



Can we trust climate models?

J. C. Hargreaves* and J. D. Annan

What are the predictions of climate models, should we believe them, and are they falsifiable? Probably the most iconic and influential result arising from climate models is the prediction that, dependent on the rate of increase of CO₂ emissions, global and annual mean temperature will rise by around 2–4°C over the 21st century. We argue that this result is indeed credible, as are the supplementary predictions that the land will on average warm by around 50% more than the oceans, high latitudes more than the tropics, and that the hydrological cycle will generally intensify. Beyond these and similar broad statements, however, we presently find little evidence of trustworthy predictions at fine spatial scale and annual to decadal timescale from climate models. © 2014 The Authors. *WIREs Climate Change* published by John Wiley & Sons, Ltd.

How to cite this article:

WIREs Clim Change 2014, 5:435–440. doi: 10.1002/wcc.288

INTRODUCTION

Climate models are the primary tool by which we create knowledge about the future impact of human activities on the global climate system. Possibly the most iconic and influential result arising from climate models is the prediction that, dependent on the rate of increase of CO₂ emissions, global and annual mean temperature will (with high probability) rise by around 2–4°C over the 21st century (see Ref 1, Fig SPM.5). Supplementary predictions of a similar status are that the land will on average warm by around 50% more than the oceans, high latitudes more than the tropics, and nights more than days. Besides surface temperature changes, the hydrological cycle is expected to generally intensify by a few percentage points per degree Celsius of warming, and the stratosphere will cool. These results are broadly agreed upon by all global climate models (GCMs) which have contributed to the Climate Model

Inter-comparison Project (CMIP) experiments² over the past several decades, although the magnitudes of the expected changes remain somewhat uncertain. Many more detailed predictions can, in principle, be made, for example on a regional basis, but uncertainty tends to increase substantially as the spatial scale decreases. For example, even the sign of the change in precipitation is uncertain in many areas over the coming decades.

But should we believe any or all of these predictions? And if so, which ones, and why? These are the fundamental questions which we hope to address in this article. We start by exploring the origins of the models and considering the nature of the knowledge that they impart. We overview the strengths and weaknesses of the models and then consider to what extent these models may be falsifiable or considered trustworthy.

THE ORIGINS OF CLIMATE MODELS

Scientific models are simplifications of nature built to gain better understanding of how nature works. They are of course, trivially false in the sense that all models are approximations to reality. Climate models are no different in this sense, and numerous

*Correspondence to: jules@blueskiesresearch.org.uk

Uncertainty Estimation Research Group, RIGC/JAMSTEC, Yokohama, Japan

Conflict of interest: The authors have declared no conflicts of interest for this article.

limitations are immediately apparent under any close examination. Therefore, we approach the question in a more general sense, considering the ensemble of models as a whole: Are the models sufficiently wrong that we should anticipate reality falling outside the range of model results?

While the underlying motivation was for improved weather forecasting, a primary goal of atmospheric research in the 1940s and 1950s was to understand the general circulation of the atmosphere,³ and the first numerical model of the atmosphere was built for that purpose.⁴ With the addition of representations of the thermal structure and hydrology, the first true climate models were born.⁵ It was not long until those models were being used to make predictions of temperature change caused by increasing CO₂ in the atmosphere.⁶ Such models were further developed and used as a basis for the original 1.5–4.5°C estimate of climate sensitivity in 1979.⁷ This estimate, of the equilibrium global temperature change under a doubling of CO₂, has been repeatedly re-confirmed by both newer climate models, and a wide range of observational studies (see Ref 8 and references therein).

As model development has continued over the decades the model resolution and the range of processes modeled has increased considerably. Many models now include representations of the carbon cycle and atmospheric chemistry, and may be referred to as earth system models (ESMs) rather than General Circulation or Global Climate Models. Thanks to increased computer power, both the complexity of the parameterizations and the spatial resolution of the models have increased. However, as new components have been added, the fundamental physical responses of the modeled climate system have remained consistent with the early simpler models, and the coupling process has not uncovered any major errors in the pre-existing models, which are encouraging results in themselves.

Thus, the models started out as tools for helping us understand the general features of the atmosphere climate of the Earth, but they are now expected to predict future changes over the whole earth system at a variety of temporal and spatial scales. The actual level of performance of today's models across the range of spatial and temporal scales is not clear *prima facie*, so bears further scrutiny.

INTERPRETATION AND EVALUATION OF THE ENSEMBLE

The interpretation of climate model consensus has, over recent years, become the focus of increasing

attention. It cannot be argued on a rigorous basis that climate model agreement necessarily implies correctness,⁹ largely because of the *ad hoc* origins of the ensemble members, and unclear characterization of their inter-relationships. Models have shared code and ideas according to their origins,¹⁰ and therefore rather than considering them as independent sources of evidence concerning the climate system, it may be more realistic (albeit still perhaps optimistic) to interpret the ensemble collectively as representing (at least approximately) our range of beliefs and uncertainties regarding the behavior of the climate system.¹¹

As simulators of the present-day climate, the models have considerable success. As well as elucidating many of the organizing principles of the large-scale circulation, models reproduce many details of the behavior of the climate system, including the large-scale temperature and precipitation patterns, regular annual and daily cycles, and quasi-periodic internal variability such as El Niño, with a performance which has steadily increased over time.^{12,13} It is arguable that some of this fit to observations could be due merely to (over-)tuning,¹⁴ which could mislead as to model performance in predicting of future climate change. However, tuning such complex models to fit multiple criteria is a rather *ad hoc* and difficult process, and therefore it seems reasonable to conclude that the models have also improved in a more fundamental sense, more closely resembling the climate system generally. This is particularly the case when models are found to represent, or predict, phenomena that were not previously recognized, as discussed in some detail in Ref 15.

Models built on the same principles as climate models have decades of strong results in numerical weather prediction, which have also improved over time,¹⁶ with model improvement being a significant factor in this. While there are strong arguments that climate model ensembles cannot provide probabilistically perfect predictions,¹⁷ this is a very demanding level of performance. Here we consider a more appropriate goal to be merely that the model consensus (as outlined previously) does not mislead us.

An ability to simulate features of the current climate, while encouraging, does not imply that the models can accurately predict future climate changes, such as those expected under increasing levels of carbon dioxide. The broad scale predictions, which are robust within the model ensemble, are, however, also supported by physical understanding. As one example of this, polar amplification has long been anticipated¹⁸ and can be explained as largely due to a combination of albedo and water vapor feedbacks, although

research continues into the finer details.¹⁹ Our understanding of anthropogenically forced climate change is not based wholly on complex, incomprehensible, and possibly unverifiable computer models. Rather, the models provide one strand of support for (and quantification of) effects that have a broader underpinning.

To gain more insight on the performance of the predictions from current models, it is useful to consider information from earlier generations. In 1984, James Hansen appeared before a US Senate Committee and provided a forecast of continuing warming on the global scale (later published in Ref 20). This forecast has turned out to show significant skill, although the limited archiving of output has precluded a detailed analysis.²¹ More recent models appear to be doing slightly better,²² but it must be noted that opportunities for true validation of past decadal and longer forecasts are extremely limited, such that it is hard to draw general conclusions about climate model performance. Looking further back in time, paleoclimate simulations provide an alternative out of sample test of the models, since these simulations are invariably considered a low priority at the institutes where state of the art climate models are developed, and are only attempted after the conclusion of a phase of model development. As part of the most recent CMIP5 experiment, simulations of the Last Glacial Maximum (LGM) provide the opportunity to test the models' ability to represent quasi-equilibrium changes in the climate system due to large external forcings. Comparison with proxy data provides strong support for the models qualitatively and quantitatively reproducing the changes on the broadest scales, including the spatial patterns of temperature highlighted at the start of this article. On the regional scale there are, however, substantial disagreements in magnitude and pattern of temperature anomalies both between models and data, and also within the model ensemble.²³ Therefore, we cannot expect precise predictions from current climate models.

In fact, models are very far from being perfect. They struggle to generate robust simulations of recent climate changes on regional scales, even when run at the highest resolutions available.²⁴ There are numerous reasons for this, including both a reduction in signal-to-noise ratio and errors in the representation of the physics that can easily lead to displacement in the position of features of the climate system. One hope for the future is in the development of methods that attempt to correct such position errors.²⁵ Recently, a concerted effort has been put into developing prediction systems aiming at the decadal timescales. This timescale falls awkwardly between the

short-term initial value problem of weather prediction (up to seasonal duration) where strong results have been obtained, and estimating the long-term response to external forcing. A recent analysis suggests that dynamical forecasts based on climate models perform clearly worse than empirical methods.²⁶ It seems that genuinely useful climate forecasting on the multianual to decadal timescale may be still some way away at this time.^{27,28} Thus it is clear that the models can currently only be relied upon for a broad picture of future climate changes.

FALSIFIABILITY OF CLIMATE KNOWLEDGE

One fundamental requirement for a hypothesis to be considered scientifically valid is that it is in principle amenable to falsifiability. The hypotheses arising from model consensus (described above) are trivially falsifiable in principle, by the process of waiting for 100 years and observing the resulting climate changes. If anthropogenic emissions were to be very different from the assumed scenarios, then it might be necessary to re-run the models with appropriate forcing, but this is a technical detail. Far more problematic, is that we are unwilling to wait 100 years before learning about climate models, and cannot wait before making today's decisions. It might not take as long as 100 years, and indeed recent evidence does hint at the models modestly overestimating the rate of climate change,²⁹ but there is certainly not yet sufficient evidence to overturn the major paradigms of today's models. Therefore, it could be argued that predictions of long-term climate change are in a practical sense unfalsifiable. However, an alternative route to challenging the underlying assumptions of the models would be to create competing climate models based on different hypotheses, which reproduce existing phenomena with acceptable skill, but generate substantively different behavior in the future. It is therefore interesting to consider why there are no such models. There are two possible, mutually exclusive answers. Either models agree because their consensus represents a reality that any plausible depiction of the climate system will exhibit, or alternatively, no-one has created a different model because there has been insufficient pursuit of alternative hypothesis regarding the climate system.

Generally in scientific research, models based on alternative hypotheses and theories exist wherever there is sufficient uncertainty for them not to be ruled out. In climate models, the details of the parameterizations in the model code may vary, but the principles of fluid dynamics, thermodynamics, and radiation that

lead to the primary results of global warming under increasing atmospheric carbon dioxide are common to all climate models. Conversely, if we step outside the purely physical realm and consider biological components of the earth system (which are now increasingly incorporated into climate models), the underlying processes operating are less clear and there is not as much consensus concerning the underlying principles. Here it is possible to find a wider range of hypotheses underpinning the models, and new models based on alternative hypotheses are under development.³⁰ There is also, within climate science, a strong incentive on the individual scientists to produce novelties in their representations of the climate system. Any improved formulation may be widely cited and adopted, and will bear the researcher's name for years to come. Furthermore, there is nothing to prevent researchers from constructing and developing radically new formulations of the climate system, especially within the university system. Such models could first be implemented as rather simple models, and only grow in complexity as they demonstrate success. Indeed there does exist a wide range of models of intermediate complexity (called EMICs), but none of these are based on substantively different hypotheses of the underlying processes. The lack of credible alternative models is, in our opinion, evidence that such models are not sufficiently successful for them to have progressed. Some researchers have attempted to generate a range of results by varying parameter values within a GCM.^{31–33} These experiments can be viewed as an attempt to search for a wider range of behaviors than previously exhibited. While some extreme behavior can be found, these models are generally only subjected to rather rudimentary tests of performance, far less intensive than the suite of experiments that is standard for a GCM. More detailed analysis may reveal more substantial problems with their behavior.³⁴ Furthermore, a large majority of all these alternative models (and typically the best performers out of the candidate ensemble) generate behaviors that are highly compatible with existing knowledge. It is debatable whether these models with different parameter values can sample as broad a range of uncertainty as is already achieved by the GCM ensemble due to its structural diversity,³⁵ but this would seem to lend weight to the argument that the main results of the GCM ensemble are indeed robust.

Nevertheless, one could argue that some combination of social pressure, and convenience, results in models sharing too much theory and even code, to the extent that they are little more than replicates. The validity of these competing arguments can hardly be decided on the basis of rhetoric, but there is yet

little progress on how they can be assessed by analysis of the ensemble or other methods. Thus we consider this to be a particularly important area for future research. If it is the case that the pressure to conform in climate science has led to a serious disruption of the scientific process, then attention should indeed be focused toward building alternative models based on fundamentally contrasting physical hypotheses that perform equally well or better for the modern and past climates. However, proponents of such an exercise should note that while encouraging diversity in model design seems a laudable goal, alternative ideas cannot easily be conjured up from nothing, but are instead typically provoked by failures of the existing paradigm.³⁶

Despite understanding the basic processes underlying the physics of the climate system, it is clear that the state-of-the-art climate models are not 'good enough', if we desire high resolution predictions with high temporal and spatial resolutions over coming decades. Thus far we seem to have only built sufficient confidence in the broad scale response of temperature and precipitation. The large-scale understanding of the physics seems to be sufficient, but the details are either not well understood, or are not being sufficiently well approximated by the model code. Given the spread of model results at the local scale, the issue is not so much one of falsification, but rather that current models do not provide much of a guide as to future climate change. Research to address this deficit in the models is required in order for the models to become truly trustworthy, but it is not clear when, if ever, this will be achieved.

CONCLUSIONS

Approaches for interpreting climate model consensus and reliability are still limited, and this is an important area for further research. While it is universally accepted that scientific knowledge is always provisional and imperfect, we see little reason to anticipate any major re-evaluation that could undermine current understanding, which although supported (and in some cases initially produced) by complex climate models, is also amenable to a more qualitative level of explanation. We can therefore be confident that the broad features of the climate system response to anthropogenic forcing are reasonably represented by current models. However, the credibility of model outputs is clearly limited when we focus on the finer scales at which knowledge is desired by stakeholders. Increases in model resolution have generated more spatially detailed, but not necessarily more accurate, predictions, and sub-continental scale performance

remains poor. While prediction of climate variation on the decadal timescale appears theoretically possible, there are as yet few results showing a useful degree of skill. Recent investment in these areas of research

has been substantial, but the benefits have been limited and it is not clear how best to make progress in closing the gap between actual and potential (or perhaps desired) performance.

ACKNOWLEDGMENTS

We are grateful to the reviewers and the editor for their constructive comments. This research was supported by the Environment Research and Technology Development Fund (S-10) of the Ministry of the Environment, Japan. We would like to acknowledge Ayako Abe-Ouchi and Seita Emori for their continual encouragement over the last 12 years, and for financially supporting the publication of this paper.

REFERENCES

- Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York: Cambridge University Press; 2007, 996 pp.
- Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE. The WCRP CMIP3 multimodel dataset. *Bull Am Meteorol Soc* 2007, 88:1383–1394.
- Lewis JM. Clarifying the dynamics of the general circulation: Phillips's 1956 experiment. *AMS* 1998, 79:39–60.
- Phillips NA. The general circulation of the atmosphere: a numerical experiment. *Q J Roy Meteorol Soc* 1956, 82:123–164.
- Manabe S, Smagorinsky J, Strickler RF. Simulated climatology of a general circulation model with a hydrological cycle. *Mon Weather Rev* 1965, 93:769–798.
- Manabe S. The dependence of atmospheric temperature on the concentration of carbon dioxide. In: Singer SF, ed. *Global Effects of Environmental Pollution*. Dordrecht, the Netherlands: D. Reidel Publishing Co.; 1968, 25–29.
- Charney JG. *Carbon Dioxide and Climate: A Scientific Assessment*. Washington, DC: NAS; 1979.
- Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, et al. Global climate projections. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York: Cambridge University Press; 2007.
- Parker WS. When climate models agree: the significance of robust model predictions. *Philos Sci* 2011, 78:579–600.
- Masson D, Knutti R. Climate model genealogy. *Geophys Res Lett* 2011, 38:L08703.
- Yokohata T, Annan JD, Collins M, Jackson CS, Tobis M, Webb MJ, Hargreaves JC. Reliability of multi-model and structurally different single-model ensembles. *Clim Dyn* 2011, 39:599–616.
- Randall DA, Wood RA, Bony S, Colman R, Fichetef T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, et al. Climate models and their evaluation. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, eds. *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK and New York: Cambridge University Press; 2007.
- Reichler T, Kim J. How well do coupled models simulate today's climate? *Bull Am Meteorol Soc* 2008, 89:303–311.
- Kiehl JT. Twentieth century climate model response and climate sensitivity. *Geophys Res Lett* 2007, 34:L22710.
- Lloyd EA. Confirmation and robustness of climate models. *Philos Sci* 2010, 77:971–984.
- Kalnay E, Lord SJ, McPherson RD. Maturity of operational numerical weather prediction: medium range. *Bull Am Meteorol Soc* 1998, 79:2753–2769.
- Smith L. What might we learn from climate forecasts? *Proc Natl Acad Sci USA* 2002, 99:2487–2492.
- Arrhenius S. On the influence of carbonic acid in the air upon the temperature of the ground. *Philos Mag* 1896, 41:237–276.
- Yoshimori M, Watanabe M, Abe-Ouchi A, Shiogama H, Ogura T. Relative contribution of feedback processes to Arctic amplification of temperature change in MIROC GCM. *Clim Dyn* 2013:1–18.
- Hansen J, Fung I, Lacis A, Rind D, Lebedeff S, Ruedy R, Russell G, Stone P. Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. *J Geophys Res* 1988, 93:9341–9364. doi: 10.1029/JD093iD08p09341.

21. Hargreaves JC. Skill and uncertainty in climate models. *WIREs Clim Change* 2010, 1:556–564. doi: 10.1002/wcc.58.
22. Allen MR, Mitchell JFB, Stott PA. Test of a decadal climate forecast. *Nat Geosci* 2013, 6:243–244.
23. Hargreaves JC, Annan JD, Ohgaito R, Paul A, Abe-Ouchi A. Skill and reliability of climate model ensembles at the Last Glacial Maximum and mid-Holocene. *Clim Past* 2013, 9:811–823.
24. Racherla PN, Shindell DT, Faluvegi GS. The added value to global model projections of climate change by dynamical downscaling: a case study over the continental U.S. using the GISS-ModelE2 and WRF models. *J Geophys Res* 2012, 117:D20118. doi: 10.1029/2012JD018091.
25. Levy AAL, Ingram W, Jenkinson M, Huntingford C, Lambert FH, Allen M. Can correcting feature location in simulated mean climate improve agreement on projected changes? *Geophys Res Lett* 2013, 40:354–358.
26. Suckling EB, Smith LA. An evaluation of decadal probability forecasts from state-of-the-art climate models. *J Clim* 2013, 26:9334–9347.
27. Tollefson J. Climate change: the forecast for 2018 is cloudy with record heat. *Nature* 2013, 499:139–141.
28. Keenlyside NS, Ba J. Prospects for decadal climate prediction. *WIREs Clim Change* 2010, 1:627–635. doi: 10.1002/wcc.69.
29. Stott P, Good P, Jones G, Gillett N, Hawkins E. The upper end of climate model temperature projections is inconsistent with past warming. *Environ Res Lett* 2013, 8:014024.
30. Smith SL, Pahlow M, Merico A, Wirtz KW. Optimality-based modeling of planktonic organisms. *Limnol Oceanogr* 2011, 56:2080–2094.
31. Murphy JM, Sexton DMH, Barnett DN, Jones GS, Webb MJ, Collins M, Stainforth DA. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 2004, 430:768–772.
32. Annan JD, Hargreaves JC, Ohgaito R, Abe-Ouchi A, Emori S. Efficiently constraining climate sensitivity with paleoclimate simulations. *Sci Online Lett Atmos* 2005, 1:181–184.
33. Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, et al. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 2005, 433:403–406.
34. Rodwell MJ, Palmer TN. Using numerical weather prediction to assess climate models. *Q J Roy Meteorol Soc* 2007, 133:129–146.
35. Yokohata T, Annan JD, Collins M, Jackson CS, Shiogama H, Watanabe M, Emori S, Yoshimori M, Abe M, Webb MJ, et al. Reliability and importance of structural diversity of climate model ensembles. *Clim Dyn* 2013, 41:2745–2763.
36. Kuhn TS. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press; 1962.