0.2} m_3](A) \\ &\approx 0.4339 \end{aligned} \tag{8}$$

$$\begin{aligned} m_{(12)3}^{KRC}(B) &= [(m_1 \oplus_{0.2} m_2) \oplus_{0.2} m_3](B) \\ &\approx 0.2711 \end{aligned} \tag{9}$$

$$\begin{aligned} m_{(12)3}^{KRC}(A \cup B) &= [(m_1 \oplus_{0.2} m_2) \oplus_{0.2} m_3](A \cup B) \\ &\approx 0.2950 \end{aligned} \tag{10}$$

#### B. Derivation of $m_1 \oplus_{0.2} (m_2 \oplus_{0.2} m_3)$

For the combination of  $m_2$  with  $m_3$  we have

$$\begin{aligned} K_{23} &= m_2(A)m_3(B) + m_2(B)m_3(A) \\ &= 0.8 \cdot 0.3 + 0.1 \cdot 0.4 = 0.28 \end{aligned}$$

The results of the conjunctive fusion of  $m_2$  with  $m_3$  for  $A$  and  $B$  are

$$\begin{aligned} m_{23}(A) &= m_2(A)m_3(A) + m_2(A)m_3(A \cup B) \\ &\quad + m_3(A)m_2(A \cup B) \\ &= 0.8 \cdot 0.4 + 0.8 \cdot 0.3 + 0.4 \cdot 0.1 = 0.60 \end{aligned}$$

$$\begin{aligned} m_{23}(B) &= m_2(B)m_3(B) + m_2(B)m_3(A \cup B) \\ &\quad + m_3(B)m_2(A \cup B) \\ &= 0.1 \cdot 0.3 + 0.1 \cdot 0.3 + 0.3 \cdot 0.1 = 0.09 \end{aligned}$$

For KRC of  $m_2$  with  $m_3$  we get (taking  $\lambda = 0.2$ )  $m_{23}^{KRC}(\emptyset) = 0$  and

$$\begin{aligned} m_{23}^{KRC}(A) &= [m_2 \oplus_\lambda m_3](A) \\ &= \frac{m_{23}(A)}{1 - \lambda K_{23}} = \frac{0.60}{1 - 0.2 \cdot 0.28} \approx 0.6356 \end{aligned}$$

$$\begin{aligned} m_{23}^{KRC}(B) &= [m_2 \oplus_\lambda m_3](B) \\ &= \frac{m_{23}(B)}{1 - \lambda K_{23}} = \frac{0.09}{1 - 0.2 \cdot 0.28} \approx 0.0953 \end{aligned}$$

$$\begin{aligned} m_{23}^{KRC}(A \cup B) &= [m_2 \oplus_\lambda m_3](A \cup B) \\ &\approx 1 - 0.6356 - 0.0953 \approx 0.2691 \end{aligned}$$

For the combination of  $m_1$  with  $m_{23}^{KRC} = m_2 \oplus_\lambda m_3$  we have the degree of conflict

$$\begin{aligned} K_{1(23)} &= m_{23}^{KRC}(A)m_1(B) + m_{23}^{KRC}(B)m_1(A) \\ &\approx 0.63563 \cdot 0.7 + 0.0953 \cdot 0.2 \approx 0.4640 \end{aligned}$$

The results of the conjunctive fusion of  $m_1$  with  $m_{23}^{KRC}$  for  $A$  and  $B$  are

$$\begin{aligned} m_{1(23)}(A) &= m_{23}^{KRC}(A)m_1(A) + m_{23}^{KRC}(A)m_1(A \cup B) \\ &\quad + m_1(A)m_{23}^{KRC}(A \cup B) \\ &\approx 0.6356 \cdot 0.2 + 0.6356 \cdot 0.1 + 0.2 \cdot 0.2691 \approx 0.2445 \end{aligned}$$

$$\begin{aligned} m_{1(23)}(B) &= m_{23}^{KRC}(B)m_1(B) + m_{23}^{KRC}(B)m_1(A \cup B) \\ &\quad + m_1(B)m_{23}^{KRC}(A \cup B) \\ &\approx 0.0953 \cdot 0.7 + 0.0953 \cdot 0.1 + 0.7 \cdot 0.2691 \approx 0.2646 \end{aligned}$$

Therefore, the KRC of  $m_1$  with  $m_{23}^{KRC}$  yields  $m_{1(23)}^{KRC}(\emptyset) = 0$  and

$$m_{1(23)}^{KRC}(A) = [m_1 \oplus_{\lambda} m_{23}^{KRC}](A) = \frac{m_{1(23)}(A)}{1 - \lambda K_{1(23)}} = \frac{0.2445}{1 - 0.2 \cdot 0.4640} \approx 0.2695$$

$$m_{1(23)}^{KRC}(B) = [m_1 \oplus_{\lambda} m_{23}^{KRC}](B) = \frac{m_{1(23)}(B)}{1 - \lambda K_{1(23)}} = \frac{0.2646}{1 - 0.2 \cdot 0.4640} \approx 0.2917$$

$$m_{1(23)}^{KRC}(A \cup B) = [m_1 \oplus_{\lambda} m_{23}^{KRC}](A \cup B) \approx 1 - 0.2695 - 0.2917 \approx 0.4388$$

Hence for the fusion  $m_1 \oplus_{0.2} (m_2 \oplus_{0.2} m_3)$  we get finally

$$m_{1(23)}^{KRC}(A) = [m_1 \oplus_{0.2} (m_2 \oplus_{0.2} m_3)](A) \approx 0.2695 \tag{11}$$

$$m_{1(23)}^{KRC}(B) = [m_1 \oplus_{0.2} (m_2 \oplus_{0.2} m_3)](B) \approx 0.2917 \tag{12}$$

$$m_{1(23)}^{KRC}(A \cup B) = [m_1 \oplus_{0.2} (m_2 \oplus_{0.2} m_3)](A \cup B) \approx 0.4388 \tag{13}$$

We see clearly that KRC is not associative because  $(m_1 \oplus_{\lambda} m_2) \oplus_{\lambda} m_3 \neq m_1 \oplus_{\lambda} (m_2 \oplus_{\lambda} m_3)$  as reported in the Table II.

Table II  
COUNTER-EXAMPLE OF ASSOCIATIVITY OF KRC WITH  $\lambda = 0.2$ .

Elements	$(m_1 \oplus_{\lambda} m_2) \oplus_{\lambda} m_3$	$m_1 \oplus_{\lambda} (m_2 \oplus_{\lambda} m_3)$
$\emptyset$	0	0
A	0.4339	0.2695
B	0.2711	0.2917
$A \cup B$	0.2950	0.4388

C. Comment on decision-making method used by Kenn et al.

For our simple example we get with the sequential fusion  $(m_1 \oplus_{0.2} m_2) \oplus_{0.2} m_3$  the following belief intervals

$$[Bel_{(12)3}(\emptyset), Pl_{(12)3}(\emptyset)] = [0, 0]$$

$$[Bel_{(12)3}(A), Pl_{(12)3}(A)] = [0.4339, 0.7289]$$

$$[Bel_{(12)3}(B), Pl_{(12)3}(B)] = [0.2711, 0.5661]$$

$$[Bel_{(12)3}(A \cup B), Pl_{(12)3}(A \cup B)] = [1, 1]$$

and with the sequential fusion  $m_1 \oplus_{0.2} (m_2 \oplus_{0.2} m_3)$

$$[Bel_{1(23)}(\emptyset), Pl_{1(23)}(\emptyset)] = [0, 0]$$

$$[Bel_{1(23)}(A), Pl_{1(23)}(A)] = [0.2695, 0.7083]$$

$$[Bel_{1(23)}(B), Pl_{1(23)}(B)] = [0.2917, 0.7305]$$

$$[Bel_{1(23)}(A \cup B), Pl_{1(23)}(A \cup B)] = [1, 1]$$

Based on these results and the decision-making method used by Kenn et al. (see section 3.4 of [1]) it is clear that no decision for  $A$  or for  $B$  can be made using the sequential fusion  $(m_1 \oplus_{0.2} m_2) \oplus_{0.2} m_3$  because we have neither

$Bel_{(12)3}(A) > Pl_{(12)3}(B)$ , nor  $Bel_{(12)3}(B) > Pl_{(12)3}(A)$ . Similarly, no decision can be drawn for  $A$  or for  $B$  from the result of the sequential fusion  $m_1 \oplus_{0.2} (m_2 \oplus_{0.2} m_3)$  because we have neither  $Bel_{1(23)}(A) > Pl_{1(23)}(B)$ , nor  $Bel_{1(23)}(B) > Pl_{1(23)}(A)$ . In fact we just could always take as final decision based on Kenn's decision-making method the whole frame of discernment because one always has  $(Bel_{(12)3}(A \cup B) = 1) > (Pl_{(12)3}(\emptyset) = 0)$  and  $(Bel_{1(23)}(A \cup B) = 1) > (Pl_{1(23)}(\emptyset) = 0)$  but such type of decision is obviously not useful at all for the applications because it does not help to make a clear choice between  $A$  and  $B$ . So, the decision-making method used by Kenn et al. does not work for all cases of BBA distributions as shown in this very simple example, and that is why it is not judicious and not recommended for applications.

V. CONCLUSION

The consequence of non-associativity of the method presented by Kenn et al. in [1] can have a strong impact on the results and on decision-making in general if the KRC is applied sequentially for information fusion as it is proposed by the authors in their paper (see formula (9) page 7 of [1]). Because of this problem, we have a serious concern about the interest and the effectiveness of the method presented by Kenn et al.. We warn the potential users of this approach about the high risk of wrong decisions (when they are possible which is not always the case as shown in our counter-example) based on this method. This could have dramatical therapy consequences. If the authors want to use this KRC-based approach we think they should at least better consider a global information fusion processing than a sequential one, and they should adopt a better decision-making strategy. They also should compare their results with other advanced rules of combination and use the same decision strategy to make comparisons to show the real advantages of this approach, if any. The measure of the performances of the method with real open data sets for breast cancer therapy application and ground truth is also recommended.

REFERENCES

- [1] M. Kenn, R. Karch, C.F Singer, G. Dorfner, W. Schreiner, *Flexible Risk Evidence Combination Rules in Breast Cancer Precision Therapy*, J. Pers. Med. 2023, 13, 119. <https://doi.org/10.3390/jpm13010119>
- [2] <https://www.onera.fr/fr/staff/jean-dezert/references>
- [3] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [4] L.A. Zadeh, *A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination*, The AI Magazine, Vol. 7 (2), pp. 85–90, 1986.
- [5] J. Dezert, P. Wang, A. Tchamova, *On The Validity of Dempster-Shafer Theory*, in Proc. of Fusion 2012, Singapore, July 2012.

# Weighted Fusion of Multiple Classifiers for Human Activity Recognition

Kezhu Zuo<sup>a</sup>, Xinde Li<sup>b</sup>, Jean Dezert<sup>c</sup>, Yilin Dong<sup>d</sup>

<sup>a</sup>School of Cyber Science and Engineering, Southeast University, Nanjing, China.

<sup>b</sup>Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing, China.

<sup>c</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

<sup>d</sup>College of Computer Science and Engineering, Shanghai Maritime University, Shanghai, China.

Emails: kezhuzuo@seu.edu.cn, xindeli@seu.edu.cn, jean.dezert@onera.fr, yldong@shmtu.edu.cn

Originally published as: K. Zuo, X. Li, J. Dezert, Y. Dong, *Weighted Fusion of Multiple Classifiers for Human Activity Recognition*, in Proc. of Int. Conf. on Information Fusion (Fusion 2023), Charleston, SC, USA, June 2023, and reprinted with permission.

**Abstract**—Human Activity Recognition (HAR) based on wearable device has become a hot topic of research due to its wide range of applications in health-care, fitness and smart homes. However, the classification of some activities with similar sensor readings, such as standing and sitting, is usually more challenging for the design of efficient activity recognition algorithms. Considering the inconsistent performance of different classifiers, which can provide information complementary for individual classifier, we propose a novel multi-classifier fusion method based on belief functions (BFs) theory for HAR. Specifically, at first, four classifiers are trained using time-domain and frequency-domain features to obtain basic belief assignments (BBA) of activity, respectively. Then, three assessment criteria are utilized to evaluate the reliability of the classifiers and a scoring matrix is constructed. Next, the algorithm of Belief Function based the Technique for Order Preference by Similarity to Ideal Solution (BF-TOPSIS) is employed to calculate the weighting coefficients for each classifier. Finally, the discounting and Dempster's rules are adopted to combine the multiple classifiers and further decision making. Several experiments were conducted to illustrate the performance of the proposed method using the UCI smartphone dataset, and the results show that the proposed method is more accurate than the state-of-art methods.

**Index Terms**—Belief functions theory, multiple classifiers fusion, BF-TOPSIS, human activity recognition.

## I. INTRODUCTION

With the booming development of micro-sensor technology, Human Activity Recognition (HAR) based on wearable sensors has become one of the hot research topics [1], [2]. Data of daily activities can be well collected in an all-round and non invasive discrete manner using accelerometers, gyroscopes and other such portable wearable devices, so as to accomplish the work of assisted living and health monitoring while effectively protecting the privacy of users [3]. Obviously, it has certain advantages compared to traditional vision-based methods. However, the accuracy of HAR based on wearable devices is affected by many factors, such as the number and the deployment location of sensors, the complexity of activities [4], and so on. Due to the uncertainty, diversity and individual differences of activities [5], many scholars took the perspective of multi-sensor information fusion to achieve higher accuracy of HAR. For example, Dong et al. [6] developed the kernel density estimation models to fit the multi-sensor data to obtain the basic belief assignments

(BBAs), and then Dezert-Smarandache theory (DSMT) was adopted to combine the acquired BBAs. Uddin et al. [7] fused data from different multimodal sensors with statistical features of different orders and then trained a deep recurrent neural network (RNN) for activity recognition. Although they achieve good accuracy, it is still difficult to accurately identify some activities with high similarity of sensor readings such as sitting and standing. Furthermore, the reliability of activity recognition can be significantly compromised when sensor readings are missing or disturbed by noise without additional sensor information.

Recently, the multi-classifier fusion has been applied in pattern recognition [8], information fusion [9], [10] and other fields, especially for classification problems in complex environments. Different classifiers can learn different feature information, and multiple classifiers can provide complementary information compared with any individual classifier, which can help identify similar human activities such as sitting and standing. By using multi-classifier fusion, we expect the improvement of the classification accuracy, which brings the possibility of high precision HAR. On the other hand, multiple classifiers can be seen as multiple sources of evidence, and we fuse the basic belief assignments (BBAs) of the human activity categories output by the classifiers.

The multi-classifier fusion usually consists of generating membership classifiers, applying combination rules, and make a decision about the positioning of the patient. Various approaches have been proposed for membership classifier generation, for example, using different training samples, different features and different types of classifiers [11]. Common classifier fusion methods include voting method [12], naive Bayes [13], Dempster-Shafer (DS) rule in Dempster-Shafer theory (DST) [14], and so on. In the fusion process, the classifiers may have different reliabilities (weights) and their decision results may be contradicting, which inevitably brings conflict issues. In order to improve classification accuracy, it becomes particularly important to evaluate the reliability of classifiers before combining them. For instance, Liu et al. employed contextual reliability evaluation based on inner reliability and relative reliability concepts [10]. Dong et al. [15] took two classes of criteria into account to evaluate the classifiers. The

first class is the conflict between the classifiers and the second class is the imprecision of the information provided by each classifier. The effective evaluation of the reliability of multiple classifiers and their fusion is a challenging problem for HAR tasks.

In this article, we propose a novel Weighted Fusion of Multiple Classifiers (WFMC) method for HAR based on BFs theory. Our main contributions are summarized as follows:

- Four classifiers including support vector machine (SVM), random forest (RF), multi-layer perceptron (MLP) and logistic regression (LR) are trained by same training dataset for acquiring BBAs of human activities. To improve the multi-classifier fusion accuracy, Belief Jensen–Shannon (BJS) divergence, Interval distance function and belief entropy are considered to measure the reliability of the classifiers and a scoring matrix is constructed.
- The BF-TOPSIS<sup>1</sup> multi-criteria decision-making algorithm is employed to calculate the weighting coefficients for each classifier, and multiple classifiers are fused using discounting technique and DS rule in this work, the final decision is made based on the maximum belief mass of all involved single focal elements.
- We evaluate the performance of our proposed method on the widely used UCI Smart-phone public dataset.

The rest of this article is organized as follows. Section II presents the basic concepts of BFs theory, discounting technique and pignistic probability transformation. Section III provides a detailed description of the new proposed multi-classifier fusion strategy for HAR. Section IV presents the detailed experimental results and discussions. The final section V gives concluding remarks with some perspectives of this work.

## II. PRELIMINARIES

### A. Belief Functions Theory

BFs theory (known also as DST) has been widely used in multi-sensor information fusion due to its ability to deal with uncertain and imprecise information [17]. The basic concepts are introduced in this section based on [14]. Let  $\Theta$  be a finite set of elements denoted by

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}. \quad (1)$$

The set  $\Theta$  is called a frame of discernment (FoD), which consists of exhaustive and exclusive hypotheses. Information sources distribute mass of belief to elements of the power set of the FoD, denoted by  $2^\Theta$ . For example, if  $\Theta = \{\theta_1, \theta_2\}$ , then

$$2^\Theta = \{\emptyset, \theta_1, \theta_2, \theta_1 \cup \theta_2\}. \quad (2)$$

A BBA, called a mass function, is defined by the mapping  $m(\cdot) : 2^\Theta \mapsto [0, 1]$ , which satisfies  $m(\emptyset) = 0$  and

$$\sum_{A \in 2^\Theta} m(A) = 1. \quad (3)$$

<sup>1</sup>BF-TOPSIS is an extension of the technique for order preference by similarity to ideal solution (TOPSIS) based on belief functions (BF) [16].

For a proposition  $A \subseteq \Theta$ , the belief function is defined as:

$$Bel(A) = \sum_{B \subseteq A, B \in 2^\Theta} m(B). \quad (4)$$

The plausibility function is defined as:

$$Pl(A) = \sum_{B \cap A \neq \emptyset, B \in 2^\Theta} m(B). \quad (5)$$

If the focal elements of BBA are all singletons, the BBA is called Bayesian BBA [14]. In pattern classification,  $m(A)$  represents the support degree of the object associated with class. For example, if  $A$  is a set of classes (e.g.,  $A = \{\theta_1, \theta_2\}$ ),  $m(A)$  denotes the possibility of classification among the class  $\theta_1$  and  $\theta_2$  with respect to the object. In DST, the classical Dempster's rule (also called Dempster-Shafer rule, or just DS rule) is used to combine two (or more<sup>2</sup>) independent Sources of Evidence (SoEs), which is denoted as  $m_1 \oplus m_2$  and defined as follows [14]: for  $\forall A \in 2^\Theta, A \neq \emptyset$ ,

$$(m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B, C \in 2^\Theta | B \cap C = A} m_1(B) m_2(C) \quad (6)$$

with

$$k = \sum_{B, C \in 2^\Theta | B \cap C = \emptyset} m_1(B) m_2(C) \quad (7)$$

where  $k$  represents the total conflict degree. If  $k = 1$ , it implies that the two SoEs are in total conflict, and the DS rule cannot be applied because of division by zero.

### B. Classical Discounting Technique

The SoEs may have varying degrees of reliability due to their different abilities of classification. The discounting operations are frequently conducted by using a discounting factor  $\alpha$  for each source of evidence. A particular discounting operation has been introduced by Shafer [14] for the combination of SoEs with different degrees of reliability. Shafer discounts the masses of all focal elements by a discounting (weighting) factor  $\alpha \in [0, 1]$  to the total ignorance. Each discounted BBA characterizing each discounted source of evidence is used in the fusion process. More precisely, for  $\forall A \in 2^\Theta \setminus \{\emptyset\}$ , the discounted mass of discounted source of evidence is defined as follows:

$$\begin{cases} m^\alpha(A) = \alpha \cdot m(A) \\ m^\alpha(\Theta) = 1 - \alpha + \alpha \cdot m(\Theta) \end{cases} \quad (8)$$

where  $\alpha = 1$  means that the SoE is completely reliable, and  $\alpha = 0$  means that the SoE is completely unreliable.

### C. Pignistic Probability Transformation

When multi-source information is combined, there may be disjunctive focal elements with strictly positive mass of belief. It is worth noting that the final decision is made only among singleton focal elements. Classically, a BBA is usually transformed into a (possibly subjective) probability measure

<sup>2</sup>To keep the presentation as simple as possible, we present DS rule for only two BBAs, see [14] for its generalization.

for decision making. The Pignistic Probability Transformation (PPT, or *BetP* transform) proposed by Smets in [18], [19] is generally considered as a reasonable in-between decisional attitude between the max of  $Bel(\cdot)$  (pessimistic attitude) and max of  $Pl(\cdot)$  (optimistic attitude). The betting probability  $BetP(\theta_i)$  of any singleton focal element  $\theta_i$  of the FoD is defined by

$$BetP(\theta_i) = \sum_{\theta_i \in X, X \in 2^\Theta} \frac{m(X)}{|X|} \quad (9)$$

where  $|X|$  refers to the cardinality of a subset  $X$ . One clearly sees that the BetP transform evenly distributes the belief assignment of disjunctive focal element to the singleton focal element it contains.

### III. WEIGHTED FUSION OF MULTIPLE CLASSIFIERS

#### A. Classifiers for HAR

In this article, we use classical machine learning classifiers [20] such as SVM, RF, MLP and LR to generate BBAs of human activity, and these classifiers can only give the mass of belief for singleton focal elements (i.e. we work with Bayesian BBAs). In the fusion of multiple classifiers, a BBA can be represented by the output of each classifier. It is worth noting that we should choose different types of classifiers as far as possible. In general, when the diversity between the multiple classifiers is larger, the advantages will be more obvious. At the same time, we need to guarantee the individual prediction accuracy of each classifier, which is the basis for the high accuracy of our WFMC algorithm. Furthermore, we train each classifier separately using the same training dataset. Once multiple classifiers are trained, we can obtain the corresponding BBAs for each category of human activity.

#### B. Assessment Criteria

After acquiring multiple BBAs of the human activity to be identified, we can use DS rule to fuse these BBAs and further make decisions. In this article we work with DS rule mainly because of its simplicity even we are aware of its well-known disputable dictatorial behavior in some cases, and that is why we use discounting techniques. We will evaluate the performances of alternative fusion rules in our future works. From the perspective of conflicts between multiple BBAs or uncertain information, the reliability of multiple BBAs should be evaluated before combination, its goal is to eliminate and reduce the negative influence of unreliable BBAs on the final recognition accuracy. For this reason, the appropriate assessment criterion need to be chosen in advance. In this article, we have selected three assessment criteria, described as follows:

*a) Divergence degree:* The Belief Jensen–Shannon (BJS) divergence measure was presented by Xiao [21] to measure the divergence between belief functions in DST. It is the generalization of the Jensen–Shannon divergence [22] where the probability distribution is replaced with belief mass functions.

Let  $m_1$  and  $m_2$  be two BBAs on the same FoD, containing  $n$  mutually exclusive and exhaustive hypotheses. The BJS divergence between  $m_1$  and  $m_2$  is denoted as:

$$BJS(m_1, m_2) = \frac{1}{2} \left[ \sum_{i=1}^{2^n-1} m_1(A_i) \log \left( \frac{2m_1(A_i)}{m_1(A_i) + m_2(A_i)} \right) + \sum_{i=1}^{2^n-1} m_2(A_i) \log \left( \frac{2m_2(A_i)}{m_1(A_i) + m_2(A_i)} \right) \right] \quad (10)$$

where  $A_i$  is a non empty element of the power-set  $2^\Theta$ , and  $\sum_{i=1}^{2^n-1} m_1(A_i) = 1$ ,  $\sum_{i=1}^{2^n-1} m_2(A_i) = 1$ . The lower and upper bounds of the BJS divergence measure are respectively equal to zero and one. When  $m_1$  has the same BBAs as  $m_2$ , the BJS divergence between  $m_1$  and  $m_2$  is 0. When two BBAs are completely different, the BJS divergence value is 1. In this article, the average BJS divergence of a BBA can be calculated by

$$\widetilde{BJS}(m) = \frac{1}{N-1} \sum_{j=1}^N BJS(m, m_j) \quad (11)$$

where  $N$  indicates the number of classifiers.

*b) Distance degree:* The smaller the distance between a pair of BBAs, the closer their belief values are, and the better for our decision-making. In this article, the interval distance [23] is an excellent metric, as it considers the belief intervals using the belief and plausibility functions of each focal element to describe the closeness between BBAs. The interval distance is defined as follows:

$$d_{BI}^{Ec}(m_1, m_2) \triangleq \sqrt{\frac{1}{2^{n-1}} \cdot \sum_{i=1}^{2^n-1} [d_{BI}(BI_1(A_i), BI_2(A_i))]^2} \quad (12)$$

with

$$BI(A_i) = [Bel(A_i), Pl(A_i)] \quad (13)$$

$$d_{BI}([a_1, b_1], [a_2, b_2]) =$$

$$\sqrt{\left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2} \quad (14)$$

The average interval distance of one set of BBAs can be calculated by

$$\widetilde{d}_{BI}^{Ec}(m) = \frac{1}{N-1} \sum_{j=1}^N d_{BI}^{Ec}(m, m_j) \quad (15)$$

where  $N$  indicates the number of classifiers. The larger the value of the interval distance, the greater the degree of conflict between the current BBA and other BBAs, the less reliable it will be, and vice versa.

*c) Uncertain degree:* A novel effective measure of uncertainty (i.e. entropy) of BBAs is proposed by Dezert [24], this new continuous measure is effective in the sense that it satisfies

a small number of very natural and essential desiderata. The new entropy measure is defined by

$$U(m) = \sum_{X \in 2^\Theta} s(X) \quad (16)$$

with

$$s(X) \triangleq -(1 - u(X))m(X) \log(m(X)) + u(X)(1 - m(X)) \quad (17)$$

$$u(X) \triangleq Pl(X) - Bel(X). \quad (18)$$

$s(X)$  is the uncertainty contribution of  $X$  in  $U(m)$ . This measure of uncertainty coincides with Shannon entropy for any Bayesian BBA, it can be also interpreted as an effective generalization of Shannon entropy. We always have  $U(m) \geq 0$ , and  $U(m) < U(m_v)$  if the BBA  $m(\cdot)$  is different of the vacuous BBA  $m_v(\cdot)$  defined by  $m_v(\Theta) = 1$ . It is worth noting that it is possible that a non-Bayesian BBA can have an entropy value  $U(m)$  smaller than the maximum of Shannon entropy given by  $\log(|\Theta|)$ . When  $X$  is a single focal element and satisfies  $m(X) = 1$ ,  $U(m)$  has a minimum value of 0, which indicates that the source of evidence is completely certain and it plays an important role in the final combination.

### C. Reliability Evaluation of Classifiers

In this article, each classifier can be regarded as a evidence source. We obtain the reliability of one classifier by evaluating its output, as follows:

a) *Construction of scoring matrix*: Supposing that there exists  $N$  classifiers over the same FoD, and their BBAs composition are as follows:

	$A_1$	$A_2$	$\dots$	$A_M$
$C_1$	$m_1(A_1)$	$m_1(A_2)$	$\dots$	$m_1(A_M)$
$C_2$	$m_2(A_1)$	$m_2(A_2)$	$\dots$	$m_2(A_M)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$C_N$	$m_N(A_1)$	$m_N(A_2)$	$\dots$	$m_N(A_M)$

(19)

where  $A_i \in 2^\Theta$ , and  $C_j, j = 1, 2, \dots, N$  represents the  $j$ th classifier. Then we calculate the scores of each classifier according to the assessment criteria  $Crit_\eta, \eta = 1, 2, \dots, q$  and the scoring matrix  $S$  can be generated as follows:

	$C_1$	$C_2$	$\dots$	$C_j$	$\dots$	$C_N$
$Crit_1$	$S_{11}$	$S_{12}$	$\dots$	$S_{1j}$	$\dots$	$S_{1N}$
$Crit_2$	$S_{21}$	$S_{22}$	$\dots$	$S_{2j}$	$\dots$	$S_{2N}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$Crit_q$	$S_{q1}$	$S_{q2}$	$\dots$	$S_{qj}$	$\dots$	$S_{qN}$

(20)

In this article,  $q = 3$ , that is:  $Crit_1 \triangleq BJS(\cdot)$ ,  $Crit_2 \triangleq d_{BI}^{Ec}(\cdot)$  and  $Crit_3 \triangleq U(\cdot)$ .

b) *Construction of local BBAs for classifiers*: Considering the assessment criteria and their corresponding evaluation vectors, we can calculate the positive support degree  $Sup_\eta(C_j)$  and negative support degree  $Inf_\eta(C_j)$  for each classifier by the following equations (see [16] for details)

$$Sup_\eta(C_j) \triangleq \sum_{\kappa \in \{1, \dots, N\} | S_{\eta\kappa} \leq S_{\eta j}} |S_{\eta j} - S_{\eta\kappa}|. \quad (21)$$

$$Inf_\eta(C_j) \triangleq - \sum_{\kappa \in \{1, \dots, N\} | S_{\eta\kappa} \geq S_{\eta j}} |S_{\eta j} - S_{\eta\kappa}|. \quad (22)$$

Then, the maximum value  $C_{\max}^\eta$  and minimum value  $C_{\min}^\eta$  of the classifier  $C_j$  under the assessment criteria  $Crit_\eta$  can be obtained by the following equations.

$$C_{\max}^\eta \triangleq max_j Sup_\eta(C_j) \quad (23)$$

$$C_{\min}^\eta \triangleq min_j Inf_\eta(C_j). \quad (24)$$

Next, the construction of local BBAs is based on the method presented in [16] and defined as follows:

$$\begin{cases} m_{j-\eta}(C_j) \triangleq Bel_\eta(C_j) \\ m_{j-\eta}(\bar{C}_j) \triangleq 1 - Pl_\eta(C_j) \\ m_{j-\eta}(C_j \cup \bar{C}_j) \triangleq Pl_\eta(C_j) - Bel_\eta(C_j) \end{cases} \quad (25)$$

with

$$\begin{cases} Bel_\eta(C_j) \triangleq \frac{Sup_\eta(C_j)}{C_{\max}^\eta} \\ Bel_\eta(\bar{C}_j) \triangleq \frac{Inf_\eta(C_j)}{C_{\min}^\eta} \\ Pl_\eta(C_j) \triangleq 1 - \frac{Inf_\eta(C_j)}{C_{\min}^\eta} \end{cases} \quad (26)$$

where  $m_{j-\eta}(C_j)$ ,  $m_{j-\eta}(\bar{C}_j)$  and  $m_{j-\eta}(C_j \cup \bar{C}_j)$  respectively represent the positive support belief, negative support belief and uncertainty belief of the classifier  $C_j$  based on the assessment criteria  $Crit_\eta$ .

c) *Calculation of weight factors*: We employ the BF-TOPSIS algorithm [16] to calculate the weight factors for each classifier and the specific steps are as follows.

- Step 1 Calculate the local BBAs  $m_{j-\eta}(C_j)$ ,  $m_{j-\eta}(\bar{C}_j)$  and  $m_{j-\eta}(C_j \cup \bar{C}_j)$  of each classifier according to the scoring matrix.
- Step 2 For each classifier, calculate  $d_{BI}^{Ec}(m_{j-\eta}, m_\eta^{best})$  and  $d_{BI}^{Ec}(m_{j-\eta}, m_\eta^{worst})$ ,  $m_\eta^{best}$  and  $m_\eta^{worst}$  represent the best and the worst ideal BBAs based on the assessment criteria  $Crit_\eta$ , respectively, where  $m_\eta^{best}(C_j) = 1$  and  $m_\eta^{worst}(\bar{C}_j) = 1$ .
- Step 3 Calculate the weighted average distance  $d^{best}(C_j)$  and  $d^{worst}(C_j)$  of classifier, where

$$d^{best}(C_j) \triangleq \sum_{\eta=1}^N v(Crit_\eta) \cdot d_{BI}^{Ec}(m_{j-\eta}, m_\eta^{best}) \quad (27)$$

$$d^{worst}(C_j) \triangleq \sum_{\eta=1}^N v(Crit_\eta) \cdot d_{BI}^{Ec}(m_{j-\eta}, m_\eta^{worst}) \quad (28)$$

where  $v(Crit_\eta)$  represents the weight of assessment criteria  $Crit_\eta$ . In this article,  $v(Crit_1) = v(Crit_2) = v(Crit_3) = 1/3$ .

- Step 4 The final weight of the classifier  $C_j$  is defined as follows:

$$\omega(C_j) = \frac{d^{worst}(C_j)}{d^{worst}(C_j) + d^{best}(C_j)} \quad (29)$$

In the proposed WFCM algorithm, when a classifier is in complete conflict with other classifiers, it will be supported to a small degree. According to the reliability evaluation algorithm, the classifier will receive a small weighting factor, which discounts the masses of all focal elements to the total ignorance. This reduces the total conflict between classifiers in the fusion process, making the total conflict in the proposed WFCM algorithm always less than 1, thus improving the reliability of the fusion results.

After obtaining the weight factors for each classifier, multiple classifiers can be fused using the classical (i.e. Shafer's) discounting technique and DS rule, and decision can be made based on the maximum BetP probability value. For the convenience of implementation, the brief framework of the WFCM method is given in Fig. 1.

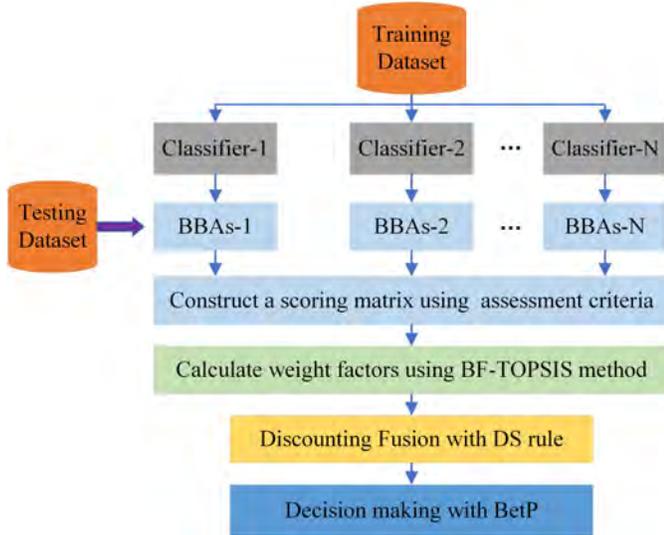


Fig. 1. The framework of WFCM method.

As we can see in Fig. 1, the proposed WFCM algorithm includes four main steps:

- **Step 1 (Classifiers trained):** Multiple classifiers of different types are trained by the same training dataset for acquiring BBAs.
- **Step 2 (Classifiers evaluation):** For each BBA generated by the classifier, a reliability evaluation is performed using three criteria and a scoring matrix is constructed.
- **Step 3 (Calculation of weight factors):** The BF-TOPSIS algorithm is employed to calculate the weight factors for each classifier based on the scoring matrix.
- **Step 4 (Discounting fusion):** Multiple classifiers are combined sequentially using the classical discounting technique and DS rule, the final decision can be made based on the maximum BetP probability.

## IV. EXPERIMENTS AND DISCUSSIONS

### A. UCI Smartphone Dataset

In this article, the UCI Smartphone dataset is considered for experimental verification. In UCI Smartphone dataset, the experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (walking, walking upstairs, walking downstairs, sitting, standing and laying) wearing a smartphone on the waist. Three-axial linear acceleration and three-axial angular velocity at a constant rate of 50Hz were captured by using its embedded accelerometer and gyroscope. The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50 percent overlap (128 readings/window). More descriptions of the UCI Smart-phone dataset can be found in [25].

### B. Example

In order to show how our WFCM method works, an example is given to illustrate its specific procedures. Firstly, the focal element in BFs theory can be applied to mathematically represent human activities. Specifically  $\theta_1 \triangleq$  walking,  $\theta_2 \triangleq$  walking upstairs,  $\theta_3 \triangleq$  walking downstairs,  $\theta_4 \triangleq$  sitting,  $\theta_5 \triangleq$  standing,  $\theta_6 \triangleq$  laying. For the 480th sample data with a true label of (standing) in the test dataset, the corresponding BBAs generated by four classifiers are shown in Table I. According to the principle of maximum probability, it can be seen that SVM and RF support  $\theta_4$  (sitting) while MLP and LR support  $\theta_5$  (standing), which causes trouble to make decisions. We utilize DS rule to combine the four classifiers and the fusion results have the maximum belief value of 0.567 to support  $\theta_4$  (sitting), which is not what we want.

Next, we use WFCM algorithm for testing. The scoring matrix is acquired based on (9), (13) and (14), as shown in Table II. Then, we can get the positive support and negative support degree of each classifier according to (17) and (18), which are given in Table III and Table IV. It can be seen that  $BJS(\cdot)$  has the highest support for RF, while  $U(\cdot)$  has the highest support for SVM, and  $d_{BI}^{Ec}(\cdot)$  supports both RF and MLP. After that, the derived local BBAs of each classifier can be also obtained using (21) shown in Table V, Table VI and Table VII. And then by using step 2 and step 3 in BF-TOPSIS algorithm, we can obtain distance  $d^{best}(C_j)$  and  $d^{worst}(C_j)$  of classifiers. The weight coefficients of each classifier can be further obtained based on (27), as shown in Table VIII. It can be seen that SVM acquires the smallest weighting factor, while MLP gets the largest weighting factor and RF has a similar weighting factor to MLP, which indicates that MLP has the highest reliability for the current activity. Finally, four classifiers are combined using DS rule (6) generalized<sup>3</sup> for four BBAs, and the probability values for each category of activity are obtained based on (9), as shown in Table IX. We

<sup>3</sup>Because DS rule is associative, the four BBAs can also be fused sequentially and the sequential order of DS fusion does not impact the final result.

can see that  $\theta_5$  (standing) has the maximum BetP probability value, which is consistent with the true label.

TABLE I  
BBAS OF THE 480TH TEST SAMPLE.

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	Result
SVM	0.0	0.0	0.0	0.844	0.156	0.0	$\theta_4$
RF	0.0	0.0	0.0	0.573	0.427	0.0	$\theta_4$
MLP	0.0	0.0	0.0	0.349	0.651	0.0	$\theta_5$
LR	0.0	0.0	0.0	0.251	0.749	0.0	$\theta_5$
DS rule	0.0	0.0	0.0	0.567	0.433	0.0	$\theta_4$

TABLE II  
SCORING MATRIX OF FOUR CLASSIFIERS.

	SVM	RF	MLP	LR
$BJS(\cdot)$	0.218	0.043	0.056	0.088
$d_{BI}^{Ec}(\cdot)$	0.113	0.068	0.068	0.084
$U(\cdot)$	0.628	0.984	0.935	0.8162

TABLE III  
POSITIVE SUPPORT DEGREE  $Sup_{\eta}(\cdot)$  OF FOUR CLASSIFIERS.

$Sup_{\eta}(\cdot)$	SVM	RF	MLP	LR
$BJS(\cdot)$	0.0	0.232	0.194	0.130
$d_{BI}^{Ec}(\cdot)$	0.0	0.061	0.061	0.029
$U(\cdot)$	0.852	0.0	0.05	0.287

TABLE IV  
NEGATIVE SUPPORT DEGREE  $Inf_{\eta}(\cdot)$  OF FOUR CLASSIFIERS.

$Inf_{\eta}(\cdot)$	SVM	RF	MLP	LR
$BJS(\cdot)$	-0.466	0.0	-0.013	-0.077
$d_{BI}^{Ec}(\cdot)$	-0.119	0.0	0.0	-0.033
$U(\cdot)$	0.0	-0.575	-0.425	-0.188

TABLE V  
LOCAL BBAS OF FOUR CLASSIFIERS ON  $BJS(\cdot)$ .

	SVM	RF	MLP	LR
$m_{BJS(\cdot)}(C_j)$	0.0	1.0	0.835	0.560
$m_{BJS(\cdot)}(\bar{C}_j)$	1.0	0.0	0.027	0.164
$m_{BJS(\cdot)}(C_j \cup \bar{C}_j)$	0.0	0.0	0.138	0.276

TABLE VI  
LOCAL BBAS OF FOUR CLASSIFIERS ON  $d_{BI}^{Ec}(\cdot)$ .

	SVM	RF	MLP	LR
$m_{d_{BI}^{Ec}(\cdot)}(C_j)$	0.0	1.0	1.0	0.470
$m_{d_{BI}^{Ec}(\cdot)}(\bar{C}_j)$	1.0	0.0	0.0	0.273
$m_{d_{BI}^{Ec}(\cdot)}(C_j \cup \bar{C}_j)$	0.0	0.0	0.0	0.257

### C. Measure of Performances

The classical Accuracy is applied to measure the performance of our proposed method. The specific definitions are as follows:

$$Accuracy = \frac{1}{n} \sum_{i=1}^n \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \quad (30)$$

TABLE VII  
LOCAL BBAS OF FOUR CLASSIFIERS ON  $U(\cdot)$ .

	SVM	RF	MLP	LR
$m_{U(\cdot)}(C_j)$	1.0	0.0	0.059	0.337
$m_{U(\cdot)}(\bar{C}_j)$	0.0	1.0	0.740	0.327
$m_{U(\cdot)}(C_j \cup \bar{C}_j)$	0.0	0.0	0.202	0.336

TABLE VIII  
WEIGHTED COEFFICIENTS OF FOUR CLASSIFIERS.

	$d^{best}(C_j)$	$d^{worst}(C_j)$	$\omega(C_j)$
SVM	0.471	0.236	0.333
RF	0.236	0.471	0.667
MLP	0.238	0.506	0.680
LR	0.334	0.461	0.580

TABLE IX  
RESULTS OF THE WFCM METHOD.

	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	$\Theta$
Weighted fusion	0.0	0.0	0.0	0.389	0.548	0.0	0.063
$BetP(\cdot)$	0.01	0.01	0.01	0.399	0.558	0.01	0.0

where  $i$  denotes class index and  $n$  is the number of classes.  $TP_i$ ,  $TN_i$ ,  $FN_i$  and  $FP_i$  are respectively True Positives, True Negatives, False Positives and False Negatives.

### D. Experimental Results and Analysis

According to the specific steps described in Fig. 1, we first train four classifiers using 7352 samples, including a SVM, a RF, a MLP and a LR. For the parameters of SVM, the sigmoid function is selected as kernel function, and the penalty parameter is set to 1.0. For the parameters of RF, the number of trees in the forest is set to 150. For the parameters of MLP, the number of hidden layers is set to 300. For the parameters of LR, the penalty is set to L1. Default parameters are selected for the remaining parameters of four classifiers. In this article, features are extracted from raw sensor data for model training, including 11 time-domain and 6 frequency-domain features as shown in Table II. Then the trained four classifiers are employed to predict the testing dataset containing 2947 samples. Furthermore, we fuse the four classifiers using the DS rule and the proposed WFCM algorithm, respectively, the results are shown in the Table III, and the related confusion matrixs are shown in Fig. 2 and Fig. 3. We can find that LR has the highest accuracy among the individual classifier with 93.52%, which is weaker than the DS rule approach. It indicates that individual classifier has limited classification ability. Moreover, we can clearly see that the performance of the proposed WFCM method is significantly better than other mentioned method, which shows the effectiveness of our strategy.

Compared to the approach of traditional DS rule, the proposed method effectively improves the recognition accuracy. The misclassification where sitting was incorrectly recognized as standing is reduced from 12.4% to 8.4% and the misclassification where walking downstairs was incorrectly recognized

TABLE X  
FEATURE EXTRACTION

Domain	Features
Time	Mean value, Standard deviation, Median absolute value, Maximum, Minimum, Signal magnitude area, Average sum of the squares, Interquartile range, Signal entropy, Autoregression coefficients, Correlation
Frequency	Largest frequency component, Weighted average, Skewness, Kurtosis, Energy of a frequency interval, Angle between two vectors

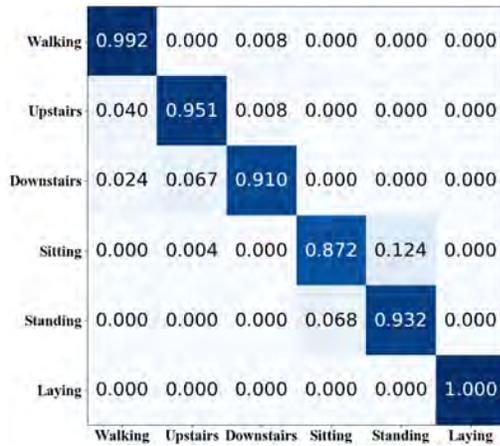


Fig. 2. Confusion matrix on UCI smartphone dataset by DS rule.

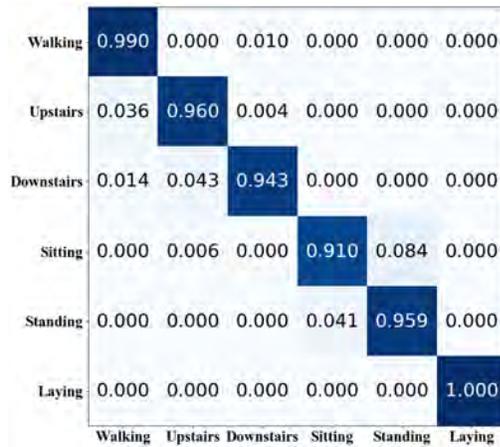


Fig. 3. Confusion matrix on UCI smartphone dataset by the WFMC method.

as walking upstairs is reduced from 6.7% to 4.3%. This is due to the fact that the three evaluation criteria we have given are a good measure of the conflict between multiple classifiers and their own uncertainty, and the BF-TOPSIS algorithm efficiently calculates the weight coefficients for each classifier, which improves the accuracy of the multi-classifier fusion.

Furthermore, we compare with some state-of-the-art approaches in literatures to demonstrate the superiority of our method, including Activity Graph Based Convolutional Neural Network [26], DSmT-Based Kernel Density Estimation [6], Sensor fusion and deep recurrent neural network-based [7], Two-stream Transformer Network [27], Hesitant Fuzzy Belief

TABLE XI  
COMPARISON OF WFMC METHOD WITH TRADITIONAL METHODS ON THE UCI SMARTPHONE DATASET.

Method	Accuracy	Time (s)
SVM	91.75%	9.08
RF	92.94%	2.06
MLP	92.53%	1.95
LR	93.52%	1.97
DS rule	94.43%	26.01
<b>WFMC</b>	<b>96.20%</b>	<b>33.10</b>

Structure Based Fused Extreme Learning Machine [28]. As we can see in Table IV, our method outperforms these state-of-the-art methods in terms of accuracy.

TABLE XII  
COMPARISON OF WFMC METHOD WITH STATE-OF-THE-ART METHODS ON THE UCI SMARTPHONE DATASET.

Method	Accuracy	Time(s)
Activity Graph CNN-Based [26]	90.17%	11.34
DSmT-Based Kernel Density Estimation [6]	93.05%	24.46
Sensor fusion and deep RNN-based [7]	94.27%	15.95
Two-stream Transformer Network [27]	94.12%	20.19
Hesitant Fuzzy Belief Based ELM [28]	95.20%	23.78
<b>WFMC</b>	<b>96.20%</b>	<b>33.10</b>

In terms of time consumption, our WFMC method was programming in Python 3.7 with a hardware of Intel Core i7-8700 CPU at 3.20 GHz and 16 GB RAM. We use 2947 test samples and counted the total time consumed by each method. As can be seen, traditional machine learning algorithms have the advantage of being fast. As our WFMC algorithm is developed based on DST, it inevitably increases the computational burden. Nevertheless, the average elapsed time per test sample is about 11ms, which is sufficient for practical applications.

## V. CONCLUSION

In this article, we have proposed a novel weighted fusion of multiple classifiers based on belief functions theory for human activity recognition. Firstly, we train four classical machine learning classifiers by using time-domain and frequency-domain features to obtain basic belief assignments of human activities. Secondly, we evaluate the outputs of four classifiers using three criteria and construct a scoring matrix. Thirdly, we use the multi-criteria BF-TOPSIS algorithm to calculate the weight coefficients of each classifier. Finally, we adopt a discounting technique and DS rule to combine the four classifiers, and make decisions thanks to the pignistic probability values. Several experiments have been conducted based on the UCI Smartphone dataset. The experimental results prove that our WFMC approach can significantly improve the classification accuracy with respect to several classical and state-of-the-art methods.

In our future works, we will evaluate a better measure of divergence between belief functions based on a more effective definition of relative entropy and cross-entropy. We will also explore the possibility to adapt our Stable Preference Ordering Towards Ideal Solution (SPOTIS) rank reversal multi-criteria

method for HAR instead using the BF-TOPSIS method which is not robust to rank reversal. We will test and compare an other decision-making technique based on belief-interval distance, and work on how to reduce the complexity of multi-classifier fusion for HAR in order to apply it to an online real activity recognition system.

#### ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 62233003 and 62073072, and in part by the Key Projects of Key Program of Jiangsu Province under Grant BE2020006 and Grant BE2020006-1, and in part by Shenzhen Natural Science Foundation under Grant JCYJ20210324132202005 and JCYJ20220818101206014.

#### REFERENCES

- [1] R. Gravina, P. Alinia, H. Ghasemzadeh, G. Fortino, *Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges*, Information Fusion, Vol. 35, pp. 68–80, May 2017.
- [2] Y.J. Hao, R. Zheng, B.Y. Wang, *Invariant Feature Learning for Sensor-Based Human Activity Recognition*, IEEE Trans. on Mobile Computing, Vol. 21(11), pp. 4013–4024, 2022.
- [3] Y. Tang, L. Zhang, F. Min, J. He, *Multiscale Deep Feature Learning for Human Activity Recognition Using Wearable Sensor*, IEEE Trans. on Indus. Elect., Vol. 70 (2), pp. 2106–2116, Feb. 2023.
- [4] Y. Dong, R. Zhou, C. Zhu, L. Cao, X. Li, *Hierarchical Activity Recognition Based on Belief Functions Theory in Body Sensor Networks*, IEEE Sensors Journal, Vol. 22(15), pp. 15211–15221, Aug. 2022.
- [5] X. Li, X. Huang, J. Dezert, F. Smarandache, *Enrichment of qualitative beliefs for reasoning under uncertainty*, in Proc. of 10th Int. Conf. on Information Fusion, Québec, Canada, July 2007.
- [6] Y. Dong, X. Li, J. Dezert, M.O. Khyam, Md. Noor-A-Rahim, S.S. Ge, *Dezert-Smarandache Theory-Based Fusion for Human Activity Recognition in Body Sensor Networks*, IEEE Trans. on Industrial Informatics, Vol. 16(11), pp. 7138–7149, Nov. 2020.
- [7] Md.Z. Uddin, M. Hassan, A. Alsanad, C. Savaglio, *A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare*, Information Fusion, Vol. 55, pp. 105–115, March 2020.
- [8] Y. Zhao, X. Wang, Z. Lin, Z. Huang, *Multi-Classifer Fusion for Open-Set Specific Emitter Identification*, Remote Sensing, Vol. 14(9), May 2022.
- [9] X. Li, W. Yang, X. Wu, and J. Dezert, *A fast approximate reasoning method in hierarchical DSMT(B)*, Acta Electronica Sinica, vol. 39, no. S1, pp. 31–36, 2011.
- [10] Z.G. Liu, Q. Pan, J. Dezert, J.W. Han, Y. He, *Classifier Fusion With Contextual Reliability Evaluation*, IEEE Trans. on Cybernetics, Vol. 48(5), pp. 1605–1618, 2018.
- [11] S. Liu, X. Li, C. Hu, J. Yao, Junping, X. Han, J. Wang, *Spammer detection using multi-classifier information fusion based on evidential reasoning rule*, Scientific Reports, Vol. 12(1), July 2022.
- [12] Z.G. Liu, Q. Pan, J. Dezert, A. Martin, *Combination of Classifiers With Optimal Weight Based on Evidential Reasoning*, IEEE Trans on Fuzzy Systems, Vol. 26(3), pp. 1217–1230, June 2018.
- [13] X. Li, J. Dezert, X. Huang, Z. Meng, and X. Wu, *A fast approximate reasoning method in hierarchical DSMT(A)*, Acta Electronica Sinica, vol. 38, no. 11, pp. 2566–2572, 2010.
- [14] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, NJ, USA, 1976.
- [15] Y. Dong, X. Li, J. Dezert, S.S. Ge, *A Novel Multi-Criteria Discounting Combination Approach for Multi-Sensor Fusion*, IEEE Sensors Journal, Vol. 19(20), pp. 9411–9421, Oct. 2019.
- [16] J. Dezert, D. Han, H. Yin, *A New Belief Function Based Approach for Multi-Criteria Decision-Making Support*, Proc. of 19th Int. Conf. on Information Fusion, pp. 782–789, Heidelberg, Germany, July 5–8, 2016.
- [17] Z.G. Liu, X. Zhang, J. Niu, J. Dezert, *Combination of Classifiers With Different Frames of Discernment Based on Belief Functions*, IEEE Trans on Fuzzy Systems, Vol. 29(7), pp. 1764–1774, July 2021.
- [18] P. Smets, *Constructing the pignistic probability function in a context of uncertainty*, Uncertainty in AI, Vol. 5, pp. 29–39, 1990.
- [19] P. Smets, *Decision making in the TBM: the necessity of the pignistic transformation*, Int. Jour. of Approx. Reasoning, Vol. 38, 2005.
- [20] L. Breiman, *Random forests*, Machine Learning, Vol. 45(1), pp. 5–32, 2001.
- [21] F. Xiao, *Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy*, Information Fusion, Vol. 46, pp. 23–32, March 2019.
- [22] [https://en.wikipedia.org/wiki/Jensen-Shannon\\_divergence](https://en.wikipedia.org/wiki/Jensen-Shannon_divergence).
- [23] D. Han, J. Dezert, Y. Yang, *Belief Interval-Based Distance Measures in the Theory of Belief Functions*, IEEE Trans. on SMCS, Vol. 48(6), pp. 833–850, June 2018.
- [24] J. Dezert, *An Effective Measure of Uncertainty of Basic Belief Assignments*, in Proc. of 25th Int. Conf. on Information Fusion, Linköping, Sweden, July 4–7, 2022.
- [25] D. Thakur, S. Biswas, *An Integration of feature extraction and Guided Regularized Random Forest feature selection for Smartphone based Human Activity Recognition*, J. of Network and Computer Applications, Vol. 204, Aug. 2022. <https://doi.org/10.1016/j.jnca.2022.103417>.
- [26] P. Yang, C. Yang, V. Lanfranchi, F. Ciravegna, *Activity Graph Based Convolutional Neural Network for Human Activity Recognition Using Acceleration and Gyroscope Data*, IEEE Trans. on Industrial Informatics, Vol. 18(10), pp. 6619–6630, Oct. 2022.
- [27] S. Xiao, S. Wang, Z. Huang, Y. Wang, H. Jiang, *Two-stream transformer network for sensor-based human activity recognition*, Neurocomputing, Vol. 512, pp. 253–268, Nov. 2022.
- [28] Y. Dong, X. Li, J. Dezert, R. Zhou, C. Zhu, L. Wei, S.S. Ge, *Evidential Reasoning With Hesitant Fuzzy Belief Structures for Human Activity Recognition*, IEEE Trans. on Fuzzy Systems, Vol. 29(12), pp. 3607–3619, Dec. 2021.

# Imbalanced Data Classification Based on Belief Functions Theory

Jiawei Niu, Zhunga Liu

School of Automation, Northwestern Polytechnical University, Xi'an, China.

Emails: niujiawei@mail.nwpu.edu.cn, liuzhunga@nwpu.edu.cn

**New contribution:** J. Niu, Z. Liu, *Imbalanced Data Classification Based on Belief Functions Theory*.

**Abstract**—Imbalanced data is an important research for the classification and there are multiple techniques to deal with this problem. Each method has its own advantage for solving imbalanced data. To improve the classification accuracy, these strategies are combined in decision level via an appropriate way to fully take advantages of the complementary information among different methods. Thus a new method is proposed as imbalanced data classification based on belief functions theory (IDCBF). The classification result generated by different strategies (i.e., hybrid-sampling, over-sampling, or under-sampling) may have different reliabilities for query patterns. So an appropriate quality evaluation rule is created to estimate the credibility of each classification result based on the close neighborhoods. The revised classification results from different strategies are then combined by Dempster's rule to reduce the ignorant information and to generate the final classification result. Multiple experiments are used to test the performance of the IDCBF method, and the results show that IDCBF can efficiently improve the classification accuracy with respect to other related methods.

**Keywords:** pattern classification, belief functions, evidence theory, imbalanced data.

## I. INTRODUCTION

Traditional classification methods [1]–[3] usually assume that each category in a dataset contains the same number of samples and the misclassification costs are equal. However, the data in the real world may have imbalanced distributions [4]–[7]. A class with fewer instances is known as a positive class or a minority class, and a class with more examples is called a negative class or a majority class. The minority class is more important than the majority class in the real world, and the cost of misclassification is also higher. Nowadays, imbalanced classification is widely used in information security [8] and software prediction [9]. In such a way, the imbalanced data classification has attracted extensive interest from many researchers. This paper is an extension of our works presented in [10] and [11].

The imbalanced data classification methods are divided into three kinds: data preprocessing level [12]–[14], feature selection level [15]–[17], and classification methods improvement level [18], [19]. In this work, we attempt to solve the problem at data preprocessing level, which decreases the imbalanced ratio of the dataset via creating minority data or deleting majority data. It focuses on under-sampling [20], over-sampling [21] and hybrid-sampling [22] methods

to minimize the imbalanced ratio by redistributing the data. In under-sampling technique, it deletes the majority data to increase the classification accuracy of minority classes such as the Nearmiss [20] method. In the over-sampling method, it creates the minority data by the Euclidean distance to balance the sample ratio such as Synthetic Minority Oversampling Techniques (SMOTE) [21]. The hybrid-sampling method is linked with under-sampling and over-sampling techniques such as SmoteTomek [23] method.

These methods have their own advantages and drawbacks when they are utilized to deal with the imbalanced data classification. Over-sampling methods allow to generate minority data but they may cause the over-fitting problems. Under-sampling techniques remove majority data which may discard potentially important information. Hybrid-sampling algorithms are conducted with the connection of under-sampling and over-sampling methods. Each technique has its own particular benefits. To better improve the classification accuracy, we will propose a new method at decision level to combine these three algorithms via making full use of their complementary information.

Belief functions theory provides an essential decision-level information fusion tool, and it is able to well combine the uncertain information. It has been already applied in data fusion and pattern classification fields [24] [25]. In this paper, we want to propose a new method called imbalanced data classification based on belief functions theory (IDCBF). The output classification results generated by different methods (i.e., hybrid-sampling, over-sampling, and under-sampling) may have different qualities/reliabilities. A reliability matrix through the neighborhood of the object is proposed to make a refined reliability evaluation. The classification outputs by different techniques will be cautiously revised utilizing the reliability matrix. Finally, the corrected classification results are combined by the evidence combination rule for making the final decision.

The remainder of this paper is organized as follows. Section II describes the proposed method in detail. The experimental applications are presented to test the performance of IDCBF in Section III. Section IV provides the conclusions.

## II. IMBALANCED DATA CLASSIFICATION BASED ON BELIEF FUNCTIONS THEORY

The three imbalanced data classification methods (hybrid-sampling, over-sampling, and under-sampling) have their own advantages and drawbacks. To better improve the classification accuracy, these three methods are combined through an appropriate way for taking fully advantages of the complementary information among these methods. Belief functions theory also called as evidence theory, which provides an efficient tool to combine the uncertain information at decision level. Thus, the belief functions theory will be utilized here to combine these three techniques. A new method called imbalanced data classification based on belief functions theory (IDCBF) is proposed here to revise the classifier. We can obtain three pieces of classification results represented by evidence with three classifiers (i.e., hybrid-sampling, over-sampling, and under-sampling), and we will combine these classification results under the belief functions framework efficiently.

The classification results of different data sampling method may have different reliabilities, and it may be harmful for the combination if the result with low reliability. So it is essential to evaluate the reliability of each classification output properly, and then revising the result based on the evaluation to improve the combination performance. We propose to estimate a refined reliability matrix to represent the qualities of each classification result. Such reliability matrix will be estimated based on the neighborhoods of objects in training dataset space, and it will show the possibility of the object misclassified to other classes. After that, the classification results are able to revised according to this matrix in a cautious way under belief functions framework. The three corrected classification results are combined by belief functions theory for predicting the class of object.

### A. Basics of belief functions theory

Belief functions theory, also called evidence theory or Dempster-Shafer Theory (DST), provides an efficient tool to combine the uncertain information at decision level. In belief functions theory, the mass function  $m$ , also called the basic belief assignment (BBA) is defined over the frame of discernment denoted by  $\Omega = \{\omega_i, i = 1, 2, \dots, c\}$ , consisting of  $c$  exhaustive and exclusive hypotheses (classes)  $\omega_i$ ,  $i \in \{1, 2, \dots, c\}$ . The power-set  $2^\Omega$  is composed by all the subsets of  $\Omega$ . A BBA is a mapping  $m(\cdot)$  from  $2^\Omega$  to  $[0, 1]$  which satisfies  $m(\emptyset) = 0$  and

$$\sum_{A \in 2^\Omega} m(A) = 1, \quad (1)$$

$A$  is called a focal element of  $m(\cdot)$  which satisfy  $m(A) > 0$ . The BBA is called Bayesian BBA if the focal elements of BBA are all singleton classes. In this paper, we mainly assume that combining the classification results in form of BBAs.

Dempster's rule (DS rule) is usually utilized to combine the multiple classification results represented by BBA. DS rule for the combination of two BBA as  $\mathbf{m} = \mathbf{m}_1 \oplus \mathbf{m}_2$  over  $2^\Omega$  is

defined by  $m(\emptyset) = 0$ , and  $\forall A \neq \emptyset \in 2^\Omega$  with the following formula,

$$\mathbf{m} = \mathbf{m}_1 \oplus \mathbf{m}_2 = \begin{cases} \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - m_{12}(\emptyset)}, & \forall A \in 2^\Omega \setminus \{\emptyset\} \\ 0, & \text{if } A = \emptyset \end{cases} \quad (2)$$

where  $m_{12}(\emptyset) \triangleq \sum_{B, C \in 2^\Omega | B \cap C = \emptyset} m_1(B)m_2(C)$  is the total conjunctive conflicting masses. DS rule is associative, the combination results are not influenced by the combination order for multiple BBA.

In reality, the classification result by different classification methods (hybrid-sampling, over-sampling, and under-sampling) may have different reliabilities. It is essential to evaluate the reliabilities of classification results and revised the results based on the evaluation before combination.

### B. Evidence reliability evaluation

The classification results by different classifier may have different qualities. The under-sampling method deletes the majority data which may change the distribution of data to affect the classification accuracy. The over-sampling technique generates fake instance for minority class which may also has bad influence on classification result. The classification result of different methods can be seen as the evidence (BBA). The three classifiers (hybrid-sampling, over-sampling, and under-sampling) are denoted by three classifiers as  $C_1, C_2, C_3$  here. The object is classified over the frame of discernment  $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ , and  $\omega_i$  represents the class label. Assume a training set of  $S$  labelled patterns is available. For each classifier  $C_l$ ,  $l \in \{1, 2, 3\}$ , the classification result for the training data  $x_i$ ,  $i \in \{1, 2, \dots, S\}$ , is denoted by  $\mathbf{p}_i = \{p_{i,1}, p_{i,2}, \dots, p_{i,c}\}$ , where  $p_{i,j}$  represents the probability which  $x_i$  belongs to  $\omega_j$ . The true classification result of training data is  $t_i(\omega_j) = 1$  and  $t_i(\omega_g) = 0$ ,  $\omega_j \neq \omega_g$  when the true label of  $x_i$  is  $\omega_j$ . Given a test pattern  $y$ , the classification result of  $y$  by different classifier can be shown as a BBA  $\mathbf{m}_l$ ,  $l \in \{1, 2, 3\}$ . The final label of  $y$  is calculated by the combination of these BBAs.

For each classifier  $C_l$ ,  $l \in \{1, 2, 3\}$ , it often shows close performance to close neighborhoods, and the close neighbors of object in dataset can be used to evaluate the quality of each classification result [26], [27]. The classification results of training data are given by  $\mathbf{p}_i$ , and the true label of training data is also known. So the bias error of classifier can be computed by comparing the classifier output and the true label. Thus we can estimate the quality of the classification result of the  $y$  based on these neighbors.

How to select the suitable neighbors is an essential rule in reliability evaluation of each classification result. If we seek the close neighbors according to the attribute data, the selected neighbors seem near from the object, but the classification result of these neighbors as  $\mathbf{p}_i$  may not close to the object of  $\mathbf{m}_i$ . These neighbors are not very useful to efficiently evaluate the quality of the classification result.

However, if we seek the close neighbors according to the distance of classification results. The selected neighbors with

the similar classification results to the object may quite differ from the object in attribute data space. If so, these neighbors cannot provide necessary knowledge on the reliability evaluation of the classification result of the object. Thus, we propose a new way to select the neighbors of the object based on both the attribute data and the classifier output. This ensures the selected neighbors with close attribute and the classification results to the object.

We seek  $N$  nearest neighbors of the object  $y$  using the attribute information at first, and the selected ones are represented by  $x_1, x_2, \dots, x_N$ . In these selected  $N$  neighbors, there may exist some of them whose classification results  $\mathbf{p}_i$  are quite different from the classification result of the object as  $\mathbf{m}_l$ . These neighbors may not be beneficial and even harmful for the refined reliability evaluation of  $\mathbf{m}_i$ . Thus, we will choose  $K$  neighbors from the previous  $N$  ones according to the distance of classification results between  $\mathbf{p}_i$  and  $\mathbf{m}_l$ . Eventually, the classification results of the  $K$  selected neighbors are given by  $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_K$ , and the corresponding class label are known as  $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K$ . These  $K$  neighbors can provide important prior knowledge for reliability evaluation.

Since these chosen neighbors are not totally similar to the object, they cannot be completely trusted during the reliability evaluation. The confidence factor mainly depends on the difference between the object and the selected neighbors, and both distance of attribute as well as the classifier output are considered to compute the difference. The beliefs in the classification result are divided into two parts. One will enter the correction process based on reliability evaluation, and the other will be preserved in the original result.

The attribute data is normalized by the general linear normalization method as eq. (3) to make the value in  $[0, 1]$ .

$$a_j^* = \frac{a_j - a_{\min}}{a_{\max} - a_{\min}}, \quad (3)$$

where  $a_j$  represents the attribute value in dimension  $j$ , and  $a_j^*$  is the normalized value.

The confidence factor  $\alpha_l$ ,  $l \in \{1, 2, 3\}$  is computed to the average distance between the object and these neighbors. The neighbors are very similar to the object when their distance is small, and the confidence on the reliability evaluation is high, and vice versa.

$$\alpha_l = e^{-\beta_l d_l}, \quad (4)$$

$$d_l = \frac{1}{2K} \left( \frac{1}{\bar{d}^A} \sum_{k=1}^K d_{yk}^A + \frac{1}{\bar{d}^P} \sum_{k=1}^K d_{yk}^P \right), \quad (5)$$

$$\bar{d}^A = \frac{1}{KS} \sum_{i=1}^S \sum_{k=1}^K d_{xk}^A, \quad \bar{d}^P = \frac{1}{KS} \sum_{i=1}^S \sum_{k=1}^K d_{xk}^P. \quad (6)$$

For each classifier  $C_l$ ,  $\beta_l$  is a parameter used to adjust the influence of attribute distance and classifier output distance ratio on the confidence factor.  $d_l$  is the average distance between the object and its  $K$  neighbors in regard to the attribute and the classification result.  $d_{yk}^A = \|y - x_k\|$  represents the Euclidean distance between the object and the  $K$  neighbors.

$d_{yk}^P = \|\mathbf{m}_l - \mathbf{p}_k\|$  represents the Euclidean distance between the classification result of the object and the  $K$  neighbors.  $\bar{d}^A$  is the mean value of the average distance from each training data to its  $K$  neighbors.  $\bar{d}^P$  represents the mean value of the average distance from the classification result for the training data to its  $K$  neighbors.

If the confidence factor is high, it means that these neighbors are quite similar to the object, and we are likely to get important knowledge from these neighbors to correct the classification result of the object. In such case, a large amount of beliefs will be allowed to enter the correction process. However, if the confidence factor is low, it means that these neighbors are not quite close to the object, and we are not very confident of the reliability evaluation from these neighbors. So most beliefs will be kept in the original classification results, and only a few will be redistributed in the sequel.

The beliefs on the classification results of the object are divided into two parts. One part will be redistributed in the correction process on the basis of the reliability evaluation, and the amount of beliefs to be redistributed  $\mathbf{m}_{lr}$  is determined by eq. (7). The other is still preserved on each class as in the original classification result, and the amount of beliefs in this part  $\mathbf{m}_{lo}$  is given by eq. (8).

$$\mathbf{m}_{lr} = \alpha_l \mathbf{m}_l, \quad (7)$$

$$\mathbf{m}_{lo} = (1 - \alpha_l) \mathbf{m}_l. \quad (8)$$

### C. Classification result correction

The quality of the classification result of the object will be evaluated in a refined way based on  $K$  neighbors, and then the classifier output will be revised according to the evaluation.

A reliability matrix  $\Phi$  reflects the information about the misclassification error of the object, and the element  $\phi_{ij}$  is the probability of the object classified to  $\omega_i$  but the ground truth is  $\omega_j$ . Now we will see how to calculate the value of  $\phi_{ij}$  using these  $K$  selected neighbors.

We have the class label  $\mathbf{t}$  of these neighbors as training data, and the classification results  $\mathbf{p}_i$  of these neighbors by the given classifier are also known. So we can estimate the possibility (i.e.,  $w_{ji}$ ) of the object classified to  $\omega_i$  when it truly belongs to  $\omega_j$ . It is defined by the sum of the probabilities committed to  $\omega_i$  for the neighbors with the ground truth  $\omega_j$  (i.e.,  $t_{kj} = 1$ ).

$$w_{ji} = \sum_{k=1|t_{kj}=1}^K e^{-\lambda_l \tilde{d}_k} p_k(\omega_i), \quad (9)$$

where  $\tilde{d}_k = \frac{1}{2} \left[ \frac{d_{yk}^A}{\min_k d_{yk}^A} + \frac{d_{yk}^P}{\min_k d_{yk}^P} \right]$  is the relative distance between the object and the  $K$  neighbors.  $\lambda_l > 0$  is a tuning parameter to control the influence of the distance here.

In eq. (9), one can see the neighbors close to the object will play an essential role in the calculation of  $w_{ji}$ , because these neighbors can provide more useful prior knowledge on classification for the object. The neighbor far from the object has little influence on the reliability evaluation. Therefore, our proposed method is robust to the  $K$  number of selected neighbors to a certain extent.

The probability of the object classified to  $\omega_i$  when it belongs to  $\omega_j$  can be calculated by Bayesian rule,

$$\phi_{ij} = \frac{w_{ji}}{\sum_g w_{gi}}, \quad (10)$$

where  $\sum_{i=1}^c \phi_{ij} = 1$ ,  $c$  is the number of the classes. We can similarly calculate the reliability matrix for each object with the given classification method (i.e., hybrid-sampling, over-sampling, and under-sampling).

The classification result of the object can be revised by this matrix. The reliability matrix provides the prior knowledge about the conditional probability of the object belonging to one class when it is classified to another class. We can get the belief of the object belonging to each class  $\omega_j$ ,  $j \in \{1, 2, \dots, c\}$ , as follows,

$$\tilde{m}_{lr}(\omega_j) = \sum_{i=1}^c \phi_{ij} m_{lr}(\omega_i). \quad (11)$$

Thus we can obtain the evidence as

$$\begin{aligned} \hat{m}_l(\omega_j) &= m_{lo}(\omega_j) + \tilde{m}_{lr}(\omega_j) \\ &= (1 - \alpha_l) m_l(\omega_j) + \sum_{i=1}^c \phi_{ij} m_{lr}(\omega_i). \end{aligned} \quad (12)$$

In our method, each piece of evidence produced by different imbalanced data classification methods (i.e., hybrid-sampling, over-sampling, and under-sampling) can be corrected similarly as explained above. DS rule as eq. (2) is employed here to combine these updated pieces of evidence to acquire the final classification result,

$$\mathbf{m}^f = \mathbf{m}_1 \oplus \mathbf{m}_2 \oplus \mathbf{m}_3, \quad (13)$$

where  $\oplus$  denotes the DS combination operation.

#### D. Parameter Optimization

Our proposed method requires two tuning parameters:  $\beta$ , and  $\lambda$ . The parameter  $\beta$  is used to determine the confidence factor  $\alpha$  by eq. (4). It can control the influence of distance on the confidence factor. The bigger  $\beta$ , the smaller confidence factor  $\alpha$ . If  $\beta$  is too big, it will make a few beliefs entering the correction process, which is not efficient for improving the classification performance. If  $\beta$  is too small, most beliefs will be redistributed even when the neighbors are not very close to the object, which may increase the risk of belief redistribution. The parameter  $\lambda$  is involved in calculating the reliability matrix  $\phi$  by eq. (9). Because the normalization operation is used to calculate the reliability matrix by eq. (10), this matrix is usually not very sensitive to the tuning of  $\lambda$  to some extent.

The optimal parameter is sought by minimizing an error criteria defined by the sum of distances between combined classifier result  $\mathbf{m}^f$  and the true label  $\mathbf{t}$ . In Matlab<sup>TM</sup>, the function *fmincon* is used to deal with this optimization problem,

$$\{\beta, \lambda\} = \arg \min_{\beta, \lambda} \sum_{i=1}^S \left\| \mathbf{m}_i^f - \mathbf{t}_i \right\|, \quad \beta \in [0.5, 1.5], \lambda > 0, \quad (14)$$

where  $\|\cdot\|$  is the Euclidean distance, and  $S$  is the number of the training dataset.  $\mathbf{m}_i^f$  is the result of combining evidence concerning the  $i$ th training data, and  $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{ic}]$ .  $t_{ij} = 1$  means the true label of  $x_i$  is  $\omega_j$ .

### III. EXPERIMENTAL APPLICATION

In this section, we will test the performance of our proposed IDCBF method with some benchmark datasets by comparing with other related imbalanced data classification methods and information fusion methods such as Smote, Nearmiss, SmoteTomek, and averaging fusion (AF) [28].

#### A. Base classifier

In our experiments, the Random Forest (RF) [29], and the K-nearest neighbors (KNN) [30], [31] are employed as base classifiers to classify the imbalanced datasets. The RF is an ensemble tool to build a decision tree. It creates multiple trees and merges them to obtain a better prediction result through maximum voting from a panel of independent judges. The KNN predicts the result by majority rule with the major class of its  $k$  most similar training data in the feature space. In the KNN, the weight of distance is set to “distance”. In all these base classifiers, the optimal parameter values (the number of trees, the maximum number of features, the minimum sample leaf size in RF, the distance in KNN) can be determined by grid search on the training data.

#### B. Benchmark datasets

Some imbalanced datasets are selected from UCI<sup>1</sup> and KEEL<sup>2</sup> dataset repository. The basic information of these data sets, including instance, attribute, class, majority instances, minority instances and imbalanced ratio, are shown in Table I. The imbalanced ratio is the ratio of the sample size of the majority data and that of the minority data, which is calculated as  $\frac{N_{major}}{N_{minor}}$ . Different from other existing binary class imbalanced data handling methods, this work contains multi-class data sets with several classes as high as ten classes in the case of the Penbased dataset. Classes that include only one example have been removed from the datasets because the Smote method cannot generate instances in only one data.

#### C. Performance evaluation metrics

We evaluate the model sensitivity towards the minority class using the Area under Curve (AUC) [32] method. In our experiment, we use the one-vs-rest strategy [33], which is also widely applied in multi-classification problem. This method computes the average AUC values for each class against the rest of the other classes. Each class generates one AUC value, and the weight of each AUC is computed by the reference class’s prevalence in the data set. The large AUC means that the classifier has high accuracy.

<sup>1</sup><http://archive.ics.uci.edu/ml>

<sup>2</sup><https://sci2s.ugr.es/keel/datasets.php>

Table I  
IMBALANCED DATASET DESCRIPTION OF THE UCI AND KEEL.

Data	Instance	Attribute	Class	Majority instances	Minority instances	Imbalanced ratio
Abalone	2560	8	3	1000	200	5
Dermatology	366	34	6	77	14	5.55
Ecoli	336	7	8	100	7	71.5
Genus	2880	2	3	2000	100	20
Pageblocks0	5472	10	2	3932	438	8.79
Penbased	1100	16	10	80	75	1.06
Pima	768	8	2	400	215	1.87
Shuttle	2175	9	4	1200	9	853
Thyroid	720	21	3	533	13	36.94
Yeast	1484	8	7	400	20	23.15
Yeast4	1484	8	2	1150	40	28.1
Yeast5	1484	8	2	971	29	32.73

Table II  
THE AUC VALUES USING RF CLASSIFIER.

Datasets	Original	Smote	Nearmiss	Smotetomek	Voting	average	IDCBF
Abalone	66.96	76.48	70.32	76.65	66.51	77.05	<b>77.71</b>
Dermatology	86.52	95.34	87.94	96.84	94.41	96.19	<b>98.08</b>
Ecoli	85.14	97.31	93.67	97.01	92.07	96.47	<b>97.75</b>
Genus	78.61	89.02	80.67	87.28	84.06	87.43	<b>88.54</b>
Pageblocks0	94.48	89.69	98.93	98.76	94.79	98.00	<b>98.96</b>
Penbased	91.07	96.55	91.26	96.17	96.23	95.66	<b>99.14</b>
Pima	71.24	81.79	84.24	83.69	75.84	84.16	<b>84.34</b>
Shuttle	88.53	96.99	89.87	97.02	93.54	97.21	<b>99.99</b>
Thyroid	99.01	<b>99.37</b>	98.96	99.37	98.43	98.38	99.01
Yeast	60.21	91.07	86.15	90.63	76.97	90.65	<b>91.16</b>
Yeast4	<b>88.96</b>	87.64	72.64	88.64	66.79	86.05	87.73
Yeast5	97.55	95.11	89.56	98.99	86.02	98.40	<b>99.41</b>
<b>Average</b>	84.02	91.36	87.01	92.58	85.47	92.13	<b>93.48</b>

Table III  
THE AUC VALUES USING KNN CLASSIFIER.

Datasets	Original	Smote	Nearmiss	Smotetomek	Voting	average	IDCBF
Abalone	59.41	68.12	65.65	71.22	<b>71.47</b>	70.29	69.62
Dermatology	57.42	60.37	64.77	65.35	57.27	68.08	<b>68.67</b>
Ecoli	81.19	93.31	93.35	93.29	90.38	95.09	<b>95.41</b>
Genus	43.57	51.32	54.32	50.67	49.67	53.69	<b>55.69</b>
Pageblocks0	90.09	93.36	92.38	95.81	90.89	95.33	<b>97.12</b>
Penbased	96.61	99.15	98.63	99.15	98.11	99.16	<b>99.19</b>
Pima	66.67	72.54	75.63	74.81	69.07	71.26	<b>75.84</b>
Shuttle	89.37	97.78	86.64	98.58	96.69	98.99	<b>99.96</b>
Thyroid	86.11	93.75	62.18	91.38	80.21	93.96	<b>95.38</b>
Yeast	80.53	84.42	84.56	87.42	76.78	<b>89.62</b>	89.61
Yeast4	62.93	66.06	79.54	81.86	78.46	82.15	<b>86.71</b>
Yeast5	79.46	86.02	96.84	92.51	86.02	98.83	<b>98.91</b>
<b>Average</b>	74.44	80.51	79.54	83.50	78.51	84.70	<b>86.00</b>

#### D. Experimental results and evaluation

The AUC values of imbalanced datasets using different classifiers are summarized in Tables II-III. The maximum AUC value is marked in boldface type. Imbalanced data classification based on belief functions theory (IDCBF) method generally produces higher accuracy than single data sampling methods. This indicates that the complementary information among different techniques is very useful for improving classification performance. We can also find that the proposed IDCBF method typically yields the highest accuracy comparing with the other combination methods. In IDCBF method, the reliability is evaluated based on the close neighbors in a refined way, and then the classifier output is cautiously revised

to improve the quality. Moreover, the involved parameter in IDCBF is automatically optimized by minimizing an error criteria. Thus, IDCBF is able to produce the best classification accuracy in general.

#### IV. CONCLUSION

In this paper, we have proposed a new method for combination of classifiers to solve the imbalanced data classification. Imbalanced data classification based on belief functions theory (IDCBF) method is able to take advantage of essential complementary information among different data sampling techniques to improve classification performance. Multiple imbalanced datasets are used to validate the performance of the proposed method. The experimental results show that

the IDCBF method is able to improve classification result comparing with other data sampling techniques and fusion methods.

## REFERENCES

- [1] M. Mohammadi, et al., *A comprehensive survey and taxonomy of the SVM-based intrusion detection systems*, Journal of Network and Computer Applications, Vol. 178,102983, 2021.
- [2] D.A. Otchere, et al., *Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models*, Journal of Petroleum Science and Engineering, Vol. 200, 108182, 2021.
- [3] D. Bhatt, et al., *CNN variants for computer vision: History, architecture, application, challenges and future scope*, Electronics, Vol. 10(20), 2470, 2021.
- [4] Y.L. Tian, et al., *Rethinking few-shot image classification: a good embedding is all you need*, in Proc. 16th European Conference of Computer Vision (ECCV 2020), Glasgow, UK, August 23-28, 2020.
- [5] K.H. Tang, J.Q. Huang, H.W. Zhang, *Long-tailed classification by keeping the good and removing the bad momentum causal effect*, Advances in Neural Information Processing Systems, pp. 1513–1524, 2020.
- [6] K.H. Tang, et al., *Invariant feature learning for generalized long-tailed classification*, in Proc. of European Conference on Computer Vision (ECCV 2022), Cham: Springer Nature Switzerland, 2022.
- [7] C.J. Feng, Y.J. Zhong, W.L. Huang, *Exploring classification equilibrium in long-tailed object detection*, in Proc. of the IEEE/CVF International conference on computer vision, 2021.
- [8] W. Xiong, Q. Gu, B. Li, et al, *Collaborative web service QOS prediction via location-aware matrix factorization and unbalanced distribution*, Journal of Internet Technology, Vol. 19(4), pp. 1063–1074, 2018.
- [9] S. Pouyanfar, S.C. Chen, *Automatic Video Event Detection for Imbalance Data Using Enhanced Ensemble Deep Learning*, International Journal of Semantic Computing, Vol. 11(01), pp. 85–109, 2017.
- [10] J. Niu, Z. Liu, *Imbalance Data Classification Based on Belief Function Theory*, in Proc. of 6th Int. Conf. on Belief Functions (Belief 2021), Springer LNAI 12915, Shanghai, China, Oct. 15-19, 2021.
- [11] J. Niu, Z. Liu, Y. Lu, Z. Wen, *Evidential Combination of Classifiers for Imbalanced Data*, IEEE Trans. on SMC: Systems, Vol. 52(12), December 2022.
- [12] X. Yang, Q. Kuang, W. Zhang, et al, *AMDO: an Over-Sampling Technique for Multi-Class Imbalanced Problems*, IEEE Trans. on Knowledge and Data Engineering, Vol. 99, 2017.
- [13] H. Ali, et al., *A review on data preprocessing methods for class imbalance problem*, International Journal of Engineering & Technology, Vol. 8, pp. 390–397, 2019.
- [14] P. Zyblewski, S. Robert, W. Michał, *Preprocessed dynamic classifier ensemble selection for highly imbalanced drifted data streams*, Information Fusion, Vol. 66, pp. 138–154, 2021.
- [15] Y. Xiao, W.U. Zhefu, T. Zhang, et al., *Feature Selection Based Classification Algorithm with Imbalanced Data*, Journal of Integration Technology, Vol. 5(1), pp. 68–74, 2016.
- [16] H.Y. Liu, M.C. Zhou, Q. Liu, *An embedded feature selection method for imbalanced data classification*, IEEE/CAA Journal of Automatica Sinica, Vol. 6(3), pp. 703–715, 2019.
- [17] H.M. Chen, et al., *Feature selection for imbalanced data based on neighborhood rough sets*, Information sciences, Vol. 483, pp. 1–20, 2019.
- [18] S. Dhar, V. Cherkassky, *Development and evaluation of cost-sensitive universum-SVM*, IEEE Trans on Cybernetics, Vol. 45(4), pp. 806–818, 2017.
- [19] A.N. Tarekegn, G. Mario, M. Krzysztof, *A review of methods for imbalanced multi-label classification*, Pattern Recognition, Vol. 118, 107965, 2021.
- [20] M. Chen, L.U. Xiaoyong, *Three random under-sampling based ensemble classifiers for Web spam detection*, Journal of Computer Applications, 2017.
- [21] N.V. Chawla, K.W. Bowyer, L.O. Hall, et al., *SMOTE: Synthetic Minority Over-sampling Technique*, 2011.
- [22] C. Seiffert, T.M. Khoshgoftaar, J.V. Hulse, *Hybrid sampling for imbalanced data*, in Proc. of IEEE International Conference on Information Reuse & Integration, IEEE, 2008.
- [23] D. Devi, S.K. Biswas, B. Purkayastha, *Redundancy-driven modified Tomek-link based undersampling: A solution to class imbalance*, Pattern Recognition Letters, Vol. 93, pp. 3–12, July 2017.
- [24] L.A. Zadeh, *Review of A Mathematical Theory of Evidence*, AI Magazine, Vol. 5(3), pp. 81–83, Fall 1984.
- [25] A. Skowron, J. Grzymala-Busse, *From Rough Sets Theory to Evidence Theory*, Advances in the Dempster-Shafer Theory of Evidence, John Wiley & Sons, 1994.
- [26] Z. Liu, Q. Pan, J. Dezert, J. Han, Y. He, *Classifier Fusion With Contextual Reliability Evaluation*, IEEE Trans. on Cybernetics., Vol. 48(5), pp. 1605–1618, 2018.
- [27] Z. Liu, Y. Liu, J. Dezert, F. Cuzzolin, *Evidence Combination Based on Credal Belief Redistribution for Pattern Classification*, IEEE Trans. on Fuzzy Systems, Vol. 28(4), pp. 618–631, 2019.
- [28] B.M. Rodriguez, G.L. Peterson, S.S. Agaian, *Multi-Class Classification Averaging Fusion for Detecting Steganography*, in Proc. of 2007 IEEE Int. Conf. on System of Systems Engineering, pp. 505–512, 2007.
- [29] M.W. Ahmad, M. Mourshed, Y. Rezugui, *Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption*, Energy and Building, Vol. 147, July 2017.
- [30] T.M. Cover, P.E. Hart, *Nearest neighbor pattern classification*, IEEE Trans. on Information Theory, Vol. 13(1), pp. 21–27, January 1967.
- [31] I. Tomek, *Experiment with Edited Nearest-Neighbor Rule*, IEEE Trans. on Syst., Man, Cybern., Syst., Vol. 6(6), pp. 448–452, January 1976.
- [32] P. Bradley, *The use of the area under the ROC curve in the evaluation of machine learning algorithms*, Pattern Recognition, Vol. 30(7), pp. 1145–1159, July 1997.
- [33] F. Provost, P. Domingos, *Well-trained PETs: Improving probability estimation trees*, CeDER Working Paper, 2001.

# Decision Based on Belief Interval for Multi-class Obstacle Perception of Self-driving Cars

Dănuț-Vasile Giurgi<sup>a</sup>, Mihreteab Negash Geletu<sup>a</sup>, Thomas Josso-Laurain<sup>a</sup>,  
Maxime Devanne<sup>a</sup>, Jean-Philippe Lauffenburger<sup>a</sup>, Jean Dezert<sup>b</sup>

<sup>a</sup>IRIMAS-UR7499, Université de Haute-Alsace, Mulhouse, France.

<sup>b</sup>The French Aerospace Lab, ONERA, Palaiseau, France.

Emails: vasile.giurgi@uha.fr, mihreteab-negash.geletu@uha.fr, thomas.josso-laurain@uha.fr,  
maxime.devanne@uha.fr, jean-philippe.lauffenburger@uha.fr, jean.dezert@onera.fr

**Abstract**—The self-driving cars face important challenges in the applications of perception tasks. The driving surrounding areas are often chaotic and the conditions of the weather vary significantly. In terms of sensors, the capacities have increased, raising interest in big data fields such as artificial intelligence. Alongside, fusion techniques allow accurate coupling of information from different sources. Neural networks have proved good performance, but limitations in complex situations still occur. In the decision-making world, evidence theory increases the robustness of decisions and handles conflict management. In this research, a deep learning architecture based on a camera-lidar cross-fusion technique is proposed to achieve semantic segmentation capabilities. The model is coupled with evidence theory and serves for semantic segmentation tasks. The architecture has the decision-making part relying on the decision belief interval and the information that can be represented through belief functions. The evidence theory is versatile and contributes to understand better imprecise data and to achieve more efficient predictions. The KITTI dataset is used in this work. The results highlight the interest in integrating belief theory functions into deep learning architecture fusing information from two heterogeneous sensors.

**Keywords:** intelligent vehicles, environment perception, evidence theory, belief functions, deep learning.

## I. INTRODUCTION

### A. Self-driving cars

Autonomous cars grasp big improvements thanks to approaches based on neural networks. The goal of self-driving cars is to offer safe driving and efficiency, minimizing the routine tasks of humans and putting forward better transportation. To facilitate quicker levels of autonomy, the cars are equipped with various sensors like cameras, point cloud devices, and more. In this way, perception can benefit from multi-modal sensor fusion from different sources of information to increase the robustness of the decisions. Alongside with, self-driving cars rely on both reference generation (path and trajectory planning) and control theory techniques [1]. The intelligent vehicles provide therefore localization and environment understanding about the object and traffic participants, so that, the navigable area can be projected and followed using control algorithms.

### B. Motivation

This paper focuses on the first hierarchical step of autonomous driving: perception, and particularly surrounding environment perception. In scene analysis, there are various particularities to be considered. For example, on the path planning side, the interval distance between a car and a sidewalk has a different sensibility than the interval distance between two cars. In the risk analysis state of the art, rules for safe minimal distance are referred to. These protocols differ with respect to the situation, for instance, at least 1 meter distance between two cars longitudinally on the lane is required. However, there is no specified rule related to the tolerated lateral distance between a car and a sidewalk, where the environment is prone to involve pedestrians, or just the interval between two cars, where there is a risk of crashing as well [2]. Within these situations, imprecise information from the perception system could result in inaccurate control actions. Consequently, decision-making approaches that can represent efficient information are required, so that, the perception and path-planning control chain are well-assured.

Neural networks approaches, particularly prominent deep learning methods [3]–[5] have been slightly designed for environment perception features to bring value in detection, classification, and segmentation tasks. Besides neural networks, fusion methods broadcast meaningful information from various sensors to empower the robustness of decisions.

The benefits introduced by deep learning facilitate perception tasks in the environment and help understand the self-driving car needs. However, since the field is sensitive to the quality of decisions, confidence in decision-making is required. A relevant way to both represent and trust information is to use the belief theory. This approach is a well-known framework utilized in the world of probabilities and information reasoning.

### C. Belief functions theory

The theory of evidence, also known as Belief Functions (BF) theory or Dempster-Shafer Theory (DST), was proposed by Shafer in 1976 [6] based on previous works of Dempster [7]. It represents evidence elements (i.e. beliefs) for uncertain models. The evidence theory key features are the following:

*generality* (it extends both propositional logic and probabilistic reasoning), *operationality* (works with elementary pieces of evidence coupled with Dempster-Shafer's rule<sup>1</sup> of combination), and *scalability* (evidential reasoning can be addressed to more complicated problems in terms of decisions and uncertainties from different sources of information), making it richer than the theory of probabilities for dealing with epistemic uncertainty. In autonomous driving tasks, such as obstacle avoidance, BF proves to deliver accurate performances, for example, the occupancy grid map of a LiDAR sensor by expressing conflict in more representative ways [8]. In the scope of pedestrian detector tasks, evidential combination rules fetched considerable performances over the Bayesian approach [9].

Moreover, in multi-model perception, evidential theory handles missing information, imprecision, and ignorance. In [10], KITTI semantic segmentation images from various sensors, cameras, and different layers of LiDAR are embodied together, which allows for enlarging the object classes or the number of sensors. The approach seeks to improve performances for a better understanding of the drivable area.

In the deep-learning field, Cappelier et al. [11] propose a neural network architecture based on MLP (Multi-Layer Perceptron) to classify arbitrary LiDAR objects for perception. Their model replaces the probabilistic output with an evidential inference method, inspired by the generalized logistic classifier of Denœux [12].

Thus, frameworks based on belief functions evince promising results in perception systems for both road segmentation and multi-object detection tasks, which are the main topics addressed in this work. The main goal is then to provide an evidential deep-learning model that fuses information from different sensors to achieve autonomous driving capabilities.

The approaches based on belief functions delivered significant improvements for different axes of research such as decision-making, conflict management, and fusion [13]. The partial or total ignorance produced by the evidence theory are very appealing to model uncertainties. However, combining efficiently several independent belief functions and agreeing on a final decision from a belief function are challenging tasks, especially the decision-making under uncertainties for defense and security applications, like autonomous vehicles. In [14], Dezert et al. proposed a new decision-making method based on a belief interval distance that helps the decision-making process under uncertainty. This decision-making approach provides a judgment by selecting the best focal element (i.e. object class) for which the minimal distance with respect to the piece of evidence under concern is obtained. This decision-making approach also provides the calculation of a quality (or confidence factor) characterizing the quality of decision (i.e. the final judgment) for a future action.

## II. BASICS OF BELIEF FUNCTIONS

Evidence theory is a formalism for reasoning and making a decision with uncertainty. The classical approach of evidence

theory is based on Dempster-Shafer rule of combination. A more detailed discussion can be found in [6], [15], which is adopted in this work.

Let  $\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_n\}$  be a finite set of mutually exclusive elements, called the *frame of discernment* (FoD), and the mutually exclusive elements of single cardinality are called *singletons*. A *basic belief assignment* (BBA), or mass function  $m(\cdot)$ , is a mapping  $m : 2^\Theta \rightarrow [0, 1]$  such that:

$$m(\emptyset) = 0 \quad (1)$$

$$\sum_{X \subseteq \Theta} m(X) = 1 \quad (2)$$

The quantity  $m(X)$ , known as the mass of element (i.e. subset)  $X$  of  $\Theta$ , measures the belief that one commits exactly to  $X$ ; and (1) indicates closed world assumption. The subset  $X$  is called a *focal element* of  $m(\cdot)$  if and only if  $m(X) > 0$ .

Given a BBA  $m(\cdot)$ , two concepts can be defined, a *belief function* (*Bel*) and a *plausibility function* (*Pl*) using the following expressions:

$$Bel(X) = \sum_{Y \subseteq X} m(Y) \quad (3)$$

$$Pl(X) = \sum_{Y \cap X \neq \emptyset} m(Y) = 1 - Bel(\bar{X}) \quad (4)$$

$Bel(X)$  can be interpreted as the degree of total support to  $X$ , whereas  $Pl(X)$  is the degree one fails to doubt  $X$ .

If the BBA  $m(\cdot)$  is only focalized on the whole set  $\Theta$ , i.e.  $m(\Theta) = 1$ , the BBA  $m(\cdot)$  is called the *vacuous* BBA, which models the total *ignorance*.

In DST, two BBAs  $m_1$  and  $m_2$  representing independent pieces of evidence are combined by Dempster's rule defined by:

$$(m_1 \oplus m_2)(X) = \frac{1}{1 - K} \sum_{Y \cap Z = X} m_1(Y)m_2(Z) \quad (5)$$

For all  $X \subseteq \Theta$ ,  $X \neq \emptyset$ , and  $(m_1 \oplus m_2)(\emptyset) = 0$ . The degree of conflict between the two BBAs, denoted by  $K$ , is given by:

$$K \triangleq \sum_{Y \cap Z = \emptyset} m_1(Y)m_2(Z) \quad (6)$$

This DS rule of combination can be easily generalized for the combination of more than two sources of evidence. DS rule is commutative and associative which is very appealing for its implementation because the fusion of several sources can be done sequentially and the sequential fusion order does not matter. In the vehicle perception applications developed so far in the IRIMAS lab, the DS rule produces generally good outcomes, but because the DS rule is not always exempt from leading to some decision issues due to its dictatorial behavior in some cases [16], [17], alternative research works and comparative analysis with other fusion rules are also under consideration.

<sup>1</sup>also referred as DS rule in the literature.

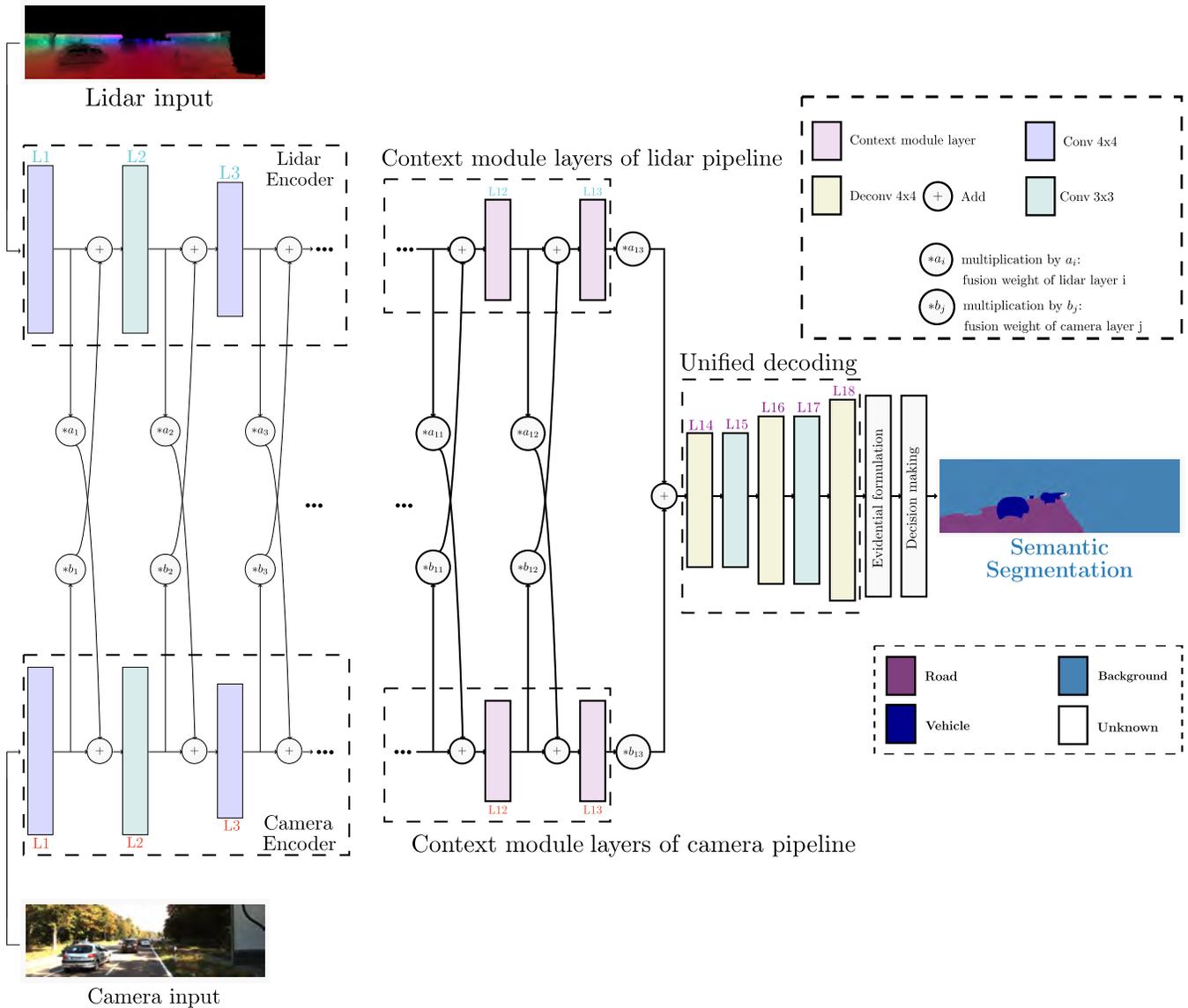


Figure 1: Evidential Lite-CF Architecture: Semantic Segmentation.

### III. PROPOSED METHOD FOR AUTOMATIC PERCEPTION

#### A. Evidential Cross-Fusion Architecture

In a previous work, an innovative camera-lidar fusion method known as Lite Cross-fusion was proposed, featuring a fully convolutional neural network for road recognition as detailed in [18]. This cross-fusion architecture exhibited significant results compared with early or late fusion approaches. Consequently, the network was subsequently integrated in another work [19], ensuing a reduction in the computational complexity, by 15%. The architecture is founded upon an encoder-decoder model that leverages dilated convolution for strengthening image resolution conservation. This architecture represents the baseline of this work in terms of neural networks and multimodal fusion. Regarding evidence theory, [11], [20] propose the evidential classifier for classification tasks. The decision-making approach is based on the distance to pro-

totypes method as a substitute for the conventional softmax decision layer.

Taking into account the two previous methods, specifically the cross-fusion road detection (Lite-CF) and the architecture based on evidential classifiers, this work, introduces a combination of these two methodologies. The resulting architecture, referred to as Lite CF-Evi, combines the strengths of the two frameworks designed for semantic segmentation tasks. An overview of the complete architecture for evidential Lite-CF is illustrated in Fig. 1.

The standard neural network architecture produces probability distributions from logits using a softmax layer. The evidential Lite-CF produces mass functions (BBAs) rather than probabilities to represent the prediction of imprecise data. It has an encoder-decoder-based network, evidential formulation layer, and a decision-making unit. The encoding section has

two processing pipelines of 13 layers each, one for feeding LiDAR input and the other for the camera frames. At each layer level, information from one modality is combined with the corresponding layer from the other modality through a trainable weighted sum operation ( $*b_i$  and  $*a_i$  respectively, where  $i$  is the layer number). These fusion weights are adaptable, allowing the fusion's position and its extent to be fixed by the data.

Once the LiDAR and camera inputs are transformed into Basic Belief Assignments (BBAs) within the evidential formulation layer, decisions can be rendered concerning specific elements within the power set<sup>2</sup> denoted by  $2^\Theta$ . In the semantic segmentation task, this power set encompasses elements such as "road", "vehicle", and "background" elements in the probabilistic version, and additionally "unknown" area in the evidential formulation. Consequently, the evidential approach enables having an uncertain prediction represented by the disjunction of single classes and interpreted as an "unknown" class.

### B. Decision-Making. Distance to Prototypes

The evidential formulation layer uses as its input, the feature maps generated by the decoding section. When the decoder reaches its maximum resolution, BBAs are generated by computing the distances between their corresponding feature maps (i.e., L18 in Fig. 1) and the propagated prototypes, which are learned automatically. The technique is called distance to prototypes. An illustration of the approach is shown in Fig. 2.

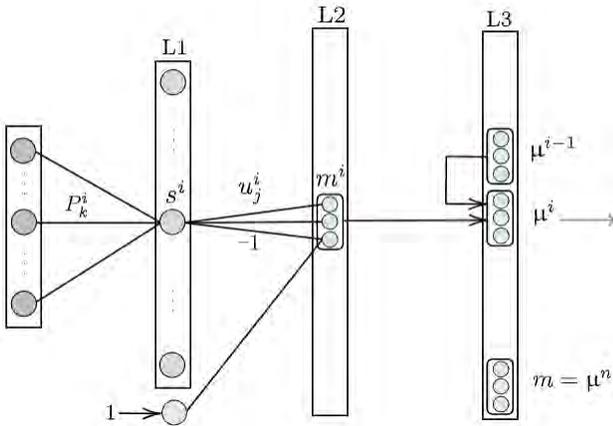


Figure 2: Distance to Prototypes [21].

The distance to prototypes methodology can be described in three steps procedure as it follows (more details in [21]):

Step 1: Distance to prototype: Let  $x$  be a feature vector representing features of a pixel to be classified possibly as  $class_1$  ( $\theta_1$ ) or  $class_2$  ( $\theta_2$ ) (i.e., the

FoD  $\Theta = \{\theta_1, \theta_2\}$ ). The Euclidean distance  $d^i$  is determined between  $x$  and each prototype  $p^i$ :

$$d^i = \|x - p^i\| \quad i = 1, \dots, n. \quad (7)$$

Step 2: Establish the correspondence of mass functions to prototypes and their interference: Each prototype  $p^i$  has a degree of membership  $u_j^i$  to each class  $\theta_j$ , with a constraint  $u_1^i + u_2^i = 1$ . Using the class membership  $u_j^i$  and the distance  $d^i$ , a BBA  $m^i$  is constructed as:

$$\begin{aligned} m^i(\{\theta_j\}) &= \alpha^i u_j^i \phi^i(d^i), \quad j = 1, 2 \\ m^i(\Theta) &= 1 - \alpha^i \phi^i(d^i), \end{aligned} \quad (8)$$

ensuring that the cumulative mass sum equals 1, as indicated in the subsequent formula:

$$\sum_{X \subseteq \Omega} m^i(X) = \sum_{j=1}^2 m^i(\{\theta_j\}) + m^i(\Theta) = 1 \quad (9)$$

where, in the expression (8),  $0 < \alpha^i < 1$ , and the decreasing function  $\phi^i$  are defined as:

$$\phi^i(d^i) = \exp(-\gamma^i (d^i)^2), \quad \gamma^i > 0 \quad (10)$$

Step 3: Combination: The BBAs from step 2 are combined using Dempster's rule (5). The resulting combined BBA represents the evidence to make a decision on the pixel class.

The parameters associated with the prototype  $p^i$  (i.e.,  $\alpha^i$ ,  $u_j^i$ , and  $\gamma^i$ ), are incorporated into the evidential deep learning-based architectures as weighting factors.

However, the learnable weights are not inherently restricted. Therefore, they are redefined and implemented in terms of some real number valued variables  $\eta^i$ ,  $\xi^i$ , and  $\beta_j^i$ :

$$\gamma^i = (\eta^i)^2 \quad (11)$$

$$\alpha^i = \frac{1}{1 + \exp\{-\xi^i\}} \quad (12)$$

$$u_j^i = \frac{(\beta_j^i)^2 + \epsilon}{\sum_{k=1}^2 ((\beta_k^i)^2 + \epsilon)} \quad (13)$$

Equation (13) is slightly modified from the expression given in [21]. To avoid the membership values  $u_j^i$  from becoming zero, a small positive term denoted as  $\epsilon$  is introduced. This precautionary measure handles conflict limitations that could occur in Dempster's *total conflict* (i.e. to prevent the case with  $K = 1$ ).

Following the previous steps and the Fig. 2, a simple case with only two classes can be exemplified. A frame of discernment such as the following FoD:  $\Theta = \{R, V\}$  (where R stands for road, V for vehicle) is considered. Thus, for feature vector  $x$  the neural network model outputs some pixel characteristic values corresponding to road and vehicle. Firstly in level 1 (L1) distances to prototypes ( $p_i$ ) are calculated. In this case, the number of prototypes has to be defined and is tunable. Moreover, it influences directly the model's complexity and

<sup>2</sup>The power set of  $\Theta$  is the set of all the subsets of  $\Theta$ , the empty set  $\emptyset$  and  $\Theta$  included.

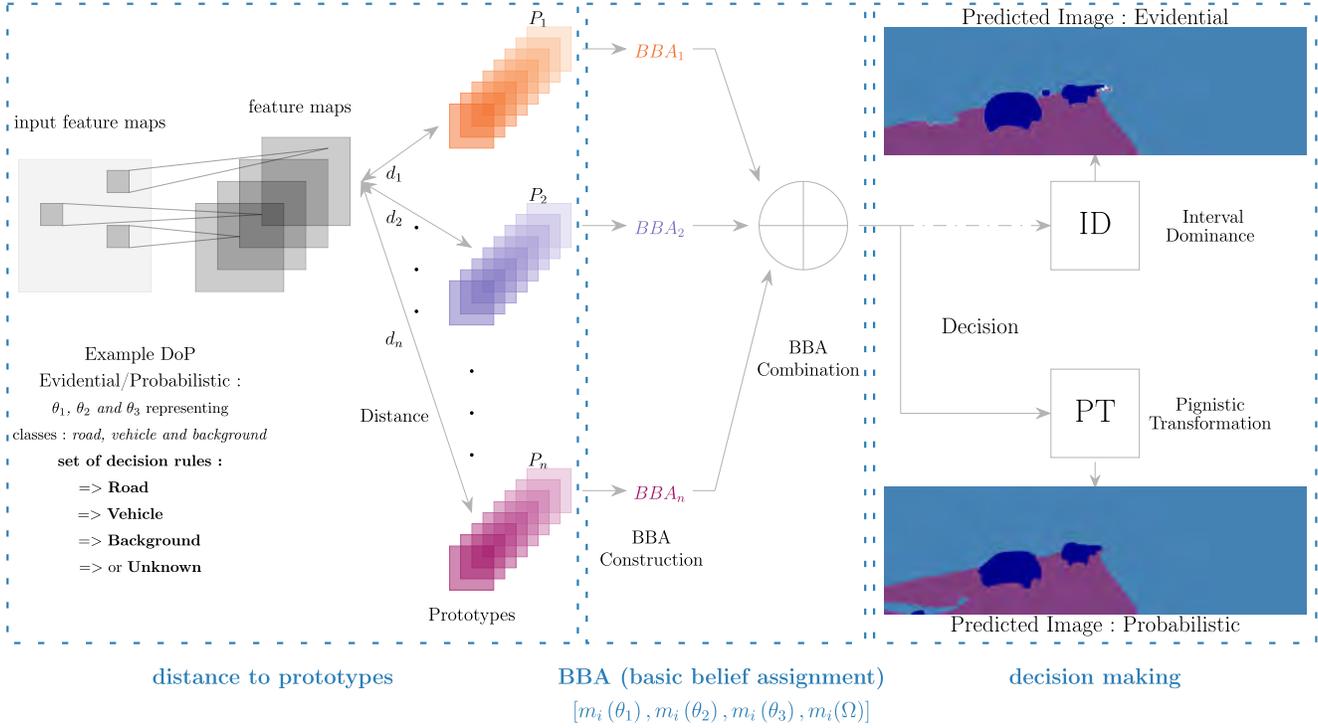


Figure 3: Evidential Formulation - Decision making using BBAs.

it has to be at least equal to the number of classes. The more prototypes are injected, the more complex the model is. Continuously, the example with 3 steps approaches is reproduced as:

- Step 1 (L1): Distance to prototype  
 $p^1 = [-1.5 \ -2.5 \ -3.5]^T$ ,  $\gamma^1 = 0.1$ ,  $\alpha^1 = 0.5$  and  
 $p^2 = [2.3 \ 1.5 \ 3.1]^T$ ,  $\gamma^2 = 0.1$ ,  $\alpha^2 = 0.5$   
 Then expressing equation (7), the following distances are computed:  
 $d^1 = 5.10$  and  $d^2 = 3.70$   
 Continuously, with respect to expression (10), the activations  $s^i$  have the following values:  
 $s^1 = 0.04$  and  $s^2 = 0.13$

Once the Euclidean distances based on prototypes and their activations are set, within the level 2, mass functions are constructed:

- Step 2 (L2): BBA construction  
 Classes  $\{R, V\}$  have the membership degrees (13)  $u^i$ :  
 $u^1 = [0.8 \ 0.2]^T$  and  $u^2 = [0.3 \ 0.7]^T$   
 The BBAs  $m^i$  are computed following expression (8):  
 $m^1 : m^1(R) = 0.03$ ,  $m^1(V) = 0.01$ ,  $m^1(\Theta) = 0.96$   
 $m^2 : m^2(R) = 0.04$ ,  $m^2(V) = 0.09$ ,  $m^2(\Theta) = 0.87$

Continuously, for the last level, the mass functions from the previous step are calculated with DS rule (5):

- Step 3 (L3): DS combination of  $m^1$  with  $m^2$ . Mass function values after computing the conjunctive rule:  
 $m(R) = 0.0657$ ,  $m(V) = 0.096$ ,  $m(\Theta) = 0.8352$ ,  
 $m(\emptyset) = 0.0031(K)$

Finally, the BBAs after Shafer's normalization by  $1 - K$ ,  $m_{DS}(\cdot) = m(\cdot)/(1 - K)$  are:  
 $m_{DS}(R) = 0.0659$ ,  $m_{DS}(V) = 0.0963$ ,  $m_{DS}(\Theta) = 0.8378$ ,  $m_{DS}(\emptyset) = 0$ , and their total sum is 1.

In this case, the conflict is very low, the highest value of the mass function is the vacuous  $m(\Theta)$ , meaning that model does not have enough evidence to support road or vehicle.

A scheme that illustrates how the Distance to Prototypes approach works is presented in Fig. 3. Since this illustration represents the perception application, it shows two options for the decision-making, both evidential and probabilistic for a semantic segmentation task with three classes: road, vehicle, and background.

The probabilistic approach relies on Bayesian formulation for decision-making and is based on the classical pignistic transformation. The evidential formulation part is based on interval dominance for the decision-making before the final prediction, but it can be substituted with another method such as Jousselme's distance [22] or the Decision based on Belief Interval (dBI) method [14]. The latter one represents the main baseline for the contribution of this work. The method is implemented and applied to semantic segmentation tasks to help the decision-making part of the cross-fused evidential deep learning model.

### C. Decision based on Belief Interval

Han et al. present different decision-making using belief functions in [23], and they propose an Euclidean belief interval distance  $d_{BI}(m_1, m_2)$  between two mass functions  $m_1(\cdot)$  and

$X \in 2^\Theta$	$m(R)$	$m(V)$	$m_3(R \cup V)$	$m(B)$	$m(R \cup B)$	$m(V \cup B)$	$m(R \cup V \cup B)$
$m(\cdot)$	0.83	0.17	0	0	0	0	0
$d_{BI}^{(m_i, m_X)}$	0.1700	0.8300	0.4384	0.9268	0.5265	0.779	0.5991
$d_{BI}^{(m_i, m_X)}$	0.1700	-	-	-	-	-	-
$q(\hat{X})$	0.9118	0.5692	0	0.5190	0	0	0

 Table I: Example decision belief interval and quality of decision  $q(\hat{X})$ .

$m_2(\cdot)$  represented on the power set for the aforementioned FoD  $\Theta = \{\theta_1, \theta_2, \theta_3, \dots, \theta_n\}$ . This  $d_{BI}$  distance is defined by:

$$d_{BI}(m_1, m_2) \triangleq \sqrt{N_c \cdot \sum_{X \in 2^\Theta} d_W^2(BI_1(X), BI_2(X))} \quad (14)$$

$N_c = 1/2^{n-1}$  represents a normalization factor to have  $d_{BI}(m_1, m_2) \in [0, 1]$ , and  $d_W(BI_1(X), BI_2(X))$  is the Wasserstein's distance, [24], between belief intervals  $BI_1(X) \triangleq [Bel_1(X), Pl_1(X)] = [a_1, b_1]$  and  $BI_2(X) \triangleq [Bel_2(X), Pl_2(X)] = [a_2, b_2]$ .

Wasserstein's distance is concisely noted as:

$$d_W = d_W([a_1, b_1], [a_2, b_2]) \quad (15)$$

and it explicitly has the following expression:

$$d_W \triangleq \sqrt{\left[ \frac{a_1 + b_1}{2} - \frac{a_2 + b_2}{2} \right]^2 + \frac{1}{3} \left[ \frac{b_1 - a_1}{2} - \frac{b_2 - a_2}{2} \right]^2} \quad (16)$$

Continuously,  $m_X$  represents the categorical BBA, that contains only  $X$  as the focal element, while  $\hat{X}$  is the final decision. The final decision is characterized by the minimum between the mass functions  $m(\cdot)$  and  $m_X$ ,  $X \in 2^\Theta \setminus \{\emptyset\}$ .  $\hat{X}$  is defined as:

$$\hat{X} = \arg \min_{X \in 2^\Theta \setminus \{\emptyset\}} d_{BI}(m, m_X) \quad (17)$$

where  $d_{BI}(m, m_X)$  is calculated according to expression (14). Here,  $m(\cdot)$  represents the mass functions under consideration, and  $m_X(\cdot)$  is the categorical BBA focused on a chosen focal element  $X$ .

The decision-making methodology is scalable implying the advantages of the model to consider a decision among all the power-set  $2^\Theta$  possibilities. Consequently, decision space can encapsulate more elements than the singletons of the frame of discernment, including unions (i.e. disjunctions) of these ( $singleton_1 \cup singleton_2$ ). More details can be found in [14]. Alongside the final decision  $\hat{X}$ , the method proposed in [14] is also able to assess how good (or trustable) the decision  $\hat{X}$  is, considering other alternatives. Thus, the quality indicator (confidence factor)  $q(\hat{X})$  is defined by:

$$q(\hat{X}) \triangleq 1 - \frac{d_{BI}(m, m_{\hat{X}})}{\sum_{X \in 2^\Theta \setminus \{\emptyset\}} d_{BI}(m, m_X)} \quad (18)$$

Undoubtedly, the value of the quality indicator  $q(\hat{X})$  will be bigger (and close to 1) when the model is more confident in making the decision  $\hat{X}$ .

Once the BBAs representing the pieces of evidence in the corresponding pixels are evaluated, a final task remains to

decide the classes from pixels. Therefore, given the previous statement (17) as presented in [14], the decision rule can be expressed by three classes for the semantic segmentation task as following:

Case i) The decision is constrained to singletons: The possible judgment elements are  $\theta_1$  (road),  $\theta_2$  (vehicle) and  $\theta_3$  (background). In this situation, the expression (17) becomes:

$$\hat{X} = \arg \min_{X \in \{R, V, B\}} d_{BI}(m, m_X) \quad (19)$$

Case ii) The decision is not constrained: It might be interesting to allow the assignment of ambiguous pixels to imprecise classes like  $\Theta$ . This can reduce classification errors by avoiding decisions that have more of an arbitrary nature.

Recalling the example of distance to prototypes approach with two classes, road, and vehicle, here an example with the following FoD:  $\Theta = \{R, V, B\}$  is presented. These three elements are representative of the goal of this work. Therefore,  $R$  stands for road,  $V$  for vehicle,  $B$  for background, and the decision space includes all singletons, and their disjunctions, that is  $2^\Theta \setminus \{\emptyset\} = \{R, V, R \cup V, B, R \cup B, V \cup B, R \cup V \cup B\}$  which appears in formula (17).

In this case, and considering Fig. 3, the decision-based on belief interval distance method is positioned in the middle phase, within the BBA construction and it represents the final decision for the prediction. The  $d_{BI}$  approach reflects another way of combining efficiently the mass functions to handle uncertain predictions.

For the given example, both the decision based on interval and its quality indicator are explained and highlighted in the above Table I. In this case, the mass functions have assigned the following values:

- road:  $m(R) = 0.83$ , vehicle:  $m(V) = 0.17$
- background:  $m(B) = 0$
- unions of classes:
  - road-vehicle:  $m(R \cup V) = 0$
  - road-background:  $m(R \cup B) = 0$
  - vehicle-background:  $m(V \cup B) = 0$
  - road-vehicle-background:  $m(R \cup V \cup B) = 0$ .

Continuously, the distances based on the  $d_{BI}$  method are calculated. As expected in this situation, the distance corresponding to the element  $R$  (road) is minimal (0.17), because  $R$  has the highest value of the mass function. Given this, the minimal distance will follow as the final decision. Quality indicators are computed accordingly. As highlighted in the table I, the  $q(\hat{X})$  values show that the model is pretty confident when considering the element  $R$ , with a confidence of 91.18%.

In the realm of decision-making methodology, it is important to get accurate judgments, and to assess how efficient and trustable is the decision made for the application under concern.

Perception systems based on neural networks are more sensitive in terms of making a decision and encountering uncertainties, therefore confidence in a judgment is crucial. Correspondingly, this work highlights perception applications using multi-sensors, fusion, and neural networks, evidence theory-based that provides decision-making features and assesses the quality of decisions.

#### IV. EXPERIMENTAL RESULTS

This section presents the dataset used for the application and the obtained results. The Lite-CF-Evi architecture is assessed for scene segmentation tasks against the KITTI semantic segmentation database. Firstly the probabilistic and evidential results are exemplified. Alternatively, the decision belief interval and quality indicators are computed for a set of frames.

##### A. Dataset

The semantic KITTI dataset has originally only 200 camera images. The dataset is similar to KITTI Stereo and KITTI Flow 2012/2015 datasets. Since the KITTI semantic has no LiDAR frames (like the road dataset for instance), the corresponding 3D point-cloud points of the existing camera frames have to be identified in the big original KITTI raw dataset [25], which contains the data for all tasks. Hence, for 127 out of the 200 camera images, LiDAR frames have been successfully projected and up-sampled to create dense depth images. A 3D LiDAR point  $x$  is mapped into a point  $y$  in the camera plane according to the KITTI projection  $P$ , rectification  $R$  and translation  $T$  matrices:

$$y = P R T x \quad (20)$$

As the projected LiDAR scan is sparse, up-sampling is employed to generate a dense depth map, as depicted in Fig. 4. The up-sampling process is implemented following the method outlined in [19] and [26].

##### B. Semantic Segmentation

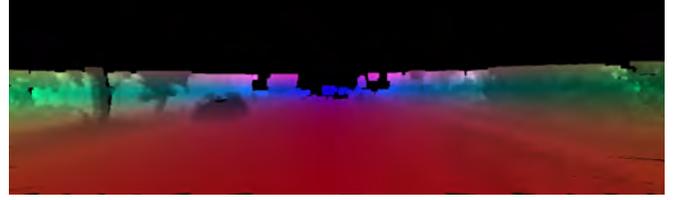
After the up-sampling process, the newly constructed dense depth images from LiDAR are integrated into the Lite-CF-Evi model in parallel with the camera images to feed the two pipeline inputs of the architecture.

Concerning the ground truth, the masks are simplified to 3 classes: road (magenta), vehicle (dark blue), and background (blue), according to the original annotation. The road class is preserved, however, the vehicle class incorporates annotations such as car, truck, and bus of the original ground truth. In turn, the background class encapsulates all the other classes, except for the above-mentioned ones.

Fig. 5a shows an illustration with an example of the original ground truth, while Fig. 5b describes the simplified ground truth.



(a) LiDAR projection.



(b) LiDAR up-sampling.

Figure 4: LiDAR pre-processing.



(a) Original ground truth.



(b) Simplified ground truth, 3 classes: road, vehicle, background.

Figure 5: Dataset pre-processing.

The dataset consists of 127 images: 114 for training and 13 for validation. This method has been exclusively assessed using the specially reconstructed KITTI semantic dataset, which includes the added LiDAR frames for the evidential cross-fusion architecture. To the best of the author's knowledge, this dataset has not yet been examined by any other methods, unless their own work, since the point clouds were identified in the raw dataset and included. The ground-truth masks are one-hot encoded and class weight is applied to address the unbalanced data. Consecutively, the model is trained for 500 epochs using mean squared error loss and Adam optimizer.

To measure the performances, the model is evaluated using the intersection-over-union metric, denoted as  $IoU$ , in accordance with the PASCAL VOC benchmark [27]:

$$IoU = \frac{TP}{TP + FP + FN} \quad (21)$$

with TP, FP, and FN, respectively, true positive, false positive, and false negative.

The Lite-CF-Evi is evaluated for 3 classes in a probabilistic manner. It can be observed that the global mean  $IoU$ , 92.707%, in the evidential architecture, is higher than 92.384% for the probabilistic model. Individually over each class, the evidential model outperforms the probabilistic one (Table II), and visually the results are better for the Lite-Cf Evi.

IoU \ Arch.	Probab. Lite-CF	Lite-CF Evi.
mean IoU	92.384%	92.707%
mean $IoU_{road}$	92.713%	93.163%
mean $IoU_{vehicle}$	87.118%	87.446%
mean $IoU_{background}$	97.322%	97.513%

Table II: Mean Intersection over Union/class.

One interesting part of the evidential formulation is that the decision-making can be adapted to derive from a fixed number of classes (equal to the number of singletons) to the maximum number of acts,  $|2^{\Theta}| - 1$ . However, often the desired decision elements are considered only the singletons and the uncertain predictions.

The image of Fig. 6a represents the predicted image with the probabilistic model. The image of Fig. 6b represents the predicted image with the evidential model (Lite-CF-Evi). It can be observed that classes road, vehicle, and background exhibit slightly higher accuracy in their predictions, with road class being notably precise. Furthermore, an additional class, denoted as the “unknown” (depicted in white) area, effectively captures pixels associated with uncertain predictions. This approach prevents the misclassification of uncertain pixels into incorrect categories, a scenario that may arise when utilizing a probabilistic approach.

The “unknown” primarily manifests itself at the class boundaries, where the model frequently provides errors in its predictions. Likewise, pixels from distant objects often lack sufficient information, suggesting that the model encounters challenges in classifying them due to data uncertainty. Consequently, these pixels are classified as “unknown”, offering improved comprehension and demonstrating the effectiveness of evidential reasoning in managing uncertainties.

The image of Fig. 6c represents an illustration, where the so-called partial ignorance is highlighted. That means that the previous class “unknown” is now distributed into the unions. This occurs as a result of changing the decision space.

In the previous examples, results with predictions have been shown, when considering different decision space elements.

For the considered perception task in the world of autonomous driving, the frame of discernment containing road ( $R$ ), vehicle ( $V$ ) and background ( $B$ ) is considered:

- $FoD = \{R, V, B\}$   
while the decision space is represented by more elements, singletons, or the disjunctions (i.e. unions) of singletons:
- decision =  $\{R, V, R \cup V, B, R \cup B, V \cup B, R \cup V \cup B\}$

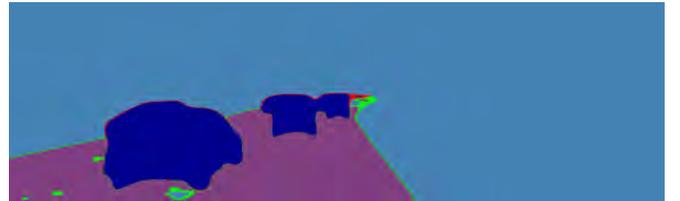
Therefore, the scalability of Dempster-Shafer’s theory is highlighted: the model is flexible, such that, by changing the value of decision elements, it directly influences the number of predicted classes. In this place, classes are represented by



(a) Probabilistic Prediction Lite-CF: road, vehicle, background.



(b) Evidential Prediction with Lite-CF-Evi: road, vehicle, background, unknown (white).



(c) Partial ignorance: singletons + unions (road  $\cup$  background (green), car  $\cup$  background (red), etc).

Figure 6: Semantic segmentation results.

elements of the non-zero power-set elements. Similar with the case of only one additional class unknown, in the case of unions, last image from the group of three, respectively Fig. 6c, the disjunctions occur preponderantly in the regions with boundaries between classes, where the model is more vulnerable to wrong predictions, which makes sense.

### C. Quality of Decision

To evaluate the decision-making part, the quality of the decision method is pursued over the predictions. In this way, the confidence in making decisions can be evaluated both for each pixel of the image (each decision that has been effectuated for defining the prediction), similarly as it can be computed like the average decision for each class (e.g. the average quality of decision for the road class).

The quality indicator recalled in equation (18) is used. Thus,  $\hat{X}$  represents the final decision and  $q(\hat{X})$  the quality indicator (whenever the model decides which class should be predicted, based on the minimal belief interval distance as in expression (14). After applying the mathematical formulas, corresponding to the previously mentioned expressions, two aspects are illustrated in the next images. Two situations are considered: the best solution with the minimal distance and the second best result, with the second smallest distance. One is the quality decision indicator for each pixel, respectively, the second shows the average quality indicator for each class.

Therefore, the first image, Fig. 7a shows the  $q(\hat{X})$  for each decision of the model, basically how accurate the judgment

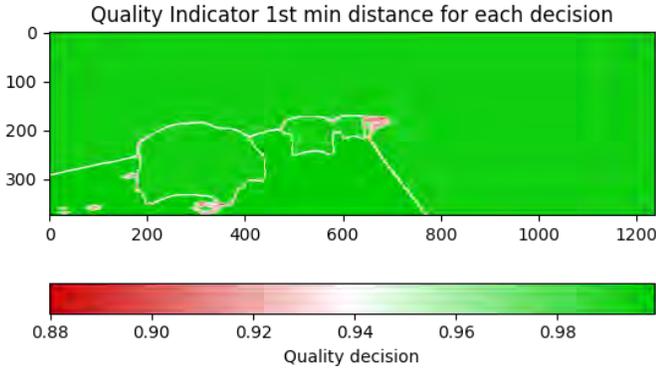
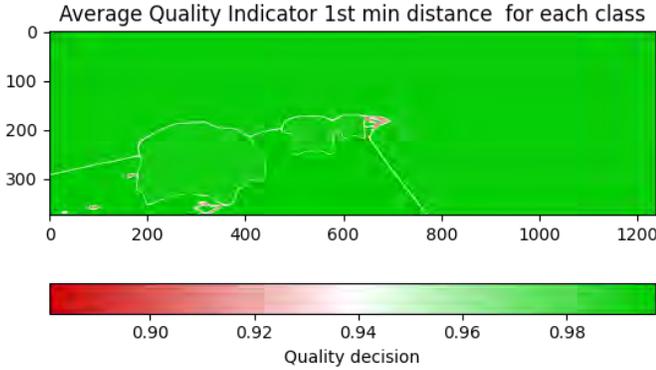

 (a)  $q(\hat{X})$  for each pixel.

 (b) Average  $q(\hat{X})$  for each class.

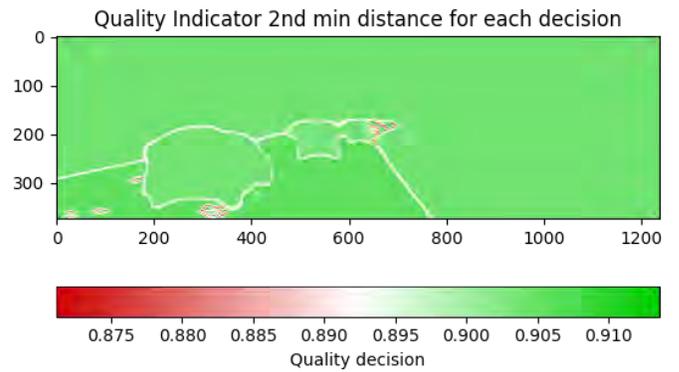
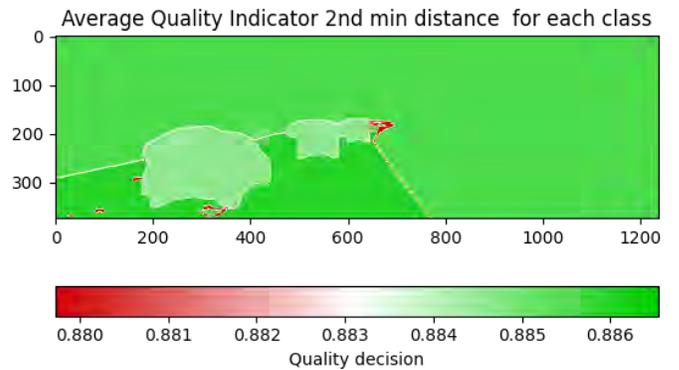
 Figure 7: Quality Indicator. 1<sup>st</sup> best solution.

 (a)  $q(\hat{X})$  for each pixel.

 (b) Average  $q(\hat{X})$  for each class.

 Figure 8: Quality Indicator. 2<sup>nd</sup> best solution.

is for each pixel. Most of these values are close to 1, which represents high confidence in decision-making. In this case, the colorbar represents how confident the model is, the more green it is, the more sure about prediction is, and the opposite when it is red. The edges between classes and the boundaries are more prone to be sensitive when considering a decision. Therefore, the values for these decisions are slightly around 0.94 between green and red. The second image, Fig. 7b, is realized in the same way, but it does illustrate the average quality indicator for each class. Hence, the judgment parameter for this distribution is reduced to 7 values for each element of the decision space, and this judgment value represents the quality indicator of each class.

Similar to the previous images, a group of two other predictions are assessed against quality indicators. Likewise in the first set, Fig. 8a stands for the quality of each decision of the model, while the second image, Fig. 8b represents the average quality indicator for the second best solution.

The range of confidence has slightly changed. The quality indicator values have decreased as well. The figures resemble the first set of images but with a lower confidence. In the next table, there is a short comparison between the two solutions. The table shows the average quality indicator for each element of the decision space in the two scenarios, when the minimum distance is calculated (first best solution), and when the second minimum belief interval distance is considered. As expected,

the model is more confident in the first case.

Element	Average $q(\hat{X})$	1 <sup>st</sup> solution	2 <sup>nd</sup> solution
$R$		0.9970	0.8802
$V$		0.9928	0.8816
$R \cup V$		0.8806	0.8797
$B$		0.9950	0.8807
$R \cup B$		0.9149	0.8865
$V \cup B$		0.9085	0.8855
$R \cup V \cup B$		0	0

 Table III: 1<sup>st</sup> best solution vs 2<sup>nd</sup> best solution. Average quality indicator/class.

The uncertain predictions are spread between unions of two singletons, therefore the last line of the table III for the union of all the singletons is equal to zero.

## V. CONCLUSION

In this work, a camera-lidar fusion has been proposed by using a deep learning architecture combined with evidence theory for intelligent vehicle perception. The combination is realized at the very last level, replacing the softmax decision with a decision calculation by the distance to prototypes approach. The introduction of an unknown class as a decision element further improves efficiency. Hence, distant points and ambiguous features can be categorized as “unknown” rather than being erroneously assigned to specific predictions.

The conducted experiments show different results depending on the chosen decision space that can include unions of singletons (i.e. disjunctions). For the decision-making part, the application adopts the decision based on the belief interval distance which avoids a loss of information with respect to the classical maximum of pignistic probability decision-making approach. Moreover, the quality of predictions are evaluated thanks to the quality-decision indicator (i.e. confidence factor) presented in the results section. The results of performed simulation show that the obtained confidence factors are very high (within  $[0.9, 1.0]$  interval) for all decisions taken, which shows a good behavior of the decision-making method used for this application. As perspectives, the plan is to enhance the Lite CF-Evi model for various class configurations and more intricate tasks while maintaining the computational efficiency needed for real-time applications. Additionally, a more in-depth examination of the distribution and impact of “unknown” predictions is intended to be explored, as well as advanced rules of combination based on the proportional conflict redistribution principle.

## REFERENCES

- [1] B. Paden, M. Cap, S.Z. Yong, D. Yershov, E. Frazzoli, *A survey of motion planning and control techniques for self-driving urban vehicles*, IEEE Trans. on Intelligent Vehicles, Vol. 1, pp. 33–55, 2016.
- [2] A.E. Bull, G.T. Cooperation, *Traffic congestion: the problem and how to deal with it*, 2003.  
<https://api.semanticscholar.org/CorpusID:166441818>
- [3] A. Krizhevsky, I. Sutskever, G.E. Hinton, *Imagenet classification with deep convolutional neural networks*, Communications of the ACM, Vol. 60(6), pp. 84–90, 2017.
- [4] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, *You only look once: Unified, real-time object detection*, in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 779–788, 2016.
- [5] O. Ronneberger, P. Fischer, T. Brox, *U-net: Convolutional networks for biomedical image segmentation*, in International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241, 2015.
- [6] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, NJ, USA, 1976.
- [7] A.P. Dempster, *A generalization of bayesian inference*, Journal of the Royal Statistical Society: Series B (Methodological), Vol. 30(2), pp. 205–232, 1968.
- [8] H. Laghmar, M.-T. Boudali, T. Laurain, J. Ledy, R. Orjuela, J.-P. Lauffenburger, M. Basset, *Obstacle avoidance, path planning and control for autonomous vehicles*, in Proc. of 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 529–534, 2019.
- [9] P. Xu, F. Davoine, T. Denœux, *Evidential combination of pedestrian detectors*, in Proc. of British Machine Vision Conference, pp. 1–14, 2014.
- [10] P. Xu, F. Davoine, J.-B. Bordes, H. Zhao, T. Denœux, *Multimodal information fusion for urban scene understanding*, Machine Vision and Applications, Vol. 27(3), pp. 331–349, 2016.
- [11] E. Capellier, F. Davoine, V. Cherfaoui, Y. Li, *A deep evidential learning for arbitrary lidar object classification in the context of autonomous driving*, in Proc. of 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 1304–1311, 2019.
- [12] T. Denœux, *Logistic regression, neural networks and Dempster-Shafer theory: A new perspective*, Knowledge-Based Systems, Vol. 176, pp. 54–67, 2019.
- [13] P. Smets, *Practical uses of belief functions*, in Proc. of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI1999), 1999.  
<https://arxiv.org/ftp/arxiv/papers/1301/1301.6741.pdf>
- [14] J. Dezert, D. Han, J.-M. Tacnet, S. Carladous, Y. Yang, *Decision-making with belief interval distance*, in Proc. of Belief 2016 Int. Conf., Prague, CZ, September 21–23, 2016.
- [15] T. Denœux, D. Dubois, H. Prade, *Representations of uncertainty in ai: beyond probability and possibility*, in A guided tour of artificial intelligence research, Springer, pp. 119–150, 2020.
- [16] J. Dezert, P. Wang, A. Tchamova, *On The Validity of Dempster-Shafer Theory*, in Proc. of Fusion 2012, Singapore, July 2012.
- [17] A. Tchamova, J. Dezert, *On the Behavior of Dempster’s Rule of Combination and the Foundations of Dempster-Shafer Theory*, IEEE IS’2012, Sofia, Bulgaria, Sept. 6–8, 2012.
- [18] L. Caltagirone, M. Bellone, L. Svensson, M. Wahde, *LIDAR-camera fusion for road detection using fully convolutional neural networks*, in Robotics and Autonomous Systems, Volume 111, 2019, Pages 125–131, ISSN 0921-8890, <https://doi.org/10.1016/j.robot.2018.11.002>.
- [19] M.N. Geletu, T. Josso-Laurain, M. Devanne, M.M. Wogari, J.-P. Lauffenburger, *Deep learning based architecture reduction on camera-lidar fusion for autonomous vehicles*, in Proc. of 2022 2nd International Conference on Computers and Automation (CompAuto), IEEE, pp. 25–31, 2022.
- [20] Z. Tong, P. Xu, T. Denœux, *An evidential classifier based on Dempster-Shafer theory and deep learning*, Neurocomputing, Vol. 450, pp. 275–293, Aug. 2021.  
<https://doi.org/10.1016%2Fj.neucom.2021.03.066>
- [21] T. Denœux, *A neural network classifier based on Dempster-Shafer theory*, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, Vol. 30(2), pp. 131–150, 2000.
- [22] A.-L. Jousselme, D. Grenier, É. Bossé, *A new distance between two bodies of evidence*, Information Fusion, Vol. 2(2), pp. 91–101, 2001.  
<https://www.sciencedirect.com/science/article/pii/S1566253501000264>
- [23] D. Han, J. Dezert, Y. Yang, *Belief interval-based distance measures in the theory of belief functions*, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 48(6), pp. 833–850, 2018.
- [24] A. Irpino, R. Verde, *Dynamic clustering of interval data using a Wasserstein-based distance*, Pattern Recognit. Lett., Vol. 29(11), pp. 1648–1658, 2008.
- [25] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, *Vision meets robotics: The kitti dataset*, International Journal of Robotics Research (IJRR), 2013.
- [26] C. Premebida, J. Carreira, J. Batista, U. Nunes, *Pedestrian detection combining rgb and dense lidar data*, in Proc. of 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.4112–4117, 2014.
- [27] M. Everingham, S. Eslami, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, *The pascal visual object classes challenge: A retrospective*, International journal of computer vision, Vol. 111(1), pp. 98–136, 2015.

# An Evidential Deep Network Based on Dempster-Shafer Theory for Large Dataset

Lucas Deregnaucourt<sup>a</sup>, Alexis Lechervy<sup>b</sup>, Hind Laghmara<sup>a</sup>, Samia Ainouz<sup>a</sup>

<sup>a</sup>LITIS, Normandie Univ, INSA Rouen, *UNIROUEN, UNIHAVRE*, Rouen, France.

<sup>b</sup>GREYC, Normandie Univ, UMR CNRS 6072, *UNICAEN, ENSICAEN*, Caen, France.

Emails: lucas.deregnaucourt@insa-rouen.fr, alexis.lechervy@unicaen.fr,

hind.laghmara@insa-rouen.fr, samia.ainouz@insa-rouen.fr

**Abstract**—We introduce a novel deep neural network architecture based on Dempster-Shafer theory capable of handling large image datasets with numerous classes, such as ImageNet. Our approach involves analyzing images through multiple experts, composed of convolutional deep neural networks that predict mass functions. These experts are then merged using Dempster’s rule, thereby returning a set of potential classes by selecting the best expected utility based on the previously computed mass functions. Our innovative algorithm can identify the best set of classes among the  $2^K$  possible sets for  $K$  classes while maintaining a complexity of  $O(K \log(K))$ . To illustrate our approach, we apply it to an out-of-distribution example search problem, demonstrating its efficiency.

**Keywords:** Dempster-Shafer theory, evidence theory, belief function, deep learning, out-of-distribution.

## I. INTRODUCTION

In recent years, image classification has made remarkable strides with the advent of deep neural networks (DNNs). However, high ambiguity in the feature vector may lead to missclassification due to the fact that multiple classes share similar expected probabilities. Moreover, a model only trained for precise classification may struggle to detect out-of-distribution (OOD) data.

One promising solution to this problem is set-valued classification [1], [2]. This method allows the model to assign a new data to a non-empty set of classes, particularly when uncertainty is high and precise classification is challenging.

In the context of Out-of-Distribution (OOD) detection, a prevalent approach is the utilization of a classification method with a reject option [3], [4], which can be seen as a special case of set-valued classification. Rejection is defined by assigning a data to the set of all possible classes, indicating a state of high uncertainty.

Recently, several works have sought to integrate the Dempster-Shafer theory (DST) into deep neural networks, aiming to leverage the power of evidential reasoning [5]–[7]. However, these attempts have been confined to relatively small and well-structured datasets such as MNIST [8] or CIFAR-10 [9]. The primary impediment has been the algorithmic complexity of DST, which scales exponentially with the size of the frame of discernment  $\Omega$ , containing  $2^K$  subsets where  $K = |\Omega|$ .

Based on [10], [11] proposed an end-to-end deep evidential neural network that allocates mass values only to singletons and  $\Omega$ . This method addresses this computational bottleneck, effectively reducing the spatial complexity from  $O(2^K)$  to  $O(K + 1)$  for the training phase. Nevertheless, the decision-making process for set-valued classification during the evaluation phase remains a computationally expensive task, requiring an exhaustive selection from all possible subsets of  $\Omega$ , still operating at  $O(2^K)$  complexity. Thus, they selected the possible subsets of  $\Omega$  based on the distance between the classes derived from the confusion matrix.

We propose in this work an algorithmic solution to mitigate the  $O(2^K)$  complexity, making set-valued decisions derived from a mass function output by a Convolutional Neural Network (CNN) feasible with linear complexity without intermediate steps to restrict the number of subsets. Additionally, we introduce mathematical optimizations to enhance numerical computations, enabling scalable implementation of set-valued classification evidential models. These contributions pave the way for the application of the DST theoretical framework to high-dimensional real-world datasets with many classes. They offer significant potential for improving the reliability of deep learning models in various applications such as OOD detection.

The remaining parts of this work are organized as follows. In section II we recall basics of Dempster-Shafer theory. In section III, we present the evidential neural network architecture we use and the algorithmic solution we propose to make set-valued decision in linear complexity. The experiments and preliminary results on large datasets are presented in section IV. Finally, we conclude in section V.

## II. BELIEF THEORY

### A. Background on belief functions

Belief function theory, called also Evidence theory or Dempster-Shafer theory [12], [13], is able to model and reason about imprecise and uncertain problems, and has more obvious advantages in the representation and combination of uncertain information.

To represent partial knowledge in the belief function theory, let consider the *frame of discernment*  $\Omega$  as a finite set of

variables  $\omega$  which refers to  $K$  elementary events to a given problem ( $\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ ).

The power set of  $\Omega$  is the set of all the  $2^K$  possible subsets. It is presented as follows:

$$2^\Omega = \{\emptyset, \{\omega_1\}, \dots, \{\omega_k\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \dots, \Omega\}, \quad (1)$$

where the  $\{\omega_i\}$  elements are titled as singletons and  $\emptyset$  denotes the empty set.

The key point of Dempster-Shafer theory is the basic belief assignment (*bba*) which represents the partial knowledge about the value of  $w$ . A *bba* is a function from  $2^\Omega$  to  $[0, 1]$  defined as follows:

$$\begin{aligned} m : 2^\Omega &\rightarrow [0, 1] \\ A &\mapsto m(A) \end{aligned} \quad (2)$$

where  $m$  satisfies the following constraint:

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (3)$$

An element  $A$  of  $\Omega$  is called a *focal element* when  $m(A) > 0$ , and the set containing all these elements is called a *body of evidence* (BOE). When each element in BOE is a singleton,  $m$  is named a *Bayesian bba*. On the other hand, when BOE contains only  $\Omega$  as a focal element, we are in the *complete ignorance* situation and  $m$  is called vacuous belief function. However, when it contains only one singleton of  $\Omega$  as a focal element,  $m$  is presented as a *Certain mass function*.

A *bba* function is normalized when the mass given to the empty set is constrained to be zero ( $m(\emptyset) = 0$ ). In that case, it corresponds to the *closed-world assumption* [13]. A contrary explanation is that the frame of discernment  $\Omega$  can be incomplete and the value of  $w$  can be taken out of  $\Omega$ . Accordingly, the mass of belief that is not linked to  $\Omega$  can allowed to be strictly positive ( $m(\emptyset) > 0$ ). That case corresponds to the *open world assumption* [14].

### B. Information fusion

The most common way to combine two *bbas*  $m_1$  and  $m_2$  defined on the same frame of discernment  $\Omega$  is the Dempster's rule [13], denoted as  $\oplus$ . It is defined by  $m_{DS}(\emptyset) = 0$  and  $\forall A \in 2^\Omega \setminus \{\emptyset\}$  by

$$m_{DS}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1 - \kappa} \sum_{\substack{B \cap C = A \\ B, C \in 2^\Omega}} m_1(B)m_2(C) \quad (4)$$

where  $\kappa$  is the degree of conflict between the two sources of evidence defined by:

$$\kappa = \sum_{\substack{B \cap C = \emptyset \\ B, C \in 2^\Omega}} m_1(B)m_2(C).$$

This fusion can be seen as the normalized version of the conjunctive rule which is defined by:

$$m_\cap(A) = \sum_{\substack{B \cap C = A \\ B, C \in 2^\Omega}} m_1(B)m_2(C). \quad (5)$$

### C. Decision-making

The most common way of making decisions with belief functions is to apply the pignistic transformation [15] to obtain a probability vector of size  $K$ , then the predicted class corresponds to the argmax of this vector. However, such a strategy does not allow the model to predict a set of classes. To this end, [16] defines the lower and upper expected utilities of selecting  $A \subseteq \Omega$  as follows:

$$\overline{\mathbb{E}}(f_A) = \sum_{B \subseteq \Omega} m(B) \max_{\omega_j \in B} u_{A,j} \quad (6)$$

$$\underline{\mathbb{E}}(f_A) = \sum_{B \subseteq \Omega} m(B) \min_{\omega_j \in B} u_{A,j} \quad (7)$$

where  $u_{A,j} \in [0, 1]$  designates the utility of the act of selecting  $A \subseteq \Omega$  denoted as  $f_A$  when the ground truth is  $\omega_j$ . The utility matrix  $U_{2^{|\Omega|} \times K}$  is computed following [17], [18] with a parameter  $\gamma \in [0.5, 1]$  that represents the imprecision tolerance degree. If the true class is  $\omega_j$ , the utility of assigning a sample to set  $A$  is calculated as an Ordered Weighted Average (OWA) aggregation [18] of the individual utilities associated with each precise assignment within  $A$  as follows:

$$u_{A,j} = g_{|A|} 1_{\{\omega_j \in A\}} \quad (8)$$

where  $g \in \mathbb{R}^{|A|}$  is a weight vector whose elements represent the decision making strategy's tolerance to imprecision, and  $1_{\{\omega_j \in A\}} = 1$  if  $\omega_j \in A$  for  $A \subseteq \Omega$ , and 0 otherwise. For example if  $g = (1, 0, \dots, 0)$ , then the decision making's strategy will be totally intolerant to imprecision, thus forcing the model to output only one class.

Following [17] and [19], this weight vector is obtained by maximizing the following entropy:

$$Ent(g) = \sum_{k=1}^{|A|} \log g_k \quad (9)$$

subject to constraints  $\sum_{k=1}^{|A|} g_k = 1$ ,  $\sum_{k=1}^{|A|} \frac{|A| - k}{|A| - 1} g_k = \gamma$  and  $g_k \geq 0$  where  $\gamma$  is a parameter representing the tolerance to imprecision. An example of a utility matrix with  $\gamma = 0.9$  and  $\Omega = \{\omega_1, \omega_2, \omega_3\}$  is shown in Table I. As we can see, the values in the utility matrix are the same according to the cardinality of the selected set. This means that instead of computing every values of the utility matrix, we only need to compute a value  $U_k$  for each possible cardinality of the subsets of  $\Omega$ . In this example, we have  $U_1 = 1$ ,  $U_2 = 0.9$  and  $U_3 = 0.8263$ .

Since we have:

$$\min_{\omega_j \in A} u_{A,j} = \begin{cases} U_k & \text{if } A = \Omega \\ 0 & \text{else} \end{cases} \quad (10)$$

and

$$\max_{\omega_j \in A} u_{A,j} = U_{|A|} \quad (11)$$

the equations (6) and (7) can be simplified as illustrated in section III-C.

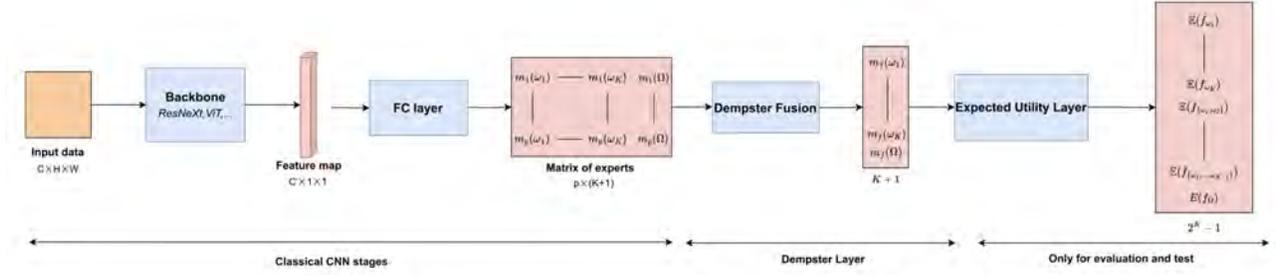


Figure 1. Architecture of an evidential deep neural network.

	$\omega_1$	$\omega_2$	$\omega_3$
$f_{\{\omega_1\}}$	1	0	0
$f_{\{\omega_2\}}$	0	1	0
$f_{\{\omega_3\}}$	0	0	1
$f_{\{\omega_1, \omega_2\}}$	0.9	0.9	0
$f_{\{\omega_1, \omega_3\}}$	0.9	0	0.9
$f_{\{\omega_2, \omega_3\}}$	0	0.9	0.9
$f_{\{\Omega\}}$	0.8263	0.8263	0.8263

 Table I  
 UTILITY MATRIX WITH  $\gamma = 0.9$  AND  $K = 3$ .

The expected utility is then obtained using the generalized Hurwicz decision criterion [20], [21] as follows:

$$\mathbb{E}(f_A) = \nu \underline{\mathbb{E}}(f_A) + (1 - \nu) \overline{\mathbb{E}}(f_A). \quad (12)$$

Where  $\nu \in [0, 1]$  is the pessimism index.

When  $\gamma = 0.5$ , the decision-making strategy is totally intolerant to imprecision so that  $u_{ij} = 1$  if  $\omega_i = \omega_j$ , else  $u_{Aj} = 0$ . In this sense, we can see the expected utility as a generalized accuracy. The other extreme strategy is totally tolerant, which is achieved when  $\gamma = 1$  where  $u_{Aj} = 1$  if  $\omega_j \in A$ , else  $u_{Aj} = 0$  so that a model that always outputs  $\Omega$  will get an expected utility of 1.

We chose this decision-making strategy among all those proposed in [16] since it is the most general form of decision criterion resulting from Jaffray's axioms [21]. Moreover, the expression of the expected utility leads to interesting simplifications in the restricted framework where we only consider the singletons and  $\Omega$ .

### III. SCALABLE EVIDENTIAL NEURAL NETWORK

In this section, we present how the DST framework can be incorporated into a deep neural network architecture. Based on some assumptions on the structure of *bbas*, we propose an algorithmic solution to make set-valued decision in linear complexity along with mathematical optimizations for a more scalable implementation.

#### A. Evidential deep neural network

As depicted in Figure 1, the proposed evidential neural network architecture is very similar to a probabilistic one. Our architecture is based on the evidential deep neural network architecture introduced in [11]. The main difference resides in the construction of the mass function. The given image

of size  $(C \times H \times W)$  first passes through the backbone of a convolutional neural network, resulting in a feature map of size  $(C' \times 1 \times 1)$ . This feature map captures the data's latent representation.

In the work presented in [11], the construction of mass functions involves the use of a distance-based layer. The classifier is composed of  $p$  prototypes  $t_i$  in  $\mathbb{R}^P$ , where  $P$  is the dimension of the feature map. In their method, the first step is to compute the distance-based support between the feature map  $x$  of a data and each prototype  $t_i$ . For the second step, the mass function  $m_i$  associated to  $t_i$  is computed by multiplying the distance-based support  $s_i$  by a weight  $h_{ij}$  which characterizes the degree of membership of prototype  $t_i$  to the class  $\omega_i$ .

Our method for constructing the mass functions is more computer vision oriented and is inspired by mixture of experts approaches [22]. Instead of considering prototypes, we consider  $p$  experts that see the feature map of a data from different points of view. For this purpose, the classical fully connected layer is replaced by a depthwise convolution [23] with a kernel of size  $(1 \times 1)$  and  $p$  groups. For a given feature map and a given number of experts  $p$ , the depthwise convolution will output a matrix of size  $(p \times (K + 1))$ , namely one mass function per expert. Each mass function holds  $|\Omega| + 1$  values, with one value dedicated to each singleton and another one for the entire set  $\Omega$ . This vector is then reshaped into a matrix of experts of size  $p \times (|\Omega| + 1)$ . We apply a softmax activation to satisfy the equation (3). In this matrix, the  $i$ -th row represents the mass function associated with expert  $p_i$ . The *bbas* of this matrix are then fused with Dempster's rule to obtain a final *bba* of size  $|\Omega| + 1$  which we will present in the next section.

#### B. Computational optimization of Dempster's rule

As seen in the previous section, since our network is only considering the masses assigned to singletons and  $\Omega$ , the expression of the conjunctive rule simplifies to formula (13)  $\forall A \in \Omega$

$$\begin{aligned} m_{\cap}(A) &= \sum_{\substack{B \cap C = A \\ B, C \in 2^{\Omega}}} m_1(B) m_2(C) \\ &= m_1(A) m_2(A) + m_1(A) m_2(\Omega) + m_1(\Omega) m_2(A) \end{aligned} \quad (13)$$

This brings us to an iterative algorithm for performing Dempster's rule as shown by the Algorithm 1. We define  $\mu_1 = m_1$  and  $\mu_{i+1} = m_{\cap}(\mu_i, m_i)$  where  $\mu_i$  represents the mass function obtained by the fusion of the  $i$  first expert's mass functions by the conjunctive rule.

---

**Algorithm 1** Iterative Dempster's rule
 

---

**Require:**  $p$  mass functions  $m_1, \dots, m_p$

$\mu_1 \leftarrow m_1$

**for**  $i = 2, \dots, p$  **do**

**for**  $j = 1, \dots, K$  **do**

$$\begin{aligned} \mu_i(\{\omega_j\}) &= \mu_{i-1}(\{\omega_j\})m_i(\{\omega_j\}) \\ &\quad + \mu_{i-1}(\{\omega_j\})m_i(\Omega) + \mu_{i-1}(\Omega)m_i(\{\omega_j\}) \end{aligned}$$

**end for**

$$\mu_i(\Omega) = \mu_{i-1}(\Omega)m_i(\Omega)$$

**end for**

**return**  $\mu_p/Z$

where  $Z$  is a normalization term.

---

The expression of  $\mu_i(\{\omega_j\})$  can be rewritten as follows:

$$\begin{aligned} \mu_i(\{\omega_j\}) &= \mu_{i-1}(\{\omega_j\})m_i(\{\omega_j\}) + \mu_{i-1}(\{\omega_j\})m_i(\Omega) \\ &\quad + \mu_{i-1}(\Omega)m_i(\{\omega_j\}) \\ &= (\mu_{i-1}(\{\omega_j\}) + \mu_{i-1}(\Omega)) \times (m_i(\{\omega_j\}) + m_i(\Omega)) \\ &\quad - \mu_{i-1}(\Omega)m_i(\Omega) \end{aligned} \quad (14)$$

which leads to an improved algorithm that only iterates on the number of classes  $K$  as presented in the Algorithm 2.

---

**Algorithm 2** Scalable Dempster's rule
 

---

**Require:**  $p$  mass functions  $m_1, \dots, m_p$

$$\mu_p(\Omega) = \prod_{i=1}^p m_i(\Omega)$$

**for**  $j = 1, \dots, K$  **do**

$$\mu_p(\{\omega_j\}) = \prod_{i=1}^p (m_i(\{\omega_j\}) + m_i(\Omega)) - \mu_p(\Omega)$$

**end for**

**return**  $\mu_p/Z$  where  $Z$  is a normalization term.

---

The algorithm 2 is highly parallelizable and each element of the loop can be calculated independently of the others, unlike the algorithm 1 where each element depends on the previous iteration. In practice, this second algorithm provides a very fast implementation of Dempster's rule in the restricted framework chosen where we only consider singletons and  $\Omega$  as focal elements.

### C. Scalable decision making

Since we only consider the singletons and  $\Omega$  for the construction of the mass function, we can simplify the equations (6) and (7) as follows:

$$\overline{\mathbb{E}}(f_A) = \sum_{\omega_i \in \Omega} (m(\{\omega_i\})u_{A,i}) + m(\Omega) \max_{\omega_k \in \Omega} u_{A,k}, \quad (15)$$

$$\underline{\mathbb{E}}(f_A) = \sum_{\omega_i \in \Omega} (m(\{\omega_i\})u_{A,i}) + m(\Omega) \min_{\omega_k \in \Omega} u_{A,k}. \quad (16)$$

During the training phase, we want  $f_A$  to be a singleton. That's to say  $u_{ii} = 1$  and  $u_{ij} = 0 \forall i \neq j$  which can be seen as the classical accuracy metric. Under those hypotheses, we can simplify the equations (15) and (16) as follows:

$$\overline{\mathbb{E}}(f_{\omega_i}) = m(\{\omega_i\}) + m(\Omega) \quad (17)$$

$$\underline{\mathbb{E}}(f_{\omega_i}) = m(\{\omega_i\}) \quad (18)$$

leading to this simplified expression of the expected utility:

$$\begin{aligned} \mathbb{E}(f_{\omega_i}) &= \nu m(\{\omega_i\}) + (1 - \nu) (m(\{\omega_i\}) + m(\Omega)) \\ &= m(\{\omega_i\}) + (1 - \nu)m(\Omega). \end{aligned} \quad (19)$$

This expression can be considered as a rewriting of the pignistic transformation in our restricted framework. Indeed, taking  $\nu = 1 - \frac{1}{|\Omega|}$  in equation (19) leads to the pignistic probability expression when  $m(A) = 0 \forall A \subset \Omega$  such that  $|A| \geq 2$ .

We propose to use the cross-entropy loss on the expected utilities vector for training our network:

$$- \sum_{i=1}^n \sum_{k=1}^K y_{i,k} \log(\mathbb{E}(f_{\omega_k}(x_i))) \quad (20)$$

with  $n$  is size of training dataset,  $y_{i,k}$  is 1 if the label of example  $x_i$  is  $\omega_k$  and 0 otherwise.

For decision-making during the evaluation and test phase, we want our network to be able to output a subset of  $\Omega$ . The main obstacle is the algorithmic complexity since it would require to compute  $2^{|\Omega|}$  expected utilities to choose the subset that maximizes it. To solve this issue, [11] proposes to compute the confusion matrix from the training set generated by an evidential deep neural network as explained above. Based on the distance between the classes, they only keep the classes and groups of classes that are similar enough by thresholding. Although in practice this strategy reduces the number of expected utilities to be computed, it remains in  $2^{|\Omega|}$  in the worst case (when the result is to be attributed to the  $\Omega$  set). Furthermore, we are not convinced that this strategy is sufficient to scale to databases with a large number of classes such as ImageNet [24] where  $|\Omega|=1000$ . Moreover, it requires a costly intermediate step between the training phase and the evaluation and test phases.

To this end, we propose a very simple and computationally efficient iterative algorithm (3) to determine the argmax between all subsets of  $\Omega$  without any *a priori* about the correlation between the classes nor intermediate step to restrict the number of subsets of  $\Omega$ . The first step is to compute the expected utilities of singletons using the equation (19) and to sort them in a decreasing order. We then compare the higher singleton expected utility with the expected utility of the subset composed of the two best singletons using the equations (12),(15),(16) and so on until adding a new singleton to the subset decreases the expected utility. Let's consider  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$  with  $\mathbb{E}(\omega_1) > \mathbb{E}(\omega_2) > \mathbb{E}(\omega_3) > \mathbb{E}(\omega_4)$ . We

then compute  $\mathbb{E}(\{\omega_1, \omega_2\})$  and compare it with  $\mathbb{E}(\omega_1)$ . Let's suppose that  $\mathbb{E}(\{\omega_1, \omega_2\})$  is effectively higher than  $\mathbb{E}(\omega_1)$ , we now have to compute  $\mathbb{E}(\{\omega_1, \omega_2, \omega_3\})$ . By considering that  $\mathbb{E}(\{\omega_1, \omega_2\}) > \mathbb{E}(\{\omega_1, \omega_2, \omega_3\})$ , we obtain  $A^* = \{\omega_1, \omega_2\}$ . If  $\mathbb{E}(A^*) > \mathbb{E}(\Omega)$  then the model outputs  $A^*$ , else it outputs  $\Omega$ .

---

**Algorithm 3** Argmax of the Expected Utility
 

---

**Require:** sorted singletons expected utilities  $\mathbb{E}(\{\omega_{\alpha_1}\}) \geq \mathbb{E}(\{\omega_{\alpha_2}\}) \geq \dots \geq \mathbb{E}(\{\omega_{\alpha_K}\})$ .

```

A* ← ωα1
for i = 2, ..., K do
    A*temp ← {A*, ωαi}
    if  $\mathbb{E}(A^*_{temp}) > \mathbb{E}(A^*)$  then A* ← A*temp
end if
end for
return A*
    
```

---

This strategy allows the model to output a set of classes among all the possible subsets of  $\Omega$  while maintaining a complexity of  $O(K \log(K))$  without requiring any limitations on the number of subsets of  $\Omega$  to compare their expected utilities.

#### IV. EXPERIMENTS

To demonstrate the relevance of our model, we conducted several experiments. Firstly, we carry out a study on the impact of the various parameters on our model. Secondly, we sought to demonstrate the ability of our model to process large databases containing a large number of classes and compare our model with a standard probabilistic model for classification problem. Finally, we demonstrated the superiority of our approach over the standard probabilistic model for an OOD detection task.

In all our experiments, we assume that the backbone used is of type ResNext50 [25]. This applies both to our model and to the probabilistic models to which the comparison is conducted.

##### A. Datasets

We conducted our experiments using the following 3 databases: CIFAR-100, ImageNet and SVHN dataset.

CIFAR-100 [26] is a database of low-resolution  $28 \times 28$  images. It contains 60,000 images divided into 100 classes with 600 images per class.

ImageNet [24] contains 1.5 million images of  $224 \times 224$  resolution, manually annotated in 1,000 categories. The annotation is based on the WordNet hierarchical object categorization structure (augmented by 120 dog categories).

The SVHN (Street View House Numbers) database [27] is a collection of  $32 \times 32$  digital images that includes handwritten digits from photos of house numbers taken in street scenes. The database contains 10 classes, corresponding to digits from 0 to 9.

##### B. Ablation study

In this section, we present some experiments designed to measure the impact of the various parameters of our approach on its performances. We measure two metrics: *expected utility* and *average cardinality*.

Given that the accuracy is obtained by fixing the imprecision tolerance degree  $\gamma$  to 0.5 while computing the expected utility, we propose to evaluate the *expected utilities* across a range of  $\gamma$  values from 0.5 to 0.95.

We compute the *average cardinality* of the predictions according to  $\gamma$  as follows:

$$AC(T) = \frac{1}{|T|} \sum_{i=1}^{|T|} |A(i)| \quad (21)$$

where  $T = \{x_1, \dots, x_{|T|}\}$  is the test set and  $A(i)$  is the set-valued output for the data  $x_i \in T$ . It is clear that for  $\gamma = 1$ , the model will always output  $f_\Omega$  since  $\mathbb{E}(\Omega) = 1$  and the *average cardinality* will be equal to the number of elements in  $\Omega$ .

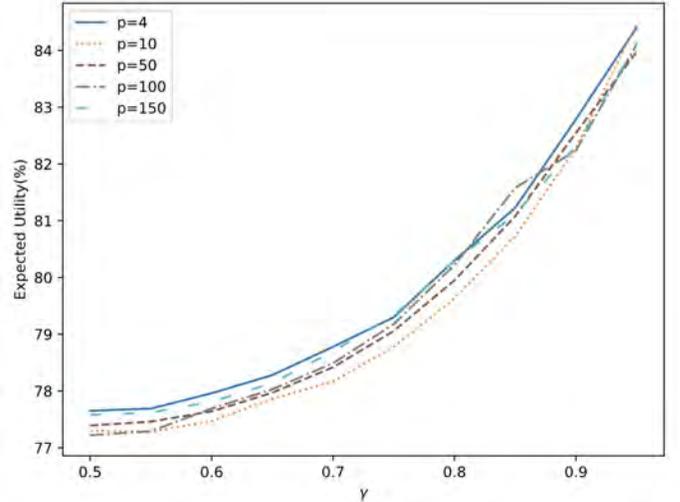


Figure 2. Expected Utility according to the number of experts on CIFAR-100.

Firstly, we need to determine the hyperparameters of our model, namely the number of experts  $p$  and the degree of pessimism  $\nu$ . Since this search process is quite time-intensive, we restrict it to the CIFAR-100 dataset. To identify the optimal number of experts, we fix  $\nu$  to 0.99 so that the equation(19) corresponds to the pignistic probability. As shown on Figure 2, the impact of the number of experts does not appear to be significant. This is mainly because there is no guarantee that the experts simulated by the fully connected layer will be independent. So we choose  $p = 4$  as there is no need for a lot of experts. Then we search for the optimal  $\nu$  by setting the number of experts  $p = 4$ . As depicted in Figure 3, the model learns in a similar way, independently of  $\nu$ . Indeed, the model always outputs a value very close to zero for  $m(\Omega)$  for precise classification task, so the impact of  $\nu$  is not significant during the training phase. Consequently, we have selected  $\nu = \frac{1}{|\Omega|}$ , namely  $\nu = 0.99$  for CIFAR-100 and  $\nu = 0.999$  for ImageNet.

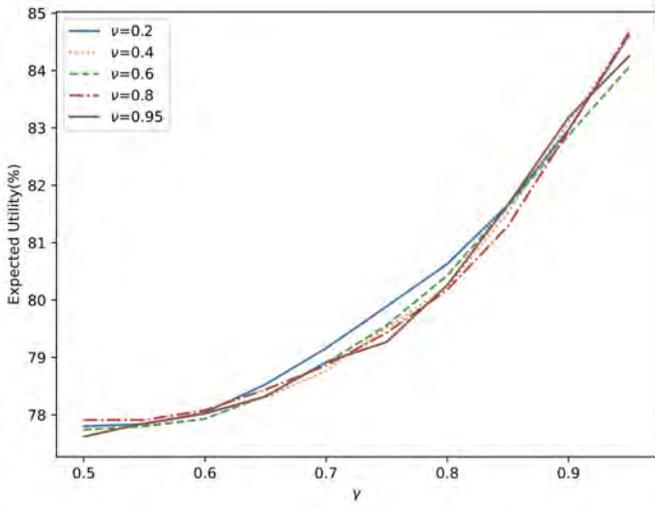


Figure 3. Expected Utility according to  $\nu$  on CIFAR-100.

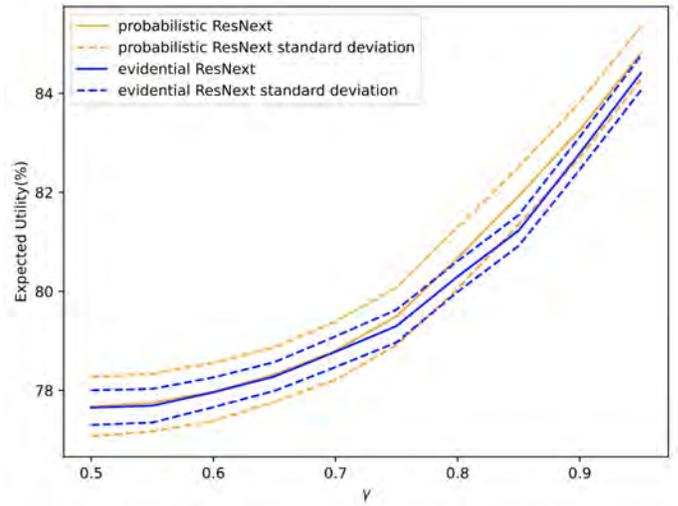


Figure 4. Expected Utility on CIFAR-100.

C. Comparison with probabilistic approaches for image classification

Now that we have fixed the model hyperparameters, we can compare the evidential neural network with the probabilistic one on precision classification. As mentioned previously, the probabilistic model used corresponds to a ResNext50 type backbone. This is followed by a fully connected layer and a softmax.

For fair comparison between our method and the probabilistic approach, we have to allow the probabilistic network to output set-valued predictions in order to compute the expected utility. To do so, we consider the probability vector output by the model as a mass function with  $m(\Omega) = 0$  and  $m(\{\omega_j\}) = p(\omega_j) \forall j = 1, \dots, K$ .

The Expected Utility and Cardinality curves over 10 runs on CIFAR-100 are respectively presented in Figure 4 and Figure 5. The Expected Utility and Cardinality curves on ImageNet are respectively presented in Figure 6 and Figure 7. Due to the size of the database, we limited the ImageNet experiments to a single run and were therefore unable to calculate standard deviations. For both experiments, we can see that there is almost no difference between the two models from  $\gamma = 0.5$  to  $\gamma = 0.7$  where the decision-making strategy is quite intolerant to uncertainty, forcing the model to output one or two classes. For  $\gamma = 0.75$  to  $\gamma = 0.95$  the evidential model is less confident than the probabilistic one and outputs sets with a higher cardinality, which decreases the Expected Utility. On Imagenet the performance of the probabilistic model is 77.77% in accuracy against 77.65%. The difference in performance is relatively small.

D. OOD detection

For OOD detection task, we want to evaluate the capability of the network to output  $\Omega$  if, and only if, the data does not belong to the classes from the training set. For this purpose,

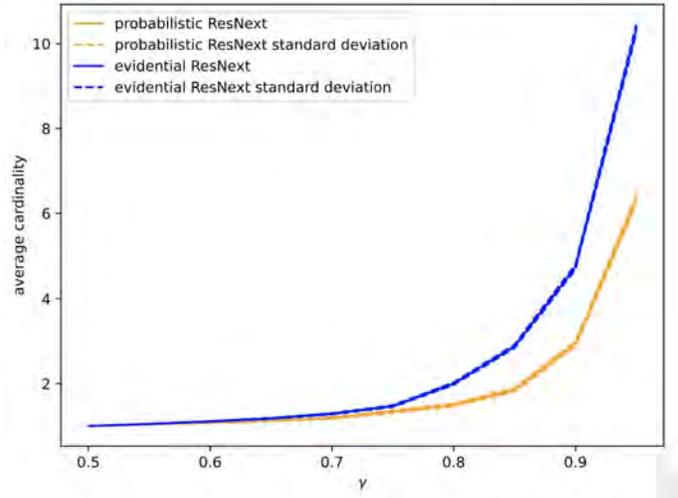


Figure 5. Average Cardinality on CIFAR-100.

we evaluate the rate of  $f_\Omega$  by varying  $\gamma$  from 0.5 to 0.95. A good model has to get a high rate of  $f_\Omega$  on out-of-distribution data and a low rate of  $f_\Omega$  on in-distribution data. For  $\gamma = 1$ , the model will always predict  $\Omega$  since all the non-zero values in the utility matrix will be equal to 1. So the  $f_\Omega$  rate will always be equal to 100%.

The results on the OOD detection task for the models trained on CIFAR-100 and ImageNet are respectively presented in Figure 8 and Figure 9. As expected, the  $f_\Omega$  rate is very low for the evidential and the probabilistic models on in-distribution test set. However, it is clear that the evidential network outperforms the probabilistic network for OOD detection task when we evaluate them on the SVHN dataset.

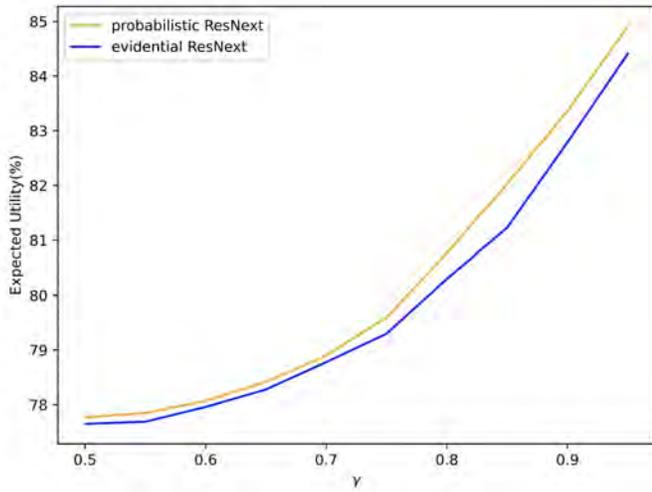


Figure 6. Expected Utility on ImageNet.

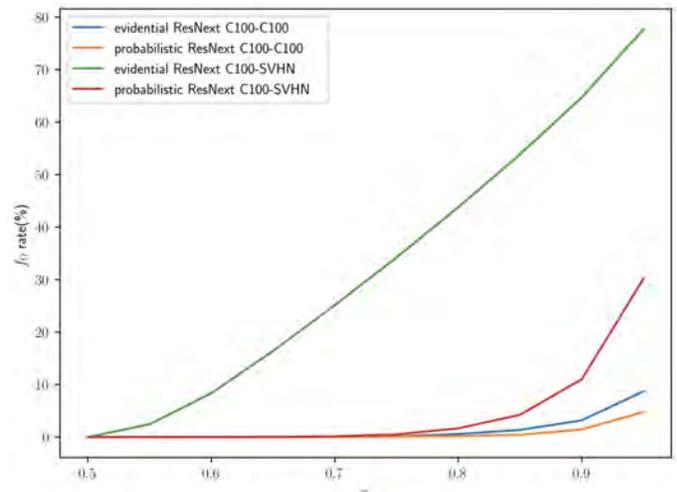


Figure 8.  $f_{\Omega}$  rate for OOD detection, CIFAR-100.

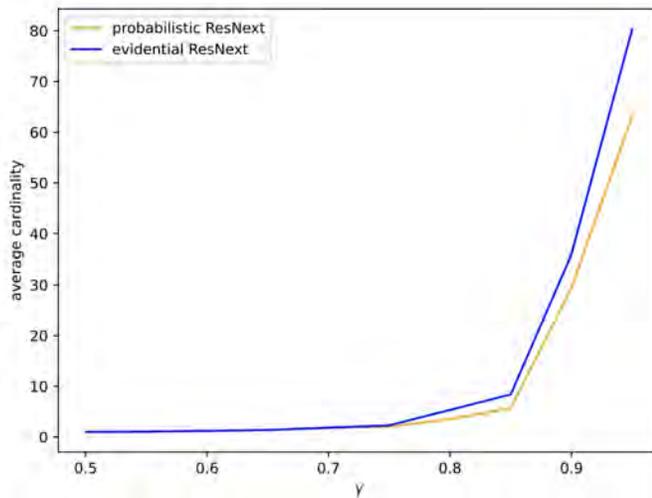


Figure 7. Average Cardinality on ImageNet.

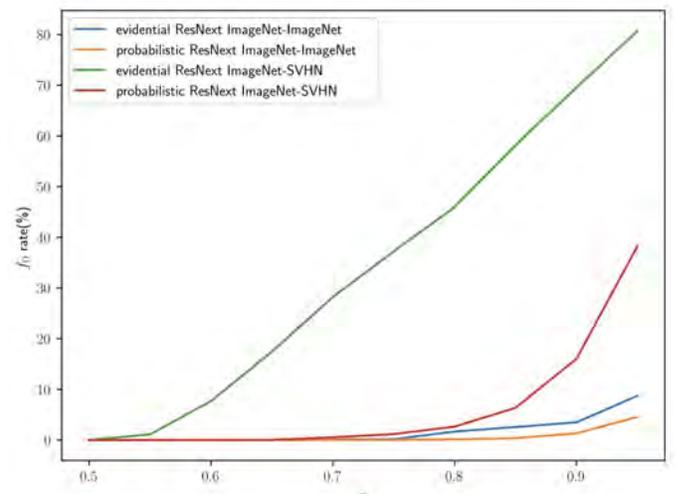


Figure 9.  $f_{\Omega}$  rate for OOD detection, ImageNet.

## V. DISCUSSIONS AND CONCLUSIONS

In this work, we have presented a novel deep neural network based on Dempster-Shafer theory capable of handling large datasets for image classification. Furthermore, we have introduced mathematical optimizations to improve numerical computations, facilitating a scalable implementation of evidential models for set-valued classification. This approach makes it possible to obtain results on databases with a large number of classes, while avoiding the problem of traversing the  $2^K$  subset of possible classes.

The proposed evidential neural network shows similar results to the probabilistic one for precise classification task. One way to improve it can be to ensure the independence of the experts with a Deep Ensemble approach [28], [29].

However, our network clearly outperforms the probabilistic one for OOD detection task regarding the  $f_{\Omega}$  rate. This

illustrates that the proposed method overcomes one of the main problems of neural networks, namely the overconfidence even if the data is out-of-distribution. Of course, the scope of our method does not limit itself to image classification. We can adapt it to other computer vision tasks such as semantic segmentation and instance segmentation.

Another way of improving our method would be to also take into account the partial ignorance of the experts when fusing the mass functions and making a decision. This would require to overcome computational bottlenecks but would open the doors for other decision-making strategies and more optimal fusion rules.

## ACKNOWLEDGEMENTS

The first author is supported by the French National Research Agency (ANR) and Region Normandie under grant

HAISCoDe. We also thank our colleagues from the Criann who provided us some computation resources with Myria and Austral and by so enabled us to get our results efficiently and fast.

## REFERENCES

- [1] E. Chzhen, C. Denis, M. Hebiri, T. Lorieul, *Set-valued classification – overview with a unified framework*, arXiv preprint, arXiv:2102.12318, 2021.
- [2] E. Grycko, *Classification with set-valued decision functions*, Information and Classification (O. Opitz, B. Lausen and R. Klar, Eds.), pp. 218–224, 1993.
- [3] C. Chow, *On optimum recognition error and reject tradeoff*, IEEE Trans. on Information Theory, Vol. 16(1), pp. 41–46, 1970.
- [4] M. Pimentel, D. Clifton, L. Clifton, L. Tarassenko, *A review of novelty detection*, Signal Processing, Vol. 99, pp. 215–249, 2014.
- [5] S. Xu, Y. Chen, C. Ma, X. Yue, *Deep evidential fusion network for medical image classification*, Int. J. of Approximate Reasoning, Vol. 150, pp. 188–198, 2022.
- [6] M. Sensoy, L. Kaplan, M. Kandemir, *Evidential deep learning to quantify classification uncertainty*, Advances in Neural Information Processing Systems, Vol. 31, pp. 3179–3189, 2018.
- [7] Z. Guo, Z. Wan, Q. Zhang, X. Zhao, Q. Zhang, L. Kaplan, A. Jøsang, D. Jeong, F. Chen, J.-H. Cho, *A survey on uncertainty reasoning and quantification in belief theory and its application to deep learning*, Information Fusion, Vol.101, 2024.
- [8] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, *Gradient-based learning applied to document recognition*, Proceedings of the IEEE, Vol. 86(11), pp. 2278–2323, 1998.
- [9] A. Krizhevsky, *Learning multiple layers of features from tiny images*, Ph.D. Dissertation, 2009.
- [10] T. Denœux, *A neural network classifier based on Dempster-Shafer theory*, IEEE Trans. on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 30(2), pp. 131–150, 2000.
- [11] Z. Tong, P. Xu, T. Denœux, *An evidential classifier based on Dempster-Shafer theory and deep learning*, Neurocomputing, Vol. 450, pp. 275–293, 2021.
- [12] A.P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The Annals of Mathematical Statistics, Vol.38(2), pp. 325–339, 1967.
- [13] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
- [14] P. Smets, *The combination of evidence in the transferable belief model*, Trans. on Pattern Analysis and Machine Intelligence, Vol. 12(2), pp. 447–458, 1990.
- [15] P. Smets, R. Kennes, *The transferable belief model*, Artificial Intelligence, Vol. 66, pp. 191–234, 1994.
- [16] T. Denœux, *Decision-making with belief functions: a review*, Int. J. of Approximate Reasoning, Vol. 109, pp. 87–110, 2019.
- [17] L. Ma, T. Denœux, *Partial classification in the belief function framework*, Knowledge-Based Systems, Vol. 214, p. 106742, 2021.
- [18] R. Yager, *On ordered weighted averaging aggregation operators in multicriteria decision-making*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 18, pp. 183–190, 1988.
- [19] M. O’Hagan, *Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic*, in Twenty-Second Asilomar Conference on Signals, Systems and Computers, Vol. 2, pp. 681–689, 1988.
- [20] L. Hurwicz, *The generalized bayes minimax principle: a criterion for decision making under uncertainty*, Cowles Commission Discussion Paper Statistics, Vol. 355, 1951.
- [21] J.-Y. Jaffray, *Linear utility theory for belief functions*, Operations Research Letters, Vol. 8(2), pp. 107–112, 1989.
- [22] T. Baldacchino, E.J. Cross, K. Worden, J. Rowson, *Variational bayesian mixture of experts models and sensitivity analysis for nonlinear dynamical systems*, Mechanical Systems and Signal Processing, Vol.66–67, pp. 178–200, 2016.
- [23] F. Chollet, *Xception: Deep learning with depthwise separable convolutions*, in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.
- [24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, *Imagenet: A large-scale hierarchical image database*, IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, 2009.
- [25] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, *Aggregated residual transformations for deep neural networks*, in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5987–5995, 2017.
- [26] A. Krizhevsky, G. Hinton, *Learning multiple layers of features from tiny images*, Technical report, University of Toronto, Toronto, Ontario, Tech. Rep. 0, 2009, backup Publisher: University of Toronto, 2009.
- [27] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A.Y. Ng, *Reading digits in natural images with unsupervised feature learning*, in NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2011.
- [28] B. Lakshminarayanan, A. Pritzel, C. Blundell, *Simple and scalable predictive uncertainty estimation using deep ensembles*, in Advances in Neural Information Processing Systems, I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, and R. Ferg, Eds., Vol. 30, Curran Associates, Inc., 2017.
- [29] O. Laurent, A. Lafage, E. Tartaglione, G. Daniel, J.-M. Martinez, A. Bursuc, G. Franchi, *Packed-ensembles for efficient uncertainty estimation*, in ICLR, 2023.

## BIOGRAPHIES OF CONTRIBUTORS

**Joaquín ABELLÁN** received the Ph.D. degree in mathematics science from the University of Granada, Spain, in 2003. Currently, he is a Full Professor with the University of Granada. His current research interests include representation of the information, principally via imprecise probabilities and its quantification, and also general applications in the data mining area related to the information theory. He has been the principal research head of some national and regional research projects.  
*Email:* jabellan@decsai.ugr.es

**Samia AINOUZ** is a full Professor at INSA Rouen Normandy, France. She was the Head of the team Intelligent Transportation systems until September 2023. She is currently the co-head of the doctorate school MIIS of Normandy. Her research area is around Multimodality for intelligent vehicle navigation including data fusion, 3D reconstruction, VSLAM. She supervised 10 Ph.D. students around road scene analysis and autonomous navigation. Recently, she focused her research toward non-conventional imaging for autonomous navigation in adverse weather conditions using computer vision and data-fusion methods. She is currently the head of the ANR projects INARI and previously ICUB dealing with road scene analysis in adverse weather conditions with collaboration with CEREMA, UGE, Peugeot PSA and Stereolabs and several academic partners (IRIMAS, Heudyasic, ImVia, ...).  
*Email:* samia.ainouz@insa-rouen.fr

**Sonda AMMAR BOUHAMED** received her Informatique and Multimedia diploma in 2009, the Master Degree from Rouen University, France, in 2010 and the Ph.D. Degree from the National Engineering School of Sfax, Tunisia (2014). Now she is assistant professor in National School of Electronics and Telecommunication of Sfax (Enet-Com). Till 2022, she was a member of the Control and Energy Management Laboratory (CEMLab) and a member of the Cybernics Team at Sfax University, Tunisia. She is an associate researcher in Image & Information Processing (iTi) Department at IMT Atlantique Brest, France. Her research interests range from uncertainty modeling, data and information quality, machine learning, image and signal processing, analytics and information fusion technologies to apply in pattern recognition, smart real-time systems for obstacle detection and recognition, and decision support systems.  
*Email:* sonda.ammar@gmail.com

**Hanan ANZID** received her Ph.D. from the University Ibn Zohr, Agadir, Morocco and the university of Bourgogne Franche-Comté. She is a member of the Laboratory IRF-SIC, and her current researches are focused on processing/analysis of multimodal imaging processing/analysis and its interactions with cultural heritage and media technology.  
*Email:* hanananzid@gmail.com

**Samia BARHOUMI** M.Sc., received her Computer Science diploma in 2009 and her Master's Degree in Industrial and automatic IT from the Engineering School of Sfax (ENIS) in 2012. She is now a fourth-year Ph.D. student and a member of the Control and Energy Management Laboratory (CEMLab) at National Engineering School of Sfax (ENIS) and a member of Cybernics Team within CEMLab at Sfax University. Her thesis project is about uncertainty management within the Dempster-Shafer framework. Her main research interests center on uncertain information modeling and information fusion.  
*Email:* semya.barhoumi@gmail.com

**Mireille BATTON-HUBERT** is Professor in the field of Mathematical Modelling applied to Sciences of Engineering at Mines de Saint-Étienne - Henri Fayol Institute - UMR CNRS 6158 LIMOS Laboratory of Informatics, Modelling and Optimization of the Systems. Her research work deals with multicriteria optimisation/analysis methods applied to geo-referenced data dealing with continuous phenomena (e.g. torrential discharge, efficiency of structures, and more recently energy and district heat network) for controlling environmental efficiency (risk analysis) taking into account the uncertainty and the imprecision in the mathematical formulization and their propagation effects in the modelling. She is in charge of the Mathematical and Industrial Engineering Department.  
*Email:* batton@emse.fr

**Aissam BEKKARI** received his Ph.D. from the University Ibn Zohr, Agadir Morocco in 2014. He is now Associate Professor at the university of Cadi Ayyad, Marrakech, Morocco, and member of the Laboratory IRF-SIC, where he is leading research in the various research themes of computer science such as remote sensing, image processing, and 3D pattern recognition.  
*Email:* a\_bekkari@yahoo.fr

**Kaouthar BENAMEUR** completed her Ph.D. on *Control Strategies for an active vision system* from McGill University, Canada on 1997. As an International Cooperation Officer in charge of North America, Kaouthar Benameur's role is to promote The French Aerospace Lab (ONERA) research activities and to develop, coordinate, and negotiate international agreements for collaboration on aeronautics and space activities. Kaouthar also develops and provides briefing materials and policy guidance to ONERA officials to benefit from international cooperation activities. Kaouthar is a member of the advisory team for IFAR (International Forum for Aviation Research) and a member of ACARE (Advisory Council for Aviation Research and Innovation in Europe) INCO group. Before joining the International Affairs Directorate in 2018, Kaouthar was doing research in the data fusion domain. Her research focus areas were on target tracking, data association and sensor management. She worked on many DGA (Directorate General of Armaments) projects with different applications on passive tracking and data fusion.

*Email:* kaouthar.benameur@onera.fr

**Erik P. BLASCH** Ph.D./MBA, is a program officer with the Air Force Research Laboratory. He received his B.S. in Mechanical Engineering from the Massachusetts Institute of Technology and Ph.D. in Electrical Eng. from Wright State University in addition to numerous Master's Degrees in Mech. Eng., Ind. Eng., Elect. Eng., Medicine, Military Studies, and Economics. Additionally, his assignments include Col (ret) USAF reserves, adjunct Associate Professor, and President of professional societies. His areas of research include information-fusion performance evaluation, image fusion, multi-domain avionics, and human-machine integration; compiling over 129 journal papers, 500+ peer-reviewed publications, 54 patents, and 11 books. His contributions include: physics-based and human-derived information fusion (PHIF), game-theoretic space situational awareness, and evidential reasoning simultaneous tracking and identification (STID) filters. He is a Fellow of AIAA (astronautics), IEEE (electrical), MSS (sensing), RAeS (aerospace), and SPIE (optical) societies.

*Email:* erik.blasch@gmail.com

**Nadejda BOCHEVA** is an Associate Professor at the Institute of Neurobiology, Bulgarian Academy of Sciences, and Head of the Department of Sensory Neurobiology. She received a BS in solid-state physics and an MS in biophysics from Sofia University in 1979 and a Ph. Degree from the Institute of Physiology, Bulgarian Academy of Sciences in 1986. In 2006, she became an Associate Professor at the Institute of Neurobiology. Her research interests include human information processing, sensory neuroscience, spatial vision, decision-making, aging, and developmental disorders. She is a Fulbright Scholar and has a Fogarty International collaborative award. She is a member of the Bulgarian Physiological Society and the American Psychological Association.

*Email:* nbbocheva@hotmail.com

**Éloi BOSSÉ** received the B.A.Sc., M.Sc., and Ph.D. degrees from the Université Laval, QC, Canada, in 1979, 1981, and 1990, respectively, all in electrical engineering. In 1981, he joined the Communications Research Centre, Ottawa, Canada, where he worked on signal processing and high resolution spectral analysis. In 1988, he was transferred to the Defence Research Establishment Ottawa to work on radar target tracking in multipath. In 1992, he moved to Defence Research and Development Canada Valcartier (DRDC Valcartier) to lead a group of defense scientists on analytics, information fusion, and decision support. From 1993 to 2015, he has held an Adjunct Professor positions at several universities at the Université Laval, the University of Calgary, and McMaster University. He headed the C2 Decision Support Systems Section, DRDC Valcartier, from 1998 to 2011. He represented Canada as a DRDC member in numerous international research fora under the various cooperation research programs (NATO, TTCP, bi, and tri-laterals) in his area of expertise. He has coauthored and a Co-Editor of 5-6 books on information fusion. He retired from DRDC in September 2011. Since 2011, he conducted some research activities under NATO Peace and Security Programme, as a Researcher with the Mathematics and Industrial Engineering Department, Polytechnic of Montreal, as an Associate Researcher at IMT Atlantique, Brest, France, as a Researcher at McMaster University, Canada, and since 2015, has been the President of Expertise Parafuse, Inc., a consultant research firm on analytics and information fusion technologies. He has published more than 200 articles in journals, book chapters, conference proceedings, and technical reports. His research interests are mainly in analytics and machine learning, information fusion, decision support, and complex systems.

*Email:* ebosse861@gmail.com

**Aurélie BOUCHARD** is Engineer on atmospheric physics, holding a Ph.D. on physical methods on remote sensing. She is working, since 2010, at the ONERA office, in the Physics Instrumentation Environment Space Department. In the framework of the air safety, she is developing diagnostic tools for thunderstorm nowcasting from very to medium short range time, using the fusion of different kinds of observations.

*Email:* Aurelie.Bouchard@onera.fr

**Mohammed BOUMEDIENE** received the M.Eng. degree in computer science at the University of Oran 1 - Ahmed Ben Bella - Algeria in 2004, and the Ph.D. degree in image processing at the University of sciences and technologies of Oran - Mohamed Boudiaf - Algeria in 2015. He is currently an Associate Professor at the National Higher School of Telecommunications and ICT - Abdelhafid Boussouf - Algeria. His main research interests include autonomous vehicles perception, multi-object tracking, decision-making using belief functions.  
*Email:* mboumediene@inttic.dz

**Dănuț BUCUR** (Ph.D. Eng.) is a dynamic and accomplished Research Assistant and Senior Software Engineer with a strong background in Robotics and Intelligent Control, coupled with extensive experience in the automotive industry. Throughout his career, he has made significant contributions to both research and practical applications. As a Research Assistant at the Institute of Solid Mechanics of the Romanian Academy, Dănuț excelled in developing mathematical and physical models for robotic systems and conducting simulations and experimentation using cutting-edge technologies like 3D design, image processing, neural networks, fuzzy logic, and genetic algorithms. He has published 12 scientific papers in national and international conferences and journals, showcasing his dedication to advance the field of Robotics and Intelligent Control. In the automotive domain, Dănuț has played key roles as a Senior Software Engineer in reputable companies. He has led projects involving model-based design for the Door Control Unit and function development for diesel engines. Dănuț's expertise in MATLAB/Simulink has enabled him to develop different tools and optimize software modules. With his unique blend of research acumen and practical skills in the automotive domain, Dănuț is a valuable asset who continually drives innovation and advances the automotive industry's technological frontiers.  
*Email:* xdan.bucur@yahoo.com

**Magalie BUGUET** completed her Ph.D. on the study of links between lightning activity and microphysics/dynamics of thunderstorms from Meteorage lightning network and Météo-France radar network in 2012. Since 2013, she is working at ONERA on the analysis of electric field mill data developed by ONERA and the forecast of thunderstorm hazard for aeronautics.  
*Email:* magalie.buguet@onera.fr

**Lei CAO** received the Ph.D. degree in computer software and theory from Tongji University, Shanghai, China, in 2016. He was a Lecturer with Shanghai Maritime University, Shanghai. In 2013, he was a Visit Researcher with the Institute of Medical Psychology and Behavioral Neurobiology, Tuebingen University, Tuebingen, Germany. He has authored or coauthored about 20 journal and conference papers. His research interests include EEG signal processing, brain-computer interface, machine learning, and medical image analysis.  
*Email:* lcao@shmtu.edu.cn

**Simon CARLADOUS** is an agricultural and environmental engineer from the ENGEES (Strasbourg, France), with an M2 in geography for natural risk management from the Université Paul Valéry (Montpellier, France) and a PhD in Environmental Sciences and Engineering from the Mines of Saint-Étienne, France. For 8 years, his professional career was devoted to operational project management, followed by expertise in the analysis and management of natural hazards within the mountain land restoration services of the Office National des Forêts (ONF-RTM). He developed a solid technical foundation in the context of a PhD, defended in 2017, in decision sciences, combining methods from dependability analysis and Evidential Reasoning based-decision-aiding methods, using fuzzy sets and possibility theories. Since 2017, the desire to apply and transpose this methodological framework to the practice of ONF-RTM and State services has earned him recognition as an international expert by the Ministry of Agriculture in 2022. During this last year, he has broadened his skills to include team management and information systems, drawing on his profile as a complex project manager and technical expert.  
*Email:* simon.carladous@onf.fr

**Snehashish CHAKRAVERTY** is a Higher Administrative Grade Professor in Dept. of Maths. (Appl. Maths. Group) & Dean (Academic). Earlier was with CSIR-Central Building Research Institute, Roorkee, did Graduation from St. Columba's College (Ranchi Univ.) M. Sc. (Appl. Maths.) & M. Phil. (Comp. Applications) from Univ. of Roorkee (Now, IIT Roorkee) securing First positions in the Univ. Ph. D. from IIT Roorkee in 1993 & Post-Doctoral research at Inst. of Sound and Vibration Res., Univ. of Southampton, U.K. and at Faculty of Eng & Computer Sc., Concordia Univ., Canada. He was visiting Professor at Concordia & McGill Universities., Canada and at Univ. of Johannesburg, South Africa. He has authored/co-authored/edited 32 books, published 437 research papers in journals and conferences. He is in the Editorial Boards of various Int. Journals, Book Series and Conferences. Prof. Chakraverty is the Chief Editor of International Journal of Fuzzy Computation and Modelling, Inderscience Pub. (<http://www.inderscience.com/ijfcm>), Associate Editors of *Computational Methods in Structural Engineering*, *Frontiers in Built Environment* and *Curved and Layered Structures* (De Gruyter). He was the President of

the Section of Math. sciences (including Statistics) of Indian Science Congress (2015-2016) & the Vice President Orissa Math. Society (2011–2013). Recipient of prestigious awards viz. Indian National Science Academy nomination under Int. Collaboration/Bilateral Exchange Program (with Czech Republic), Platinum Jubilee ISCA Lecture Award (2014), CSIR Young Scientist (1997), BOYSCAST (DST), UCOST Young Scientist (2007, 2008), Golden Jubilee Director's (CBRI) Award (2001), Roorkee Univ. gold Medals (1987, 1988) for first positions in M. Sc. & M. Phil. etc. He is in the list of 2% top world scientists (2020, 2021) in Artificial Intelligence & Image Processing category. IOP Publishing Top Cited Paper Award for consecutive two terms from India (2018 to 2020) and (2019 to 2021). He has already guided 22 Ph.D. students and undertaken 16 research projects as PI.

*Email:* sne\_chak@yahoo.com

**Frédéric DAMBREVILLE** studied mathematics, logic, signal and image processing. He received the Ph.D. degree in signal processing and optimization, from the Univ. of Rennes, France, in 2001, and received the habilitation to supervise research in 2011. He enjoyed a stay in California, U.S.A., and worked as a postdoctorate in the Naval Postgraduate School at Monterey in 2001/2002. In 2002, he joined the dept. image, perception and robotic of the CTA Lab (french ministry of Defense); in 2020, he joined the ONERA Lab (French Aerospace Lab). His main interests are in optimization, optimal planning, data & sensor fusion, Markov models & Bayesian networks, logic & conditional logic, simulation methods, Bayesian optimization, machine learning. His recent works are about rare event simulation applied to meta-modelling and application to nonlinear Bayesian optimization, optimal decision with partial observation, machine learning and belief functions, including software implementations and architectures.

*Email:* frederic.dambreville@onera.fr

**Lucas DEREGNAUCOURT** received his Applied Mathematics diploma in 2022 from the INSA Rouen Normandie in France, and the Master's Degree in Data Science and Engineering from the University of Rouen in 2022. He is currently a Ph.D. candidate in the LITIS laboratory, France. His main research interests are evidence theory, sensor and data fusion, deep learning and intelligent vehicles.

*Email:* lucas.deregnaucourt@insa-rouen.fr

**Maxime DEVANNE** received his Ph.D. Degree in Computer Science from both Université de Lille, France and Università di Firenze, Italy, in 2015. He then was a postdoctoral researcher at Université de Lille, France from 2015 to 2016, and at IMT Atlantique, France from 2016 to 2019. Since 2019, he is an Associate Professor of Computer Science at Université de Haute Alsace, in Mulhouse, France. His research interests are mainly focused on computer vision and machine learning through the analysis of depth images, 3D videos, time series and human body motion. He works on various application fields like medicine, robotics, ambient assisted living and autonomous vehicles.

*Email:* maxime.devanne@uha.fr

**Jean DEZERT** graduated from EFREI Engineering School in Paris in 1985, and obtained the Ph.D. degree in automatic control and signal processing from the University Paris XI, Orsay, France in 1990. During 1986–1990, he did research in multi-sensor multi-target tracking (MS-MTT) at the French Aerospace Research Lab (ONERA), Châtillon, France. During 1991–1992, he visited the ESE Department at UConn University in Connecticut USA as Post-doc Research Fellow under supervision of Professor Bar-Shalom. During 1992–1993, he was a Teaching Assistant in Electrical Engineering Department at Orléans University, France. He joined ONERA in 1993, where he is a Maître de Recherches and Senior Research Scientist at ONERA, Palaiseau, France. His current research interests include estimation theory with applications to MS-MTT, information fusion, plausible reasoning under uncertainty, and decision-making support. He is a reviewer for several international journals. He has worked for the development of the International Society of Information Fusion ([www.isif.org](http://www.isif.org)) and served in its Board. He was President of ISIF in 2016. Dr. Dezert has been involved in the Technical Program Committees of several International Conferences on Information Fusion. He gave several seminars, lectures, and tutorials on information fusion and tracking in Europe, USA, Canada, Australia, and China during the last three decades. He has published several book chapters, around 170 papers in conferences, and 90 papers in journals. He has also co-edited several books with Professor Smarandache on information fusion related to DSMT for reasoning about uncertainty and decision-making support.

*Webpage:* [www.onera.fr/fr/staff/jean-dezert](http://www.onera.fr/fr/staff/jean-dezert)

*Emails:* jean.dezert@onera.fr, jdezert@gmail.com

**Théo DEZERT** is a Postdoctoral Researcher, under the supervision of Professor Leif Lia and Associate Professor Fjola Gudrun Sigtryggisdottir at the Norwegian University of Science and Technology (Norway). His current research focuses on embankment dams' resistance to overtopping process. He obtained a bachelor's degree in Earth and Environmental sciences from Tours University in 2014 and a master's in planetary geosciences from Nantes University in 2016. During his Ph.D.,

obtained in 2019, working with CEREMA and IFSTTAR French institutes, he developed a fusion methodology to combine spatialized geophysical and punctual geotechnical data to improve dike diagnosis. This methodology relies on the use of belief functions. He has collaborated with EDF (French Electric utility company) to apply this methodology to their hydraulic works during a postdoc from March 2020 to May 2021. In 2021, he co-organized the francophone First National Information Fusion Seminar for the Characterization of Porous Media: Applications to Geosciences and Civil Engineering with the French INRAE Institute and Clermont-Auvergne University. He has published eight international articles and participated in over twenty international and national conferences, as well as reviewing geophysical articles for international journals and chairing sessions in international congresses. He recently joined the start-up FI-NDT, a spin-off of the Gustave Eiffel University in Bouguenais, France.

*Email:* theodezert@yahoo.fr

**Valentina DRAGOS** is a research scientist working for ONERA, The French Aerospace Lab in Palaiseau, France. Valentina holds Ph.D. from Paris Descartes University and HDR from University of Paris. Her research interests include semantics interoperability and uncertainty analysis for defense and security systems. Valentina is also a member of the advisory board of ENISA, the European Agency for Cybersecurity.

*Email:* valentina.dragos@onera.fr

**Yilin DONG** was born in Yancheng, China, in 1990. He received the Ph.D. degree in control science and engineering from the School of Automation, Southeast University, Nanjing, China, in 2020. He was with the College of Computer Science and Engineering, Shanghai Maritime University, Shanghai, China. His research interests include belief function theory, information fusion, and pattern recognition.

*Email:* yldong@shmtu.edu.cn

**Yuan-Wei DU** received the B.S. degree in engineering management from Jilin University, Changchun, China, and the M.S. degree and the Ph.D. degree in Management Science and Engineering from Jilin University in 2004, 2007, and 2010, respectively. He previously worked at Kunming University of Science and Technology and Ocean University of China. Now, he is with the School of Business and Tourism Management at Yunnan University, Kunming, China. He is the author or co-author of over 100 papers in management, engineering journals, and conferences. He has been involved in more than 10 projects supported by the Major Program of National Social Science Foundation of China, the National Natural Science Foundation of China (NSFC), the Special Funds of Taishan Scholars Project of Shandong Province, and other essential foundations. His research interests are in decision-making, information fusion, etc.

*Email:* duyuanwei@foxmail.com

**Zhansheng DUAN** received the B.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1999 and 2005, respectively, and the Ph.D. degree from The University of New Orleans, New Orleans, LA, USA, in 2010, all in electrical engineering. From January 2010 to April 2010, he was an Assistant Professor/Research with the Department of Computer Science, the University of New Orleans. In July 2010, he joined the Center for Information Engineering Science Research, Xi'an Jiaotong University, where he is currently a Full Professor. He has authored or coauthored two books entitled Multisource Information Fusion (Beijing, China: Tsinghua University Publishing House, 2006, 2010, and 2022) and Principal Component Analysis Networks and Algorithms (Springer Singapore, 2017), one book chapter, and 106 journal and conference proceedings papers. His research interests include estimation and detection theory, target tracking, information fusion, nonlinear filtering, and performance evaluation. Dr. Duan was a member of the Board of Directors of the International Society of Information Fusion from 2018 to 2020 and is a member of the Honor Society of Eta Kappa Nu, and IEEE Member.

*Email:* zsduan@mail.xjtu.edu.cn

**Xiaojing FAN** was born in Taiyuan, China, in 1990. She received the bachelor's degree in electrical engineering and automation from Hebei Normal University, Shijiazhuang, China, in 2010 and the master's degree in control engineering from Northeast Forestry University, Harbin, China, in 2017. She received the Ph.D. degree in control science and engineering at the School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an, China, in 2022. Her research interests include evidence theory, information fusion, and pattern classification.

*Email:* jjingxiulian@126.com

**Yannick FARGIER** graduated from the Institut de Physique du Globe de Paris (IPGP) in 2008 and obtained a Ph.D. in geophysics from École Centrale de Nantes in 2011, collaborating with the Laboratoire Central des Ponts et Chaussées (LCPC). From 2008 to 2011, he served as a research engineer at EDF R&D, and from 2012 to 2017 as the Head of a research laboratory at Cerema, a public institution dedicated to supporting public policies, supervised by the Ministry for

Ecological Transition and Regional Cohesion. He focused his research on using geophysical and geotechnical methods to assess natural hazards and civil engineering infrastructures. Since 2017, he has been a senior researcher at the Gustave Eiffel University, where his current research revolves around the development of data processing methods that utilize information fusion, artificial intelligence, and inversion techniques to gain a deeper understanding of physical phenomena and evaluate civil engineering infrastructures. He has been involved in organizing several international conferences and has published more than sixty articles published in conferences and journals. He currently facilitates the transfer of research and technologies to the start-up FI-NDT, a spin-off of the Gustave Eiffel University in France.

*Webpage:* <https://pagespro.univ-gustave-eiffel.fr/yannick-fargier>

*Emails:* [yannick.fargier@univ-eiffel.fr](mailto:yannick.fargier@univ-eiffel.fr), [Yannick.fargier@fi-ndt.com](mailto:Yannick.fargier@fi-ndt.com)

**Marek FIDALI** is an Associate Professor in the Department of Fundamentals of Machinery Design at the Faculty of Mechanical Engineering of the Silesian University of Technology in Gliwice, Poland since 2015. He received his Master of Science and Ph.D. degrees in Mechanical Engineering from Silesian University of Technology in 1997 and 2003, respectively. His research interests lie in technical diagnostics in the broad sense, infrared thermography, image and signal processing methods, as well as modal analysis, measurement systems, and acoustics.

*Email:* [marek.fidali@polsl.pl](mailto:marek.fidali@polsl.pl)

**Stefka FIDANOVA** is Professor of Computer Science at Institute of Information and Communication Technologies, Bulgarian Academy of Sciences. Her research interests include theory, methods, applications of combinatorial optimization and parallel algorithms. She heads the research group of Parallel Algorithms and Machine Learning. She has authored over 200 refereed journal, proceedings and collection papers, edited 13 proceedings, collections and special issues and written a 2 monograph. She belongs to the editorial boards of several international journals. She has received Career Award 2018 of Marie Curie Alumni Association of EU.

*Email:* [stefka.fidanova@gmail.com](mailto:stefka.fidanova@gmail.com)

**Ionel-Alexandru GAL** is Secretary of the Scientific Council of the Institute of Solid Mechanics of the Romanian Academy, PhD Eng. Ionel-Alexandru Gal is a distinguished researcher and innovator in the field of Robotics and Mechatronics. He boasts an extensive work experience, progressively assuming roles as a Scientific Researcher, Technical Leader, and Chief Technology Officer. Ionel-Alexandru Gal has made significant contributions to Robotics and Mechatronics. Notably, he has been deeply involved in the development of control laws and methods for mobile robots, leveraging technologies like Matlab and LabView. Furthermore, he has demonstrated remarkable proficiency in designing and implementing software for virtual environments, utilizing cutting-edge tools such as Unity, C#, .Net, and others. In addition to his technical expertise, Ionel-Alexandru Gal has established noteworthy research affiliations and participated in prestigious projects, collaborating with esteemed institutions and universities. His active involvement in 7 various national and European research grants has played a vital role in advancing cutting-edge projects related to rescue robots for fire-fighting, water quality monitoring, and smart city waste management. Moreover, Ionel-Alexandru Gal's academic prowess is reflected in his publication record, with over 70 papers and a significant contribution as the author of the book "Hybrid force-position control strategies in robotics". His expertise extends to state-of-the-art technologies, such as DSMT Decision-Making Algorithms, virtual simulation techniques, and sliding motion control with bond graph modeling, solidifying his position as a respected figure within the Robotics and Mechatronics community. Ionel-Alexandru Gal's dedication to scientific research, technical innovation, and active involvement in professional organizations stand as a testament to his profound impact on the field. His contributions have shaped the landscape of Robotics and Mechatronics, cementing his reputation as a prominent and influential figure in the domain.

*Email:* [alexandru.gal@imsar.ro](mailto:alexandru.gal@imsar.ro)

**Shuzhi Sam GE** received the B.Sc. and M.Sc. degrees in control engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1986 and 1988, respectively, and the Ph.D. degree in mechanical/electrical engineering from the Imperial College of Science, Technology and Medicine, University of London, London, U.K., in 1993. He is currently the Founding Director of the Robotics Institute and the Institute of Intelligent Systems and Information Technology, University of Electronic Science and Technology of China, Chengdu, China. He is also the Founding Director of the Social Robotics Laboratory, Interactive Digital Media Institute, National University of Singapore, Singapore, where he is a Professor with the Department of Electrical and Computer Engineering. Dr. Ge is the Editor-in-Chief for the International Journal of Social Robotics. He has served as an Associate Editor for a number of flagship journals. He has also served as the Vice-President for Technical Activities from 2009 to 2010, Membership Activities from 2011 to 2012, and the IEEE Control Systems Society.

*Email:* [samge@nus.edu.sg](mailto:samge@nus.edu.sg)

**Mihreteab Negash GELETU** received his M.Sc. in 2013 in Electrical Engineering under Control Engineering specialization from Addis Ababa University (AAU), Addis Ababa, Ethiopia. He is currently a Ph.D. candidate in Institut de Recherche en Informatique, Mathématiques, Automatique et Signal (IRIMAS) of Université de Haute-Alsace in Mulhouse, France, and School of Electrical and Computer Engineering of AAU. His main research interests are perception, deep learning, sensor and data fusion, evidence theory, and intelligent vehicles.

*Email:* mihreteab-negash.geletu@uha.fr

**Bilyana Zaharieva GENOVA** - M.Sci. in Physics-Meteorology, Sofia University “St. Kliment Ohridski”, Bulgaria; Ph.D. in Psychophysiology; Assistant Professor; Research Team *Mechanisms of Perception*, Department of Sensory Neurobiology, Institute of Neurobiology (former Institute of Physiology), Bulgarian Academy of Sciences. Main research topics: Visual perception of motion and changes in motion, Simple and choice reaction times to changes in speed and direction of visual motion, Age-related changes in visual motion perception. Methods: Psychophysics - threshold procedures, matching procedures, Response time measurements, Eye movements, EEG. Scientific activity: over 20 scientific publications and 25 participations in international and national conferences. Projects: participation in 6 national and 1 international research projects.

*Email:* b.genova@abv.bg

**Ghalem Kamel GHANEM** is Associate Professor at Higher School of Electrical Engineering and Energetic, Oran, Algeria. He obtained his undergraduate degree in Electronics in 2009 from University of Science and Technology of Oran - Mohamed Boudiaf - Algeria. Ghalem received his Magister Degree in 2013 and Ph.D. in 2018 in Electronics from the same University under the supervision of Professor Hendel Fatiha. He was promoted to Associate Professor in 2022. He is member of Complex System Laboratory at Higher School of Electrical Engineering and Energetic, Oran, Algeria. His research interest includes: Image processing, Biometrics, Iris recognition, Machine Learning.

*Emails:* ghalem\_kamel\_ghanem@esgee-oran.dz, ghalem.kamel@live.fr

**Dănuț-Vasile GIURGI** works as a Ph.D. researcher at the Université de Haute-Alsace, Mulhouse, France, in the IRIMAS laboratory. His current research orientations are englobing evidential deep learning and data fusion for autonomous driving capabilities. He received his double MSc Degree in Image and Signal Processing (TSI, ISIS) from the Technical University of Cluj-Napoca (TUCN), and respectively University of Bordeaux. He finished his BSc Degree in Automation and Control Engineering at TUCN with a thesis exchange in Robotics at TU Delft. His current interests are related to deep learning, belief functions, decision-making under uncertainty, conflict management, and multi-modal sensor fusion for the autonomous driving industry.

*Email:* vasile.giurgi@uha.fr

**Jean-François GRANDIN** was with Thales Airborne Systems, Élancourt, France. He did Experimental Analysis of Time Deviation on Passive Localization Systems, and conducted many projects in defense and surveillance systems for Thales during his career.

*Email:* jean-francois.grandin@fr.thalesgroup.com

**Enguang GUAN** is affiliated with the School of Logistics Engineering, Shanghai Maritime University. He has previously published research papers on topics such as linearity of the force leverage mechanism based on flexure hinges.

*Email:* egguan@shmtu.edu.cn

**Deqiang HAN** was born in Xi'an, China, in 1980. He received the bachelor's degree in Communication and Control Engineering and the master's and Ph.D. degrees in Control Science and Engineering from Xi'an Jiaotong University, China, in 2001, 2004, and 2008, respectively. He is currently a Professor with Xi'an Jiaotong University and the Vice-Dean of School of Automation, Xi'an Jiaotong University. His current research interests include evidence theory, information fusion, and pattern classification. Dr. Han is an International Society of Information Fusion (ISIF) Member and Technical Program Committee Member for the 1st–11th Chinese Conference on Information Fusion in 2009–2023.

*Email:* deqhan@xjtu.edu.cn

**Fatiha HENDEL** is a Professor at USTO-MB, Oran, Algeria, and member of Laboratory of Research in Intelligent Systems (LARESIS). Her research area includes: mobile robot applications, biometrics, pattern recognition, optimization, and machine learning.

*Email:* fa\_hendel@yahoo.fr

**Wojciech JAMROZIK** is Assistant professor in the Department of Fundamentals of Machinery Design at the Faculty of Mechanical Engineering of the Silesian University of Technology in Gliwice, Poland. He received the Ph.D. degree in Mechanical Engineering from Silesian University of Technology in 2012. His main research areas are technical diagnostics, termography, image processing, and analysis, as well as aggregation of incomplete and uncertain information in technical diagnostics.

*Email:* wojciech.jamrozik@polsl.pl

**Luc JAULIN** was born in Nevers, France in 1967. He received the Ph.D. degree in automatic control from the University of Orsay, France in 1993 and the habilitation a Diriger des recherches degree in 2000. He is currently full Professor of Robotics at the ENSTA (Ecole Nationale Supérieure de Techniques Avancées) Bretagne, engineering school Brest, France since 2004. He does his research on underwater and sailboat robotics using interval methods and constraint propagation in the Lab-STICC. He is the co-author of more 200 papers in international journals and congresses. He received the R.E. Moore Prize in 2012 for his work on localization and map building of underwater robots using interval constraint propagation.

*Webpage:* [www.ensta-bretagne.fr/jaulin/](http://www.ensta-bretagne.fr/jaulin/)

*Email:* lucjaulin@gmail.com

**Thomas JOSSE-LAURAIN** received his PhD at the Université de Valenciennes in 2017 where he presented Takagi-Sugeno controllers for Internal Combustion engines. Since 2018, he is Associate Professor at Université de Haute-Alsace in Mulhouse, France, where he teaches at ENSISA engineering school and doing his research at IRIMAS institute. He is project manager of ANR JCJC EviDeep (2021-2025). His research interests go from data fusion to artificial intelligence applied to autonomous vehicles, rail transport, mobility and other related things.

*Email:* thomas.josso-laurain@uha.fr

**Imene KHANFIR KALLEL** received the Diploma degree in electrical engineering, the master's degree, and the Ph.D. degree, and the HDR degree from the National Engineering School of Sfax, Tunisia, in 2001, 2003, 2010, and 2019, respectively. She is currently an Associate Professor with the Biotechnology Higher Institute of Sfax, a member of the Control & Energy Management Laboratory (CEMLab), University of Sfax, an Associate Researcher with the Image & Information Processing (iTi) Department, IMT Atlantique, Brest, France. In 2019, she created the Cybernics Team, Sfax University, where she is working on the fusion of human, machine, and information systems. The team objective is to conduct applied research to develop human assistive technologies in biorobotics. She has been the Head of a research team working on decision support system based applications since 2011. Her research work has been mainly applied to biometrics and biotechnologies. She has published more than 30 articles in journals and conferences. Her research interests range from information theory, image and information processing, machine learning, classification, pattern recognition to analytics, and information fusion to computer-based decision support systems. She co-chaired CSP Conference of SSD Multiconference from 2015 to 2018 and actually co-chairs SCIS of SSD Multiconference. She is on the Editorial Board of the Journal on Digital Signal Processing (JDP) (i-manager), since February 2015.

*Email:* imen.khanfir.kallel@gmail.com

**Mohammad Omar KHYAM** received the B.Sc. degree in electronics and telecommunication engineering from the Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh, in 2010, and the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, Australia, in 2015. He is currently a Lecturer with the Central Queensland University, QLD, Australia. His research interests include signal processing and wireless communication.

*Email:* m.khyam@cqu.edu.au

**Pavlina KONSTANTINOVA** received M.Sc. and Ph.D. degrees in Computers and Computer-Aided Transient and Steady State Analysis of nonlinear Electronic Circuits from Technical University of Sofia in 1967 and 1987, respectively. She is an Associate Professor at the European University in Pernik, Bulgaria. Her research interests include sensor data processing, data association, mathematical modeling, parallel algorithms, and object-oriented programming. Her scientific activity includes more than 80 publications, with 15 indexed in Scopus, and in journals with SJR and IF.

*Email:* pavlina.konstantinova@gmail.com

**Hind LAGHMARA** received her Ph.D. in 2020 from the university of Haute-Alsace, Mulhouse in France. She is now an Associate Professor at INSA Rouen Normandie since 2021, and is a member of the LITIS laboratory. Her research interests are mainly data fusion and uncertainty modeling for safety in autonomous systems.

*Email:* hind.laghmara@insa-rouen.fr

**Jean-Philippe LAUFFENBURGER** received a Ph.D. degree in electrical and control engineering from the Université de Haute-Alsace (UHA), Mulhouse, France, in 2002. Since 2004, he has been an Assistant Professor at the UHA with the ENSISA Engineering College. He is responsible for the international relations for ENSISA. Since 2016, he has been Full Professor and Dean of the Control Engineering and Embedded Systems Department for ENSISA. His research interests include multisensor perception, data fusion, belief functions, navigation, and tracking for intelligent transportation systems. He is a member of the IFAC “Transportation Systems” TC7.4 and Associate Editor of the IEEE Transactions on Intelligent Vehicles. He is also a regular Associate Editor of IEEE and IFAC conferences.  
*Email:* jean-philippe.lauffenburger@uha.fr

**Alexis LECHERVY** received his Ph.D. in 2012 from the University of Cergy-Pontoise in France. He is now an Associate Professor at the University of Caen in France since 2013, and is a member of the Greyc laboratory (UMR CNRS 6072). His research focuses on the design of neural network architecture for multimodal data fusion and metric learning. He is also studying the possibilities of building machine learning models that are more economical in terms of resources and data.  
*Email:* alexis.lechervy@unicaen.fr

**Gaétan LE-GOIC** received his Ph.D. from the university of Savoie Mont-Blanc (Annecy, France) in 2012. Afterwards, he was postdoctoral fellow at the SYMME Laboratory. He is now Associate Professor at the university of Bourgogne Franche-Comté and member of the Laboratory IMVIA where he is leading research and industrial projects on surface topography metrology and visual appearance digitization.  
*Email:* Gaetan.Le-Goic@u-bourgogne.fr

**Xinde LI** (born in 1975.09) got his doctor degree from Huazhong University of Science and technology in 2007. After that, he joined in Southeast University. From 2012, he spent one year as a visiting scholar in Georgia Institute of Technology. In 2016, he was a Postdoc Research Fellow at the Department of ECE in National University of Singapore. Presently, he is Professor at Southeast University, distinguished researcher and Head of Department in Nanjing Applied Math Center, Vice-Director of Intelligent Robot Committee, Vice-Director of Intelligent Products and Industry Working Committee of Chinese Association for Artificial Intelligence, Vice-Director of Intelligent Manufacturing System Committee of Chinese Association of Automation, IEEE Senior Member (RAS, CS, CIS), etc. He has undertaken many projects, including 863 key projects, Key and general projects from National Natural Science Foundation of China, the 13th Five-Year Pre-research key projects, JKW163 key projects, etc. He has published more than 100 reputable papers such as IEEE Transactions Journals, etc. He also published 3 books, 31 national invention patents. He won many prizes, such as the International Award, Province Award and Association Award, etc. In addition, he also served on several editorial boards of journal, such as J. of Sys. Eng. and Elec, Int. J. Robot. Autom. Technol and Int. J. of Aero Aero Eng, etc. He has also given respectively more than 40 reports as plenary speaker, keynote speaker in international /civil conference. He has also organized and severed international/civil conferences as general chairman. Due to his outstanding achievements, He was elected as an Academician of Russian Academy of Natural Sciences in 2021, and was awarded as the cover character of Scientific Chinese (A well-known magazine, like Scientific American), and 2022, he was elected as FIET.  
*Email:* xindeli@seu.edu.cn

**Yihai LIU** is Senior Engineer at the Jiangsu Institute of Automation, Lianyungang, Jiangsu province, China. His main research interests include object detection and recognition techniques and neural network fusion methods. He has published several technical papers related to the specific areas about information fusion.  
*Email:* liuyihai@126.com

**Zhunga LIU** was born in China. He received the bachelor’s, master’s, and Ph.D. degrees in control theory and applications from Northwestern Polytechnical University (NPU), Xi’an, China, in 2007, 2010, and 2013, respectively. He also studied with Telecom Bretagne, Brest, France, for Ph.D. in 2010 and 2013. He has been a full-time Professor with the School of Automation, NPU, since 2017. His research interests mainly focus on belief functions, pattern recognition, and information fusion.  
*Email:* liuzhunga@nwpu.edu.cn

**Chaomin LUO** holds a Ph.D. degree in electrical and computer engineering from the Department of Electrical and Computer Engineering at the University of Waterloo, Canada. He also earned an M.Sc. degree in engineering systems and computing from the University of Guelph, Canada, and a B.Sc. in electrical engineering from Southeast University, Nanjing, China. Currently, he serves as an Associate Professor in the Department of Electrical and Computer Engineering at Mississippi State University. He is Associate Editor in 2019 *IEEE/RSJ International Conference on Intelligent Robots and*

*Systems* (IROS 2019). He is Tutorials Co-Chair in the 2020 IEEE Symposium Series on Computational Intelligence. Dr. Luo was the recipient of the Best Paper Awards in IEEE International Conference on Information and Automation, International Conference on Swarm Intelligence, and SWORD Conference. His research interests include Robotics, Autonomous Systems, and Control and Automation. Dr. Luo is an IEEE senior member, INFORMS, and ASEE member. He was the General Co-Chair of the 1st IEEE International Workshop on Computational Intelligence in Smart Technologies (IEEE-CIST 2015), and Journal Special Issues Chair, IEEE 2016 International Conference on Smart Technologies (IEEE-SmarTech), Cleveland, OH, USA. He was Chair and Vice Chair of IEEE SEM - Computational Intelligence Chapter and was a Chair of IEEE SEM - Computational Intelligence Chapter and Chair of Education Committee of IEEE SEM. Dr. Luo serves as Associate Editor of *IEEE Transactions on Cognitive and Developmental Systems*, *International journal of Robotics and Automation*, and Associate Editor of *International Journal of Swarm Intelligence Research (IJSIR)*.

*Email:* Chaomin.Luo@ece.msstate.edu

**Nisha Rani MAHATO** is currently working as a Senior Quantitative Analyst at the Global Risk and Research Solutions (GR&RS), Model Risk (MR) Department, CRISIL Ltd., India. She holds a Master of Science in Mathematics and a Ph.D. in Mathematics from Department of Mathematics, National Institute of Technology Rourkela, Odisha, India. She was awarded Raman Charpak Fellowship-2016 (RCF-2016) by CEFIPRA, New Delhi, India. Also, she has been awarded best paper at the 38th Annual Conference of Orissa Mathematical Society, Odisha, India in 2011 and best poster in mathematics at Research Scholar Week 2018, NIT Rourkela, Odisha, India. She has authored about 2 books, 3 book chapters and published 7 research papers in respected journals.

*Email:* nisha.mahato1@gmail.com

**Alamin MANSOURI** received his Ph.D. from the university of Bourgogne, France in 2005. Afterwards, he was postdoctoral fellow at the ColorLab laboratory in Norway. He is now full professor at the university of Bourgogne Franche-Comté and member of the Laboratory IMVIA where he is leading the thematic project of Physical and Multimodal Imaging. His current research is focused on multispectral imaging (both sides acquisition and processing/analysis) and its interactions with cultural heritage and media technology.

*Email:*alamin.mansouri@u-bourgogne.fr

**Driss MAMMASS** is professor of Higher Education at the Faculty of Sciences, University Ibn Zohr, Agadir Morocco. He received a Doctorat in Mathematics in 1988 from Paul Sabatier University (Toulouse, France) and a Doctorat sciences degrees in Mathematics and Image Processing from Faculty of Sciences, University Ibn Zohr Agadir Morocco, in 1999. He supervises several Ph.D. theses in the various research themes of mathematics and computer science such as remote sensing and GIS, digital image processing and pattern recognition, the geographic databases, knowledge management, semantic web, etc.

*Email:*mammass@uiz.ac.ma

**Julien MORAS** received the Engineering degree from the ISAE in 2008 and the Ph.D. degree in technology of information and systems from the UTC in 2013. Since 2013, he is a Research Engineer at the French Aerospace Laboratory, ONERA. His research interests include robotic embedded perception for autonomous navigation, and data fusion theory in wireless sensor networks.

*Email:* julien.moras@onera.fr

**Jiawei NIU** was born in China. She received the B.S. degree in automation from the Xi'an University of Technology, Xi'an, China, in 2016, and the M.S. degree in mechanical control and engineering from the Kyushu Institute of Technology, Kitakyushu, Japan, in 2019. She is currently pursuing the Ph.D. degree in control theory and applications with the School of Automation, Northwestern Polytechnical University, Xi'an. Her research interests include pattern recognition and belief functions.

*Email:* niujiawei@mail.nwpu.edu.cn

**Md. NOOR-A-RAHIM** received the Ph.D. degree in telecommunications from the Institute for Telecommunications Research, University of South Australia, Adelaide, SA, Australia, in 2015. He was a Postdoctoral Research Fellow with the Centre for Infocomm Technology, Nanyang Technological University, Singapore. He is currently a Senior Postdoctoral Researcher (MSCA Fellow) with the School of Computer Science and IT, University College Cork, Cork, Ireland. His research interests include control over wireless networks, intelligent transportation systems, information theory, signal processing, and DNA-based data storage. Dr. Noor-A-Rahim was the recipient of the Michael Miller Medal from the Institute for Telecommunications Research, University of South Australia, for the most Outstanding Ph.D. thesis, in 2015.

*Email:* m.rahim@cs.ucc.ie

**Benjamin PANNETIER** is a researcher in tracking and sensor fusion. He received the Ph.D. degree in automatic control and signal processing from the University of Grenoble in 2006. From 2005 to July 2021, he was a research engineer at the French Aerospace Laboratory, ONERA, working on target tracking, detection, estimation theory and data fusion for the battlefield surveillance systems for the French army. Since August 2021, he has integrated the new Research Lab from CS Group company as research engineer. His currently works are new sensor network concept and data fusion to improve countering unmanned aerial systems. In 2022, he becomes Research Director at the Polytechnic Institute from Paris and Associated Professor at the French Aerial Research Center (CREA).

*Email:* benjamin.pannetier@csgroup.eu

**Tadeusz PIETKIEWICZ** graduated from the Faculty of Cybernetics at the Military University of Technology in Warsaw (Poland) in 1976. In the years 1978-1981 he completed doctoral studies at the Faculty of Cybernetics at the Military University of Technology and obtained the PhD degree in cybernetics in 1982. In the years 1982–2003 he was the head of the modeling department at the Institute of Automation of Command Systems at the Military University of Technology. During this period, his research interests were related to the design of automated air defense command and control systems and decision support for these systems. Since 2003, he has been a research and teaching assistant professor at the Institute of Radioelectronics of the Faculty of Electronics at the Military University of Technology. His current research interests include image recognition methods, information fusion, decision support systems and radioelectronic recognition systems. He has published 83 papers in conference proceedings and 16 journal articles. He was the supervisor of 106 B.Sc. and M.Sc. theses.

*Email:* tadeusz.pietkiewicz@wat.edu.pl

**Rongjun QIN** is an Associate Professor in Geomatics and Computer Vision at The Ohio State University, U.S. He received Ph.D. from the ETH Zurich in Photogrammetry, and holds a master's degree in Photogrammetry and Remote Sensing, and bachelor's degree in Computational Mathematics from Wuhan University. His research interest lies in the intersection among Photogrammetry, Computer Vision and Remote Sensing, with specialization on 3D reality based modeling using multi-modality and multi-source data. He is the author of the RPC stereo processor (RPC), a satellite photogrammetry package that produces large-scale and high-quality digital surface models for multi-view satellite images. Rongjun Qin serves as an associate editor for the ISPRS Journal of Photogrammetry and Remote Sensing, the Photogrammetric Engineering and Remote Sensing journal, and the editorial member of the Photogrammetric Record journal. He was the chairing the ISPRS working group of "Satellite Constellation for Remote Sensing" (2016-2022), and now is the Commission Secretary supporting the ISPRS Commission II Presidents. He is also an IEEE Senior Member. His awards include the first prize of Mathematical Modeling Contest 2009, winner of IARPA Topcoder 3D challenge, winner of IEEE Data fusion contest (2019, 2020), ASPRS Talbert Abrams Grand Award, Duan Brown Senior Award, ISPRS CATCON Silver Award etc.

*Email:* qin.324@osu.edu

**Laurent RATTON** has been Engineer in THALES since 1997. He is an Expert in tracking and data fusion. He has been involved in several R&D projects regarding multisensor systems such as Air-to-Ground surveillance systems, Maritime Patrol Aircrafts, Sense and Avoid for UAV as well as Electronic Warfare (EW) and ELeCtronic INTelligence systems. Also, he has been working on numerous research projects investigating new information processing architectures and advanced technologies for target tracking and sensor management in combat system application. He is currently team leader in THALES DMS France, in Algorithm and Processing R&D. Domains of concern include Radar (SAR) imagery, EW and ELINT processing, Data Fusion and sensor management, for which the developed solutions are based on signal/image processing, statistical information processing and artificial intelligence. Laurent Ratton is SEE Senior Member. He has been author or co-author of more than twenty papers, and hold several patents on processing architecture and algorithm design.

*Email:* laurent.ratton@fr.thalesgroup.com

**Kellyn REIN** is a Research Associate at the Research Institute for Communication, Information Processing Ergonomics (FKIE) at Fraunhofer Gesellschaft, Europe's largest applied research organization. She received her Ph.D. from the University of Bonn in Linguistics, after a B.A. at Michigan State University, and two masters, one in Management and one in Computer Information Systems, both with joint conferral from Boston University (1984) and Vrije Universiteit Brussel in Belgium. Her area of research interest is in the analysis of uncertainty in natural language information, information fusion and information quality. She has been active in EU projects focused variously on crisis management, cross-border organized crime, smart cities, and CBRN terrorism. She has also been involved in numerous NATO Research Task Groups, including one on analysis and communication of uncertainty in intelligence, mission assurance and cyber risk assessment for unmanned autonomous system, and currently chairing a task group on multi-level, multi-source information fusion.

*Email:* kellyn.rein@fraunhofer.fkie.de

**Peter REINARTZ** received his Diploma (Dipl.-Phys.) in theoretical physics in 1983 from the University of Munich and his Ph.D. (Dr.-Ing) in civil engineering from the University of Hannover, in 1989. His dissertation is on optimization of classification methods for multispectral image data. Currently he is the acting co-director of the Remote Sensing Technology Institute (IMF) and department head of the department Photogrammetry and Image Analysis, IMF, at the German Aerospace Centre (DLR), Remote Sensing Technology Institute (IMF). He also holds a professorship for computer science at the University of Osnabrueck. He has more than 35 years of experience in image processing and remote sensing and over 500 publications in these fields. His main interests are in machine learning, stereo-photogrammetry and data fusion, using space borne and airborne image data, generation of digital elevation models and interpretation of very high resolution data from sensors like WorldView and Pleiades. He is also engaged in using remote sensing data for disaster management and using high frequency time series of airborne image data for real time image processing and for operational use in case of disasters as well as for traffic monitoring.

*Email:* peter.reinartz@dlr.de

**Lucile SAUSSAYE** got her Ph.D. in Materials Chemistry (Environmental geotechnics and geochemistry) in 2012. She is Head of Department at Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement (CEREMA), Blois, Région Centre, France.

*Email:* Lucile.saussaye@cerema.fr

**Florentin SMARANDACHE** is a Professor of Mathematics at the University of New Mexico, USA. He got the M.Sc. degree in mathematics and computer science from the University of Craiova, Romania, Ph.D. in mathematics from the State University of Kishinev, and postdoctoral in applied mathematics from Okayama University of Sciences, Japan. He is the Founder of Neutrosophy (generalization of dialectics), Neutrosophic Set, Logic, Probability and Statistics since 1995 and has published hundreds of papers and books on neutrosophic physics, superluminal and instantaneous physics, unmatter, quantum paradoxes, absolute theory of relativity, redshift and blueshift due to the medium gradient and refraction index besides the Doppler effect, paradoxism, outerart, neutrosophy as a new branch of philosophy, law of included multiplemiddle, multispace and multistructure, hypersoft set, degree of dependence and independence between neutrosophic components, refined neutrosophic set, neutrosophic over-under-off-set, plithogenic set, neutrosophic triplet and duplet structures, quadruple neutrosophic structures, extension of algebraic structures to NeutroAlgebras and AntiAlgebras, DSMT, and so on to many peer-reviewed international journals and many books and he presented papers and plenary lectures to many international conferences around the world. In addition, he published many books of poetry, dramas, children stories, translations, essays, novel, folklore collections, traveling memories, and art albums. For more information.

*Webpage:* fs.unm.edu/FlorentinSmarandache.htm.

*Email:* smarand@unm.edu

**Basel SOLAIMAN** received the degree in telecommunication engineering from the École Nationale Supérieure des Télécommunications de Bretagne, in 1983, and the DEA, Ph.D., and HDR degrees from the Université de Rennes-I, France, in 1983, 1988, and 1997, respectively. In 1983, he joined the Laboratoire d'Électronique de Philips (L.E.P), Paris, France, where he worked on image compression standards. In 1988, he joined the Institute of Industrial Informatics, Brest, France, as a R&D Engineer-in-Chief to work on new technologies transfer into the industrial regional tissue. In 1992, he joined the École Nationale Supérieure des Télécommunications de Bretagne (Télécom Bretagne), Brest, France, where he is currently a Full Professor and the Head of the Image and Information Processing Department. He has published several academic books, more than 180 journal articles, several book chapters, more than 230 conference proceedings, and technical reports. He has held visiting professor positions at several universities the University of Bristol, U.K.; the University of Carabobo, Venezuela; the University of Michigan, Ann Arbor, USA; the Technical University of Budapest, Hungary; E.T.S, Montreal, Canada; M.U.T, Warsaw, Poland; and the Technical University of Melaka (UTEM), Malaysia. His application domains research activities range from civilian into military ones: medical, remote sensing, underwater imaging, machine learning, information, and knowledge fusion. He chaired several international conferences focusing on his research activities and related to signal and image processing, data and information fusion, and pattern recognition.

*Email:* basel.solaiman@imt-atlantique.fr

**Miroslava STEFANOVA**, M.S. in physics, Assistant Professor in Institute of Neurobiology, Bulgarian Academy of Science, Department of Sensory Neurobiology, Research Team *Mechanisms of Perception* - Research topics: Motion perception, Real and Apparent motion, Global direction, Short-range process, Direction and speed changes. Experience in: statistical modeling, psychophysical methods, preparing experiments, organizing a subject pool. Scientific activity includes the 8 publications in journals indexed in Medline or ISI Web, 14 publications in journals without ISI, 10 participations in International Conferences

(ECVP) and 3 in Regional Balkan Meetings.

*Email:* md.stefanova@inb.bas.bg

**Jean-Marc TACNET** works in INRAE (National Research Institute for Agriculture, Food and the Environment) as a senior researcher in the Institute of Geosciences, a joined research Team between Grenoble Alpes University, CNRS, IRD, Grenoble-INP and INRAE in Grenoble, France. He received two Master Engineering Degrees, a M.Sc. degree in Applied Computer Science (Grenoble University) and a Ph.D. thesis (École Nationale Supérieure des Mines de Saint-Étienne) whose subject was: *Considering uncertainty in the natural risks expertise process using multicriteria decision analysis and information fusion*. He is involved in applied researches dealing with integrated, multifactorial, territorial decision support systems for mountain natural risks management in a context of imperfect information. He also addresses issues related to effectiveness, maintenance and resilience of risk reduction strategies and measures (from structural to nature-based solutions), including their effects on critical infrastructures and systems. His disciplinary fields concern expert assessments, multi-criteria decision making, information imperfection, information fusion, artificial intelligence (fuzzy sets, possibility, belief function theories), safety and reliability analysis, uncertainty analysis, software engineering with also a technical background in hydraulics, water management and civil engineering.

*Email:* jean-marc.tacnet@inrae.fr

**Albena TCHAMOVA** is an Associate Professor at the Institute of Information and Communication Technologies, Bulgarian Academy of Sciences. She received M.Sc. degree in Microelectronics and Ph.D. degree in Radiolocation and Radionavigation from the Technical University of Sofia, Bulgaria, in 1978 and 1998 respectively. She works as a researcher in numerous academic and industry projects (Bulgarian and international) in the area of Multisensor Multitarget Tracking (MS-MTT) and Data Fusion. Her main research interests include MS-MTT, Dezert-Smarandache Theory of Plausible and Paradoxical Reasoning for Information Fusion and Decision making under Uncertainty. The scientific activity includes more than 60 publications, with 37 indexed in Scopus and in journals with SJR and IF.

*Emails:* tchamova@bas.bg, albena.p.tchamova@gmail.com

**Jiaojiao TIAN** (Dr) is a senior research fellow at the Photogrammetry and Image Analysis department, Remote Sensing Technology Institute, German Aerospace Center, Germany, where she is currently heading the 3D and modeling team. She received her Ph.D. degree in Mathematics and Computer Sciences from Osnabrueck University in 2013. She holds IEEE senior membership and serves as co-chair of the ISPRS Commission WG I/8-Multi-sensor Modelling and Cross-modality Fusion. Her research interests include 3D Change Detection, building reconstruction, 3D point cloud segmentation, forest monitoring, and DSM assisted object extraction and classification.

*Email:* Jiaojiao.Tian@dlr.de

**Luige VLĂDĂREANU** (Ph.D.Eng.) is Professor and Director of Robotics and Mechatronics Department Romanian Academy, Institute of Solid Mechanics, Bucharest, Romania. He is the Coordinator of the Robotics and Mechatronics Department of the Institute of Solid Mechanics, Romanian Academy, and Director of the Technologic Transfer Center on Robotics and Advanced Technologies (ROBTA). His scientific work is focused on real-time control in solid mechanics, AI in robotics, multi-microprocessor systems, semi-active control of magnetorheological dissipater systems, complex industrial automation in distributed and decentralized structures, etc. He has published over 50 books and book chapters, 30 edited books, and over 300 papers in journals, proceedings, and conferences in these areas. He is the author of the virtual projection method known as the Vlădăreanu–Munteanu method, and the robot neutrosophic control (RNC) method known as the Vlădăreanu–Smarandache method. He is the winner of two awards, has achieved gold medals for Excellence in Research, SIR 2000, from the Romanian Government, and has received over twenty awards from the World Exhibition of Inventions from Geneva, Brussels, Barcelona, etc. He received the “Traian Vuia” Award from the Romanian Academy, Romania’s highest scientific research forum. He is a Corresponding Member of the American Romanian Academy and a member of the International Institute of Acoustics and Vibration (IIAV), Auburn University, USA, and is on ABI’s Research Board of Advisors, American Biographical Institute etc. He is a PhD advisor in the field of mechanical engineering and robotics at the Romanian Academy. He has served as the organizer of numerous international conferences, is the General Chair of Congress, CANWEST, 2023, Madrid, and of four WSEAS International Conferences, and gives Chair and Plenary lectures at Houston, Harvard, Boston, Penang, Malaysia, Paris, Florence, etc.

*Web:* www.vipro.edu.ro

*Emails:* luige.vladareanu@vipro.edu.ro, luigiv2007@yahoo.com.sg

**Thanuka WICKRAMARATHNE** received the B.Sc. degree in electronics and telecommunication engineering from the University of Moratuwa, Moratuwa, Sri Lanka, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Miami, Coral Gables, FL, USA. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA, USA, and directs UML's S&H Fusion Research Group. He was a Research Assistant Professor with the Departments of Computer Science and Engineering, and Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA. He also serves on the Board of Directors of IEEE Vehicular Technology Society and Administrative Committee of IEEE Sensors Council. His research interests include multi-sensor data fusion, machine intelligence and cybernetics, with an emphasis on soft and hard data fusion (i.e., human-based and physics-based data) and use of opportunistic sensing for enhancing situational awareness, and performance in human-machine teams.

*Webpage:* <https://www.uml.edu/engineering/electrical-computer/faculty/wickramarathne-thanuka.aspx>

*Email:* Thanuka\_Wickramarathne@uml.edu

**Alicja WRÓBEL** studied at the Faculty of Electronics at the Military University of Technology in Warsaw (Poland) in 2017-2022. She obtained the MSc degree in Electronics and Telecommunications in 2022. The subject of her diploma thesis was devoted to the recognition of maritime objects in infrared images and the fusion of image information.

*Email:* alaala-rybak@wp.pl

**Kunhai XU** obtained a Bachelor's degree in Network Engineering from the School of Information Engineering, Shanghai Maritime University in 2022. He is currently pursuing a Master's degree in Software Engineering at the same institution. His current research focuses on fine-grained ship image classification, particularly ship classification in remote sensing images.

*Email:* 202230310302@stu.shmtu.edu.cn

**Yi YANG** was born in Xi'an, Shaanxi, China. She received the Bachelor degree of Automation from Xi'an University of Technology, Xi'an, China in 2002. She received the Master degree and the PhD degree of Control Science and Engineering in 2005 and 2010 respectively from Xi'an Jiaotong University, Xi'an, China. She is currently an Associate Professor at School of Aerospace Engineering, at Xi'an Jiaotong University, China. Her main research interests are information fusion and image processing.

*Email:* jjafeiyy@mail.xjtu.edu.cn

**Hossni ZEBIRI** received B.S on the processing and control theory from the École Nationale Polytechnique, Algiers, Algeria in 2011. He conducted his research on the controllers order reduction and obtained his Ph.D from the IRIMAS Laboratory, Université de Haute Alsace Mulhouse, France, in 2016. He has been with Stellantis group for the development of the braking systems ESC from 2016 to 2019. He Joined Nexteer Automotive France where he has led the development of various EPS steering systems for multiple OEMs for 3 years. In 2022, he was in charge of the steering system for the solar vehicle Sono Motors GmbH in Munich, Germany. Since 2023, he is the Steer-by-wire product manager for KYB Europe GmbH.

*Email:* hossni.zebiri@gmail.com

**Zhe ZHANG** received her B.Eng. degree from Taiyuan University of Technology, China, in 2013, and her PhD degree from Xi'an Jiaotong University, China, in 2019. She is presently working in the College of Electrical and Power Engineering, Taiyuan University of Technology. Her current research concerned with image processing, computer vision and evidence theory. She studied the uncertain problems in image denoising, image registration and image segmentation, respectively, and proposed some works to deal with these uncertain problems using evidence theory. She also interested in medical image processing problems, like the detection of eye rotation angle in cataract surgery based on depth feature matching and the center localization of eye structures based on adaptive ellipse gaussian heatmap. She is currently focusing on improving the representation ability and reasoning efficiency of image restoration model using dynamic neural networks.

*Email:* zhezhang@stu.xjtu.edu.cn

**Rigui ZHOU** received the Ph.D. degree in computer science and technology from the Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronauts, Nanjing, China, in 2007. From 2008 to 2010, he was a Postdoctoral Fellow with the Tsinghua University, Beijing, China. He is currently a Professor with the College of Information Engineering, Shanghai Maritime University, Shanghai, China. His research interests include quantum image processing, quantum reversible logic, quantum genetic algorithm, etc.

*Email:* rgzhou@shmtu.edu.cn

**Wen ZHOU** received the B.S. degree from Shandong Normal University, Jinan, China, and the M.S. degree in Business Management from Ocean University of China, Qingdao, China. Her research interests are in decision-making and information fusion.

*Email:* zwzhouwen9572@163.com

**Changming ZHU** received the Ph.D. degree in computer science from the East China University of Science and Technology, Shanghai, China, in 2015. He is currently an Associate Professor with the College of Computer Science and Engineering, Shanghai Maritime University, Shanghai. His current research interests include pattern recognition and machine learning.

*Email:* cmzhu@shmtu.edu.cn

**Kezhu ZUO** was born in Anqing, Anhui Province, China, in 1992. He is currently studying for the Ph.D. degree in electronic information from School of Cyber Science and Engineering, Southeast University, Nanjing, China from 2020. His research interests include information fusion, belief functions theory, visual inspection and image processing.

*Email:* kezhuo@seu.edu.cn



This fifth volume on Advances and Applications of DSMT for Information Fusion collects theoretical and applied contributions of researchers working in different fields of applications and in mathematics, and is available in open-access. The collected contributions of this volume have either been published or presented after disseminating the fourth volume in 2015 (available at [fs.unm.edu/DSmT-book4.pdf](http://fs.unm.edu/DSmT-book4.pdf) or [www.onera.fr/sites/default/files/297/2015-DSmT-Book4.pdf](http://www.onera.fr/sites/default/files/297/2015-DSmT-Book4.pdf)) in international conferences, seminars, workshops and journals, or they are new. The contributions of each part of this volume are chronologically ordered.

First Part of this book presents some theoretical advances on DSMT, dealing mainly with modified Proportional Conflict Redistribution Rules (PCR) of combination with degree of intersection, coarsening techniques, interval calculus for PCR thanks to set inversion via interval analysis (SIVIA), rough set classifiers, canonical decomposition of dichotomous belief functions, fast PCR fusion, fast inter-criteria analysis with PCR, and improved PCR5 and PCR6 rules preserving the (quasi-)neutrality of (quasi-)vacuous belief assignment in the fusion of sources of evidence with their Matlab codes.

Because more applications of DSMT have emerged in the past years since the apparition of the fourth book of DSMT in 2015, the second part of this volume is about selected applications of DSMT mainly in building change detection, object recognition, quality of data association in tracking, perception in robotics, risk assessment for torrent protection and multi-criteria decision-making, multi-modal image fusion, coarsening techniques, recommender system, levee characterization and assessment, human heading perception, trust assessment, robotics, biometrics, failure detection, GPS systems, inter-criteria analysis, group decision, human activity recognition, storm prediction, data association for autonomous vehicles, identification of maritime vessels, fusion of support vector machines (SVM), Silx-Furtif RUST code library for information fusion including PCR rules, and network for ship classification.

Finally, the third part presents interesting contributions related to belief functions in general published or presented along the years since 2015. These contributions are related with decision-making under uncertainty, belief approximations, probability transformations, new distances between belief functions, non-classical multi-criteria decision-making problems with belief functions, generalization of Bayes theorem, image processing, data association, entropy and cross-entropy measures, fuzzy evidence numbers, negator of belief mass, human activity recognition, information fusion for breast cancer therapy, imbalanced data classification, and hybrid techniques mixing deep learning with belief functions as well.

