Focus Article

Ensemble modeling, uncertainty and robust predictions

Wendy S. Parker*

Many studies of future climate change take an ensemble modeling approach in which simulations of future conditions are produced with multiple climate models (or model versions), rather than just one. These ensemble studies are of two main types—perturbed-physics and multimodel—which investigate different sources of uncertainty about future climate change. Increasingly, methods are being applied which assign probabilities to future changes in climate on the basis of the set of projections (the ensemble) produced in a perturbed-physics or multimodel study. This has prompted debate over both the appropriate interpretation of ensembles as well as how best to communicate uncertainty about future climate change to decision makers; such communication is a primary impetus for ensemble studies. The intuition persists that agreement among ensemble members about the extent of future climate change warrants increased confidence in the projected changes, but in practice the significance of this robustness is difficult to gauge. Priority topics for future research include how to design ensemble studies that take better account of structural uncertainty, how to weight ensemble members and how to improve the process by which ensemble studies are synthesized with other information in expert assessments.

How to cite this article:
WIREs Clim Change 2013, 4:213–223. doi: 10.1002/wcc.220

INTRODUCTION

Today, many studies of future climate change take an *ensemble* modeling approach in which simulations of future conditions are produced with multiple climate models (or model versions), rather than just one. The need for multiple models stems from uncertainty about how to represent the climate system such that accurate projections of future climate change on global and regional scales will be produced. There are two main reasons for this uncertainty. First, while much has been learned about the climate system, some processes that shape the evolution of climate still are not well understood. Second, limited computing power constrains how processes can be represented in climate models. Subgrid processes like convection, for instance, must be represented in terms of larger-scale variables in a simplified way; often it is unclear which equations, and which parameter values within those equations, would be best.

Even as ensemble studies of future climate change become increasingly common, questions remain regarding the interpretation of their results. What exactly does the set of projections produced in an ensemble study tell us about current uncertainty about future climate change? Can ensembles inform us of the probabilities that we should assign to future changes in climate? Are robust projections from today's ensemble studies especially trustworthy? Such questions continue to prompt discussion and debate, in part because their answers have importance beyond the bounds of climate science: far-reaching mitigation and adaptation decisions may be influenced by what is learned about future climate change from ensemble studies.

This article provides an introduction to ensemble modeling and its use in investigating uncertainty about future climate change. The second section provides a brief historical discussion of ensemble modeling. The third section identifies and distinguishes two main types of ensemble study: perturbed-physics

*Correspondence to: parkerw@ohio.edu

Department of Philosophy, Ohio University, Athens, Ohio, USA

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

and multimodel. The fourth section discusses the nature of the uncertainty at issue as well as different approaches to quantifying this uncertainty with the help of ensembles; both probabilistic and nonprobabilistic depictions of uncertainty are discussed. The fifth section considers the significance of robust projections from ensemble studies, arguing that in general their significance is unclear. The sixth section briefly explores the value of ensemble studies in political and public spheres, highlighting their role in communicating uncertainty to decision makers. Lastly, some priority research areas are identified.

HISTORY

The ensemble approach first emerged in response to the discovery that atmospheric models display chaotic behavior.[1–3] This discovery made untenable the assumption that small errors in initial conditions would make only small differences in the weather forecasts produced. Running a forecast model with slightly different sets of initial conditions provided a means of exploring the sensitivity of the day's forecast to small errors and uncertainties in the analysis; the greater the spread of the ensemble of simulations produced, the greater the sensitivity, and the larger the recognized uncertainty about future conditions. By the early 1990s, both the National Center for Environmental Prediction in the United States and the European Center for Medium Range Weather Forecasting in the United Kingdom were employing this ensemble approach operationally, developing distinctive methods for selecting alternative sets of initial conditions.[4]

Since then, the ensemble approach in weather forecasting has become increasingly sophisticated[4,5] and has expanded to include more than just variation in initial conditions. For instance, some operational forecasting centers now also vary the values assigned to model parameters as well as the way particular processes (e.g., convection) are represented.[6] This reflects the fact that there is not only uncertainty about the initial conditions from which a weather forecast should be generated, but also uncertainty about how to model the atmosphere—within the constraints of today's computing power—such that highly accurate weather forecasts will be produced. Running different models and model versions, in combination with different sets of initial conditions, is a way to begin to take account of the latter uncertainty. The spread of an ensemble of simulations produced in this way provides a more comprehensive estimate of uncertainty about future conditions than the spread that is found when only initial conditions are varied. Still, there is no guarantee that observed conditions will fall within the ensemble spread. In fact, in practice today's ensembles tend to be underdispersive: observed weather conditions not infrequently fall outside the range defined by the spread of the forecast ensemble.[5,7,8]

In the climate context too, the ensemble approach took off in the 1990s, but serving a wider range of purposes. For instance, ensembles consisting of very long unforced simulations from different climate models were produced and analyzed to estimate the climate system's internal variability, that is, the variability in conditions that occurs even in the absence of external forcing.[9–11] Estimates of internal variability are important for detection and attribution studies but are difficult to obtain from observations. The ensemble approach was also pursued with the goal of learning why existing models disagree in some key aspects of their performance, in the hope that this would facilitate both model improvement and better understanding of climate processes. To this end, in 1989 the Program for Climate Model Diagnosis and Intercomparison (PCMDI) was established at Lawrence Livermore National Laboratory in the United States. Beginning with the Atmospheric Model Intercomparison Project (AMIP) in 1990, PCMDI has helped to coordinate model intercomparison projects (MIPs) in which participating modeling groups each perform a similar suite of simulation experiments and then deposit results in a shared database. Under the auspices of PCMDI and other scientific organizations, MIPs with various scientific aims have since been conducted—additional projects include CMIP, PMIP, SMIP, C^4MIP, CFMIP, GeoMIP, etc.—each producing ensembles that subsequently undergo extensive analysis by the research community.[12–15]

Some phases of these MIPs, including the third and fifth phases of Coupled Model Intercomparison Project (CMIP), known as CMIP3 and CMIP5, have been designed to support research relevant for IPCC scientific assessments of climate change.[16,17] This includes detection and attribution research, as well as studies of future climate change. In support of the latter, more than a dozen modeling groups participating in CMIP3 produced simulations of future conditions under specified emission scenarios and deposited results for a subset of variables in a public online database managed by PCMDI. These results helped to inform estimates of uncertainty about future climate change in the IPCC's Fourth Assessment Report (see below).[18] CMIP5, currently underway, is generating ensembles that will inform the IPCC's Fifth Assessment in a similar way.

As the examples of CMIP3 and CMIP5 illustrate, ensemble studies are sometimes undertaken as a means of investigating uncertainty about future climate change. In fact, over the last decade, this has become perhaps the most prominent use of ensembles in the climate context. Ensemble studies—including both MIPs and other types of study—are now the primary means by which uncertainty about future climate change is investigated in a quantitative way.

TYPES OF ENSEMBLE

In studies of future climate change, two types of ensemble are commonly distinguished: *perturbed physics* and *multimodel*. Ensembles of both types consist of multiple simulations of future climate under similar forcing conditions. They differ, however, in what is varied among the models used to generate the simulations and thus in the sources of uncertainty about future climate change that are explored.

Perturbed-physics studies explore how climate change projections are impacted by *parametric uncertainty*, that is, uncertainty about the values that should be assigned to a climate model's parameters.[a] Perturbed-physics ensembles are produced by running multiple versions of a single climate model, where each version incorporates a different set of parameter values. Usually, the parameter values that define a model version are chosen from ranges considered plausible on the basis of expert judgment.[b] For simple and intermediate-complexity climate models that are computationally inexpensive to run and that have a relatively small number of parameters, thousands of combinations of parameter values, chosen via formal sampling procedures, can be tried.[19–21] This is out of reach with state-of-the-art climate models, which have on the order of a hundred uncertain parameters and require significant computing time to complete even a single run.

In response to this computational roadblock, at least two alternative approaches have developed. The climateprediction.net project[22–24] takes a distributed computing approach, relying on donated idle-processing time on a large number of ordinary home computers; with the help of the public, a climateprediction.net experiment can run thousands of versions of a relatively complex climate model (see Figure 1). A second type of approach, used to produce the UKCP09 regional climate change projections for the United Kingdom, runs a limited number of versions of a relatively complex or even state-of-the-art climate model and then uses statistical methods (emulators) to estimate the set of projections that would be produced if more comprehensive sampling of parameter uncertainty in the model could be performed.[25–27]

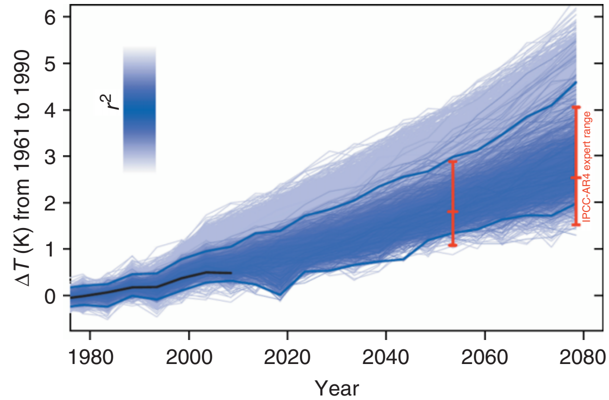

FIGURE 1 | Projected global warming under the A1B emission scenario, relative to 1961–1990 mean temperature. Results are from more than 2000 ensemble members from more than 800 model versions in the climateprediction.net BBC perturbed-physics study. Darker shading indicates better fit with observational data. The black line indicates results from observational data. The dark blue lines indicate a 'likely' range (66% confidence interval) from the ensemble (see Ref 19 for details). Red bars indicate the 'likely' ranges (66% confidence ranges) specified by IPCC experts[18] for around 2050 and 2080. (Reprinted with permission from Ref 24. Copyright 2012 Nature Publishing Group).

Multimodel studies employ more than one climate model and investigate how climate change projections are impacted by *structural uncertainty*, that is, uncertainty about the form that modeling equations should take and how they should be solved computationally. In multimodel studies, not all of the models have the same dynamical equations; some of the models might include representations of climate system processes that other models in the study do not yet include, or the models might differ in the types of parameterization schemes they employ, etc. Beyond this, there may be differences in parameter values, spatiotemporal resolution, numerical methods and computer hardware as well. Thus far, multimodel studies have employed existing climate models, and variations thereon, rather than building new sets of models from scratch for purposes of the study. The most ambitious multimodel studies conducted to date are CMIP3 and CMIP5, mentioned above. CMIP5, still underway, has collected results from approximately 60 models from nearly thirty modeling centers/groups around the world (though not all experiments are run with each model).[28]

To summarize, perturbed-physics studies investigate how projections are impacted by parametric uncertainty, and they do so by running a single climate model with alternative sets of parameter values, in some cases sampling these alternatives in an extensive and systematic way; multimodel studies investigate how projections are impacted by structural uncertainty, but they typically do so in an unsystematic way, by producing projections with a set of existing climate models rather than a set carefully designed to probe this uncertainty.

Both perturbed-physics and multimodel studies often include a limited investigation of the impacts of *initial condition uncertainty* as well: each model or model version is run more than once under a chosen emission scenario, using different initial conditions on each run. The relative contribution of initial condition uncertainty to the ensemble spread varies with the time scale, spatial scale, and quantity considered; it tends to be small for long-term changes in global temperature in scenarios with significant forcing, for instance, but large for short-term changes in precipitation.[29,30] Nevertheless, ensemble studies have tended to focus on parametric and structural uncertainty, whether or not this is justified.

Studies that combine the multimodel and perturbed-physics approaches in a systematic way—varying both parameter values and model structure according to a careful experimental design in order to thoroughly sample or bound uncertainties about future climate change—have not yet been conducted. Doing so would require a tremendous amount of computing power. But there are deeper obstacles as well. For while there are various approaches to sampling a space of parameter values, it is not clear what it would mean to sample the relevant space of alternative model structures, if that space could be specified.[26,31,32]

QUANTIFYING UNCERTAINTY ABOUT FUTURE CLIMATE CHANGE

In general terms, *uncertainty* is a lack of knowledge. There is a lack of knowledge about 21st century climate change (and beyond) in the sense that it is unclear whether particular changes in climate, such as large increases in decadal mean temperature in specific regions, will occur. This uncertainty about future climate change stems both from uncertainty about future levels of greenhouse gas emissions and other external forcings, known as *scenario uncertainty*, and from uncertainty about how the climate system would respond to a given set of forcings, known as *response uncertainty*.[c] Climate science investigates response uncertainty—uncertainty about how climate would change in the future under a specified scenario.[d]

Response uncertainty is primarily *epistemic* in origin: it stems mainly from uncertainty (lack of knowledge) about how to build adequate predictive models of the climate system, rather than from the climate system's having some irreducibly statistical or indeterministic character (an *ontic* source of uncertainty).[e] Multimodel and perturbed-physics studies investigate response uncertainty by producing projections of future climate change using alternative parameter values, modeling equations and initial conditions, as described above. But what exactly do these ensembles reveal about response uncertainty?

One view is that an ensemble indicates a lower bound on response uncertainty; the spread of an ensemble defines a range of changes (e.g., in global mean surface temperature) that cannot yet be ruled out—a 'nondiscountable climate change envelope'.[33] This interpretation requires that each model or model version whose projection is included be one that is plausibly adequate for projecting the changes of interest, and it assumes that magnitudes of change in between those projected by the ensemble members also are plausible. Suppose, for example, that each of several state-of-the-art climate models in a multimodel study is plausibly adequate for projecting 21st century changes in global mean surface temperature under an emission scenario. Then, on this view, if the changes projected by these models range from 1.2 to 3.7°C, this range will define a lower bound on uncertainty about the temperature change; a change between 1.2 and 3.7°C remains plausible for that emission scenario, but no conclusions are drawn about changes outside that range.

Why would the spread of projections provide only a lower bound on response uncertainty? Because neither multimodel nor perturbed-physics studies explore structural, parametric and initial condition uncertainty in a comprehensive way. Multimodel ensembles like those produced in the CMIP projects, for instance, are to a large extent 'ensembles of opportunity'[18,34]—produced with existing state-of-the-art models and only insofar as modeling groups are willing to participate; they are not designed to systemically sample or bound uncertainties but rather are more like a collection of best guesses.[34] Perturbed-physics studies sometimes explore parametric uncertainty quite extensively, but they typically do not account for structural uncertainty at all, insofar as they vary parameter values within a single climate model.

Nevertheless, increasingly, ensemble studies do not simply report a range of future changes in climate that cannot yet be ruled out, but instead deliver what appear to be precise probabilistic estimates of uncertainty.[18,35] That is, from the set of projections produced in the ensemble study, inferences are made about which changes in climate can be assigned more or less probability under the associated emission scenario (see Figure 2). Numerous methodologies[20,21,25–27,36–46] for producing such probabilities have been developed, many of which cast this inference task in Bayesian terms. It is beyond the scope of this article to review these methodologies, most of which are fairly technical. A fundamental assumption of many of them is that changes projected by models or model versions that better simulate past climate (on chosen metrics of performance) should be assigned more probability than changes projected by models or model versions that simulate past climate less well.

It is important to understand the nature of the probabilities produced. They are meant to indicate the degree of belief (or confidence) that one should have that a particular change in climate would occur under an emission scenario, given current information. They are not meant to indicate how frequently the change would occur in a hypothetical series of experiments in which earth-like planets were subjected to the conditions of the emission scenario. For example, a probability of 0.7 would indicate that one should be 70% confident, or have a degree of belief of 0.7, that the change would occur. Researchers sometimes emphasize that today's ensemble studies only produce 'estimates' of these probabilities. This is because the studies do not take into account all sources of uncertainty (and error) and/or because they rely on other questionable or simplistic assumptions. For instance, a study might fail to take into account uncertainty associated with climate system processes not represented in any of the models used (e.g., methane cycle feedbacks), or it might assume that projections by ensemble members are distributed around truth, each with some random error.

Especially when it comes to communicating uncertainty to decision makers, the crucial question seems to be whether these estimated probabilities would be only slightly different were all sources of uncertainty taken into account, the simplistic assumptions made more realistic, etc. Considering this question carefully, however, casts doubt on the very idea

Variable	Annual mean precipitation %					Winter mean precipitation %					Summer mean precipitation %				
Probability level	10%	50%	90%	Wider range		10%	50%	90%	Wider range		10%	50%	90%	Wider range	
North Scotland	−6	0	+5	−7	+6	+3	+13	+24	0	+26	−23	−10	+2	−23	+6
East Scotland	−4	0	+5	−5	+6	+2	+10	+20	−1	+21	−26	−12	+1	−27	+6
West Scotland	−6	0	+5	−7	+6	+5	+15	+28	0	+30	−26	−12	+1	−27	+6
Northern Ireland	−3	0	+3	−3	+3	+2	+9	+19	0	+19	−26	−12	+3	−27	+8
Isle of Man	−5	0	+4	−6	+5	+2	+16	+35	−1	+36	−31	−15	+1	−32	+8
North East England	−4	0	+5	−5	+5	+1	+11	+24	0	+26	−29	−14	+1	−30	+7
North West England	−5	0	+6	−6	+7	+3	+13	+26	0	+27	−34	−17	+1	−36	+8
Yorkshire & Humber	−3	0	+4	−4	+5	+2	+11	+24	0	+27	−35	−17	+1	−37	+9
East Midlands	−4	0	+6	−5	+6	+2	+14	+29	+1	+33	−35	−15	+6	−37	+13
West Midland	−4	0	+6	−5	+6	+2	+13	+28	+1	+31	−36	−16	+6	−38	+13
Wales	−4	0	+5	−5	+6	+2	+14	+30	0	+31	−36	−16	+6	−38	+13
East England	−4	0	+5	−4	+6	+3	+14	+31	+1	+35	−37	−16	+6	−39	+14
London	−4	0	+5	−4	+5	+2	+15	+33	0	+37	−39	−18	+7	−41	+16
South East England	−4	0	+6	−5	+6	+2	+16	+36	+1	+40	−40	−18	+7	−42	+16
South West England	−4	0	+6	−5	+6	+4	+17	+38	0	+41	−41	−19	+7	−43	+16
Channel Islands	−4	0	+3	−4	+4	+2	+15	+34	0	+38	−47	−22	+9	−49	+20

FIGURE 2 | Changes in precipitation (%) at the 10, 50 and 90% probability levels in different UK regions by the 2050s as estimated by UKCP09. Changes are relative to the 1961-1990 mean for each region, under a medium emission scenario. Wider ranges indicate the lowest and highest values of change seen across three emission scenarios and all three probability levels. (Reprinted with permission from Ref 25. Copyright 2009 Crown Copyright).

that response uncertainty can be accurately represented with precise (i.e., single-valued) probabilities, at least for many changes in climate. This is easiest to see in connection with structural uncertainty. Perturbed-physics studies typically do not take any account of structural uncertainty, and multimodel studies do so in a limited and unsystematic way. How would the distribution of modeling results, and thus the probabilities estimated in these studies, be different if structural uncertainty (and error) were taken into account more fully? The answer is itself significantly uncertain, because it is unclear which additional (not-yet-constructed) models would need to be included, much less what the projections from these models would be. This suggests that response uncertainty would be more accurately described using interval probability specifications or in some other way, rather than with precise probabilities.[32,47]

Figure 3 illustrates a shift from single-valued probabilities to interval probability specifications. Most of the solid vertical lines in the figure indicate ranges of global mean temperature change to which particular ensemble studies assigned a precise probability of 0.9 for a specified emission scenario. The uncertainty estimate ultimately given by the IPCC in its Fourth Assessment Report,[18] however, is denoted by the associated wide gray bar in the background, which indicates the range to which the IPCC assigned an interval probability of *at least* 0.66 on the basis of expert judgment. That is, after considering not only the available results from ensemble studies, but also the known or suspected limitations of these studies as well as other available information, the IPCC experts reported that their confidence (or degree of belief) was *at least* 66% that the temperature change would be in that range. Note that this 'likely' range, which was defined by adding 60% and subtracting 40% from the mean of the temperature changes projected by the CMIP3 models, is broader than many of the ranges to which individual ensemble studies assigned a precise probability of 0.9. This suggests that the IPCC experts judged that, had structural, parametric, and initial condition uncertainty been more thoroughly explored (or accounted for) in these ensemble studies, the range of projected changes would have been wider, though how much wider cannot be precisely specified.

Thus, at least three ways of connecting ensembles with conclusions about response uncertainty have emerged. One approach takes the ensemble spread to define a lower bound on response uncertainty, indicating changes in climate that cannot yet be ruled out; no probabilities are assigned. A second approach uses formal methods to assign single-valued probabilities to changes in future climate on the basis of an ensemble of projections; these probabilities are described as 'estimates' conditional on the models and methods used. A third approach views ensemble studies as imperfect investigations of response uncertainty whose results should be considered alongside all other available information; this approach, which often relies on expert judgment, recognizes that response uncertainty is often 'deeper' than single-valued probabilities would imply and, in those cases, characterizes response uncertainty using interval probability specifications or in other ways (e.g., sets of probability density functions, order of magnitude estimates, etc.).

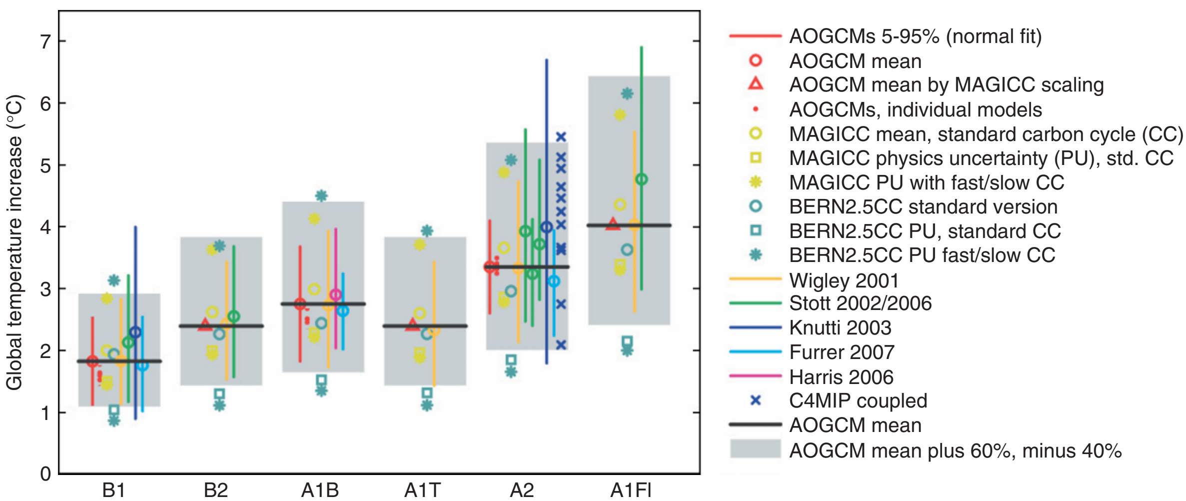

FIGURE 3 | Projections and uncertainties for global mean temperature change in 2090–2099 relative to the 1980–1999 average, under several emission scenarios and from numerous studies. The 5–95% ranges from probabilistic studies are indicated by solid vertical lines. The IPCC 'likely' ranges are indicated by wide gray bars. (Reprinted with permission from Ref 18. Copyright 2007 Cambridge University Press).

In principle, these three approaches need not conflict with one another in either their assumptions or their products: a lower bound defined by the ensemble spread can be consistent with a precise probabilistic 'estimate' of uncertainty obtained using formal methods, which in turn can be consistent with a characterization of uncertainty in terms of an interval probability specification that is reached by synthesizing ensemble results with other information. In practice, however, they often are seen as competing approaches, and there is disagreement about the reasonableness of their assumptions, the limitations of their methodologies and the extent to which their products are informative and useful for decision makers (see below).

ROBUST PREDICTIONS

Some projected changes from ensemble studies are found to be relatively *robust*.[f] That is, though different models and methods are used, similar predictions are produced.[18,48] For instance, all simulations produced in a multimodel study might agree that mean surface temperature in a region would increase by more than 2°C by the end of the 21st century under a particular emission scenario. Or several ensemble studies producing probabilistic projections via different methods might agree that an increase in global mean surface temperature of more than 1.5°C over the next several decades is very unlikely, regardless of the emission scenario considered.[18]

It is tempting to conclude that such robust predictions are particularly reliable or trustworthy. But this is not necessarily the case. Members of an ensemble might agree that a change in climate would not exceed some threshold of interest, but they might agree because all of the models fail to capture a significant driver of the change, and it might even be known that they fail to capture this driver. Moreover, if enough predictive variables and magnitudes of change are considered, cases of agreement eventually will be found, just as a matter of chance. It is robustness in combination with additional information that is significant, not robustness on its own.

For instance, suppose all ensemble members in a multimodel study agree that summer precipitation in a region would decrease over a particular period under an emission scenario. Should scientists' confidence that a decrease would occur be substantially higher after learning this result than it was before? It depends on, among other things, whether it is significantly more likely that the models all would get the sign of the precipitation change right than that they all would get it wrong.[49] The problem is that, in many cases, there is little basis for conclusions about such likelihoods. While weather forecasting models (and sets of such models) can build substantial track records of performance, opportunities are much more limited in the climate context, where predictions concern longer-term changes under novel forcing conditions. In principle, it might be argued on physical grounds that a case of agreement is unlikely to be spurious—it might be argued, for instance, that the processes that will drive precipitation change in the region are well known and well represented in the models—but in practice these arguments are often out of reach. (Moreover, if the individual models are thought to be highly reliable, then agreement among their results may not increase one's confidence much beyond what a single modeling result would do.)

A number of ensemble methods currently in use do assign higher probabilities to changes in climate as more and more members agree in predicting those changes, in effect assuming that model agreement is grounds for increased confidence. Many of these methods either implicitly or explicitly treat different models as independent sources of information. In general, however, this assumption appears unjustified, even in the case of multimodel studies like CMIP3. The CMIP3 models were not all developed independently—a single modeling center might contribute several models that were built in a similar way, even using some of the same code—and biases in their performance have been found to be correlated, especially for models developed at the same centers; some analyses have suggested that the two dozen or so CMIP3 models behave like a set of only 5–10 statistically independent models.[50–54] How to gauge and account for model dependence when interpreting ensembles has not been resolved.[34,52,55] Failing to account for it can lead to overconfidence that particular changes in climate would occur.

Thus, while the intuition persists that agreement among climate change projections warrants increased confidence in the projected changes, the extent to which this is the case is very often unclear.

SIGNIFICANCE IN POLITICAL AND PUBLIC SPHERES

Efforts to develop and improve ensemble methods for investigating uncertainty about future climate change have steered climate science in new and valuable directions. Yet the ultimate motivation for these ensemble studies is not a scientific one, but a practical one: providing policymakers and other decision makers with information about the extent of current uncertainty about future climate change, in order to facilitate their decision making. As noted above, the largest multimodel studies conducted to date, CMIP3 and CMIP5, have been undertaken expressly in support of IPCC assessments. Likewise, the UKCP09 regional climate change projections, produced in the most sophisticated perturbed-physics study conducted thus far, are intended to support adaptation decisions across a wide range of sectors in the United Kingdom.[25]

Given the aim of informing policy and other decision making, however, there has been debate about how ensembles—and especially conclusions about uncertainty derived from them—should be presented. Some have objected to the practice of transforming ensembles into precise probabilistic estimates of uncertainty.[32,56] They worry that decision makers will take these 'estimated' probabilities at face value and, as a consequence, make poorer decisions (relative to their own goals) than they would have made had the second-order uncertainty associated with these estimates been made more salient. Simply offering a lower bound on uncertainty, that is, a nondiscountable climate change envelope, garners criticism for failing to convey to decision makers whether some changes are more plausible/probable than others.[57] Imprecise probability specifications can avoid both complaints but, insofar as expert judgment plays a significant role in their production, they are sometimes perceived as lacking in rigor; this concern might be mitigated by making clearer the basis for the judgments involved.

Beyond their roles in decision support, ensemble studies of future climate change also are noteworthy as sites of coordinated international participation in climate science. In projects like CMIP5, modeling groups around the world invest time and effort to help address specific questions about future climate change; they each contribute to the ensembles that are produced.[17] Likewise, the climateprediction.net project has engaged tens of thousands of members of the public, from dozens of countries, in producing perturbed-physics simulations with their home computers.[22–24] In both cases, the ensembles are community products, created with the help of many parties who, through their participation, come to have an additional stake in the findings.

FUTURE DIRECTIONS

Despite recent progress in developing and deploying ensemble methodologies for investigating uncertainty about future climate change, much work remains to be done. Current research priorities include:

Structural uncertainty. More work is needed on how to design ensemble studies to better take account of structural uncertainty. Multimodel studies explore structural uncertainty to a limited extent but, as noted above, their results are more like a collection of best guesses than a collection designed with the goal of spanning or bounding uncertainty about future climate change. Perturbed-physics studies often ignore structural uncertainty. A recent exception was the UKCP09 study, which accounted for structural uncertainty in a preliminary way via a discrepancy term.[25,58]

Weighting and metrics. Many ensemble studies that assign probabilities to future changes in climate already perform differential weighting of projections. But questions remain about whether and how such weighting should be done, especially for multimodel studies like CMIP5.[52,55,59] Which metrics of performance[60] are most relevant when evaluating model quality for a given predictive purpose? When is qualitative weighting—such as merely highlighting a subset of projections—more appropriate than quantitative weighting? These and related questions merit further attention. At least one recent study has demonstrated that improper weighting can easily result in greater loss of skill than equal weighting.[61]

Knowledge synthesis. Taking a broader view, very little attention has been given to the process by which experts synthesize the results of ensemble studies with other background knowledge (including knowledge of the limitations of those ensemble studies), in order to arrive at uncertainty estimates that are based on all available information. It is worth considering how this process could be structured to minimize cognitive biases, such as double-counting, anchoring, etc., as well as how it could be made more transparent, so that uncertainty estimates are both more accurate and more accountable.

NOTES

[a] A parameter is a constant term in an equation, whose value is set by the modeler. An example would be a parameter representing the rate at which air is entrained from the environment in a particular kind of cloud. Parameter values are held constant during a simulation, even as the values of model variables (e.g., temperature, pressure, etc.) change from one time step to the next.

[b] This should be uncertainty about which set of parameter values will give the most accurate projections, for a specified measure of accuracy, not uncertainty about the true values of the parameters. The physical meaning of some parameters is unclear and, for those that do have clear physical meaning, the true values may not give the most accurate projections, since the structure of the model is imperfect.[32]

[c]I assume that climate is a distribution of weather conditions, not the average of those conditions. Response uncertainty—uncertainty about the change in climate under a scenario—thus includes uncertainty due to internal variability. Other authors[62] distinguish internal variability from response uncertainty.
[d]This is why simulations of future conditions often are described as *projections*; they are predictions made conditional on assumptions about future emissions and other external forcings.
[e]It is sometimes suggested that, if a system is chaotic, then uncertainty about its future state is at least partly ontic in character.[63] Since chaos is deterministic, this requires a definition of ontic uncertainty that includes more than uncertainty due to indeterminism, though exactly what else it should include is unclear, especially if overlap between ontic and epistemic uncertainty is to be avoided. Perhaps other ways of classifying sources of uncertainty would be more useful here, for example, reducible in practice vs. reducible in principle vs. irreducible.
[f]There are several senses in which predictions can be robust. Here, the focus is on robustness as mere agreement in predictions from different models (or model versions) available at a given point in time. Predictions can also be robust in the sense that they do not change much even as new generations of models are developed or new information becomes available.

REFERENCES

1. Lorenz EN. Deterministic nonperiodic flow. *J Atmos Sci* 1963, 20:130–141.
2. Lorenz EN. Predictability of a 28-variable atmospheric model. *Tellus A* 1965, 17:321–333.
3. Leith CE. Theoretical skill of Monte-Carlo forecasts. *Month Weather Rev* 1974, 102:409–418.
4. Kalnay E. *Atmospheric Modeling, Data Assimilation and Predictability*. New York: Cambridge University Press; 2003.
5. Buizza R, Houtekamer PL, Toth Z, Pellerin G, Wei M, Zhu Y. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Month Weather Rev* 2005, 133:1076–1097.
6. Environment Canada. Ensemble Forecasts. Available at: http://www.weatheroffice.gc.ca/ensemble/verifs/model_e.html. (Accessed August 20, 2012).
7. Eckel FA, Mass C. Aspects of effective mesoscale, short-range ensemble forecasting. *Weather Forecast* 2005, 20:328–350.
8. Wilks DS, Hamill TM. Comparison of ensemble-MOS methods using GFS reforecasts. *Monthly Weather Rev* 2007, 135:2379–2390.
9. Santer BD, Wigley TML, Barnett TP, Anyamba E, Bloomfield P, Cook ER, Covey C, Crowley TJ, Delworth T, Gates WL, et al. Detection of climate change and attribution of causes. In: Houghton JT, Meira Filho LG, Callandar BA, Harris N, Kattenberg A, Maskell K, eds. *Climate Change 1995: The Science of Climate Change*. New York: Cambridge University Press; 1996, 407–443.
10. Stouffer RJ, Hegerl GC, Tett SFB. A comparison of surface air temperature variability in three 1000-year coupled ocean-atmosphere model integrations. *J Clim* 1999, 13:513–537.
11. Mitchell JFB, Karoly DJ, Hegerl GC, Zwiers FW, Allen MR, Marengo J, Barros V, Berliner M, Boer G, Crowley T, et al. Detection of climate change and attribution of causes. In: Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Da X, Maskell K, Johnson CA, eds. *Climate Change 2001: The Scientific Basis*. New York: Cambridge University Press; 2001, 695–738.
12. Gates WL, Boyle J, Covey C, Dease C, Doutriaux C, Drach R, Fiorino M, Gleckler P, Hnilo J, Marlais S, et al. An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull Am Meteorol Soc* 1998, 73:1962–1970.
13. Meehl GA, Boer GJ, Covey C, Latif M, Stouffer RJ. The coupled model intercomparison project (CMIP). *Bull Am Meteorol Soc* 2000, 81:313–318.
14. Crucifix M, Braconnot P, Harrison SP, Otto-Bliesner B. Second phase of paleoclimate modelling intercomparison project. *EOS* 2005, 86:264.
15. Friedlingstein P, Cox P, Betts R, Bopp L, von Bloh W, Brovkin V, Cadule P, Doney S, Eby M, Fung I. Climate–carbon cycle feedback analysis: results from the C^{4}MIP model intercomparison. *J Clim* 2006, 19:3337–3353.
16. Meehl GA, Covey C, Delworth T, Latif M, McAvaney B, Mitchell JFB, Stouffer RJ, Taylor KE. The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull Am Meteorol Soc* 2007, 88:1383–1394.
17. Taylor KE, Stouffer RJ, Meehl GA. An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 2012, 93:485–498. doi:10.1175/BAMS-D-11–00094.1.
18. Meehl GA, Stocker TF, Collins WD, Friedlingstein P, Gaye AT, Gregory J, Kito 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*

2007: The Physical Science Basis. New York: Cambridge University Press; 2007, 747–845.

19. Wigley TML, Raper SCB. Interpretation of high projections for global-mean warming. *Science* 2001, 293:451–454.
20. Forest CE, Stone PH, Sokolov AP, Allen MR, Webter MD. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* 2002, 295:113–117.
21. Knutti R, Stocker TF, Joos F, Plattner G-K. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* 2002, 416:719–723.
22. Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 2005, 433:403–406.
23. Frame DJ, Aina T, Christensen CM, Faull NE, Knight SHE, Piani C. The climateprediction.net BBC climate change experiment: design of the coupled model ensemble. *Phil Trans R Soc A* 2009, 367:855–870.
24. Rowlands DJ, Frame DJ, Ackerley D, Aina T, Booth BBB, Christensen C, Collins M, Faull N, Forest CE, Grandey BS, et al. Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nat Geosci* 2012, 5:256–260. doi:10.1038/NGEO1430.
25. Murphy JM, Sexton DMH, Jenkins GJ, Boorman PM, Booth BBB, Brown CC, Clark RT, Collins M, Harris GR, Kendon EJ, et al. *UK Climate Projections Science Report: Climate Change Projections*. Exeter: Met Office Hadley Centre; 2009. Available at: http://ukclimateprojections.defra.gov.uk/media.jsp?mediaid=87894&filetype=pdf. (Accessed July 31, 2012).
26. Murphy JM, Booth BBB, Collins M, Harris GR, Sexton DMH, Webb MJ. A methodology for probabilistic predictions of climate change from perturbed physics ensembles. *Phil Trans R Soc A* 2007, 365:1993–2028.
27. Sexton DM, Murphy JM, Collins M, Webb MJ. Multivariate probabilistic climate prediction using imperfect climate models part I: outline of methodology. *Clim Dyn* 2012, 38:2513–2542. doi:10.1007/s00382-011-1208-9.
28. Program for Climate Model Diagnosis and Intercomparison (PCMDI). CMIP5 – Data Access – Availability. Available at: http://cmip-pcmdi.llnl.gov/cmip5/availability.html. (Accessed February 22, 2013).
29. Deser C, Phillips A, Bourdette V, Teng H. Uncertainty in climate change projections: the role of internal variability. *Clim Dyn* 2012, 38:527–546.
30. Tebaldi C, Arblaster JM, Knutti R. Mapping model agreement on future climate projections. *Geophys Res Lett* 2011, 38:L23701. doi:10.1029/2011GL049863.
31. Smith LA. Disentangling uncertainty and error: on the predictability of nonlinear systems. In: Mees A, ed. *Nonlinear Dynamics and Statistics*. Birkhauser; 2000, 31–64.
32. Parker WS. Whose probabilities? Predicting climate change with ensembles of models. *Phil Sci* 2010, 77:985–997.
33. Stainforth DA, Downing TE, Washington R, Lopez A, New M. Issues in the interpretation of climate model ensembles to inform decisions. *Phil Trans R Soc A* 2007, 365:2163–2177.
34. Tebaldi C, Knutti R. The use of the multi-model ensemble in probabilistic climate projections. *Phil Trans R Soc A* 2007, 365:2053–2075.
35. Collins M. Ensembles and probabilities: a new era in the prediction of climate change. *Phil Trans R Soc A* 2007, 365:1957–1970.
36. Räisänen J, Palmer TN. A probability and decision-model analysis of a multimodel ensemble of climate change simulations. *J Clim* 2001, 14:3212–3226. doi:10.1175/1520-0442.
37. Giorgi F, Mearns LO. Probability of regional climate change based on the reliability ensemble averaging (REA) method. *Geophys Res Lett* 2003, 30:1629. doi:10.1029/2003GL017130.
38. Knutti R, Stocker TF, Joos F, Plattner G-K. Probabilistic climate change projections using neural networks. *Clim Dyn* 2003, 21:257–272. doi:10.1007/s00382-003-0345-1.
39. 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, 429:768–772. doi:10.1038/nature02771.
40. Tebaldi C, Mearns LO, Nychka D, Smith R. Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations. *Geophys Res Lett* 2004, 31:L24213. doi:10.1029/2004GL021276.
41. Knutti R, Joos F, Müller SA, Plattner G-K, Stocker TF. Probabilistic climate change projections for CO_2 stabilization profiles. *Geophys Res Lett* 2005, 32:L20707. doi:10.1029/2005GL023294.
42. Greene A, Goddard L, Lall U. Probabilistic multimodel regional temperature change projections. *J Clim* 2006, 19:4326–4343. doi:10.1175/JCLI3864.1.
43. Harris G, Sexton D, Booth B, Collins M, Murphy J, Webb M. Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations. *Clim Dyn* 2006, 27:357–375.
44. Stott PA, Mitchell JFB, Allen MR, Delworth TL, Gregory JM, Meehl GA, Santer BD. Observational constraints on past attributable warming and predictions of future global warming. *J Clim* 2006, 19:3055–3069.
45. Furrer R, Sain SR, Nychka D, Meehl GA. Multivariate Bayesian analysis of atmosphere-ocean general circulation models. *Environ Ecol Stat* 2007, 14:249–266. doi:10.1007/s10651-007-0018-z.

46. Xu Y, Xuejie G, Giorgi F. Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. *Clim Res* 2010, 41:61–81.

47. Kandlikar M, Risbey J, Dessai S. Representing and communicating deep uncertainty in climate change assessments. *Comp Rendus Geosci* 2005, 337:443–455.

48. Knutti R, Sedláček J. Robustness and uncertainties in the new CMIP5 climate model projections. *Nat Clim Chan* 2012, 3:369–373. doi:10.1038/nclimate1716.

49. Parker WS. When climate models agree: the significance of robust model predictions. *Phil Sci* 2011, 78:579–600.

50. Jun M, Knutti R, Nychka DW. Spatial analysis to quantify numerical model bias and dependence: how many climate models are there? *J Am Stat Assoc* 2008, 103:934–947. doi:10.1198/016214507000001265.

51. Jun M, Knutti R, Nychka DW. Local eigenvalue analysis of CMIP3 climate model errors. *Tellus* 2008, 60:992–1000. doi:10.1111/j1600-0870.2008.00356.x.

52. Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA. Challenges in combining projections from multiple climate models. *J Clim* 2010, 23:2739–2758.

53. Pennell T, Reichler C. On the effective number of climate models. *J Clim* 2011, doi:10.1175/2010JCLI3814.1.

54. Masson D, Knutti R. Climate model genealogy. *Geophys Res Lett* 2011, 38:L08703. doi:10.1029/2011GL046864.

55. Knutti R. The end of model democracy? *Clim Change* 2010, 102:395–404. doi:10.1007/s10584-010-9800-2.

56. Stainforth DA, Allen MR, Tredger ER, Smith LA. Confidence, uncertainty and decision-support relevance in climate prediction. *Phil Trans R Soc A* 2007, 365:2145–2161.

57. Frame DJ, Faull NE, Joshi MM, Allen MR. Probabilistic climate forecasts and inductive problems. *Phil Trans R Soc A* 2007, 365:1971–1992.

58. Rougier J. Probabilistic inference for future climate using an ensemble of climate model evaluations. *Clim Change* 2007, 81:247–264.

59. Knutti R, Abramowitz G, Collins M, Eyring V, Gleckler PJ, Hewitson B, Mearns L. Good practice guidance paper on assessing and combining multi model climate projections. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Midgley PM, eds. *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections*. Switzerland: IPCC Working Group I Technical Support Unit, University of Bern, Bern; 2010.

60. Gleckler PJ, Taylor KE, Doutriaux C. Performance metrics for climate models. *J Geophys Res* 2008, 113:D06104. doi:10.1029/2007JD008972.

61. Weigel A, Knutti R, Liniger MA, Appenzeller C. Risks of model weighting in multimodel climate projections. *J Clim* 2010, 23:4175–4191.

62. Hawkins E, Sutton R. The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 2009. doi:10.1175/2009BAMS2607.1.

63. Petersen A. *Simulating nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice*. 2nd ed. Boca Raton: Chapman and Hall/CRC Press; 2012.

