CHANGING HOW EARTH SYSTEM MODELING IS DONE TO PROVIDE MORE USEFUL INFORMATION FOR DECISION MAKING, SCIENCE, AND SOCIETY

by Matthew J. Smith, Paul I. Palmer, Drew W. Purves, Mark C. Vanderwel, Vassily Lyutsarev, Ben Calderhead, Lucas N. Joppa, Christopher M. Bishop, and Stephen Emmott

A new mode of development for Earth system models is needed to enable better targeted and more informative projections for both decision makers and scientists.

The Earth system is the thin layer of the Earth that contains and supports life. It ultimately governs most processes vital to human health and wellbeing: from food and water availability to disease spread and global economics. It is the canonical

AFFILIATIONS: Smith, Purves, Lyutsarev, Joppa, Bishop, and Emmott—Microsoft Research, Cambridge, United Kingdom; Palmer—Microsoft Research, Cambridge, and School of GeoSciences, University of Edinburgh, Edinburgh, United Kingdom; Vanderwel—Microsoft Research, Cambridge, United Kingdom, and Department of Biology, University of Florida, Gainesville, Florida; Calderhead—Department of Statistical Science, University College London, London, United Kingdom
CORRESPONDING AUTHOR: Matthew J. Smith, Computational Science Laboratory, Microsoft Research, 21 Station Road, Cambridge CB1 2FB, United Kingdom
Email: matthew.smith@microsoft.com

example of a complex system, in which its dynamics, resulting from interacting multiscale and nonlinear processes, cannot be predicted from understanding any of its isolated components. Attempts to understand the Earth system and how it will change in the future therefore depend on computational models that represent, with varying levels of abstraction, physical, chemical, and biological components of the Earth system and their interactions (Randall et al. 2007; Edwards 2010).

Decades of research using such models have resulted in advances in the understanding of many Earth system processes, including the impacts of humans on climate. Models have also produced projections, combining current knowledge of the underlying science with a set of plausible future societal change scenarios to provide information to guide climate change mitigation policy. But what confidence can be assigned to the projections? Confidence about a particular climate projection is often judged by the agreement between different climate models, with greater confidence assigned to projected changes for which there is close agreement

(Pachauri and Reisinger 2007), although other considerations, such as why the models disagree, are also taken into account. Agreements between climate models have predominantly occurred for physical phenomena occurring over large spatial scales. All models agree that the world will become warmer on average if CO_2 levels continue to increase (Stainforth et al. 2005; Pachauri and Reisinger 2007; Oreskes et al. 2010; Kerr 2011; Rowlands et al. 2012), and they all agree that the increases will be greater at higher latitudes (Pachauri and Reisinger 2007). However, they disagree in other aspects, such as by how much the world will warm (Oreskes et al. 2010; Bretherton et al. 2012), and these disagreements become more pronounced at finer, regional spatial scales (Pachauri and Reisinger 2007).

It is now widely recognized that differences in projections not only fail to recognize a variety of additional sources of uncertainty, but also inevitably become increasingly uncertain the further into the future they are extended. For instance, projections still rarely incorporate estimates of uncertainty in many model parameters and uncertainty arising from internal model variability, and so typically underestimate the uncertainty (Stainforth et al. 2007a,b; Brekke et al. 2008; Fischer et al. 2011; Bretherton et al. 2012). However, any model will also always imperfectly represent the dynamics of the real world by failing to account for all the factors determining the dynamics by deliberately not incorporating known processes and by obviously not accounting for unknown processes. Thus, projections will inevitably become less reliable, and thus uncertain, the further into the future they are extended (Smith 2002; Parker 2011; Smith and Stern 2011).

Despite their limitations, model projections are used by governments, businesses, and scientists to help make decisions. However, the lack of clarity about their uncertainty limits the extent to which they can be treated with any more sophistication than simply a collection of plausible outcomes (Cox and Stephenson 2007; Moss et al. 2010; Oreskes et al. 2010; Kerr 2011; Maslin and Austin 2012). For example, land use managers wishing to assess how precipitation might change in the future are typically confronted with a wide range of predictions about the direction and timing of change (Stainforth et al. 2007b; New et al. 2007; Stainforth 2010; Maslin and Austin 2012). Decision makers do not depend on consistent or confident projections in order to make decisions (Polasky et al. 2011; Kunreuther et al. 2013) but analyses of uncertainty and estimates of confidence in projections provide a much a clearer understanding of the need

for, and likely consequences of, different decisions (Weaver et al. 2013; Lemos et al. 2012).

Of course, we do not suggest that Earth system modeling has not been useful in informing decision making. However, it has followed an approach that is better suited to exploring the plausible rather than identifying the probable. How can this situation be improved? How can we improve how model projections are made to provide clearer information to decision makers? We believe this can be facilitated by pursuing an alternative mode of model development; one that has the central aim of enabling the balance of models to be adjusted to allow balances of detail to be found that provide useful information for specific decisions.

DIFFERENT PERSPECTIVES ON THE FUTURE EVOLUTION OF MODELS.
There is a diverse variety of models of the Earth system (Fig. 1) because the level of detail has evolved over time to address different scientific questions. The predominant direction of model development to date has been the addition of more details, simulating an increasing number of different processes. In so doing, they have increased our understanding of the Earth system. They have also evolved to simulate processes at increasingly finer spatial resolutions (Claussen et al. 2002; Randall et al. 2007; Slingo et al. 2009), enabling phenomena to be simulated that only begin to occur at finer spatial scales (such as hurricanes). However, it is important to recognize that the same advances have also brought costs that can reduce predictive accuracy (by "accuracy" we mean the degree to which model predictions are centered on the dynamics and states of real-world phenomena rather than, for example, the number of real-world processes that the models depict). We will detail these costs below but, as an example, efforts have been biased toward adding details to individual models that are already technically unwieldy and intractable (Held 2005), rather than enabling uncertainty in the different aspects to be assessed, quantified, and incorporated into predictions and projections ("predictions": estimates of how the Earth system will change; "projections": estimates of how the Earth system might change under different scenarios; Weaver et al. 2013).

Perspectives differ on how much time and resources should be spent on adding yet more details. One perspective is that this is likely to be worthwhile because the model projections, incorporating more processes and at finer spatial resolutions, will become more realistic (Slingo et al. 2009; Gent et al. 2009; Slingo 2010). However, this is true only if our understanding of those new processes, as expressed in

model formulations and parameter values, is sufficient to enable the projections to reliably predict the dynamics of the system under future scenarios. For example, Oppenheimer et al. (2008) show that the continual refinement of model details can actually lead to "negative learning": where confidence improves over time to an answer that is different from the truth.

An alternative perspective is that continually adding details will unlikely deliver the desired improvements in decision-making capabilities (Dessai et al. 2009). Instead, it is proposed that the focus of making climate change decisions using projections should change from one that awaits sufficiently high confidence in what *will* happen before acting ("predict then act") to one that uses the projections as a set of plausible examples of what *might* happen to decide on how best to act in light of that knowledge and uncertainty (Stainforth et al. 2007b; Dessai et al. 2009; Kunreuther et al. 2013). Toward this goal, studies have investigated improving the process of decision making to make more robust decisions while incorporating information with different and diverse sources of uncertainty (Lempert and Collins 2007; Stakhiv 2011; Kunreuther et al. 2013; Weaver et al. 2013). These have led to the refinement of robust decision-making (RDM) methods within climate change decision making (Weaver et al. 2013). Existing model projections can already inform decision making under such frameworks, but for a limited range of scenarios and scales of spatial and temporal resolution. But they are not necessarily best suited for this purpose. To examine a range of scenarios, it would be more convenient to use models that could easily be simulated under many different scenarios (Weaver et al. 2013; Bretherton et al. 2012). Another focus has therefore been on how to redesign methodological frameworks for producing climate change projections so that they can be better targeted toward the needs of users (Weaver et al. 2013; Bretherton et al. 2012).

However, even if more robust decision-making frameworks are adopted, contemporary climate models still do not adequately convey important information that can be used to assess the confidence
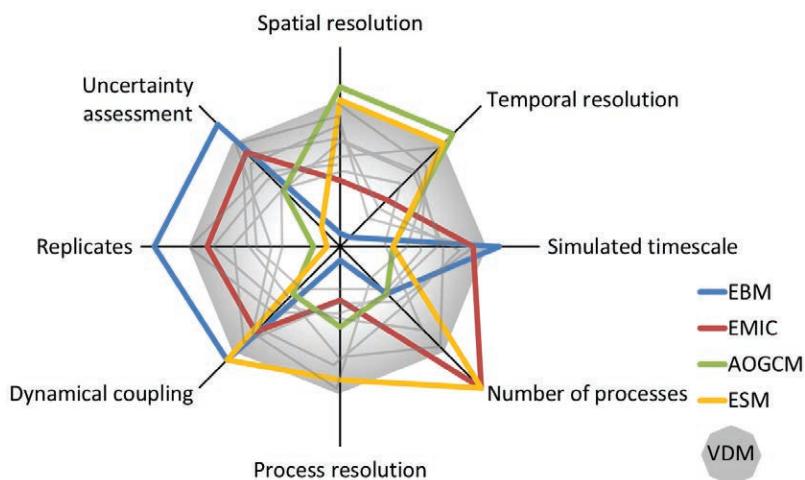

FIG. 1. **Toward more balanced models of the Earth system. There currently exists a wide spectrum of models of the Earth system (Randall et al. 2007): Earth system models (ESMs), energy balance models (EBMs) (Lenton 2000), ESMs of intermediate complexity (EMICS) (Claussen et al. 2002), and atmosphere–ocean general circulation models (AOGCMs) (Murphy et al. 2004). Each prioritizes the allocation of computational resources (the area of the gray polygon) differently. Enabling and performing fluid navigation of this model space to identify a suitable balance of details will be a key part to our new approach, termed a variable detail model (VDM), here. Gray lines represent hypothetical alternative resource allocations in the VDM.**

that can be placed in their projections. By "confidence" we mean an estimate of the probability of how the real-world system will behave. A related concept is the credibility of projections, which describes the assessment of a mechanistic model to reliably reproduce particular real-world phenomena (e.g., Brekke et al. 2008). Our perspective is that significant improvements to how climate models are developed are needed to provide more informative climate change projections. Such projections should go beyond being just a set of plausible outcomes to also convey a much more rigorous depiction of uncertainty in those projections than has been done to date. While any estimates of uncertainty will always become decreasingly reliable the further into the future they are projected (unless some proof can be given about the extent to which they truly bound real-world dynamics), they can still be seen as information-constrained predictions of the future, based on past evidence and understanding. Our focus here is on the methodological process of making the climate model projections themselves to better convey information and uncertainty relevant to the information being sought.

THE COSTS OF MODEL COMPLEXITY.
Contemporary practices are still limited in the extent to which they incorporate different sources

of uncertainty into model projections. Uncertainty arises from multiple sources: from uncertainty in the data used to initialize, parameterize, and evaluate models; from uncertainty in how adequately models represent reality; from differences in our scientific understanding of processes and how to represent them in models; from uncertainty in whether the model has been implemented correctly; and from uncertainty arising from simulated random events occurring in real-world processes (Stainforth et al. 2007a; Masson and Knutti 2011; Slingo and Palmer 2011). Yet Earth system models are currently so computationally demanding that only between 3 and 10 simulations per scenario were recommended for decadal forecasts and hindcasts to inform the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Meehl and Bony 2011). This is a very low number of replicates to characterize a distribution in something so dynamically rich as global climate. For example, Deser et al. (2012; see Daron and Stainforth 2013 for another example) estimated how many simulations were needed to detect anthropogenic changes in air temperature, precipitation, and sea level pressure predicted by a general circulation model with its between-run variations arising only from simulated random atmospheric and oceanic processes. They found that <10 simulations were insufficient to detect changes in precipitation and sea level pressure for most regions of Earth, even over multidecadal time windows, with only changes in surface air temperature being reliably detected for most regions of Earth with such few replicates. Yet this is just for one model, not even a full Earth system model, under one scenario, using one major source of variation (internal variability), and for a few global properties. Awareness of the need to understand the degree of replication required to capture internal variability is increasing and it is notable that some modeling groups conducted many more than 10 simulations per scenario for their phase 5 of the Coupled Model Intercomparison Project (CMIP5) simulations (Stainforth et al. 2005; World Climate Research Programme 2013).

There have been significant efforts to incorporate parameter uncertainty into general circulation model projections to assess how it influences uncertainty (e.g., Murphy et al. 2004; Stainforth et al. 2005; Collins et al. 2006; Knight et al. 2007; Piani et al. 2007; Murphy et al. 2009; Fischer et al. 2011; Sanderson 2011; Rowlands et al. 2012; Sexton et al. 2012; Sexton and Murphy 2012). However, these have predominantly not been extended to other components of the Earth system (e.g., biological components), to full Earth system models, or to the functional forms assumed to represent particular Earth system processes (structural uncertainty; Slingo and Palmer 2011; Maslin and Austin 2012). Quantifying this additional uncertainty and reporting its consequences is being more commonly performed for components of Earth system models but it is yet unknown whether fully incorporating these, or the additional uncertainty of whether policy recommendations will be implemented when making future projections, will swamp any anticipated improvements in predictive precision from increasing the realism of existing components (Smith and Stern 2011; Palmer 2012).

It is also currently unknown how the partitioning of detail amongst model components influences the confidence that can be placed in predictions or projections (Smith 2002; Oreskes et al. 2010). For example, increasing the spatial resolution of a specific atmospheric physics model is justifiable in order to predict atmospheric dynamics more precisely (Shaffrey et al. 2009; Palmer 2012) but if its computational requirements restrict the inclusion of other details then the model may be less accurate than had an alternative atmospheric model formulation been adopted to allow other component processes to be represented more accurately. The adequacy of a model structure, including the level of detail, can be assessed by the degree to which predictions can recapture the known (and relevant) dynamics of interest, although such assessments are not in widespread use (Judd et al. 2008; Le Bauer et al. 2013; Smith et al. 2013).

It is also worth being aware that, fundamentally, more detailed models can make worse predictions than simpler models. This could occur, for example, if mechanisms or parameters are included as if they apply generally where they are in fact applicable to a much narrower range of circumstances. This is more likely to occur if models are tested against unchanging datasets because it raises the chances of a hypothetical mechanism being found that explains the variance in the specific dataset. This is known as overfitting, in which overly detailed models make worse predictions than simpler models through being overly tuned to the specifics of the calibration or training datasets (Bishop 2006; Crout et al. 2009; Masson and Knutti 2013).

Inadequate justification of the balance of details in models of the Earth system ultimately makes it difficult to meaningfully compare the projections of different models. Intercomparison exercises [such as the current Climate Model Intercomparison Project (Doblas-Reyes et al. 2011; Meehl and Bony 2011; Stouffer et al. 2011)] illustrate the degree

of consistency between projections. However, the diverse and incompatible approaches to formulating and simulating models cause major difficulties in allowing detailed intercomparisons and the origins of differences in projections to be understood. The fact that some models used in intercomparisons are related, either by ancestry or by adoption of common formulations, means that the variation between their projections may be an overly conservative and biased estimate of the actual uncertainty (Stainforth et al. 2007a; Tebaldi and Knutti 2007; Dessai et al. 2009; Schwalm et al. 2009; Frank et al. 2010; Knutti et al. 2010; Smith and Stern 2011; Rowlands et al. 2012; Bishop and Abramowitz 2013). A rigorous analysis of the large ensemble from the general circulation model available at www.climateprediction.net reassuringly highlighted that variation due to hardware and software differences had relatively small effects on variation in model projections and that the effects arising from parameter variation were much larger (Knight et al. 2007). While extremely insightful, this analysis was performed for just one model structure and so it is largely unknown whether these findings apply more generally across a wider range of models.

A lack of detailed assessments of consistency of model components with historical observations, and the contributions of the uncertainties associated with model components to uncertainty in predictions, makes it unclear how best to proceed with future refinements. What are the most important and most reducible sources of uncertainty? Which components should be prioritized for refinement? What new data do we need to achieve this? Model development practices to date have insufficiently quantified the contributions of known sources of uncertainty to enable such questions to be addressed, although a number of model component intercomparison projects have been conducted, are planned, or are underway to help address some these issues, such as the Program for the Intercomparison of Land surface Parameterization Schemes (PILPS; Henderson-Sellers et al. 1996). Moreover, on top of the multiple necessary improvements to Earth system models, the research community will need to decide how best to make them in light of limited computational resources (Shukla et al. 2010; Palmer 2012).

AN ALTERNATIVE APPROACH. It is increasingly recognized that the current generation of model projections often do not provide decision makers, scientists, or climate model output users in general with the specific information they need (Corell et al. 2009; Dessai et al. 2009; Oreskes et al. 2010; Kerr 2011;

Lemos et al. 2012; Lemos and Rood 2012; Maslin and Austin 2012; Kunreuther et al. 2013). For example, those making decisions in relation to future land use planning might wish to understand the diversity of risks (e.g., floods) posed to potential developments (e.g., flood barriers or wind farms) as a consequence of climate change (Weaver et al. 2013) but are confronted with a wide range of projections that differ in relevance, resolution, parent model, and uncertainty (to name a few) without clear information on their credibility or uncertainty. Recent advances in decision theory have gone a long way to enabling rational decisions in light of projected climate changes, irrespective of how models are developed (Polasky et al. 2011; Liverman et al. 2010; Kunreuther et al. 2013; Weaver et al. 2013). For example, instead of decision makers awaiting confident estimates of the likelihood of particular events happening in future before acting (e.g., the chances that storm surges in a particular port will exceed 5 m), decision theory now provides robust ways of estimating the costs and benefits of acting now given the range of costs associated with different plausible events (Weaver et al. 2013). However, given the costs arising from contemporary model development practices, it is also clear that a number of changes to those practices would not only enable projections to provide more useful information for decision makers, such as providing more complete estimates of uncertainty, but also better target the needs of a much wider community of climate model users (Liverman et al. 2010; Bretherton et al. 2012). We therefore recommend here changes to model development practices to better suit the needs of those aiming to make more informative climate projections. In Table 1 we summarize the differences between the approach we recommend and contemporary practices.

Given the costs of model complexity we think there should be a greater emphasis on adopting models that are at least simpler than the current generation of extremely computationally demanding Earth system models to permit more informative uncertainty quantification. Such quantification should be conducted to reflect the sensitivity of model projections to the most important known sources of uncertainty in relation to the phenomena being targeted for prediction. Such uncertainty assessments are becoming more common in relation to parameter uncertainty and internal variability and the results of these predominantly argue for many more replicates than typically conducted for the most complex models (Stainforth et al. 2005; Sanderson 2011; Rowlands et al. 2012; Sexton and Murphy 2012; Deser et al. 2012). However,

uncertainty assessments also need to be extended to enable structural uncertainties to be assessed in more informative ways than is achieved to date through model intercomparisons. Ideally, structural uncertainty assessments would be part of the uncertainty quantification conducted by any one modeling team, incorporating the effects of alternative formulations for internal processes (e.g., alternative ways of representing vegetation fires) or even for entirely different formulations (e.g., comparing simpler models to more detailed models). Thus, when projections are served to users, they can be accompanied by a more rigorous exposition of the sensitivity of relevant model predictions to these different sources of uncertainty. However, such information should always be delivered with the caveat that any uncertainty or probability projection is increasingly likely to become misleading the further into the future it is projected.

Data assimilation and parameter inference methods will play key roles in future approaches to quantifying uncertainty in how the model reflects present day and historical phenomena. Such methods will be important for propagating uncertainty into projections and enabling assessments of the value of alternative model formulations in terms of precision, accuracy, and overall confidence in how well the model captures reality (Vrught et al. 2005; Berliner and Wikle 2007; Scholze et al. 2007; Sexton and

Murphy 2012; Le Bauer et al. 2013; Smith et al. 2013). New studies examining the tradeoffs between the level of model detail and the ability to quantify uncertainty would be informative in relation to this (Smith 2002; Ferro et al. 2012; Palmer 2012). Formal probabilistic methods (i.e., Bayesian inference) are particularly well suited for comparing models with data and making projections that incorporate estimates of uncertainty, so would be particularly attractive for our proposed approach (Kass and Raftery 1995; Berger et al. 1999; Kennedy and O'Hagan 2001; Oakley and O'Hagan 2002; Berliner 2003). So far, these have proven computationally unfeasible for the most detailed models (Oreskes et al. 1994; Smith and Stern 2011; van Oijen et al. 2011; Palmer 2012), but this could be addressed on the short term in a number of ways. For instance, Bayesian emulators of detailed models could be employed to make probabilistic predictions based on limited runs of the computer code (Kennedy and O'Hagan 2001; Oakley and O'Hagan 2002), or the number of details could even be restricted to a level where their suitability could be assessed using Bayesian methods. However other, non-Bayesian methods to uncertainty quantification could also be used to provide useful information, such as the adjoint method—a popular choice for investigating the parameter sensitivity of computationally intensive models (Courtier et al. 1993).

TABLE 1. Contrasts between the traditional modes of model development and the approach we advocate here, summarizing the points made in the main text.

Current tendencies	Proposed changes
Oriented toward improving predictive precision of components independently	Focus improving components to increase overall confidence in model projections
Oriented toward precision of individual predictions over estimation of uncertainty	Bias toward accuracy of predictions, which requires uncertainty estimation
Detail incorporated because potentially important	All detail justified by relevance, empirical evidence, and accuracy metrics
Most detailed models used to make projections, with confidence assessment made using separate analyses	Focus on adopting the most informative overall balance of details, including measures of uncertainty
Most parameters and processes defined prior to model construction	Many more parameters and processes justified through data assimilation, with prior assumptions clearly stated and accessible
Lack of ability to assess distribution of uncertainty across model and identify where inconsistencies exist	Probabilistic accounting for model uncertainty and its propagation into predictions
Incompatibility between model versions and between models of different institutions	Emphasis on enabling interoperability of components to enable identification of suitable balance of details
Intercomparison more important than intercompatibility	Intercompatibility necessary for quantitative intercomparison
Better suited to scientific exploration of the plausible	Better suited to identification of the probable, with more thorough accounting for uncertainty

Climate models convey more confidence in projected phenomena when those phenomena arise in multiple different models and the reasons for them occurring are understood to be plausibly consistent with reality (Pachauri and Reisinger 2007; Held 2005). The classic example is the consistent prediction from all modes of abstraction—from simple physical principles to multiple complex climate models, that increasing greenhouse gas concentrations leads to a global warming response (of course neither guarantee this will actually occur in reality). New approaches to climate modeling to inform decision makers need improved ways to demonstrate the consistency of projections under the different sources of uncertainty described above, but also enable the reasons behind the occurrence of projected phenomena to be investigated, understood, and assessed for real-world relevance. Structuring the diversity of possible model details hierarchically is one way to facilitate this, which encouragingly was also the first recommendation of the National Academy of Sciences' "National Strategy for Advancing Climate Modelling" (Bretherton et al. 2012; "Evolve to a common national software infrastructure that supports a diverse hierarchy of different models for different purposes . . ."). The hierarchical organization of model details is particularly helpful for enabling the reasons for particular model predictions to be understood and then tested (Held 2005). Emergent phenomena can be studied at the simplest possible level and the reasons for their emergence investigated without having to also simulate and account for excessive detail. Thus, one of the most useful (and challenging) improvements that we recommend is to develop widely applicable hierarchical descriptions of Earth system processes to provide frameworks within which models of the Earth system can be formulated and characterized, both in terms of their structure and in terms of their predictions and projections (this was also recommended by Held 2005).

The balance of detail and sources of uncertainty considered relevant to a problem will obviously depend on the problem being addressed. For example, United Nations Framework Convention on Climate Change decision makers recommending global climate mitigation decisions may require models with a different balance of details (in terms of spatial resolution and numbers of processes) than water agencies aiming to plan new water supply and wastewater management systems for their region meet the demands of the next 25 years (Rogelj et al. 2013; Weaver et al. 2013). Thus, computational methods are also needed that facilitate the assessment and adjustment of the overall balance of model details relative to the questions posed, current knowledge, data, and uncertainty. Computational methods to enable the adjustment of details within the same modeling framework have already been developed for individual families of models to meet this requirement (e.g., the Met Office Unified Model; Pope et al. 2007), although these need to also be able to incorporate estimates of uncertainty in model components and parameters, so that the various costs and benefits of adopting different levels of abstraction discussed above can be assessed. These methods will also need to be extended to apply beyond an individual family of models, as described above, to allow the quantification and assessment of structural uncertainty. Adding to these challenges is the requirement (at least occasionally) to conduct assessments of the overall balance of details at a systemic level. This is for several reasons. First, because the different components are coupled through feedbacks, the coupling of different components might be necessary to assess the accuracy with which they can predict important emergent phenomena. Second, systemic assessments can enable the detection of logical inconsistencies between the predictions of different components. Third, systemic assessment can also help to avoid the development of details of any one area in such a way as to detrimentally affect the accuracy of the overall model or the assessment of its accuracy. Fourth, such assessments can be used to help identify the most important reducible sources of uncertainty. Such approaches are being developed for numerical weather prediction models, where the poorest performing model features can be identified (Judd et al. 2008).

Achieving the sort of "balanced complexity" modeling paradigm (Fig. 1) we advocate above will obviously be extremely challenging, for both sociological and technological reasons. Modeling groups adopt different methodological approaches and have differing incentives to adopt cross-institutional standards. Much model development to date is conducted in government-funded research institutions where there is typically an incentive structure for individuals and research teams to produce research findings within a period of months to years that is publishable in peer-reviewed journals. The high resource and technical requirements to build even one detailed Earth system model mean that individuals and groups are reluctant to undertake projects involving radical modifications to their modeling architectures because of the likely time and financial costs involved. However, just as the increasing recognition of the need to conduct model intercomparisons and benchmarking has promoted

standards in compatibility data and model outputs we believe that the increasing need to utilize climate model information in decision-making context will promote methods that allow assessment of the costs and benefits of adopting different balances of detail for providing useful information.

So how could modeling systems be engineered to enable the methodological improvements described? One of the first steps will be to develop new, or evolve the existing, online repositories for model components, whole models, driver and assessment data, and model outputs (e.g., the Earth system grid; Williams et al. 2009). Such repositories should enable access to components independently from their original parent models so that other research groups can assess the implications of alternative formulations for that component. Similar assessments could be made in relation to the driver and assessment datasets as well as model structures. To facilitate the exchange of model components, future components could be developed in such a way that facilitates their use within alternative model structures. This strategy was recently employed by Smith et al. (2013) to develop a global terrestrial carbon model with the intention to facilitate investigations into the costs and benefits of alternative model components and formulations for predicting global terrestrial carbon. In that study the modeling framework included code libraries that enabled model components to obtain the data they require to make outputs from online databases, from local computers, or from other model components, depending on the structural information specified. This facilitated rapid experimentation with a wide variety of model structures.

Investigations into the effects of alternative model formulations will also benefit from adopting conventions for the description of models—their structures, components, and use histories (Dunlap et al. 2008). This should also help to minimize or eliminate reducible sources of uncertainty associated with the technical implementation of models. Uncertainty in model projections also exists because of differences in datasets, algorithms, methods, models, and simulation architectures used by different research groups. The importance of these specific details cannot practically be assessed among very different models (though see Knight et al. 2007). Thus, any new approach will benefit from enabling scientists anywhere to access pools of models and datasets and verify whether or not a change was an improvement (by various measures of performance).

The comparison of different model structures and component formulations could also be aided by the adoption of programming languages that make it easier for the intentions of the code to be understood. Functional modeling languages (Pedersen and Phillips 2009) allow for succinct and functional descriptions of models. This would not only aid in conveying the intended purpose of the code but would also aid the translation of the same underlying model to different coding languages (e.g., FORTRAN versus C++). This is one promising way of allowing interoperability between the components of models written by different institutions when it is inevitable that there would be some resistance to initiatives to adopt standards in model development. Enabling models to work with data, parameters, and predictions as probability distributions, just as naturally as they use with constants today, would also greatly facilitate modeling with uncertainty. Probabilistic programming languages are a relatively recent area of research and development aimed at facilitating the use of probability distributions and machine learning in general applications (Bishop 2013). Their application in Earth system modeling could simplify the process of computing with probability distributions. Recent developments in functional probabilistic programming languages could enable modelers to combine the benefits of both functional and probabilistic programming languages (Bhat et al. 2013). Enabling the continual quantification, storage, and retrieval of uncertainty associated with model components and projections will also require much larger computer memory requirements; this could be facilitated with online data storage and retrieval capabilities.

One of the benefits of adopting a hierarchical approach to defining the relationships amongst model components is that it should facilitate model reconstruction from simple representations to avoid becoming locked into one model or modeling approach. Ideally, model developers would be able to identify all top-level components currently known to be relevant to a particular set of phenomena (one of which, for example, might be the sea level in 100 years' time) and then, starting from the simplest possible representations of each of these, critically assess and reassess the adequacy of the level of detail used to model them. Further details, in terms of new model components (Fig. 1), would be added if justifiable. This approach to model development would not only lead to better predictions for less computing time, but also tend to check the sociological imbalances inherent to current Earth system science, helping to direct intellectual effort and scientific funding toward those components that are the *least* understood and most useful in relation to the phenomena being targeted for prediction.

Any new modeling approach to get informative projections to users on demand and operate within formal or informal decision-making frameworks will also require the ability for researchers to specify, combine, and compare projections from multiple models—some perhaps projecting on demand and others obtained from archives to suit their analyses (Bretherton et al. 2012; Weaver et al. 2013). Such systems should be designed to allow a much broader community of experts to contribute to model development and use, including some that have had little influence on model development to date. It should also permit the coproduction of new climate information from climate modelers, domain experts, and decision makers—to enable a balance to be struck between providing the information that decision makers want and the information that scientists think decision makers need to know (Lemos and Rood 2012).

CONCLUSIONS. It is now time to build from the wealth of modes of abstraction of the Earth system developed so far, on the wealth of data in existence, and on advances in computation and statistics to build climate models that deliver much more predictive information for users. A key step toward this is to enable models to be built that include much more robust estimates of uncertainty, which in turn guides where scientific and computational resources need to be directed in order to reduce uncertainties further. Combining adaptive hierarchical modeling frameworks with assessments of the uncertainty in model formulations and projections will enable much better targeted explorations of model-detail space and allow urgent questions to be answered in a much more timely and reliable way.

REFERENCES

Berger, J. O., B. Liseo, and R. L. Wolpert, 1999: Integrated likelihood methods for eliminating nuisance parameters. *Stat. Sci.,* **14,** 1–28, doi:10.1214/ss/1009211804.

Berliner, M. L., 2003: Uncertainty and climate change. *Stat. Sci.,* **18,** 411–533, doi:10.1214/ss/1081443227.

——, and C. K. Wikle, 2007: A Bayesian tutorial for data assimilation. *Physica D,* **230,** 1–16, doi:10.1016/j.physd.2006.09.017.

Bhat, S., J. Borgstroem, A. D. Gordon, and C. Russo, 2013: Deriving probability density functions from probabilistic functional programs. *Tools and Algorithms for the Construction and Analysis of Systems,* N. Piterman and S. Smolka, Eds., Lecture Notes in Computer Science, Vol. 7795, Springer, 508–522.

Bishop, C. H., and G. Abramowitz, 2013: Climate model dependence and the replicate Earth paradigm. *Climate Dyn.,* **41,** 885–900, doi:10.1007/s00382-012-1610-y.

Bishop, C. M., 2006: *Pattern Recognition and Machine Learning.* Springer, 738 pp.

——, 2013: Model-based machine learning. *Philos. Trans. Roy. Soc. London,* **A371,** 1–17, doi:10.1098/rsta.2012.0222.

Brekke, L. D., M. D. Dettinger, E. P. Maurer, and M. Anderson, 2008: Significance of model credibility in estimating climate projection distributions for regional hydroclimatological risk assessments. *Climatic Change,* **89,** 371–394, doi:10.1007/s10584-007-9388-3.

Bretherton, C., and Coauthors, 2012: *A National Strategy for Advancing Climate Modeling.* The National Academies Press, 294 pp.

Claussen, M., and Coauthors, 2002: Earth system models of intermediate complexity: Closing the gap in the spectrum of climate system models. *Climate Dyn.,* **18,** 579–586, doi:10.1007/s00382-001-0200-1.

Collins, M., B. B. B. Booth, G. R. Harris, J. M. Murphy, D. M. H. Sexton, and M. J. Webb, 2006: Towards quantifying uncertainty in transient climate change. *Climate Dyn.,* **27,** 127–147, doi:10.1007/s00382-006-0121-0.

Corell, R. W., and Coauthors, 2009: *Informing Decisions in a Changing Climate.* National Academies Press, 200 pp.

Courtier, P., J. Derber, R. Errico, J. Loiuis, and T. Vukicevic, 1993: Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus,* **45,** 342–357, doi:10.1034/j.1600-0870.1993.t01-4-00002.x.

Cox, P., and D. A. Stephenson, 2007: Changing climate for prediction. *Science,* **317,** 207–208, doi:10.1126/science.1145956.

Crout, N. M. J., D. Tarsitano, and A. T. Wood, 2009: Is my model too complex? Evaluating model formulation using model reduction. *Environ. Modell. Software,* **24,** 1–7, doi:10.1016/j.envsoft.2008.06.004.

Daron, J. D., and D. A. Stainforth, 2013: On predicting climate under climate change. *Environ. Res. Lett.,* **8,** 034021, doi:10.1088/1748-9326/8/3/034021.

Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role

of internal variability. *Climate Dyn.,* **38,** 527–546, doi:10.1007/s00382-010-0977-x.

Dessai, S., M. Hulme, R. Lampert, and R. Pielke Jr., 2009: Do we need better predictions to adapt to a changing climate? *Eos, Trans. Amer. Geophys. Union,* **90,** 111–112, doi:10.1029/2009EO130003.

Doblas-Reyes, F. J., G. J. van Oldernborgh, J. Garcia-Serrano, H. Pohlmann, A. A. Scaife, and D. Smith, 2011: CMIP5 near-term climate prediction. *CLIVAR Exchanges,* No. 56, International CLIVAR Project Office, Southampton, United Kingdom, 8–11.

Dunlap, R., and Coauthors, 2008: Earth system curator: Metadata infrastructure for climate modelling. *Earth Sci. Inf.,* **1,** 131–149, doi:10.1007/s12145-008-0016-1.

Edwards, P. N., 2010: *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.* MIT Press, 552 pp.

Ferro, C. A. T., T. E. Jupp, F. H. Lambert, C. Huntingford, and P. M. Cox, 2012: Model complexity versus ensemble size: Allocating resources for climate prediction. *Philos. Trans. Roy. Soc. London,* **A370,** 1087–1099, doi:10.1098/rsta.2011.0307.

Fischer, E. M., D. M. Lawrence, and B. M. Sanderson, 2011: Quantifying uncertainties in projections of extremes—A perturbed land surface parameter experiment. *Climate Dyn.,* **37,** 1381–1398, doi:10.1007/s00382-010-0915-y.

Frank, D. C., J. Esper, C. C. Raible, U. Büntgen, V. Trouet, B. Stocker, and F. Joos, 2010: Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. *Nature,* **463,** 527–530, doi:10.1038/nature08769.

Gent, P., and Coauthors, 2009: Community Earth System Model science plan: 2009–2015. CESM Scientific Steering Committee, 31 pp.

Held, I. M., 2005: The gap between simulation and understanding in climate modeling. *Bull. Amer. Meteor. Soc.,* **86,** 1609–1614, doi:10.1175/BAMS-86-11-1609.

Henderson-Sellers, A., K. McGuffie, and A. J. Pitman, 1996: The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS): 1992 to 1995. *Climate Dyn.,* **12,** 849–859, doi:10.1007/s003820050147.

Judd, K., C. A. Reynolds, T. E. Rosmond, and L. A. Smith, 2008: The geometry of model error. *J. Atmos. Sci.,* **65,** 1749–1772, doi:10.1175/2007JAS2327.1.

Kass, R. E., and A. E. Raftery, 1995: Bayes factors. *J. Amer. Stat. Soc.,* **90,** 773–795, doi:10.1080/01621459.1995.10476572.

Kennedy, M. C., and A. O'Hagan, 2001: Bayesian calibration of computer models (with discussion). *J. Roy. Stat. Soc.,* **63B,** 425–464, doi:10.1111/1467-9868.00294.

Kerr, R. A., 2011: Time to adapt to a warming world, but where's the science? *Science,* **334,** 1052–1053, doi:10.1126/science.334.6059.1052.

Knight, C. G., and Coauthors, 2007: Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models. *Proc. Nat. Acad USA,* **104,** 12 259–12 264, doi:10.1073/pnas.0608144104.

Knutti, R., R. Furrer, J. Cermak, and G. A. Meehl, 2010: Challenges in combining projections from multiple climate models. *J. Climate,* **23,** 2739–2758, doi:10.1175/2009JCLI3361.1.

Kunreuther, H., G. Heal, M. Allen, O. Edenhofer, C. B. Field, and G. Yohe, 2013: Risk management and climate change. *Nat. Climate Change,* **3,** 447–450, doi:10.1038/nclimate1740.

Le Bauer, D. S., D. Wang, K. T. Richter, C. C. Davidson, and M. C. Dietze, 2013: Facilitating feedbacks between field measurements and ecosystem models. *Ecol. Monogr.,* **83** (2), 133–154, doi:10.1890/12-0137.1.

Lemos, M. C., and R. B. Rood, 2012: Climate projections and their impact on policy and practice. *Wiley Interdiscip. Rev.: Climatic Change,* **1,** 670–682.

——, C. J. Kirchhoff, and V. Ramprasad, 2012: Narrowing the climate information usability gap. *Nat. Climate Change,* **2,** 789–794, doi:10.1038/nclimate1614.

Lempert, R. J., and M. T. Collins, 2007: Managing the risk of uncertain threshold responses: Comparison of robust, optimum, and precautionary approaches. *Risk Anal.,* **27,** 1009–1026, doi:10.1111/j.1539-6924.2007.00940.x.

Lenton, T. M., 2000: Land and ocean carbon cycle feedback effects on global warming in a simple Earth system model. *Tellus,* **52,** 1159–1188, doi:10.1034/j.1600-0889.2000.01104.x.

Liverman, D., and Coauthors, 2010: *Informing an Effective Response to Climate Change.* The National Academies Press, 348 pp.

Maslin, M., and P. Austin, 2012: Climate models at their limit? *Nature,* **486,** 183–184, doi:10.1038/486183a.

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

——, and ——, 2013: Predictor screening, calibration, and observational constraints in climate model ensembles: An illustration using climate sensitivity. *J. Climate,* **26,** 887–898, doi:10.1175/JCLI-D-11-00540.1.

Meehl, G. A., and S. Bony, 2011: Introduction to CMIP5. *CLIVAR Exchanges,* No. 56, International CLIVAR Project Office, Southampton, United Kingdom, 4–5.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research

and assessment. *Nature,* **463,** 747–756, doi:10.1038/nature08823.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature,* **430,** 768–772, doi:10.1038/nature02771.

——, and Coauthors, 2009: UK Climate Projections science report: Climate change projections. Met Office Rep., 10 pp.

New, M., A. Lopez, S. Dessai, and R. Wilby, 2007: Challenges in using probabilistic climate change information for impact assessments: An example from the water sector. *Philos. Trans. Roy. Soc. London,* **A365,** 2117–2131, doi:10.1098/rsta.2007.2080.

Oakley, J. E., and A. O'Hagan, 2002: Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika,* **89,** 769–784, doi:10.1093/biomet/89.4.769.

Oppenheimer, M., B. C. O'Neill, and M. Webster, 2008: Negative learning. *Climatic Change,* **89,** 155–172, doi:10.1007/s10584-008-9405-1.

Oreskes, N., K. Schrader-Frechette, and K. Belitz, 1994: Verification, validation, and confirmation of numerical models in the earth sciences. *Science,* **263,** 641–646, doi:10.1126/science.263.5147.641.

——, D. A. Stainforth, and L. A. Smith, 2010: Adaptation to global warming: Do climate models tell us what we need to know? *Philos. Sci.,* **77,** 1012–1028, doi:10.1086/657428.

Pachauri, R. K., and A. Reisinger, Eds., 2007: Climate Change 2007: Synthesis report. IPCC Rep., 73 pp.

Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.,* **138,** 841–861, doi:10.1002/qj.1923.

Parker, W. S., 2011: When climate models agree: The significance of robust model predictions. *Philos. Sci.,* **78,** 579–600, doi:10.1086/661566.

Pedersen, M., and A. Phillips, 2009: Towards programming languages for genetic engineering of living cells. *J. Roy. Soc. Interface,* **6,** S437–S450, doi:10.1098/rsif.2008.0516.focus.

Piani, C., B. Sanderson, F. Giorgi, D. J. Frame, C. Christensen, and M. R. Allen, 2007: Regional probabilistic climate forecasts from a multithousand, multimodel ensemble of simulations. *J. Geophys. Res.,* **112,** D24108, doi:10.1029/2007JD008712.

Polasky, S., S. R. Carpenter, C. Folke, and B. Keeler, 2011: Decision-making under great uncertainty: Environmental management in an era of global change. *Trends Ecol. Evol.,* **26,** 398–404, doi:10.1016/j.tree.2011.04.007.

Pope, V., and Coauthors, 2007: The Met Office Hadley Centre climate modelling capability: The competing requirements for improved resolution, complexity and dealing with uncertainty. *Philos. Trans. Roy. Soc. London,* **A365,** 2635–2657, doi:10.1098/rsta.2007.2087.

Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis,* S. Solomon et al., Eds., Cambridge University Press, 589–662.

Rogelj, J., D. L. McCollum, A. Reisinger, M. Meinshausen, and K. Riahi, 2013: Probabilistic cost estimates for climate change mitigation. *Nature,* **493,** 79–83, doi:10.1038/nature11787.

Rowlands, D. J., and Coauthors, 2012: Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nat. Geosci.,* **5,** 256–260, doi:10.1038/ngeo1430.

Sanderson, B. M., 2011: A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. *J. Climate,* **24,** 1362–1377, doi:10.1175/2010JCLI3498.1.

Scholze, M., T. Kaminski, P. Rayner, W. Knorr, and R. Giering, 2007: Propagating uncertainty through prognostic carbon cycle data assimilation system simulations. *J. Geophys. Res.,* **112,** D17305, doi:10.1029/2007JD008642.

Schwalm, C. R., and Coauthors, 2009: A model-data intercomparison of CO_2 exchange across North America: Results from the North American Carbon Program site synthesis. *J. Geophys. Res.,* **115,** G00H05, doi:10.1029/2009JG001229.

Sexton, D. M. H., and J. M. Murphy, 2012: Multivariate probabilistic projections using imperfect climate models. Part II: Robustness of methodological choices and consequences for climate sensitivity. *Climate Dyn.,* **38,** 2543–2558, doi:10.1007/s00382-011-1209-8.

——, ——, M. Collins, and M. J. Webb, 2012: Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. *Climate Dyn.,* **38,** 2513–2542, doi:10.1007/s00382-011-1208-9.

Shaffrey, L. C., and Coauthors, 2009: U.K. HiGEM: The New U.K. High-Resolution Global Environment Model—Model description and basic evaluation. *J. Climate,* **22,** 1861–1896, doi:10.1175/2008JCLI2508.1.

Shukla, J., and Coauthors, 2010: Toward a new generation of world climate research and computing facilities. *Bull. Amer. Meteor. Soc.,* **91,** 1407–1412, doi:10.1175/2010BAMS2900.1.

Slingo, J., 2010: Met Office science strategy (2010–2015). Met Office, 15 pp.

——, and T. Palmer, 2011: Uncertainty in weather and climate prediction. *Philos. Trans. Roy. Soc. London,* **A369,** 4751–4767, doi:10.1098/rsta.2011.0161.

——, and Coauthors, 2009: Developing the next-generation climate system models: Challenges and achievements. *Philos. Trans. Roy. Soc. London,* **A367,** 815–831, doi:10.1098/rsta.2008.0207.

Smith, L. A., 2002: What might we learn from climate forecasts? *Proc. Natl. Acad. Sci. USA,* **99,** 2487–2492, doi:10.1073/pnas.012580599.

——, and N. Stern, 2011: Uncertainty in science and its role in climate policy. *Philos. Trans. Roy. Soc. London,* **A369,** 4818–4841.

Smith, M. J., D. W. Purves, M. C. Vanderwel, V. Lyutsarev, and S. Emmott, 2013: The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties. *Biogeosciences,* **10,** 583–606, doi:10.5194/bg-10-583-2013.

Stainforth, D., 2010: Estimating uncertainty in future climate projections. Workshop on nonstationarity, hydrologic frequency analysis, and water management, J. R. Olsen, J. Kiang, and R. Waskom, Eds., Colorado Water Institute Information Series 109, 175–182.

——, and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature,* **433,** 403–406, doi:10.1038/nature03301.

——, M. R. Allen, E. R. Tredger, and L. A. Smith, 2007a: Confidence, uncertainty and decision-support relevance in climate predictions. *Philos. Trans. Roy. Soc. London,* **A365,** 2145–2161, doi:10.1098/rsta.2007.2074.

——, T. E. Downing, R. Washington, A. Lopez, and M. New, 2007b: Issues in the interpretation of climate model ensembles to inform decisions. *Philos. Trans. Roy. Soc. London,* **A365,** 2163–2177, doi:10.1098/rsta.2007.2073.

Stakhiv, E. Z., 2011: Pragmatic approaches for water management under climate change uncertainty. *J. Amer. Water. Res. Assoc.,* **47,** 1183–1196, doi:10.1111/j.1752-1688.2011.00589.x.

Stouffer, R. J., K. E. Taylor, and G. A. Meehl, 2011: CMIP5 long-term experimental design. *CLIVAR Exchanges,* No. 56, International CLIVAR Project Office, Southampton, United Kingdom, 5–7.

Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc. London,* **A365,** 2053–2075, doi:10.1098/rsta.2007.2076.

van Oijen, M., and Coauthors, 2011: A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest. *Agric. For. Meteor.,* **151,** 1609–1621, doi:10.1016/j.agrformet.2011.06.017.

Vrught, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten, 2005: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resour. Res.,* **41,** W01017, doi:10.1029/2004WR003059.

Weaver, C. P., R. J. Lempert, C. Brown, J. A. Hall, D. Revell, and D. Sarewitz, 2013: Improving the contribution of climate model information to decision making: The value and demands of robust decision frameworks. *Wiley Interdiscip. Rev.: Climate Change,* **4,** 39–60.

Williams, D. N., and Coauthors, 2009: The Earth System Grid: Enabling access to multimodel climate simulation data. *Bull. Amer. Meteor. Soc.,* **90,** 195–205, doi:10.1175/2008BAMS2459.1.

World Climate Research Programme, cited 2013: CMIP5 - data access - data portal. [Available online at http://cmip-pcmdi.llnl.gov/cmip5/data_portal.html.]