

# Computer models and the evidence of anthropogenic climate change: An epistemology of variety-of-evidence inferences and robustness analysis



Martin A. Vezér

The Network for Sustainable Climate Risk Management (SCRiM), Earth and Environmental Systems Institute, Pennsylvania State University, United States

## ARTICLE INFO

### Article history:

Received 27 September 2015

Received in revised form

31 December 2015

Available online xxx

### Keywords:

Variety of evidence;

Robustness;

Computer models;

Epistemology;

Climate change;

Global warming

## ABSTRACT

To study climate change, scientists employ computer models, which approximate target systems with various levels of skill. Given the imperfection of climate models, how do scientists use simulations to generate knowledge about the causes of observed climate change? Addressing a similar question in the context of biological modelling, Levins (1966) proposed an account grounded in robustness analysis. Recent philosophical discussions dispute the confirmatory power of robustness, raising the question of how the results of computer modelling studies contribute to the body of evidence supporting hypotheses about climate change. Expanding on Staley's (2004) distinction between evidential strength and security, and Lloyd's (2015) argument connecting variety-of-evidence inferences and robustness analysis, I address this question with respect to recent challenges to the epistemology of robustness analysis. Applying this epistemology to case studies of climate change, I argue that, despite imperfections in climate models, and epistemic constraints on variety-of-evidence reasoning and robustness analysis, this framework accounts for the strength and security of evidence supporting climatological inferences, including the finding that global warming is occurring and its primary causes are anthropogenic.

© 2016 Elsevier Ltd. All rights reserved.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Science*

## 1. Introduction

The global climate system is large and complex, with many causal factors interacting to produce all sorts of climatic phenomena. To study climate change, scientists employ computer models, which are imperfect representations of target systems. The most detailed, high-resolution models omit representations of major features of the planet that affect the climate (e.g., mountain ranges) and contain parameterizations that simplify complex climatic processes (e.g., cloud formation). Knowing whether a given climate model provides insight into questions about a target system at various scales and about its responses to different perturbations is, therefore, often difficult to determine.

Among the findings in the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) is the

conclusion that “[i]t is *extremely likely* that human activities caused more than half of the observed increase in global mean surface temperature [GMST] from 1951 [to] 2010” (Bindoff et al., 2013, p. 869). Important sources of evidence for this conclusion are the results of computer model simulations. Given that each of the modelled worlds used to study the climate is substantially different from the earth, how do scientists use computer simulations to generate knowledge about the causes of observed climate change?

I address this question by employing a distinction between evidential strength and security (Staley, 2004), and ideas of variety-of-evidence reasoning and robustness analysis (Lloyd, 2015), focussing on the epistemic advantages of drawing on a range of observations, experiments and models, expanding upon related philosophical enquiries into this field of study (Edwards, 2010; Katzav and Parker, 2015; Lloyd, 2009, 2010; Norton & Suppe, 2001; Oreskes, 2007; Parker, 2006, 2010; Weisberg, 2006). Although other accounts also emphasize the importance of multiple sources of data, further analysis of the relationship between

E-mail addresses: [vez@psu.edu](mailto:vez@psu.edu), [martinvez@gmail.com](mailto:martinvez@gmail.com).

varieties-of-evidence and robustness arguments can both clarify the epistemology of current climate modelling research and advance philosophical debates about the structure of reliable scientific methodology. Applying this epistemology to case studies documented by the IPCC, I argue that varieties-of-evidence reasoning and robustness analysis account for the strength and security of evidence supporting important climatological inferences that make use of imperfect computer models.

The approach of the paper is as follows. Section 2 illustrates the logic of variety-of-evidence inferences by describing a set of evidential pathways that converge in support of the global warming hypothesis. This example presents a clear case of strong evidential reasoning, which is a precursor to the more complex case of addressing the role of computer model simulations in climate change attribution studies. Section 3 explains how multiple imperfect models contribute to the security of scientific knowledge because multiple independent approximations of a target system provide alternative evidential pathways to support particular hypotheses. I explain how model agreement in studies of anthropogenic climate change exemplify the confirmatory roles of varieties-of-evidence reasoning and robustness analysis and I address some of Parker's (2011) concerns about evidential security, while also acknowledging important limitations of modelling methodology. Section 4 concludes the paper by summarizing the main points of the argument.

## 2. Evidence of global warming

An evidential pathway is a collection of information that supports an inference. An evidence claim is a proposition stating that some information is evidence for a hypothesis (*H*). Kent Staley (2004) suggests that attaining multiple lines of evidence can increase the plausibility of *H* by enhancing its evidential strength and security.<sup>1</sup> While he draws a distinction between evidential strength (i.e., the degree to which evidence indicates *H*) and security (i.e., the insensitivity of an evidence claim to changes in some evidential pathway), he sets aside the question of how one can increase the strength of an inference (p. 468). As suggested by the analysis of Elisabeth Lloyd (2015), variety-of-evidence reasoning can, however, account for the way in which multiple lines of evidence strengthen an inference. According to this perspective, evidential strength is attained when multiple independent evidential pathways indicate *H* such that it would be unlikely that these various lines of evidence would agree, if *H* were incorrect. In this context, it is not the improbability of the individual lines of evidence occurring that increases evidential strength but the improbability of their agreement, if *H* were incorrect.

Variety-of-evidence reasoning is essentially the severe test criterion advocated Mayo (1996), though Staley applies it somewhat differently, as a constraint on appeals to robustness as “second order evidence.”<sup>2</sup> The question regarding this particular use of the severity criterion as a measure of evidential strength is how exactly its satisfaction is established. On the severity criterion, to show that the convergence of different modelling results is improbable, supposing *H* to be false requires considering the alternatives to *H* and the probabilities those alternatives confer on the convergence of evidence. This criterion is stronger than the formulation of William Whewell's “consilience of inductions” as the argument that “[n]o accident could give rise to such an extraordinary

coincidence” (Whewell, 1858, p. 88), which only establishes that *H* has passed a severe test if an “accident” is the only alternative to *H*. In the current context, alternative hypotheses to account for global warming which have been established as improbable include the claims that natural variability or other forcing mechanisms, such as variations in solar input, volcanic activity, and orbital cycles, are the main drivers of recent warming (Bindoff et al., 2013).

Climate science contains many examples of variety-of-evidence reasoning, the case of global warming providing a vivid illustration. The IPCC exemplifies this point in its evaluation of the Global Warming hypothesis (*GW*)—that is, the proposition that “Global Mean Surface Temperature has increased since the late 19th century” (Hartmann et al., 2013, p. 161). The evidence converging on *GW* includes observations of different interconnected components of the climate system, the collection of which would be unlikely to occur if *GW* were incorrect. Land-surface weather stations provide the most direct evidential pathway supporting *GW*, but other climate indicators include measured changes in atmospheric and oceanic temperatures at various heights and depths; in glacier mass, snow coverage, and sea ice extent; in sea level; and in atmospheric water vapour content.

Fig. 1 contains 10 graphs depicting the consilience of evidence supporting this hypothesis. Since the atmosphere and hydrosphere are interconnected fluid bodies, a warming at the earth's surface produces detectable effects at different levels of the atmosphere and ocean. Some of the energy absorbed by the climate system is stored in the oceans, and this energy uptake is detectable in global ocean heat content records going back to the 1950s. Another line of support is the change in the amount of water vapour in the atmosphere, i.e., its specific humidity, measurements of which show a positive change both over the land and the oceans. Observed sea-level rise is another line of support; warming oceans result in water expansion, leading to rising sea levels, which are further heightened by additional water input from melting glaciers and ice sheets and changes to the storage and usage of water on land. The cryosphere (i.e., the frozen parts of the planet) is also affected by changing temperatures. Snow cover, particularly during the spring, is sensitive to temperature changes. Since the 1950s, Northern Hemisphere spring snow cover has declined. Similarly, Arctic sea-ice losses are detectable in satellite records, particularly at the end of the annual melt in September, which is the time of its minimum extent. For at least the last 20 years, the amount of ice contained in glaciers globally has declined (Hartmann et al., 2013).

Since observations of these diverse phenomena are both consistent with *GW* and would be improbable if *GW* were incorrect, these findings comprise a case of variety-of-evidence reasoning. With the detection of *GW* thus well established, the next question to consider is: What has been causing this warming?

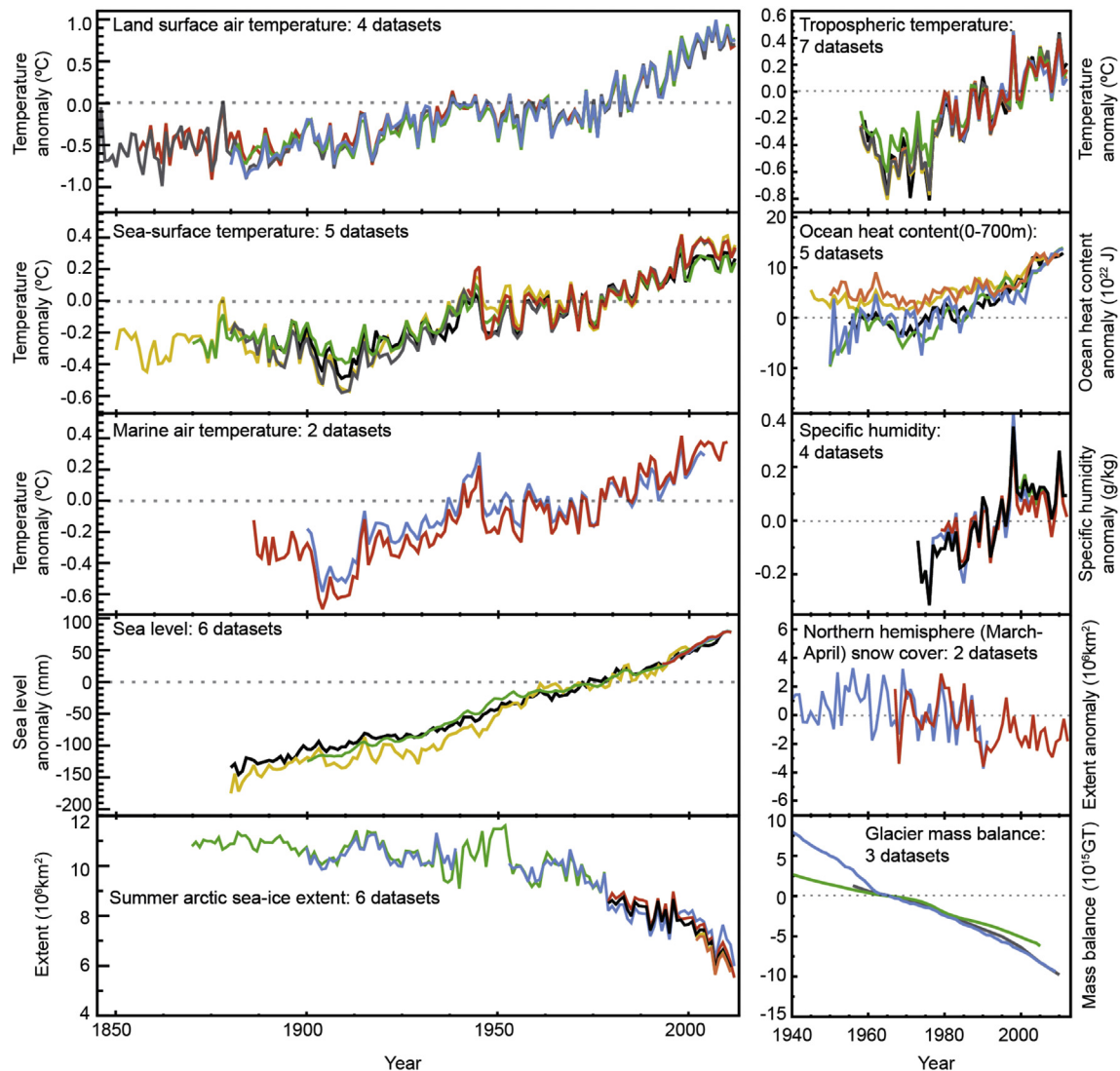
## 3. Climate modelling, robustness analysis, and anthropogenic global warming

While a diversity of evidence increases the strength of *H* by drawing on multiple sources of information, this mode of reasoning is distinct from the idea of evidential robustness. The literature on this subject distinguishes several notions of robust reasoning pertaining to theorems, phenomena, modes of detection (Calcott, 2011; Levins, 1966, 1993; Orzack and Sober, 1993; Wimsatt, 2001), inferences, measurements, derivations, causal relationships (Woodward, 2006), parameter values, mathematical structures, representation frameworks (Weisberg & Reisman, 2008), computer models and simulations (Houkes & Vaesen, 2012; Lloyd, 2015; Muldoon, 2007; Parker, 2011). Although distinctions among these ideas are philosophically interesting, for the purposes of this paper I will focus on a general sense of robust evidential reasoning, which

<sup>1</sup> Also see Staley (2011, 2012).

<sup>2</sup> Staley also links this criterion to Campbell and Fiske's (1959) criterion of “discriminant validation.”

<sup>3</sup> For related historical accounts, see Fleming (1998) and Weart (2011).



**Fig. 1.** Multiple independent indicators of a changing global climate. Each line represents an independently derived estimate of change in the climate element. In each panel all data sets have been normalized to a common period of record (Hartmann et al., 2013, p. 199) ©Cambridge University Press.

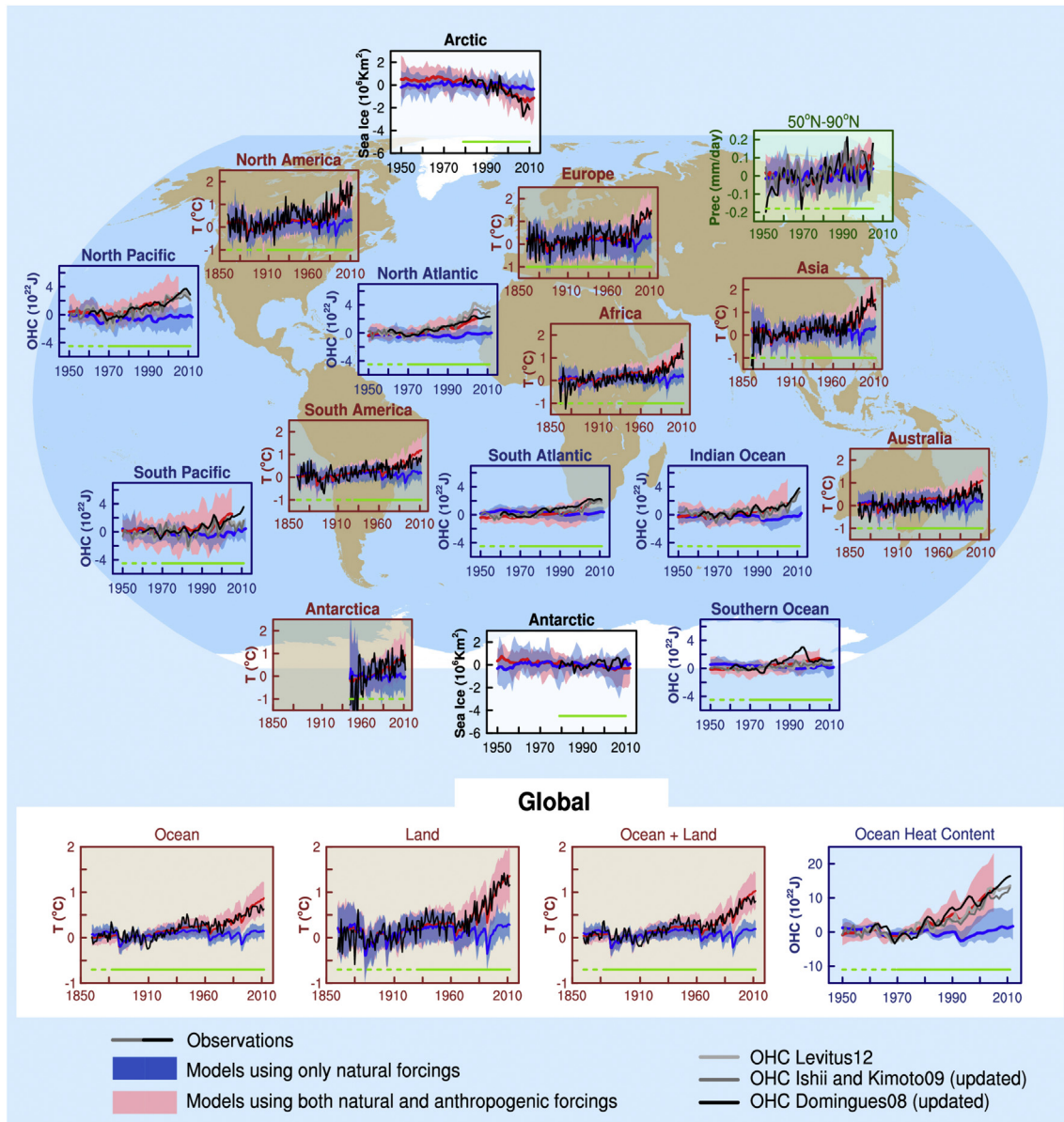
holds that multiple lines of evidence secure the plausibility of *H* by ensuring that the inference is insensitive to inaccuracies of a particular evidential pathway. The following case exemplifies how converging evidential pathways provide *H* with evidential strength and security, despite imperfections in particular lines of supporting evidence.

One of the earliest formulations of the thesis that humans are conducting a large scale ‘experiment’ on the global climate comes from the work of Guy Stewart Callendar who, in 1938, read a paper to the Royal Meteorological Society, arguing that CO<sub>2</sub> from fossil fuel consumption caused a measurable increase in the earth’s temperature. “The course of world temperatures during the next twenty years should,” he explained, “afford valuable evidence as to the accuracy of the calculated [warming] effect of atmospheric carbon dioxide” (Callendar, 1939, p. 236). Over a decade later, Gilbert Plass echoed this idea, writing that, “[i]f at the end of this century, measurements show that the carbon dioxide content of the atmosphere has risen appreciably and at the same time the temperature has continued to rise throughout the world, it will be firmly established that carbon dioxide is an important factor in causing climatic change” (Plass, 1956, p. 387). Perhaps the most

well-known statement of this thesis is that of Roger Revelle and Hans Suess:

[H]uman beings are now carrying out a large scale geophysical experiment of a kind that could not have happened in the past nor be reproduced in the future. Within a few centuries we are returning to the atmosphere and oceans the concentrated organic carbon stored in sedimentary rocks over hundreds of millions of years. This experiment, if adequately documented, may yield a far-reaching insight into the processes determining weather and climate (Revelle & Suess, 1957, pp. 19–20).<sup>3</sup>

The logic of climate change detection and attribution (D&A) studies involves a mode of induction that is more complex than simply correlating increasing anthropogenic greenhouse gas (GHG) emissions with increasing GMST. While detecting climate change involves a procedure of inferring that some climatic variable has changed in a defined statistical sense, attributing causal force to a particular variable involves a procedure of determining the components of a target system responsible for the detected change. D&A analyses investigate aspects of the climate over different scales



**Fig. 2.** Observations are shown on each panel in black or black and shades of grey. Blue shading is the model time series for natural forcing simulations and pink shading is the combined natural and anthropogenic forcings. The dark blue and dark red lines are the ensemble means from the model simulations\* (Bindoff et al., 2013, p. 930) ©Cambridge University Press. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of space (from a single grid point to the whole globe) and time (from minutes to millennia), using various methods to study the causes of climate change. Some focus on human activities that influence climatic variables such as surface air temperature, tropospheric temperature, depletion of stratospheric ozone, daily temperature extremes, precipitation patterns, upper ocean temperatures, glacier melting, ocean salinity, and ocean acidification and oxygen depletion.<sup>4</sup>

Attribution studies into the question of whether human activities are responsible for climate change employ computer modelling methods. An important modelling strategy is to examine whether a detected change is consistent with simulations (made by

skillful models) that predict a given response to anthropogenic forcing while being inconsistent with alternative (counterfactual) simulations that exclude this forcing. Accordingly, this approach has been applied to investigate the causes of GW (Fig. 2).

Although computer models and the simulations they produce are critical in scientific studies about climate change, even the most detailed, high-resolution models of the global climate omit important representations of major features of the planet that affect the climate and contain parameterizations that simplify complex climatic processes (McGuffie and Henderson-Sellers, 2014). While certain forms of scientific representations have unambiguous relations to reality, the extent of the representational relationship between numerical models and reality can be ambiguous. Richard Levins (1966) explains this point by comparing numerical models (in population biology) with traditional geographic maps. Map legends stipulate clear relations between the object

<sup>4</sup> For an overview of D&A studies investigating these and other variables, see Bindoff et al. (2013, pp. 932–939).

system (i.e., the markings on a page) and the target system (i.e., the geographical space it depicts). Continuity of lines on a map correspond to contiguity of physical features of reality; relative distances on a map indicate relative distances in space; colours and symbols indicate topographic features, and so on. Unlike a geographical map, the representational limits of mathematical models are often less clear. Whether a given climate model provides insight into questions about a target system at various scales or about its responses to various perturbations can be difficult to determine.

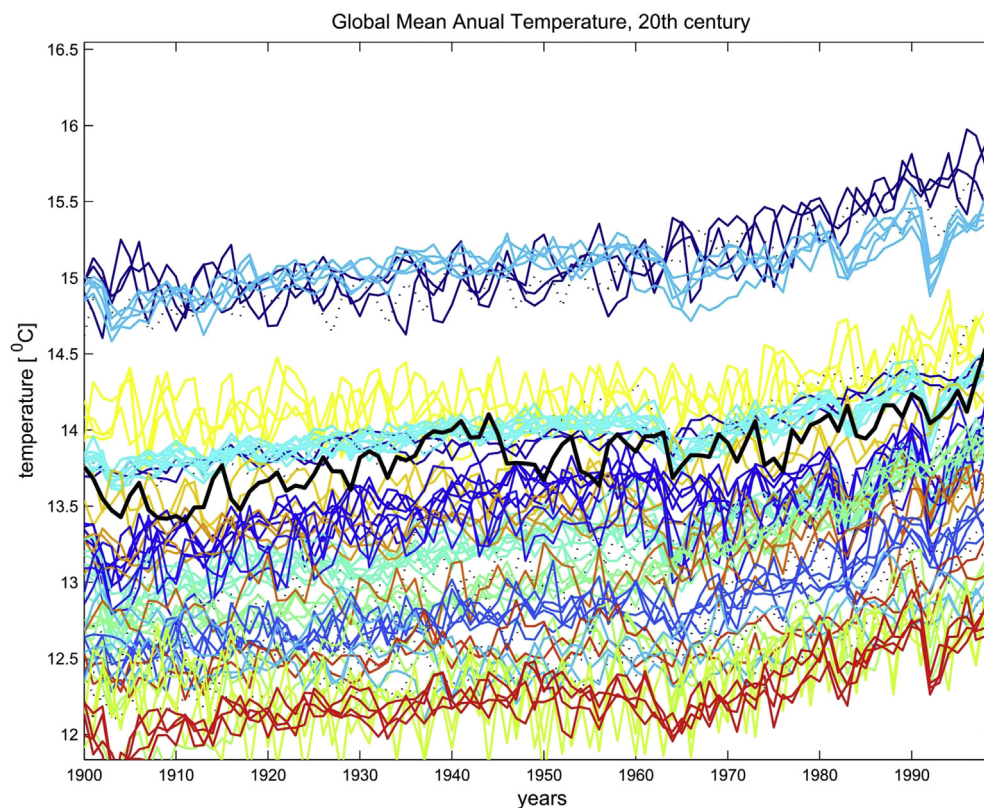
Roman Frigg, Smith, and Stainforth (2013, p. 893), for example, maintain that systematic errors in models of the Climate Model Intercomparison Project (CMIP) Phase 3 “lead to nontrivial macroscopic errors of simulation,” such that errors in the GMST “in a hindcast of the last century casts significant doubt on the viability of the informativeness assumption on a 25-km forecast to the end of this century,” particularly the downscaled regional projections of the 2009 United Kingdom Climate Projections (UKCP09) program. Fig. 3 illustrates the differences among the 24 models of the CMIP3 in GMST over the twentieth century. Whether errors in hindcasts of GMST cast doubt on the plausibility of certain regional projections, the spread of simulations and the difference between them and observed GMST raises the question of whether such models can be reliable for other purposes, such as that of attributing climate change to human activities. Should model results qualify as evidence of the causes of climate change even though they produce such a wide range of hindcasts with respect to absolute GMST change? What are the limits of how informative climate models can be about the world?

On the one hand, as indicated in Fig. 3, simulation estimates of absolute GMST change vary widely from one model to another, and between models and the instrumental record, which may limit the

informativeness of down-scaling general circulation models (GCMs) for some regional forecasts. In D&A studies, on the other hand, rather than focussing on absolute GMST change or regional projections, scientists typically concentrate on GMST anomalies with respect to the mean over some baseline period, comparing models programmed with natural and anthropogenic forcing to those that contain natural forcing only (Fig. 2). A variety of models that include anthropogenic forcing are better than those that exclude it with respect to reproducing observed estimates of climate change anomalies, which is a finding that constitutes evidence of a human influence on the target system.

The strength of variety-of-evidence reasoning and the security of robustness analysis are exemplified by the use of computer models in attribution studies supporting the anthropogenic global warming hypothesis (AGW)—that is, the proposition is that human activities caused more than half of the observed increase in GMST from 1951 to 2010 (Bindoff et al., 2013, p. 869). In testing AGW, scientists model the climate in various ways, using models of differing levels of complexity, including comparatively simple energy balance models, energy models of intermediate complexity, GCMs, and earth systems models that contain coupled interacting three-dimensional representations of different components of the climate system, integrating models of the atmosphere, the oceans, the carbon cycle, the cryosphere and the biosphere.

One can study these models and the simulations they produce individually and in climate model ensembles. Scientists use a range of methods to assess model-output, applying different statistical analyses to identify relationships between dependent and independent variables. The results of such modelling studies are also supported by a variety of other evidence, including observations of other changes in the atmosphere, the oceans and the cryosphere



**Fig. 3.** Output of Coupled Model Intercomparison Project Phase 3 models (24) simulating changes in annual global mean surface temperature over the twentieth century (coloured lines) and instrumental measurements (black line) (Frigg et al., 2013) ©Philosophy of Science Association. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Hegerl et al., 2007; Lloyd, 2009, 2010; Parker, 2010). AGW is a strong finding because multiple observations and modelling approaches of different components of the climate system are consistent with AGW and would be improbable if AGW were incorrect.

Given that each of the modelled worlds is very different from each other and from the earth with respect to their levels of complexity and the components of the climate represented, how do simulations inform scientists about the causes of global warming? The answer to this question rests in the logic of robustness analysis. In cases in which a set of various lines of evidence rely on very rough approximations of a target, if different pieces of evidence, such as those attained using different modelling methods, independently point to the same conclusion, the evidential security of the evidence indicating  $H$  can be maintained. By treating the same problem with several alternative modelling methods, scientists have established independent evidential pathways supporting AGW such that, even though each climate model is an imperfect representation of the target, each model is imperfect in a different way. “Hence,” as Levins writes, “our truth is the intersection of independent lies” (Levins, 1966, p. 423).

The convergence of multiple lines of evidence supporting AGW exemplifies the way in which robustness analysis compensates for the imperfections in models that constitute evidential pathways. Gareth Jones, Stott, and Christidis (2013), for example, investigate the causes of changes in near-surface temperatures from 1860 to 2010, utilizing the HadCRUT4 observational dataset and an ensemble of 8 coupled models from CMIP5, finding that “calculations of attributable temperature trends based on optimal detection support previous conclusions that the human-induced greenhouse gases dominate observed global warming since the mid-20th century” (p. 4001). Similarly, examining an ensemble of simulations from an energy balance model of intermediate complexity (Bern2.5D), Markus Huber and Reto Knutti (2012, p. 31) conclude that “since the mid-twentieth century, greenhouse gases contributed 0.85° of warming, ... suggesting an even higher confidence that human-induced causes dominate the observed warming.”

Further support for AGW emerges from studies of interconnected components of the climate that carry the signal of anthropogenic activities. Jan Lavšovička et al. (2006, p. 1253) find that “the anthropogenic emissions of greenhouse gases influence the atmosphere at nearly all altitudes,” including the upper atmosphere, which has cooled and contracted—a predicted result of greenhouse warming. Comparing a dataset of observed ocean temperatures with simulations from two climate models (PMC and HadCM3), Tsuga Barnett et al. (2005) identify a signal of anthropogenic forcing in the oceans: “A warming signal has penetrated into the world’s oceans over the past 40 years....[The signal] cannot be explained by natural internal climate variability or solar and volcanic forcing, but is well simulated by two anthropogenically forced climate models” (p. 284).

The evidential security that results from robustness analysis contrasts with what William Wimsatt (1981) calls “serial chain” reasoning. Serial reasoning involves drawing a conclusion by linking premises in a sequence that is only as strong as its weakest link. The problem with serial chain reasoning, both as a method of induction and as a description of scientific practice, is that, when a fault occurs in one line of evidence, any conclusion resting on it is undermined. Since reliable inferences about the natural world rely on many forms of inquiry, the fragile structure of serial reasoning is incongruent with scientific methodology. In contrast, robustness analysis involves connecting a conclusion to several independent lines of evidence, providing multiple pillars of support, the collection of which results in the insensitivity of  $H$  to alterations in some subset of the various evidential pathways.

Robustness analysis accounts for the evidential security of AGW because, even though imperfections exist in particular lines of evidence, including various modelling methods, alternative evidential pathways provide independent lines supporting the plausibility of this finding. Although each line of evidence contains imperfections, as long as these lines of evidence are imperfect in different ways, one line of imperfect evidence may compensate for the imperfections in another.

Examining several arguments for the epistemic significance of robustness arguments, Wendy Parker (2011) critiques the idea of evidential security in relation to long-term predictions using climate models, concluding that it cannot be inferred, via the arguments she considers, that robust reasoning secures a claim to have evidence for  $H$ . An evidence claim ( $E$ ) is a claim about the evidential status of some data as regards the plausibility of  $H$ . Particularly important for the present analysis is her assessment of Staley’s account, according to which the security of  $E$  is its degree of susceptibility to defeat from the failure of an auxiliary assumption of a given mode of inference (Staley, 2004, p. 468). Parker (2011, p. 596) presents the following generalized version of Staley’s argument:

1. A modeling result  $r_n$  enhances the security of an evidence claim  $E$  if
  - a)  $E$  is derivable from  $r_n$  in conjunction with a set of auxiliary assumptions,  $A_n$ , and
  - b)  $E$  is derivable from each of modeling results  $r_1 \dots r_{n-1}$ , respectively, in conjunction with sets of auxiliary assumptions  $A_1 \dots A_{n-1}$ , respectively, and
  - c)  $A_n$  is partially logically independent of each of  $A_1 \dots A_{n-1}$ .
2. 1a–1c are met in the present case.
  - ∴ The security of  $E$  is enhanced.

According to this argument, even if one method for inferring  $E$  turns out to involve a mistaken assumption (such as an oversimplification of a modelled process), the alternative assumptions involved in deriving other results provide “backup routes” to  $E$ , thus securing the inference  $H$ . Parker raises two important concerns with premises 1a and 1b. First, an individual climate model result often provides at best weak evidence for  $H$ , and only when multiple model results are taken together might they collectively constitute strong evidence. According to the above argument, however, the evidence for  $H$  would remain weak even though multiple models produce agreeing results because the claim that each model result supports is that there is weak evidence for  $H$ . This objection underscores the challenge of increasing the strength of an inference by appealing to the idea of robustness. However, one can still explain the increasing strength of evidential support for  $H$  in the case of climate modelling studies by appealing to the variety-of-evidence structure of reasoning described above (Section 2). That is to say, while each model result taken on its own may provide only weak positive support for  $H$ , since it would be unlikely that many independent model results would even weakly support  $H$  unless  $H$  were true, and such agreement does in fact occur, the multiple results taken together provide stronger support than each result taken alone. The attribution studies cited above comprise different evidential pathways supporting AGW because the sets of auxiliary assumptions in each study differ from one another in various ways. These studies include models with different structure, levels of complexity, components of the target system represented, and corresponding physical principles and empirical experiments. Even though each attribution study has its limitations, the agreement among them provides a stronger case for AGW than a given study on its own.

Parker's second concern is that the assumption that each climate model prediction has any positive evidential relevance is often unclear because of such considerations as the uncertainties regarding the roles of physical processes in the target system, technological limitations on model construction and computing power, the lack of opportunities to test model performance and, when comparisons can be made between model results and observations of the target system, problems of interpreting the significance of a fit between model and observed data. These epistemic concerns are important for many areas of climatic research, which scientists acknowledge as particularly salient in developing long-range climate projections. However, rather than being a general critique of the logic of robustness analysis, the extent of the problem of overcoming or working within these limitations depends on the context of each scientific investigation. In the context of AGW, for example, enough is known about the climate system to develop models for attribution studies. Scientists compare model output with a wide range of observed data, including those of paleoclimate reconstructions, which ground modelled approximations of the causes of climate change in a larger context of geophysical knowledge (IPCC, 2014).

#### 4. Conclusion

This account elaborates on the logic of robustness epistemology proposed by Levins to account for modelling methodology, incorporating the concepts of variety-of-evidence reasoning, and evidential strength and security, in an analysis of case studies of climate change. It borrows the ideas of evidential strength and security posited by Staley and applies them to climate studies that use computer modelling. Associating these ideas with varieties-of-evidence inferences and robustness analysis, this account expands on Lloyd's discussion of these topics, addressing the question of how climate scientists generate knowledge with the use of computer models, given the extent to which models approximate reality. Finally, it engages with some of Parker's concerns about the logic of robustness arguments, acknowledging the constraints on climate modelling methodology but defending this epistemological framework as a way to account for the success of important climatological inferences.

Studies of climate change have provided strong and secure support for GW and AGW. In these cases, the independence of different lines of evidence is due to different types of methodology, observations, models and datasets that scientists incorporate in their studies. The convergent lines of evidence supporting GW include a wide range of observations, such as measurements of changing surface atmospheric and oceanic temperatures, glaciers mass, snow cover, sea ice, sea level and atmospheric water vapour. Investigating the causes of climate change, scientists test AGW by using multiple computer modelling methods in conjunction with a variety of other lines of evidence, including observations of many components of the climate system. These various lines of evidence comprise strong support for AGW because they would be unlikely to agree if AGW were incorrect. Notwithstanding uncertainties in developing each line of evidence in climate modelling studies and despite differences between the actual climate system and computer model representations of it, models provide important sources of information when model results are robust to variations in the way they approximate the climate.

#### Acknowledgment

For helpful discussions and feedback on earlier drafts of this paper, the author thanks Wayne Myrvold, Gillian Barker, Eric Desjardins, Radoslav Dimitrov, Lisa Lloyd, Gordon McBean, Wendy

Parker, Stathis Psillos, Jamie Voogt, and the anonymous reviewers. This research was conducted with support from the Department of Philosophy and the Rotman Institute of Philosophy at Western University.

#### References

- Barnett, T., Zwiers, F., Hegerl, G., Allen, M., Crowley, T., Gillett, N., & ...Schnur, R. (2005). Detecting and attributing external influences on the climate system: A review of recent advances. *Journal of Climate*, 18(9), 1291–1314.
- Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., & ...Zhang, X. (2013). Detection and attribution of climate change: From global to regional. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, & ...P. M. Midgley (Eds.), *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge, United Kingdom and New York, NY, US: Cambridge University Press.
- Calcott, B. (2011, March). Wimsatt and the robustness family: Review of Wimsatt's re-engineering philosophy for limited beings. *Biology & Philosophy*, 26(2), 281–293.
- Callendar, G. S. (1939). The composition of the atmosphere through the ages. *Meteorological Magazine*, 74, 33–39.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81.
- Edwards, P. N. (2010, January). *A vast machine*. MIT Press.
- Fleming, J. R. (1998, August). *Historical perspectives on climate change*. Oxford University Press.
- Frigg, R., Smith, L. A., & Stainforth, D. A. (2013). The myopia of imperfect climate models: The case of UKCP09. *Philosophy of Science*, 80(5), 886–897.
- Hartmann, D., Tank, A. K., Rusticucci, M., Alexander, L., Bronnimann, S., Charabi, Y., & ...Zhai, P. (2013). Observations: Atmosphere and surface. In *Climate change: The physical science basis: Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge Univ Press.
- Hegerl, G. C., Crowley, T. J., Allen, M., Hyde, W. T., Pollack, H. N., Smerdon, J., & Zorita, E. (2007). Detection of human influence on a new, validated 1500-year temperature reconstruction. *Journal of Climate*, 20(4), 650–666.
- Houkes, W., & Vaesen, K. (2012, July). Robust! Handle with care. *Philosophy of Science*, 79(3), 345–364.
- Huber, M., & Knutti, R. (2012, January). Anthropogenic and natural warming inferred from changes in earth's energy balance. *Nature Geoscience*, 5(1), 31–36.
- IPCC. (2014). *Climate change synthesis report* (p. 132). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- Jones, G. S., Stott, P. A., & Christidis, N. (2013). Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10), 4001–4024.
- Katzav, J., & Parker, W. S. (2015, May). Introduction to assessing climate models: Knowledge, values and policy. *European Journal for Philosophy of Science*, 5(2), 141–148.
- Lavstovicka, J., Akmaev, R. A., Beig, G., Bremer, J., & Emmert, J. T. (2006). Global change in the upper atmosphere. *Science*, 314(5803), 1253–1254.
- Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54(4), 421–431.
- Levins, R. (1993, December). A response to Orzack and Sober: Formal analysis and the fluidity of science. *The Quarterly Review of Biology*, 68(4), 547–555. Article-Type: research-article/Full publication date: Dec., 1993/Copyright 1993 The University of Chicago Press.
- Lloyd, E. (2009). I—Varieties of support and confirmation of climate models. *Aristotelian Society Supplementary Volume*, 83, 213–232.
- Lloyd, E. (2010). Confirmation and robustness of climate models. *Philosophy of Science*, 77(5), 971–984.
- Lloyd, E. A. (2015, February). Model robustness as a confirmatory virtue: The case of climate science. *Studies in History and Philosophy of Science Part A*, 49, 58–68.
- Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. University of Chicago Press.
- McGuffie, K., & Henderson-Sellers, A. (2014, January). *The climate modelling Primer*. John Wiley & Sons.
- Muldoon, R. (2007, December). Robust simulations. *Philosophy of Science*, 74(5), 873–883.
- Norton, S. D., & Suppe, F. (2001). Why atmospheric modeling is good science. *Changing the atmosphere: Expert knowledge and environmental governance*, 67–105.
- Oreskes, N. (2007, September). The scientific consensus on climate change: How do we know we're not wrong?. In *Climate change: What it means for us, our children, and our grandchildren* (pp. 66–99) MIT Press. pp. 66–99.
- Orzack, S. H., & Sober, E. (1993, December). A critical assessment of Levins's the strategy of model building in population biology (1966). *The Quarterly Review of Biology*, 68(4), 533–546. Article-Type: research-article/Full publication date: Dec., 1993/Copyright 1993 The University of Chicago Press.
- Parker, W. S. (2010, December). Comparative process tracing and climate change fingerprints. *Philosophy of Science*, 77(5), 1083–1095. Article-Type: research-article/Full publication date: December 2010/Copyright 2010 Philosophy of Science Association.

- Parker, W. (2006). Understanding pluralism in climate modeling. *Foundations of Science*, 11(4), 349–368.
- Parker, W. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4), 579–600.
- Plass, G. (1956). Carbon dioxide and the climate. *American Scientist*, 44(3), 302–316.
- Revelle, R., & Suess, H. E. (1957). Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO<sub>2</sub> during the past decades. *Tellus*, 9(1), 18–27.
- Staley, K. W. (2004). Robust evidence and secure evidence claims. *Philosophy of Science*, 71(4), 467–488.
- Staley, K. W. (2011). *Philosophy of science matters: The philosophy of Peter Achinstein, chapter evidence and justification* (pp. 216–230). Oxford University Press.
- Staley, K. W. (2012). Strategies for securing evidence through model criticism. *European Journal for Philosophy of Science*, 2(1), 21–43.
- Weart, S. R. (2011, June). *Discovery of global warming (rev. and Expanded Ed.)*. Diane Publishing Company.
- Weisberg, M. (2006). Robustness analysis. *Philosophy of Science*, 73(5), 730–742.
- Weisberg, M., & Reisman, K. (2008). The robust Volterra principle. *Philosophy of Science*, 75(1), 106–131.
- Whewell, W. (1858). *Novum Organon Renovatum*. J.W. Parker and Son.
- Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination (2012). In *Characterizing the robustness of science* (pp. 61–87). Springer.
- Wimsatt, W. (2001). Richard levins as philosophical revolutionary. *Biology and Philosophy*, 16(1), 103–108.
- Woodward, J. (2006, June). Some varieties of robustness. *Journal of Economic Methodology*, 13(2), 219–240.