Confirmation and Robustness of Climate Models

Elisabeth A. Lloyd[†‡]

Recent philosophical attention to climate models has highlighted their weaknesses and uncertainties. Here I address the ways that models gain support through observational data. I review examples of model fit, variety of evidence, and independent support for aspects of the models, contrasting my analysis with that of other philosophers. I also investigate model robustness, which often emerges when comparing climate models simulating the same time period or set of conditions. Starting from Michael Weisberg's analysis of robustness, I conclude that his approach involves a version of reasoning from variety of evidence, enabling this robustness to be a confirmatory virtue

1. Introduction. There have been a few philosophical examinations of the particular problems facing climate model confirmation and evaluation (e.g., Edwards 1999, 2001; Petersen 2006; Parker 2009); this article, in contrast, outlines the underexplored strengths of global climate models. I start by reviewing a few basics of global climate modeling and then discuss several ways that global climate models gain evidential support, including model fit, variety of evidence, and independent support for aspects of the model. I also discuss the value of a certain kind of robustness.

I assume the approach of Ronald Giere, who spells out the useful

†To contact the author, please write to: History and Philosophy of Science Department, 130 Goodbody Hall, Indiana University, Bloomington, IN 47405; e-mail: ealloyd@indiana.edu.

‡I would like to thank climate researchers Caspar Ammann, William Collins, Jeffrey Kiehl, Doug Nychka, Kevin Trenberth, Tom Wigley, and especially Linda Mearns, of the National Center for Atmospheric Research, as well as Richard Somerville, of the Scripps Institute, for their assistance regarding climate models; all mistakes are of course my own. I also thank Kathryn Carter, Stephen Crowley, Brenden Fitelson, Mark Kaplan, Wendy Parker, Michael Weisberg, Sean Valles, and Eric Winsberg for their helpful comments.

Philosophy of Science, 77 (December 2010) pp. 971–984. 0031-8248/2010/7705-0036$10.00
Copyright 2010 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/657427 Published online by Cambridge University Press

relation of model to world under the semantic approach to models (or theories), in which the key claim is taken to be the theoretical hypothesis, which states "such-and-such identifiable real system is similar to a designated model in indicated respects and degrees" (1988, 81). Also important to Giere's approach is that "scientists use models to represent aspects of the world for specific purposes" (2004, 742). In his most recent work, Giere says, "Agents intend to use a model, M, to represent a part of the world, W, for some purpose, P" (2010, 269; Van Fraassen 2008, 309). Note that different parts of the model can be interpreted in different ways. For example, one aspect of the model may be interpreted more realistically or mechanistically, while another part may not have a corresponding structure or process in the real world at all. Taking this approach to models, we might consider a global climate model to represent aspects of the world for specific purposes, such as explaining the cooling of the stratosphere over the late twentieth century. We might then say that an instance of fit confirms a hypothesis about the similarity in certain respects and degrees of one element of a real system to part of this designated model. In what follows, I will use the shorthand of "models" being confirmed, instead of "theoretical hypotheses using the models" being confirmed, and will often leave off the purposes, which in the cases I am considering are often explanatory or theoretical (see Shackley 2001).[1]

There is no general theory of climate that takes all the complicating factors affecting climate into account and calculates what the effects on climate change will be on global temperature change, on precipitation, on wind, on pressure change, or on any other significant climate variable. So the climate scientists combine general pieces of theory from fluid dynamics, thermodynamics, and theories regarding radiation with detailed models of how the equations are applied to individual parts of the climate system. Details about the ice, vegetation, soil, and water vapor and the ornate interconnectedness of systems are represented in these global climate models. These sets of equations are analytically unsolvable and are instead approached through computer simulation models. Simulation modeling is thus basically an instance of theory articulation and application.

These huge general circulation models are presented in a three-dimensional spatial grid within which the model equations are solved. Processes at scales greater than this spatial resolution of the grid are represented directly in the equations, but processes smaller than this grid scale are represented only indirectly in the model, through parameters, in parts of models called "parameterizations." Examples of small-scale processes that

1. Wendy Parker is more interested in narrow predictive purposes in a recent discussion but also interprets the models in a mixed fashion (2009).

https://doi.org/10.1086/657427 Published online by Cambridge University Press

are represented by parameterizations include, for example, convection and cloud formation. The state variables of the dynamical systems in general circulation models for the atmosphere are usually velocity, pressure, temperature, and humidity. The dynamical equations consist of a set of balance equations and so on. The atmospheric, ocean, land, and ice systems have their own separate sets of equations and are linked together (Washington and Parkinson 2005).

Now let us consider how global climate models like this can be evaluated empirically. For some, the empirical satisfactoriness of models is demonstrated in terms of model "validation," and they claim that climate models have not been "validated." They invariably use vague and undefined notions of the term "validation," and worse, adherence to ideas like this has historically helped bring movement on climate policies to a standstill in the United States. Historian of science Naomi Oreskes noted many years ago that the problem with the notion of "validation" in modeling was that it seems a black-and-white issue, whereas the real issue is whether the model has *confirming evidence* to support its predictions and its assumptions—its general conformity with observational evidence—which is rather a matter of degree. But here the skeptics also have some major complaints. In a letter to Canadian Prime Minister Stephen Harper published in the *Financial Times* on April 16, 2006, which was signed by 66 scientists of various specialties, including several climate specialists, they wrote: "Observational evidence does not support today's computer climate models, so there is little reason to trust model predictions of the future."

This was especially significant because Canada had adopted policy changes on the basis of the reports of climate scientists and the 2001 report of the UN's Intergovernmental Panel on Climate Change, which relied partly on global climate model results. In any case, it is an example of the frequent concrete claim that the global models are inaccurate and cannot model present or past climate. It is true that there are still some uncertainties with global climate models, such as those with cloud presentation and sea ice. And the uncertainties multiply if the models are used for regional predictions. But it is generally untrue that the global models cannot represent present or past global climate, although that also depends on what purpose and precision are sought.

Philosophers writing on this issue have all emphasized the uncertainties involved in model building, application, and testing, as well as other perceived deficiencies of climate models. For example, Paul Edwards argues that climate models cannot be compared with observations in the usual way, and he demands a new philosophy of science: "The interdependent, even symbiotic, relationship between theory and observation in global climate science requires a different conception of the nature of

https://doi.org/10.1086/657427 Published online by Cambridge University Press

scientific work" (1999, 439). I address one of his primary concerns in section 4.

Or consider Arthur Petersen, who wrote a book detailing all the various sorts of uncertainties of general circulation models. He does offer two schemas for evaluating climate models, one statistical and one methodological, but he puts most emphasis on the great many weaknesses of the climate models that he lists, without acknowledging their strengths. He wryly notes at the beginning of his book that it could be used by climate skeptics to press their points home (2006, 13).

While not going so far as Petersen, Wendy Parker has emphasized the weaknesses of general circulation climate models, which she argues have not been confirmed for either general explanatory or specific predictive purposes (2009, 235). But hers is a black-and-white definition of confirmation and relates to applications of climate models to particular jobs at which they are weak, such as making fine-grained projections in the future or making regional predictions, rather than serving as theoretical understandings or explanations and thus is not decisive in our analysis.

Our discussion concerns, rather, the trustworthiness of models as representations of the basic processes and functions in the climate and the reliability of their representations of global phenomena in past, present, and future (see Randall et al. 2007). Since the theft of the UK climate e-mails from the University of East Anglia in 2009, there has been a real decrease in public trust in models, and there are many educated people—including well-placed politicians who have influence over future uses of models, as well as some philosophers—for whom climate models are a suspicious and unreliable means of understanding the world. Building understanding of and confidence in climate models is thus a fair and worthwhile balance to the critical philosophical approaches that have been taken so far. As Oreskes recommends, my approach to confirmation takes it as a matter of degree; models can accrue credit and trustworthiness upon being supported by empirical evidence as well as by theoretical derivation. In what follows, I offer a brief review of the various different types of evidence that modelers use to support their models and, in my analysis, to offer confirming evidence for them.

2. Model Fit. The most straightforward method of testing and confirming a model is to make a prediction using the model and then see whether the observations match the prediction. This is what I will call simply "model fit." It involves the theoretical hypothesis that the model fits the real world in certain respects and degrees, in particular, that the outcomes of the model fit the world to specified degrees, all for the specific purposes of the researcher, modeler, or agent, as Giere (2010) would put it. The simplest of these tests compare global mean temperature outcomes from

https://doi.org/10.1086/657427 Published online by Cambridge University Press

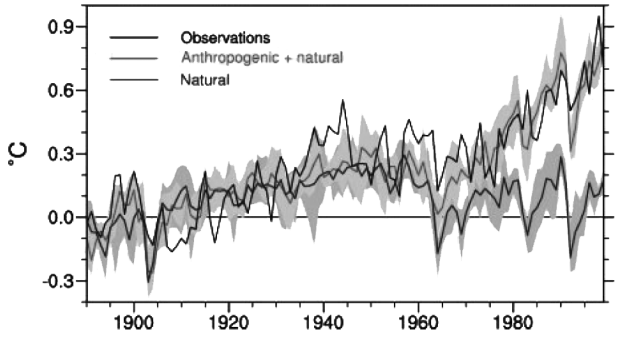


Figure 1. Source: Meehl et al. (2004).

the model against global mean temperature observations (Meehl et al. 2004; see fig. 1). The goals of the users of the model include the understanding and verification of the interactions among the causes of global warming over the twentieth century.

In figure 1, note the good fit of the in-between line of the model combining greenhouse gases and aerosol pollutants and natural causes with the most jagged line representing the observational record of actual global mean temperature; the contrast is with the pure natural causes, represented by the lowermost line. This kind of good fit has been repeated by many modeling groups, using different models but with the same basic ideas about the causes of climate changes (see also fig. 3).

It is sometimes objected that cases of fit like this are the result of model tuning (Edwards 1999; Parker 2009). Tuning involves the calibration or adjustment of the model parameters to produce a better fit with observations. This issue is especially important for a variable like mean global temperature, against which many models are adjusted during their building, so it is difficult to find an independent data set against which to test the model.

But take an especially nice example of model fit that cannot be tuned, in which Ben Santer et al. (2003; see fig. 2) simulated changes in the height of the tropopause, the transition or boundary in the atmosphere that separates the lower atmosphere from the upper atmosphere. The model results were a good match to the observed changes in tropopause height.

Looking at figure 2*A*, you can see the uppermost smooth line giving the model predictions for the situation with greenhouse gases, aerosols, and so on, and the jagged line giving the observations. There is a clear contrast between the observations and the lowermost line of figure 2*A*, which represents the predictions from natural forces alone. This success

https://doi.org/10.1086/657427 Published online by Cambridge University Press

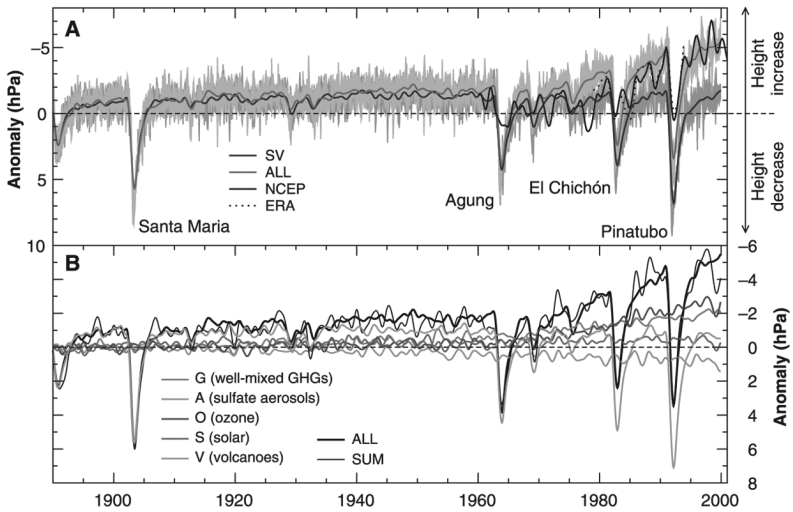


Figure 2. Changes in tropopause height: *A*, combined forcings; *B*, individual forcings. Source: Santer et al. (2003).

could not be the result of tuning since information about the change in the height of this boundary is not used in the development of these models.

3. Variety of Evidence. There are many more ways of testing model fit beyond testing whether the models can predict mean temperature accurately. Model fit or 'skill' is always measured by testing against data that were not used in model construction. It is important that general circulation models show significant skill in representing mean climate features such as large-scale distributions of the other variables of precipitation, radiation, wind, oceanic temperatures, and currents, in addition to global mean temperature (Randall et al. 2007, 600). Additionally, models can also simulate patterns of variability, a rather different sort of statistic than goodness of fit for an annual average. In this statistic, the model is compared to changes in the variable over the seasons or months.

The global models can simulate patterns of variability, such as the advance and retreat of major monsoon systems, seasonal shifts of temperatures, storm tracks, and rain belts (Randall et al. 2007, 600). Models can also be used to reproduce features of past climate and climate changes, such as the mid-Holocene warming of 6,000 years ago and the last glacial maximum of 21,000 years ago, both of which have specific spatial patterns across the globe. This success is taken to show that the forces represented

https://doi.org/10.1086/657427 Published online by Cambridge University Press

in models can handle values outside the ranges encountered recently, an important test of empirical adequacy for a global model.

The variety of evidence supporting these global models provides particularly important support against critiques or dismissals based on tuning. It may be relatively straightforward, according to critics, to tune the global model to maximize the successful fit of one variable, but when modelers attempt to adjust the model to account for another variable, or yet another, these huge and complicated models behave in unpredictable ways since the variables are interdependent and there are feedback processes among them. Thus, when a global model displays good fit in a variety of variables and features—beyond those couple that may have been tuned—this is especially good support for it.

This conclusion about variety of fit is supported by Branden Fitelson, who showed, using a Bayesian probabilistic framework, that two independent pieces of confirmatory evidence will provide stronger confirmation than either one of them provides individually (2001, S131).[2] To say that the confirmatory evidence is independent means simply that the degree to which the first evidence, or instance of fit, supports the model does not depend on whether the second instance of fit has already occurred.[3]

Thus, for example, two instances of fit of distinct variables of a general circulation model using distinct data sets considered collectively will provide stronger evidence for a model than either one of the instances considered individually. We have fulfilled these conditions, for example, when the ocean heat variable is tested against an ocean temperature observational data set and the atmospheric pressure variable at given locations is tested against the observed pressures. Thus, a model with many instances of fit is much better supported and has a higher probability under a preferred confirmation function than a model with only one or two instances. Variety of evidence is thus an important source of confirmation for the global climate models. In addition, it helps rebuff accusations of overdependence on model tuning. In section 5, we will see how model robustness is also related to variety of evidence.

4. Independent Support for Aspects of the Models. In addition to fit and

2. John Earman, with Grover Maxwell, also proved that variety of evidence provides more support for a hypothesis than does a narrow range of evidence (Earman 1992, 78–79). They also rely on the independence of the distinct trials.

3. That is, given pieces of evidence E1 and E2 and hypothesis H, E1 and E2 are mutually confirmationally independent regarding H according to c, if and only if both $c(H, E1 \mid E2) = c(H, E1)$ and $c(H, E2 \mid E1) = c(H, E2)$, where c is a confirmation function, such as a likelihood ratio measure (Fitelson 2001, S125).

https://doi.org/10.1086/657427 Published online by Cambridge University Press

a variety of cases of fit, there is also a different type of evidence supporting these global models: the direct independent support of aspects and assumptions of the model, such as parameter values and parameterizations. The simplest form of this type of support is when a parameter is measured and then fed into the model. Take the representation of water vapor in climate models. The U.S. Department of Energy developed the Atmospheric Radiation Measurement program, which uses highly instrumented measuring sites at land and ocean locations to improve understanding and representation of water vapor, clouds, and radiation. Data from these instruments and from satellites from a similar project have recently confirmed that the water vapor estimates used in the models were in the right range, thus providing independent observational support for the parameter values used in the models (Ramanathan and Inamdar 2006; Randall et al. 2007, 632).

Things become more complicated with parameterizations. Take the representation of clouds. Uncertainties arise because the physics surrounding precipitation and clouds is not fully understood. But cloud parameterizations for some models are supported by special research groups who study clouds. The scientists involved use empirical data to create small-scale models representing individual cloud elements, from which they develop improved cloud parameterizations for the general circulation models. This sort of independent empirical support for the various cloud parameters in the model has improved model performance (Randall et al. 2003, 456). This type of independent confirmation of aspects of the global models—empirical measurement of parameters or their ranges, or the supplying of parameterizations based on measurements and modeling of smaller or subgrid processes—is a particularly strong form of confirmation to complement cases of fit since it anchors the models in the world completely independently from their success in prediction.

The biggest assumption in philosophical discussion of the issue of independent support of global climate models is that derivation from theory is better than empirical support of these parts of models (Edwards 1999, 2001; Petersen 2006). The mathematical core of the model, on this view, ought not to be determined or affected by empirical data. This is a particularly problematic position when it comes to parameterizations. Philosopher Paul Edwards objects to the inclusion of what he sees as "data-laden" parameterizations into the models (such as the cloud parameterizations mentioned above), on the basis that it violates the "reductionist imperative of the physical sciences." That is, "physical science practice normally attempts to explain large-scale phenomena as an outcome of smaller-scale processes" (2001, 59). But the incorporation of parameterizations that involve empirical data in the simulations violates this reductionist mandate. It seems that Edwards thinks that climate sci-

https://doi.org/10.1086/657427 Published online by Cambridge University Press

ence must be reductionist in order to meet the standards of the physical sciences, although he never defends this supposition or explains how the earth sciences fit in his vision. Climate science, like ecology, is not necessarily restricted to deductions from higher laws. It is, rather, constituted by incomplete general theory plus the collection of models that themselves articulate the theories for these sciences.

Thus, to offer support for a climate model, there is no necessity to derive its parameterizations or parameters from a higher law. On my view, independent empirical evidence for assumptions and aspects of a model in climate science also add reasons to believe that the model is empirically adequate or realistic, which goes beyond the predictive success of the model's outcome.

5. Robustness.

5.1. Robustness and Multiple Models. Comparison of multiple models and their results performing specific experiments or trials is one of the key methods of model evaluation in climate science. Models from different modeling groups are run over the same time period or performing the same task, and the results are compared. These experiments frequently produce robust results. For example, take the fact that all available climate models produce an increase in temperature in the late twentieth century. In figure 3, we see the results of 14 models doing 58 simulations of twentieth-century warming. The smoother line is the mean of the models, and the jagged line represents the observational data.

Climate scientist Jeffrey Kiehl, in discussing the fact that all models simulate the global warming during the twentieth century with some accuracy, notes: "This is viewed as a reassuring confirmation that models to first order capture the behavior of the physical climate system and lends credence to applying the models to projecting future climates" (2007, 1). The language of "capture" appears also when Steve Lambert and George Boer discuss robust model results. When models vary only a small amount among themselves, this "supports the assumption that they are capturing the processes that govern that variable and hence its climate" (Lambert and Boer 2001, 88). But if models disagree, this indicates model "unreliability." In addition, if the mean model average disagrees with the mean observation, it suggests a systematic deficiency in the models.

There is at least one possible constraint about reasoning from robustness to the reality or "capture" of the physical processes or causes represented in models, and that has to do with the sources of the models themselves. The problem is that consistency among models may reflect a social convergence process among the institutions building models (Edwards 1999). In other words, models may agree because modelers may

https://doi.org/10.1086/657427 Published online by Cambridge University Press

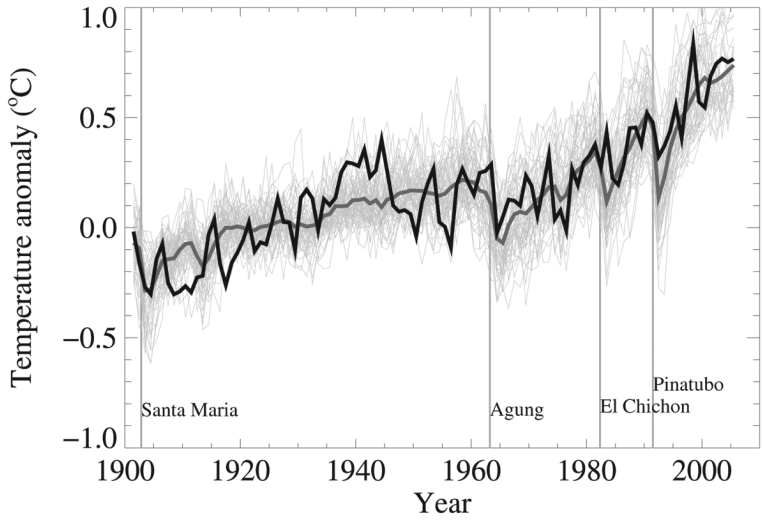


Figure 3. Global mean surface temperatures over the twentieth century from observations (*jagged black line*) and from 58 simulations produced by 14 distinct climate models that include human-caused and natural factors in climate change. Mean model is represented by the dark gray line. Source: Randall et al. (2007), 600.

share model-building conventions and practices. Nevertheless, if the models have relatively independent sources, the existence of robustness of outcomes is sometimes seen as evidence in favor of the realism or representation of specific causal processes that the models have in common.

5.2. Robustness Analysis and Variety of Evidence. Michael Weisberg's recent account of robustness in models is helpful here. His account focuses on some common outcome of the models, the "robust property," and what the models have in common to produce that outcome, the "core" structure. On his robustness analysis, the whole theorem has the form: "Ceteris paribus, if [common core (causal) structure] obtains, then [robust property] will obtain" (2006, 731).

Taking our group of 14 climate models that give global mean temperature outcomes that approximately agree, we could construct the statement: "Ceteris paribus, if [Greenhouse gases relate in lawlike interaction with the energy budget of the earth] obtains, then [increased global mean temperature] will obtain." In other words, increases in greenhouse gases are claimed to be the key causal factor in increasing the global mean

https://doi.org/10.1086/657427 Published online by Cambridge University Press

temperature, in the context of lawlike interaction with the energy budget of the earth. This is due to their representation in the core structure common to all the models. But how can we be sure that greenhouse gases are the relevant cause?

Weisberg then makes an implicit appeal to the variety of evidence. In order to infer to causes in the real world, he writes, "The key comes in ensuring that a sufficiently heterogeneous set of situations is covered in the set of models subjected to robustness analysis." If a sufficiently heterogeneous set of models for a phenomenon has the common structure, he continues, "then it is very likely that the real-world phenomenon has a corresponding causal structure." Moreover, this would allow us to infer that when we saw the robust property in a real system, "it is likely that the core structure is present, and that it is giving rise to the property" (2006, 739).

In my view, Weisberg is appealing to a variety of evidence argument here because he is explicitly appealing to a range of instances of fit of the model over different parameter values, parameter spaces, or laws.[4] It is against this background of differing model constructions that the core structure occurs and causes the robust property to appear, and it is the degree of this variety of fit for which the model has been verified that determines how confident we should be in the causal connection. Weisberg deems this to be a part of robustness analysis, but it is a step beyond the usual robustness analysis, which involves only inferences about the robust property and not about the model(s) that generated that robust property (e.g., Staley 2004; Woodward 2006). As such, it fits more naturally as a subtype of variety of evidence inferences, to which we can apply the probabilistic results demonstrating its confirmatory value.

So, continuing our application of Weisberg's analysis to our case, we do find that the models covered a wide range of assumptions and conditions, and they all have this common structure of greenhouse gas causation. Hence, on his analysis, it is very likely that the real world phenomenon has a corresponding causal structure (2006, 739; Muldoon 2007,

4. In Weisberg's recent paper with Ken Reisman (2008), they refer to a part of this range of variety as "parameter robustness analysis," in which the parameter value is varied across a given range, also known as "sensitivity analysis." The variation of laws is called "structural robustness analysis." It is possible that Weisberg does not intend that the variety of backgrounds be interpreted as a variety of empirical fit but rather only formally. If so, I would like to extend his analysis to include an empirical interpretation of the background variety.

https://doi.org/10.1086/657427 Published online by Cambridge University Press

882).[5] Therefore, we could infer that greenhouse gas concentration increases cause global warming in the real world, as the attribution studies have also shown (Hegerl et al. 2007).[6]

I would like to draw attention to how this relationship between robustness and variety of evidence helps give robustness a role in confirmation of models. Under other interpretations of robustness, philosophers have found that while it reveals important relationships among models, it does not actually confirm those models. Robustness may, for example, be used to eliminate accidents in measurement (Staley 2004), but it is not perceived as a positive confirmatory virtue. In our case, however, the core structure is compared across a variety of assumptions or model backgrounds, thus providing a variety of evidence for the core structure causing or being correlated with the model outcome. And since a variety of evidence does ordinarily give us reason to increase the degree of confirmation of a model, it does in this case as well. Thus, the type of robustness we have here embodies a confirmatory virtue (cf. Staley 2004; Woodward 2006).[7]

6. Conclusion. I have reviewed three fundamental ways that climate models are confirmed: fit, variety of evidence, and independent support for aspects of the models. I have also considered robustness. Climate models should not be judged primarily or solely on the basis of what they are weak at; if we approached other scientific theories or models this way, we would never accept any of them. While other philosophers looking at the climate models have emphasized their weaknesses and problems, I think it is vitally important to explore and understand the models' fundamental strengths. When we understand the relationships between evidence and climate models properly, global climate models appear to be much better supported than previously considered.

5. On Jim Woodward's categorization, inferring causal relationships requires manipulation and experimentation, which are generally impossible for climate, although possible for climate models (2006, 235). Our inference is also different from Woodward's "inferential robustness" since our focus is on the model structure, not the robust property (230).

6. I would emphasize that this finding depends very much on the empirical adequacy or fit of each individual model; their collection together instantiating a variety of evidence is doing the additional confirmatory work here. For worries about tuning, see Lloyd (2009).

7. Weisberg moves toward this confirmatory virtue, to the extent that he requires that the models possess "low-level confirmation," as well as the array of background variability that provides the conditions for variety of evidence (2006, 741).

https://doi.org/10.1086/657427 Published online by Cambridge University Press

REFERENCES

Earman, John. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

Edwards, Paul N. 1999. "Global Climate Science, Uncertainty and Politics: Data-Laden Models, Model-Filtered Data." *Science as Culture* 8 (4): 437–72.

———. 2001. "Representing the Global Atmosphere: Computer Models, Data and Knowledge about Climate Change." In *Changing the Atmosphere: Expert Knowledge and Environmental Governance*, ed. Clark A. Miller and Paul N. Edwards, 31–65. Cambridge, MA: MIT Press.

Fitelson, Branden. 2001. "A Bayesian Account of Independent Evidence with Applications." *Philosophy of Science* 68 (Proceedings): S123–S140.

Giere, Ronald N. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

———. 2004. "How Models Are Used to Represent Reality." *Philosophy of Science* 71:742–52.

———. 2010. "An Agent-Based Conception of Models and Scientific Representation." *Synthese* 172 (2): 269–81.

Hegerl, Gabriele C., Francis W. Zwiers, et al. 2007. "Understanding and Attributing Climate Change." In *Climate Change 2007: The Physical Science Basis; Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, ed. Susan Solomon, Dahe Qin, Martin Manning, et al., 663–745. New York: Cambridge University Press.

Kiehl, Jeffrey T. 2007. "Twentieth Century Climate Model Response and Climate Sensitivity." *Geophysical Research Letters* 34:L22710.

Lambert, Steven J., and George J. Boer. 2001. "CMIP1 Evaluation and Intercomparison of Coupled Climate Models." *Climate Dynamics* 17:83–106.

Lloyd, Elisabeth A. 2009. "Varieties of Support and Confirmation of Climate Models." *Proceedings of the Aristotelian Society* 83 (Suppl.): 217–36.

Meehl, Gerald A., Warren M. Washington, Caspar M. Ammann, Julie M. Arblaster, T. M. L. Wigley, and Claudia Tebaldi. 2004. "Combinations of Natural and Anthropogenic Forcings in Twentieth-Century Climate." *Journal of Climate* 17:3721–27.

Muldoon, Ryan. 2007. "Robust Simulations." *Philosophy of Science* 74:873–83.

Parker, Wendy. 2009. "Confirmation and Adequacy-for-Purpose in Climate Modeling." *Proceedings of the Aristotelian Society* 83 (Suppl.): 233–49.

Petersen, Arthur C. 2006. *Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice*. Amsterdam: Het Spinhuis.

Ramanathan, Veerabhadran, and Anand Inamdar. 2006. "The Radiative Forcing due to Clouds and Water Vapor." In *Frontiers of Climate Modeling*, ed. Jeffrey T. Kiehl and Veerabhadran Ramanathan, 119–51. New York: Cambridge University Press.

Randall, David A., Steven Krueger, et al. 2003. "Confronting Models with Data: The GEWEX Cloud Systems Study." *Bulletin of the American Meteorological Society* 84: 455–69.

Randall, David A., Richard A. Wood, et al. 2007. "Climate Models and Their Evaluation." In *Climate Change 2007: The Physical Science Basis; Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, ed. Susan Solomon, Dahe Qin, Martin Manning, et al., 589–662. New York: Cambridge University Press.

Santer, Ben D., Michael F. Wehner, et al. 2003. "Contributions of Anthropogenic and Natural Forcing to Recent Tropopause Height Changes." *Science* 310 (5632): 479–83.

Shackley, Simon. 2001. "Epistemic Lifestyles in Climate Change Modeling." In *Changing the Atmosphere: Expert Knowledge and Environmental Governance*, ed. Clark A. Miller and Paul N. Edwards, 107–33. Cambridge, MA: MIT Press.

Staley, Kent W. 2004. "Robust Evidence and Secure Evidence Claims." *Philosophy of Science* 71:467–88.

https://doi.org/10.1086/657427 Published online by Cambridge University Press

Van Fraassen, Bas C. 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford: Clarendon.
Washington, Warren M., and Claire L. Parkinson. 2005. *Introduction to Three-Dimensional Climate Modeling*. New York: University Science.
Weisberg, Michael. 2006. "Robustness Analysis." *Philosophy of Science* 73:730–42.
Weisberg, Michael, and Kenneth Reisman. 2008. "The Robust Volterra Principle." *Philosophy of Science* 75 (1): 106–31.
Woodward, Jim. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13 (2): 219–40.

https://doi.org/10.1086/657427 Published online by Cambridge University Press