BAMS Article

For a Pluralism of Climate Modeling Strategies

Marina Baldissera Pacchetti,[a,b] Julie Jebeile,[c,d,e] and Erica Thompson[a]

KEYWORDS:
Uncertainty;
Climate models;
General circulation;
models;
Machine learning;
Experimental
design;
Model evaluation/
performance

ABSTRACT: The continued development of general circulation models (GCMs) toward increasing resolution and complexity is a predominantly chosen strategy to advance climate science, resulting in channeling of research and funding to meet this aspiration. Yet many other modeling strategies have also been developed and can be used to understand past and present climates, to project future climates, and ultimately to support decision-making. We argue that a plurality of climate modeling strategies and an equitable distribution of funding among them would be an improvement on the current predominant strategy for informing policymaking. To support our claim, we use a philosophy of science approach to compare the increasing resolution and complexity of general circulation models with three alternate examples: the use of machine learning techniques, the physical climate storyline approach, and Earth system models of intermediate complexity. We show that each of these strategies prioritizes a particular set of methodological aims, among which are empirical agreement, realism, comprehensiveness, diversity of process representations, inclusion of the human dimension, reduction of computational expense, and intelligibility. Thus, each strategy may provide adequate information to support different specific kinds of research and decision questions. We conclude that, because climate decision-making consists of different kinds of questions, many modeling strategies are all potentially useful and can be used in a complementary way.

SIGNIFICANCE STATEMENT: The intended purpose of the paper is to argue for the simultaneous and equitable development of modeling strategies in climate science. This should replace the dominant strategy that is the continued development of resolution and complexity in general circulation models (GCMs). We argue that different modeling strategies, including storylines, machine learning techniques, and Earth system models of intermediate complexity, are complementary to inform policymaking due to the distinct values they prioritize. Importantly, this paper promotes equitable (not necessarily equal) distribution of funding among these strategies, while there is a research political tendency to prefer high-resolution complex modeling.

DOI: 10.1175/BAMS-D-23-0169.1

Corresponding author: Erica Thompson, erica.thompson@ucl.ac.uk

Manuscript received 4 July 2023, in final form 8 May 2024, accepted 22 May 2024

AFFILIATIONS: [a] Department of Science, Technology, Engineering, and Public Policy, University College of London, London, United Kingdom; [b] Barcelona Supercomputing Center, Barcelona, Spain; [c] Institute of Philosophy, University of Bern, Bern, Switzerland; [d] Oeschger Center for Climate Change Research, Bern, Switzerland; [e] CNRM UMR 3589, Météo-France/CNRS, Centre National de Recherches Météorologiques, Toulouse, France

1. Introduction

In the last 30 years, the concentration of research effort upon increasing resolution and complexity in general circulation models (GCMs) has been a predominant modeling strategy in climate science (Palmer and Stevens 2019). The hope is that this fundamental research development will both increase process understanding and, in time, lead to reliable fine-grained projections, which in turn can support decision-making and political action at all levels, from the United Nations Framework Convention on Climate Change to local climate services (Stevens et al. 2024). To support this research, funding has been channeled into the physical and computational sciences (Overland and Sovacool 2020) and particularly into the high-resolution models developed and maintained by larger modeling centers with high-performance computing facilities. In recent years, there has been more coordinated lobbying for further funding at the scale of U.S. $250 million per year to go into the development of kilometer-scale GCMs (Slingo et al. 2022; Hewitt et al. 2022). In July 2023, the Berlin Summit for Earth Virtualization Engines (EVE) called for investment of EUR300 million per center per year for three to five centers to provide "small ensembles of kilometer-scale multidecadal global climate projections" and associated outreach and capacity development activities (Stevens et al. 2024).

However, there is no universal agreement that this approach is the most effective use of resources, especially when climate research output is intended to support decision-making (Stainforth and Calel 2020; Rodrigues and Shepherd 2022; Findlater et al. 2021). While greater coordination in GCM-focused climate research is largely taken to be a good thing—allowing for intercomparison projects and better understanding of uncertainty—the prospect for providing relevant and actionable climate projections for decision support has been the subject of more debate (Lemos et al. 2012; Kirchhoff et al. 2013; Hewitt et al. 2021).

In this paper, we argue for a pluralism of modeling strategies in climate research and therefore for an equitable,[1] more diversified distribution of funding among the current strategies. There is, on the climate research market, an array of modeling strategies which, in addition to GCMs of maximum resolution and complexity, includes the construction of superparameterizations or of convection-permitting models (Fosser et al. 2020); the downscaling of GCMs via regional climate models and/or statistical models (Jacob et al. 2014; Giorgi 2019); the use of machine learning–based modules and models (Chantry et al. 2021); the deployment of narrative explanations called storylines (Shepherd et al. 2018), which diversify the evidence for narratives beyond GCMs (Baldissera Pacchetti et al. 2023); the coupling with the biosphere and the anthroposphere in Earth system models of intermediate complexity (Steffen et al. 2020); the coupling with economy and energy in integrated assessment models (Nordhaus and Boyer 2003); and the use of other non-climate-focused

[1] Articulating what is meant in practice by an equitable funding distribution requires input from a wider community, so it is beyond the scope of this paper and will be the subject of future work by the authors. Here, we only note that "equitable" does not mean "equal."

models with climatic representation or climate inputs, such as ecosystem models and climate impact models (e.g., Rosenzweig et al. 2017). We argue that given the complexity of the climate system, modeling strategies should not compete against one another to be seen as the best or the only source of information about future climate. We argue that they are compatible with each other and can be seen as complementary tools to advance our current knowledge base. Doing so, we follow the fundamental argument of Held (2005) and Balaji (2021) that hierarchies of models of differing levels of complexity can all contribute to our understanding of a system and our ability to assess the confidence we should have in our predictions. Instead of a hierarchy, however, we visualize a connected network of models which take multiple viewpoints and provide different perspectives. In this view, there is no fundamental reason to prefer one strategy over another in an absolute way: The value of each approach is mutually reinforced by investment in the others, especially where they provide different kinds of information about different aspects of the system—which may be of interest to different potential stakeholders or users.

We adopt a philosophy of science approach when comparing GCM development aimed at enhancing resolution and complexity with three other examples: machine learning techniques, the physical climate storyline (PCS) approach, and Earth system models of intermediate complexity (EMICs). With these three, we are not aiming to be comprehensive: We choose these strategies because they are already in development and/or in use and are significantly different from each other (regarding climate modeling strategies, see also Held 2005; Katzav and Parker 2015; Walmsley 2020; Balaji 2021). As such, these strategies provide examples of alternative approaches to advancing our understanding of the climate. We show that each of these strategies prioritizes a particular set of aims, among which are empirical agreement, realism, comprehensiveness, diversity of process representations, inclusion of the human dimension, reduction of computational expense, and intelligibility. More precisely, we first identify the aims of the continued increase in resolution and complexity of GCMs (section 2). We call this the "predominant approach" to reflect the continued emphasis on GCM development as described above and contrast it with the many aspects of climate change research that do not necessarily involve developing models of ever higher resolution and complexity. We then discuss the aims of three selected alternative modeling strategies (section 3). It will follow that the alternative strategies have different degrees of conflict with the aims favored by the predominant approach, a conflict which may contribute to the debate about their respective effectiveness and acceptability. We reflect this debate back onto the predominant approach itself. We suggest that for each type of approach the targeted methodological aims make the strategies more or less adapted to answer particular kinds of real-world decision questions. Thus, because climate decision-making consists of many different kinds of questions, these modeling strategies (and many others which we have not looked at here) are all useful and act in a complementary way. We conclude that we should therefore be pluralist about modeling strategies and that funding should reflect this pluralism (section 4).

2. The predominant approach: Increasing resolution and complexity of GCMs

Historically, climate science has developed directly from meteorology and numerical weather prediction, a field that has seen significant improvement in predictive skill and decision support utility due to improvements in computational power and model resolution. The scientists involved in the early stages of the development of climate science as a discipline and its standard approach shared a fundamental assumption: that having better computers and more detailed models would also drastically improve the skill of climate modeling (Edwards 2010), and indeed, there has been much improvement in this respect. Climate scientists continue to argue that still more computational power will allow for higher-resolution and more complex models that will continue to improve the predictability of climate variables (Shukla et al. 2009; Palmer 2011, 2014; Palmer and Stevens 2019; Slingo et al. 2022; Hewitt et al. 2022) and thereby improve the utility of climate prediction for decision support.

There are many nationally funded research projects that are devoting resources to the development of high-resolution climate modeling, for instance, Climate Modeling Alliance (CliMA 2021), the "Destination Earth" project (Hoffmann et al. 2023), and the U.K. Met Office Unified Model (Met Office 2021). Moreover, all of the ongoing nationally funded research centers submit climate model runs to the Coupled Model Intercomparison Project (CMIP) as contributions to the Intergovernmental Panel on Climate Change (IPCC) process. This has not been the only focus of climate research (we outline others below). But the preferred, predominant, and most-funded modeling strategy of climate science has been and still is to build increasingly high-resolution and increasingly complex simulation models of the physical climate in order to produce quantitative climate projections.[2]

We identify three methodological aims that are associated with the predominant approach. The first aim is *empirical agreement* of the model outputs. Minimally, the model reconstructions of the past and present climates must be able to qualitatively correspond to patterns in the observational data. The second aim is *realism* of the model assumptions. For a given system, there is a tendency to work toward more realistic representations, thus correcting for previous simplifications, notably through the integration of more variables and fine-grained details, or replacing previous parameterizations by explicit theory-based equations.[3] The third aim is *comprehensiveness* of the models, in terms of the sheer quantity of process representations in the models. There is a tendency to continually expand the boundaries of the modeled system and to integrate more and more processes into the models. The current emphasis on higher spatial resolution in climate modeling is connected with both realism and comprehensiveness and generally entails *increase in computational power*. It is assumed that this will also lead to better *predictability* of climate variables, which we discuss further below.

An important pragmatic consequence of prioritizing the same set of aims is that models essentially all have the same end goal, so that in principle they can be made interoperable and intercomparable and that if there were no cross dependencies, the multimodel ensemble could be interpreted as a collection of equally plausible representations of the climate system, for further statistical treatment. The Coupled Model Intercomparison Project (Eyring et al. 2016) relies on this kind of standardization of inputs and outputs in order to generate information supporting the IPCC's assessment of climatic changes and impacts over the next century. There is also a feedback effect, insofar as the IPCC framework itself helps to shape and promote the shared goals of the modeling community and to encourage this kind of intercomparability.

The outputs provided by GCMs are also conveniently shaped for ingestion into further models which elaborate the consequences of projected climatic changes (such as downscaling models, physical impact models, and economic models), although representations of uncertainty are not so convenient and often do not make it into models further down the modeling chain (see Clark et al. 2016 for a discussion of the representation of uncertainty in hydrological impacts of climate change). Limitations to computational power mean that this cascade cannot be fully explored, leading to important knowledge gaps and to what has been called the "cascade of uncertainty" (Wilby and Dessai 2010).

However, an exhaustive exploration of uncertainty and its bounds is an important component to decision-making. While there have been calls to reduce different types of uncertainty

[2] By definition, this excludes funding directed to, e.g., developing observation networks and the related technology. Here, we are currently concerned with efforts directed toward dynamical and statistical modeling and analysis of what can be defined as "the climate system." Of course, we recognize that modeling does also rely on other activities such as data collection.

[3] "Realism," here, is the term that scientists generally use to characterize their modeling assumptions, and it is not what philosophers take to be as scientific realism. We use realism instead of "accuracy" because we want to avoid any connotation of "correctness."

in climate projections (e.g., Hawkins and Sutton 2009), increasing resolution and complexity of GCMs has not always reduced model projection uncertainty (Knutti and Sedláček 2013; see especially Fig. 1). The inclusion of new processes can radically change the space of possible outcomes, such as the marine ice cliff instability contribution to uncertainty in sea level rise projections [described by Horton et al. (2020)]. Some climate scientists have called for a different type of approach to uncertainty assessment and management, such as through "counterfactual thinking" (see, e.g., Rodrigues and Shepherd 2022).

3. Alternative approaches

Let us now explore the extent to which alternative modeling strategies—machine learning techniques, storylines, and EMICs—meet the three aims of empirical agreement, realism, and comprehensiveness and promote additional methodological aims.

a. Machine learning techniques. Machine learning (ML) approaches to climate modeling offer the possibility of rapid progress on data-intensive tasks for which these methods are well suited; indeed, aspects of data assimilation, sub-grid-scale process parameterization, and model output postprocessing in weather and climate could already be classed as ML methods (Schultz et al. 2021). The potential to reduce the computation time even for a subset of tasks is attractive, given that weather and climate models operate at the limits of the available computational resource and are already implemented on high-performance computing architectures which are increasingly designed to support ML methods. We distinguish here between small-scale or "soft artificial intelligence (AI)" approaches (Chantry et al. 2021) which incorporate ML methods into small subroutines of the climate model, versus larger-scale (Chantry et al.'s 2021 "hard AI") approaches which use ML methods as climate model replacements, to predict field-scale outputs. The former constitute only an extension of the GCM approach—the subject of discussion in this section is hard AI.

With respect to *empirical agreement* with observations, ML-based models have at least the potential to perform similarly or better than simulation models, though both approaches have their limitations when the parameters of the system go beyond the data used to fit the model, as is the case for future climatic change. If the physics-based nature of a model is the underlying reason to have any confidence that it will perform as well in the future as in the past, then fully ML-based models must take some care to express intelligibly the reasons for expecting empirically derived relationships to continue to hold out of sample, but this is not in principle impossible.

Regarding *realism*, one characteristic of ML is that the data are not assumed to conform to given physical laws or regularities, instead relying on the emergence of those regularities in statistical form. In climate modeling, physical process representation has generally been the preferred form of modeling, with statistical or empirically derived parameterizations used only where physical processes are insufficiently well understood or would require a prohibitive share of the computing resource. Hard AI, or larger-scale implementation of ML methods to directly predict field-scale outputs, is in conflict with the aim of realism. The hard AI approach has made significant progress on weather-forecasting time scales, with skill levels currently similar to operational numerical weather prediction (NWP) (Lam et al. 2023). The ability to train longer-time-scale ML models is lower, due to the scarcity of data, but we note that the same caveat about calibration also applies to process-based models, since we are always subject to the inductive problem that past performance does not guarantee future success. In principle, it could be possible for a hard AI implementation to learn the laws of physics and make reliable projections without containing explicitly "realistic" representations corresponding to current human understanding of physical processes. A middle ground may be provided by new physics-informed deep learning approaches, for which the primary motivations include "more interpretable ML methods that . . . can provide accurate and physically consistent predictions, even for extrapolatory/generalization tasks" (Karniadakis et al. 2021).

ML approaches allow for *comprehensiveness* within their own architecture: for example, through a larger neural network, or the addition of more layers of deep learning, or the addition of more sources of data. But additional complexity here does not map directly to the kinds of extra complexity in terms of additional process representations usually targeted by climate model development. For a simulation model, the consideration of more physical systems necessarily implies a larger and possibly more complex model architecture, whereas for a ML-based model, the complexity of input data and complexity of the internal structure are largely independent. If the aim is to account for the effects of as many processes as possible, then a simulation-based approach necessarily meets the limits of computational resource where a ML approach can be more efficient. Balaji (2021, p. 2) argues that computational limits alone (before considering energy requirements) will force a pragmatic change to this aim: "the continual addition of detail in our simulations is something that may have to be reconsidered, given certain physical limits on the evolution of computing hardware. . . . We may be compelled down some radically different paths."

What are the additional aims that are highlighted by considering ML approaches to modeling? One possible aim is *reduction of computational expense*, something considered by the predominant perspective as a simple boundary condition. Although the computational expense of initial ML model training is significant, use for provision of climate services would likely take the form of interaction with pretrained models (analogous to the current use of CMIP data). Another aim is *intelligibility*. This is the ability of the user to generate explanations based on the model (de Regt 2017; Jebeile et al. 2021): ML models are typically opaque in this respect (Knüsel and Baumberger 2020; Jebeile et al. 2021). While simulation models are naturally intelligible in that they construct causal physical chains, Reichstein et al. (2019, p. 199) note that "given their complexity, modern Earth system models are in practice often also not easily traceable back to their assumptions, limiting their interpretability." The lesser intelligibility of ML-based models may therefore be less of a trade-off in this respect than is usually assumed. As noted above, internal representations of physical processes within ML models may or may not correspond to human understanding, but could still result in effective forecasting skill. The question is not necessarily whether skill can be achieved but whether it will be possible for us to identify it with sufficient confidence to be able to use it.

The use of ML or other data-driven approaches to modeling could also stimulate the development of new methods for uncertainty quantification, since present methods are highly model laden, tending to explore a subspace of possible physical model configurations, such as those generated by systematic perturbation of initial and boundary conditions and model parameters. While soft AI approaches do not alter this paradigm, sensitivity analysis for hard AI approaches would be interesting to compare with existing uncertainty estimates. Without a well-defined set of parameters to consider perturbing, such sensitivity analysis might take the form of retraining ML models given different sets of input data, objective functions, or calibration targets. This raises the same issues already of concern to GCM approaches, of appropriate choices of calibration targets and the trade-off in allocating computational resource to sensitivity analysis rather than the best possible single model.

b. PCSs. A PCS is defined as "a physically self-consistent unfolding of past events, or of plausible future events or pathways" (Shepherd et al. 2018), and the IPCC's Sixth Assessment Report (AR6) defines PCS as exploration of "plausible trajectories of weather and climate conditions or events, especially those related to high levels of risk" (Arias et al. 2021; Box 1). The PCS approach is still heterogeneous, and while it is not an approach that develops new models per se, instead often relying upon some kind of modeled output, it can be seen as a set of methodological guiding principles which embody different types of epistemic and nonepistemic aims (Baldissera Pacchetti et al. 2023). A common theme in this approach is that it prioritizes the *causal* representation of hazards and impact chains, by analyzing past, present, or plausible future climate conditions or events. This emphasis is important for prioritizing physical interpretation of model output and adequately interpreting probabilistic estimates of uncertainty (Shepherd 2021). The focus on conditions and events that are related to high levels of risk emphasized in the IPCC definition provides a type of question framing that aims to reduce degrees of freedom that focus on those aspects of weather and climate that are most immediately relevant to humans (Sillmann et al. 2021).[4]

While it is not possible to provide a general analysis of this approach, we here focus on some examples that show a departure from the use of increasingly complex and high-resolution GCMs as described in the "predominant approach," both in terms of how to interpret and use GCM output and in the ways they use other types of models to explore uncertainty and extreme events.

Lloyd and Shepherd (2020), for example, describe storylines as an alternative representation of causality in the context of climate change and its impacts. Their analysis focuses on different key concepts in impact assessments and attribution studies: what counts as an extreme event, how causality is conceptualized in impact assessments and attribution studies, and how this is related to the use of statistics in climate change science. Because of the difficulties tied to conceptualizing and modeling extreme events with the predominant approach (e.g., lack of sufficient historical data to validate and verify model performance for extreme events), Lloyd and Shepherd (2020) argue that the storyline approach can provide a better alternative to evaluating the causes of extreme events and how these may change under global warming by framing the scientific problem differently.

The PCS approach is not univocal in its attitude to the methodological aim of *realism*. For example, storylines that are based on weather forecast models do aim at detailed representation of the region for which the storyline is constructed, and the detailed representation of the system is generally regarded as one of the strong suits of this approach (Hazeleger et al. 2015; van den Hurk et al. 2023). On the other hand, the narrative approach of Dessai et al. (2018) focuses on reducing the degrees of freedom of a system by identifying the key components that may be most sensitive to changes in global mean temperature, as well as exploring scenarios of atmospheric circulations that would not be captured by the models, thereby not prioritizing realism in the way we have defined this aim above.

Comprehensiveness is not prioritized in most of the examples of the storyline approach, as the boundaries of the system to be represented are strictly set by the specific (e.g., a certain event or region) focus of the approach. For example, a recent commentary of event-oriented PCS (Sillmann et al. 2021) highlights how the focus of event-based climate storylines does not aim at a full causal analysis of attribution factors, but rather on the impacts of events that contribute to climate-related risks.

The approach strongly aims at *empirical agreement* by rooting PCS in observed events or NWP output, and in many cases, linear temperature scaling is subsequently used to explore counterfactual events under different thermodynamic conditions. These counterfactuals are not empirically verifiable, but their plausibility is rooted in the argument that there is less uncertainty in exploring changing thermodynamic conditions while keeping the dynamics the same (Trenberth et al. 2015).

Approaches to PCS development such as the one of Dessai et al. (2018) reduce *computational expense*, either by relying on expert elicitation to explore bounds of plausibility and/or

[4] "Storylines" are also sometimes used to describe "scenario storylines" to represent possible socioeconomic future scenarios used to force GCMs. We are not discussing that type of storyline here, but see Rounsevell and Metzger (2010), Rounsevell et al. (2021), and Baulenas et al. (2023).

uncertainty or by only exploring only key chains of counterfactual events related to climatic hazards under different thermodynamic conditions (Hazeleger et al. 2015).

Climate storylines emphasize other aims that are not prioritized by the predominant approach. Although all the storyline approaches have a focus on physics, due to the role of the (atmospheric) climatic component in storylines, the storyline approach of Hazeleger et al. (2015) and Shepherd et al. (2018) focuses on the choice of physical hazards (climate events) driven by considerations of the impacts they would generate. Thus, climate storylines typically *include the human dimension* to a greater extent. In this way, they can be the basis for an in-depth study of human and ecosystem vulnerabilities to extreme events (Sillmann et al. 2021) or allow for better integration of top-down and bottom-up decision-making approaches (Dessai and Hulme 2004; Dessai et al. 2018; Bhave et al. 2018).

Another aim that is prioritized in this approach is *intelligibility*. Risbey et al. (2002) argue that GCM output, especially when used to derive regional information, should be complemented with regional expertise to interpret model output, due to the complexity and uncertainty related to GCMs. The storyline approach avoids the proliferation of uncertainties by focusing on developing "counterfactual explanations" of plausible futures (Hazeleger et al. 2015; Sillmann et al. 2021). Similarly, the approach of Dessai et al. (2018) uses expert elicitation to derive process-based narratives that focus on key mechanisms that may drive future changes in the region of focus. Due to this focus on intelligibility, storyline approaches work not only as alternative methods to explore the uncertainty related to future climate but also as tools that can better communicate how a changing climate can impact society and ecosystems.

c. Earth system models of intermediate complexity. EMICs are being developed with the holistic ambition of capturing the climate system by representing a larger number of components than GCMs, but in correspondingly less detail. Compared to GCMs, EMICs idealize relatively well-understood components (oceans, atmosphere, and dynamics) and therefore can put greater refinement on less well-understood components which can be the source of important feedbacks or tipping points. This way, EMICs are able to investigate the influence of human elements, nonlinearities, and abrupt transitions that are the scope of Earth system studies (e.g., Kim et al. 2022).

The conceptual foundation of the Earth system science (ESS) "*one model to fit all*" (Uhrqvist 2015) strategy is the diagram of Bretherton (1985) and has been updated since (Steffen et al. 2020, p. 61, Fig. 3). Following Steffen et al. (2020), three main components interact with each other: the geosphere, the biosphere, and the anthroposphere (which cover the energy systems, science and technology, institutions and political economy, human population with their futures, values and beliefs, and production and consumption). All of these components and subcomponents interact with each other, and in this inclusive picture, only sun, volcanoes, and fossil fuel combustion are considered as external forcings. Such conceptual frameworks in ESS led to the development of policy-relevant concepts such as the Anthropocene, tipping elements, and planetary boundaries (see Steffen et al. 2020).

While "the grand challenge for ESS is to achieve a deep integration of biophysical processes and human dynamics to build a truly unified understanding of the Earth System" (Steffen et al. 2020, p. 54), this remains an asymptotic program. ESS has so far rather produced a range of comprehensive yet moderately realistic models on which we focus here.

EMICs and GCMs do not aim to study the same aspects of the climate system, and for that reason, comparing their respective *empirical agreement* is tricky: They are certainly not equally good in reproducing the same datasets. Models with more spatial detail show better empirical agreement with spatially detailed observations, but questions still remain about how to combine the variables into a consistent evaluation metric given that models do not represent the same things. EMICs do aim for empirical agreement but do not prioritize it first.

Regarding *comprehensiveness*, EMICs aim at integrating the geosphere, the biosphere, and the anthroposphere, and as such, they do prioritize comprehensiveness very highly.[5] EMICs are specifically built for including more processes than GCMs contain; what is gained in terms of computational power (by choosing lower spatial resolution and using parameterizations) is reassigned to either increasing simulation time (for paleoclimatology) and number of represented large-scale processes.

Consequently, regarding *realism*, EMICs are less realistic than complex and detailed GCMs. There is indeed an unavoidable trade-off between comprehensiveness and realism because of computational limits. EMICs are necessarily more approximate representations of the climate system, and therefore, choices are made in particular with respect to (i) what components should be actually taken into account and (ii) how realistically they should be represented within the model (i.e., with a parameterization or a specific equation).

An additional aim that the development of EMICs prioritizes is the *diversity of process representations*. The ambition of EMICs is indeed to integrate aspects of the climate system that are different in nature: from the physics of the geosphere, to the biology of the biosphere, to the socioeconomic and even psychosocial dimension of the anthroposphere. Diversity of process representations refers to the degree of heterogeneity of the natures of the interacting components included in the model. Thus, for example, an ESM which integrates a carbon cycle and a vegetation dynamics module has a higher diversity of process representations than a GCM which does not integrate those modules but describes more atmospheric and near-surface oceanic subprocesses. Diversity of process representations also includes the idea that the heterogeneous elements are interactively and dynamically coupled with each other rather than merely included in the models as prescribed boundary conditions.

Another related aim of EMICs is the *inclusion of the human dimension*. EMICs integrate the *human dimension* through diverse perspectives, i.e., social, ecosystemic, and environmental perspectives. They aim to describe interconnections and feedbacks between the geosphere, the biosphere, and the anthroposphere.

It is worth highlighting that the sacrifice of realism for the sake of quantity and diversity of process representations could be justified by the possibility that the error one makes in omitting system components of different nature may be greater than the error one makes due to lesser realism (an error that one usually tries to reduce by using spatial refinement methods like superparameterizations or soft AI approaches). Thus, the use of GCMs with increasing resolution and complexity "potentially underestimates the role of vegetation dynamics and biogeochemical cycles in affecting the climate system and overestimates the importance of high spatial resolution and comprehensiveness" (Claussen et al. 2002, p. 580). On the other hand, the tension between realism and comprehensiveness may lead one to develop only one aspect of the climate system, aiming at higher realism of physical climatic variables, while setting boundary conditions that are supposed to be representative of the other aspects. For example, anthropogenic emissions are conceived as forcing factors (or boundary conditions) within GCMs, via the definition of emission scenarios. One argument in favor of this strategy is that it is easier to add to the model things we already understand well (for a physicist, it is easier to add more physics than to learn biology or economics). In this sense, one might argue that realism (at local scale) of the physical climate system is more important than integrating more diverse components with less realism.

Finally, regarding *intelligibility*, in a sense of detailed causal relationships between processes within the model, GCMs with increasingly complex and detailed structures are more intelligible than EMICs, but in a sense of overall system-wide intelligibility, the EMICs have

[5] Having said that, we note that comprehensiveness does not have a natural definition in terms of the number of processes represented in the models (Claussen et al. 2002, p. 583) and that in a comparative sense the "comprehensiveness" is as hard to evaluate as the empirical agreement.

the potential to provide a qualitative understanding from the conceptual frameworks in a wider field which is partially outside the view of GCMs, however detailed.

4. Benefits of multiple modeling strategies

In this section, we first highlight that the predominant approach does not prioritize the additional aims uncovered by the above discussion, and as a result, some possible objectives of modeling are not fully met by the paradigm of increasing resolution and complexity in GCMs. We then argue that, because no single approach can prioritize all of these aims simultaneously and meet all possible objectives, a toolbox including all of the modeling strategies is preferable to a single one.

a. Limits of the predominant approach. First, the predominant approach tends to assume that more powerful computers are necessary to build more realistic and more comprehensive models. Therefore, the approach treats computational expense as a simple constraint: *Reduction of computational expense* is desirable, but only insofar as it allows one to make use of the saved resource elsewhere. The main trade-offs in this respect for a state-of-the-art GCM are how to allocate resources between further complexity, longer simulation time, and larger ensembles. So, the predominant approach leads to ever-increasing costs of computation (and energy, although computational efficiency is also improving): Supercomputers with price tags in hundreds of millions of dollars represent a significant proportion of the costs of existing and proposed modeling centers. In addition to the environmental impact and the effect of excluding less-resourced groups, this barrier to entry (to participation, for example, in CMIP) effectively reduces the potential ensemble size and diversity.

Improvement in modeling and modeling outputs has been heterogeneous and depends on the time scale and regions of interest. This heterogeneity may be due to a combination of availability of observations for verification, theoretical developments, and development of computational tools, among other factors. The heterogeneity of model performance is also a function of past and present sociopolitical power relations, resulting in institutions in or studying Europe and North America benefiting from higher-quality data, greater research funding, and access to more computational resources than institutions in or studying, for example, Africa. If the predominant approach is to be continued without further entrenching sociopolitical biases, efforts must be made to avoid the continuation of this pattern (James et al. 2018; Mishra et al. 2023) and indeed proposals such as EVE do emphasize the need for global participation and capacity building (Stevens et al. 2024).

Second, in the predominant approach, the *human dimension* has remained largely absent from the equations even while we became increasingly confident that humanity is the primary driver of climate change. The direct effect of humans on the climate remains separated from GCMs in economic models (the families of integrated assessment models used to generate socioeconomic scenarios), while the effects of the climate on humans are left to impact models and economic models, effectively preventing quantitative treatment of potential feedbacks or cascading consequences which could lead to social or physical “tipping points.” There is a need for more integration of different perspectives to better understand the intricacies of the human–environment interaction and increase the salience of decision-relevant science (Conway et al. 2019).

Historically, climate science has indeed been conceived as a physics-based discipline. However, the focus on physics ignores important social and methodological issues that philosophers, environmental social scientists and increasingly physical climate scientists, are recognizing as important aspects of a comprehensive evaluation of the quality of climate change–related knowledge (see, e.g., Shackley and Wynne 1995; Shackley et al. 1998; Dessai and Hulme 2004; Winsberg et al. 2020; Jebeile and Crucifix 2021). Integrative approaches which bring the human element closer to physical models include coproduction approaches (Bremer and Meisch 2017) and climate risk narratives (Jack et al. 2020) as well as storylines and EMICs.

Third, comprehensiveness can be in tension with *intelligibility*. At the small scale, physics-based GCMs are intelligible by design, in that every effect can be traced to a "physical" (within the model) cause. At the larger scale, however, more comprehensive models introduce cascading complexity and make it hard for the user to develop an intuitive understanding of the model outputs. A result of this lack of overall intelligibility is that the task of tuning a climate model has become an extremely complex undertaking (Hourdin et al. 2017; Balaji et al. 2022). Yet, intelligibility at the larger scale can be an important requirement if we consider transparent communication of scientific results to a wide audience to be of importance. In this respect, the use of storyline approaches can prune the scope of consideration sufficiently to allow simpler explanation of effects and a more targeted exploration of uncertainty.

Fourth, the *diversity of process representation* is not prioritized by GCMs, meaning that there is an imbalance in potential sources of uncertainty. Statistical methods which are used to analyze the outputs of GCM ensembles and generate probabilities about future climate are not able to take account of this imbalance, since it is unknown to what degree any errors might be independent of each other. As Katzav et al. (2021, p. 12) note, if it is desirable to provide probabilities then a more systematic attempt would need to be made to explore the range of possible model outcomes; they conclude that "Experimental designs should prioritise the maximisation of diversity within ensembles" and that sensitivity analyses alone (the standard method uses perturbed parameter ensembles, supplemented with outcomes from emulators, or at best an ensemble of alternate GCMs) are insufficient to achieve this.

Finally, we address the objection that any of the above methodological aims may only be desirable insofar as they provide some warrant that the outputs of a model will be reliable and hence useful for decision support. The methodological aims prioritized by the predominant approach of increasing resolution and complexity in GCMs (*empirical agreement*, *comprehensiveness*, and *realism*) are promoted because they are taken to be the most effective way of supporting an argument that the model outputs will offer good predictions, conditional on giving them the right input. We offer two counterpoints to this assumption: First, the increasing complexity of models means that progress toward "better" outputs is no longer (if it ever was) monotonic—physical improvements to one module or component may result in degradation of quantitative performance in other areas (Schmidt et al. 2017; Hourdin et al. 2017); and second, it is useful to distinguish between raw predictive capability and utility as decision support. Even if the quantitative reliability of projections of large-scale climate variables is improved, it is not necessarily the case that we would have more adequate information for decision support (Kirchhoff et al. 2013). On smaller scales, the provision of increasingly detailed climate projections suffers from the problem of poorly constrained large-scale uncertainties (such as sea level rise) and uncertainty cascades, including human and socioeconomic feedbacks as well as possible "climate tipping points," which could dominate the smaller uncertainties in varied GCM (or RCM) outputs. As such, it does not seem to us to be conclusively demonstrated that increasing resolution and complexity will improve utility for decision support, even if it does result in improved predictive capability for a subset of prioritized variables.

Returning to Isaac Held's concept of a "hierarchy of models," the above discussion suggests one compelling reason to reject this in favor of a "network" or "toolbox" of models and model values. There are no top-of-the-hierarchy and no single scalar dimension on which to rank either the conceptual or empirical quality of a model. Even if we could guarantee that each improvement to physical representation would result in improved predictive ability, we would still need to account for the needs of different stakeholders (user communities).

b. A toolbox for climate information. Now that we have shown that the predominant approach favors certain methodological aims while it sacrifices other possibly important aims relevant for policy support, we want to argue that having a toolbox of diverse strategies rather than a single approach will allow one to better satisfy the existing diversity of user needs. Moreover, with multiple strategies available, the range of possible stakeholder engagements with climate modeling is also increased, improving the accessibility of knowledge.

First of all, we should make it clear that favoring any modeling strategy is actually a value-laden view on how science can best inform society including the kinds of stakeholders whose information preferences are prioritized (Porter and Dessai 2017). A preference for one modeling strategy over another is strongly motivated by a certain ideal of "what good science ought to be." More precisely, it is driven by the scientific community's take on what models would produce the "best" information for decision support. However, Longino (1996) shows that scientific aims are never strictly epistemically motivated and goes further to suggest that qualities such as novelty, ontological heterogeneity, mutuality of interaction, applicability to human needs, and diffusion or decentralization of power can be taken as genuine scientific aims, although they are also tinged with values. While Longino focuses on scientific theories, where we focus on modeling strategies, and more particularly on climate modeling strategies, we use her assertion that scientific aims are driven by views about how science can best inform us about the world, and those views can partly be loaded with social, economic, and political values. Thus, in climate science, favoring any given modeling strategy is a value-laden choice, with political and social implications.

Second, each modeling strategy comes with its own set of methodological aims. It follows that each modeling strategy is more or less suitable in supporting decisions depending on the kinds of questions that are being addressed. Increasingly high-resolution and more complex GCMs provide consistent long time series of climatic variables which are of most use for large-scale quantitative decision-making such as distinguishing between options for where in Europe to locate a new wind farm. The ML approach makes a more efficient use of computational resource; it may be particularly useful for data-intensive applications at the short time scale where direct verification is feasible, such as downscaling forecasts of wind conditions to hyper-local scales. Storylines are intelligible, making them particularly suitable in public and fast communication on climate risks, for example, providing information about how the risk of a particular kind of extreme wind event is changing over time. EMICs provide one with a systemic view of interactions and feedbacks between important components of the climate system; these models are valuable in identifying and communicating larger-scale and systemic risks (blind spots for GCMs) and providing an interdisciplinary interface for climate science with fields such as energy policy and geopolitics. Other families of impact models and integrated assessment models, though we have not considered them here, target other stakeholders and other methodological aims, ingesting data and framings provided both by GCMs and by other models and scenarios of future environmental change.

Finally, we see the plural development of climate modeling strategies like adding more tools to a toolbox. With more tools, one can better satisfy the diversity of user needs as sketched in the previous paragraph. We find the metaphor of a toolbox helpful here: A screwdriver—optimized for one purpose—is no good when one needs a hammer, and a Swiss Army knife—strategy which attempts to maximize all desiderata simultaneously—will be less effective for any job than the adequate tool, but it is often said that *when all you have is a hammer, every problem looks like a nail*. When increasingly high-resolution and complex GCMs are the primary tools for provision of climate information, how might this limit our approaches to climate change adaptation? Stainforth and Calel (2020, p. 2) have argued that "Before ploughing billions into developing specialised computers and associated computer models, it would be wise to first develop a good theoretical understanding of what is necessary and sufficient to build models capable of such high-resolution predictions." We agree, and we extend their condition. It would also be wise to develop a more coherent theoretical understanding of what kind of climate modeling is necessary (though perhaps never sufficient) to inform good climate decision-making. In our view, this requires the active participation of more stakeholders than the highly quantitative organizations whose needs, wishes, and expectations are prioritized by the further development of highly complex and detailed GCMs. This could be achieved by diversifying modeling strategies, not by consolidating yet more effort onto a single strategy. A bigger and better hammer will certainly do a better job of hammering nails, but as not all things are nails that need hammering, this alone does not make it the best possible investment.

5. Concluding remarks

The predominant approach to climate modeling—making further investment in the development of increasingly complex and high-resolution GCMs at the expense of other approaches—is limiting both the kinds of climate information that can be gained and the kinds of stakeholders who can engage with and use that information. We could make better climate decisions if we had more robust information from a range of sources that was applicable to a variety of situations and relevant to more stakeholders. To do better science, we need to diversify the approaches, not double down on the existing paradigm. The currently predominant modeling strategy is also the most computationally expensive, and therefore, promoting diversity need not be a significant drain on computational resource. It does not mean we have to fund everything, distribute financial resources equally, or spread the resources too thinly to get anything useful anywhere. It does not even necessarily mean funding the limited set of complementary alternatives we have considered here. It might, however, also stimulate debate about novel forms of models and modeling approaches with even more different aims, perhaps ones centered around ecosystems, the concept of "just transition" (Heffron 2022), or participatory coproduction. The politics of science funding requires some coordination (and democratically decided priorities) to decide what projects should be funded and in which of many possible ways diversification should be achieved. The outcomes of active diversification would be to broaden the kinds of decision questions we are capable to answer, as well as to have more justified confidence in the robust core of projections, more potential input from those who will be affected by decisions, and thereby more effective consensus building for climate action.

Acknowledgments. J. J. acknowledges support from the Swiss National Science Foundation (SNF PRIMA project, Grant PR00P1_208469). E. T. was supported by a UKRI Future Leaders Fellowship (Grant MR/V024426/1). We thank Roland Séférian and Sébastien Dutreuil for helpful comments.

Data availability statement. No datasets were generated or analyzed for this paper.

References

Arias, P. A., and Coauthors, 2021: Technical summary. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 33–144, https://doi.org/10.1017/9781009157896.002.

Balaji, V., 2021: Climbing down Charney's ladder: Machine learning and the post-Dennard era of computational climate science. *Philos. Trans. Roy. Soc.*, **A379**, 20200085, https://doi.org/10.1098/rsta.2020.0085.

——, F. Couvreux, J. Deshayes, J. Gautrais, F. Hourdin, and C. Rio, 2022: Are general circulation models obsolete? *Proc. Natl. Acad. Sci. USA*, **119**, e2202075119, https://doi.org/10.1073/pnas.2202075119.

Baldissera Pacchetti, M., L. Coulter, S. Dessai, T. G. Shepherd, J. Sillmann, and B. Van Den Hurk, 2023: Varieties of approaches to constructing physical climate storylines: A review. *Wiley Interdiscip. Rev.: Climate Change*, **15**, e869, https://doi.org/10.1002/wcc.869.

Baulenas, E., G. Versteeg, M. Terrado, J. Mindlin, and D. Bojovic, 2023: Assembling the climate story: Use of storyline approaches in climate-related science. *Global Challenges*, **7**, 2200183, https://doi.org/10.1002/gch2.202200183.

Bhave, A. G., D. Conway, S. Dessai, and D. A. Stainforth, 2018: Water resource planning under future climate and socioeconomic uncertainty in the Cauvery River basin in Karnataka, India. *Water Resour. Res.*, **54**, 708–728, https://doi.org/10.1002/2017WR020970.

Bremer, S., and S. Meisch, 2017: Co-production in climate change research: Reviewing different perspectives. *Wiley Interdiscip. Rev.: Climate Change*, **8**, e482, https://doi.org/10.1002/wcc.482.

Bretherton, F. P., 1985: Earth system science and remote sensing. *Proc. IEEE*, **73**, 1118–1127, https://doi.org/10.1109/PROC.1985.13242.

Chantry, M., H. Christensen, P. Dueben, and T. Palmer, 2021: Opportunities and challenges for machine learning in weather and climate modelling: Hard, medium and soft AI. *Philos. Trans. Roy. Soc.*, **A379**, 20200083, https://doi.org/10.1098/rsta.2020.0083.

Clark, M. P., and Coauthors, 2016: Characterizing uncertainty of the hydrologic impacts of climate change. *Curr. Climate Change Rep.*, **2**, 55–64, https://doi.org/10.1007/s40641-016-0034-x.

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, https://doi.org/10.1007/s00382-001-0200-1.

CliMA, 2021: Climate modeling alliance. Accessed 30 January 2024, https://clima.caltech.edu/.

Conway, D., and Coauthors, 2019: The need for bottom-up assessments of climate risks and adaptation in climate-sensitive regions. *Nat. Climate Change*, **9**, 503–511, https://doi.org/10.1038/s41558-019-0502-0.

de Regt, H. W., 2017: *Understanding Scientific Understanding.* Oxford University Press, 320 pp.

Dessai, S., and M. Hulme, 2004: Does climate adaptation policy need probabilities? *Climate Policy*, **4**, 107–128, https://doi.org/10.1080/14693062.2004.9685515.

——, A. Bhave, C. Birch, D. Conway, L. Garcia-Carreras, J. P. Gosling, N. Mittal, and D. Stainforth, 2018: Building narratives to characterise uncertainty in regional climate change through expert elicitation. *Environ. Res. Lett.*, **13**, 074005, https://doi.org/10.1088/1748-9326/aabcdd.

Edwards, P., 2010: *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.* MIT Press, 518 pp.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Findlater, K., S. Webber, M. Kandlikar and S. Donner, 2021: Climate services promise better decisions but mainly focus on better data. *Nat. Climate Change*, **11**, 731–737, https://doi.org/10.1038/s41558-021-01125-3.

Fosser, G., E. J. Kendon, D. Stephenson, and S. Tucker, 2020: Convection-permitting models offer promise of more certain extreme rainfall projections. *Geophys. Res. Lett.*, **47**, e2020GL088151, https://doi.org/10.1029/2020GL088151.

Giorgi, F., 2019: Thirty years of regional climate modeling: Where are we and where are we going next? *J. Geophys. Res. Atmos.*, **124**, 5696–5723, https://doi.org/10.1029/2018JD030094.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1.

Hazeleger, W., B. J. J. M. van den Hurk, E. Min, G. J. van Oldenborgh, A. C. Petersen, D. A. Stainforth, E. Vasileiadou, and L. A. Smith, 2015: Tales of future weather. *Nat. Climate Change*, **5**, 107–113, https://doi.org/10.1038/nclimate2450.

Heffron, R. J., 2022: What is the "Just Transition"? *Achieving a Just Transition to a Low-Carbon Economy*, Palgrave Macmillan, 9–19, https://doi.org/10.1007/978-3-030-89460-3_2.

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

Hewitt, C. D., and Coauthors, 2021: Recommendations for future research priorities for climate modeling and climate services. *Bull. Amer. Meteor. Soc.*, **102**, E578–E588, https://doi.org/10.1175/BAMS-D-20-0103.1.

Hewitt, H., B. Fox-Kemper, B. Pearson, M. Roberts, and D. Klocke, 2022: The small scales of the ocean may hold the key to surprises. *Nat. Climate Change*, **12**, 496–499, https://doi.org/10.1038/s41558-022-01386-6.

Hoffmann, J., P. Bauer, I. Sandu, N. Wedi, T. Geenen, and D. Thiemert, 2023: Destination Earth—A digital twin in support of climate services. *Climate Serv.*, **30**, 100394, https://doi.org/10.1016/j.cliser.2023.100394.

Horton, B. P., and Coauthors, 2020: Estimating global mean sea-level rise and its uncertainties by 2100 and 2300 from an expert survey. *npj Climate Atmos. Sci.*, **3**, 18, https://doi.org/10.1038/s41612-020-0121-5.

Hourdin, F., and Coauthors, 2017: The art and science of climate model tuning. *Bull. Amer. Meteor. Soc.*, **98**, 589–602, https://doi.org/10.1175/BAMS-D-15-00135.1.

Jack, C. D., R. Jones, L. Burgin, and J. Daron, 2020: Climate risk narratives: An iterative reflective process for co-producing and integrating climate knowledge. *Climate Risk Manage.*, **29**, 100239, https://doi.org/10.1016/j.crm.2020.100239.

Jacob, D., and Coauthors, 2014: EURO-CORDEX: New high-resolution climate change projections for European impact research. *Reg. Environ. Change*, **14**, 563–578, https://doi.org/10.1007/s10113-013-0499-2.

James, R., and Coauthors, 2018: Evaluating climate models with an African lens. *Bull. Amer. Meteor. Soc.*, **99**, 313–336, https://doi.org/10.1175/BAMS-D-16-0090.1.

Jebeile, J., and M. Crucifix, 2021: Value management and model pluralism in climate science. *Stud. Hist. Philos. Sci.*, **88**, 120–127, https://doi.org/10.1016/j.shpsa.2021.06.004.

——, V. Lam, and T. Räz, 2021: Understanding climate change with statistical downscaling and machine learning. *Synthese*, **199**, 1877–1897, https://doi.org/10.1007/s11229-020-02865-z.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, 2021: Physics-informed machine learning. *Nat. Rev. Phys.*, **3**, 422–440, https://doi.org/10.1038/s42254-021-00314-5.

Katzav, J., and W. S. Parker, 2015: The future of climate modeling. *Climatic Change*, **132**, 475–487, https://doi.org/10.1007/s10584-015-1435-x.

——, E. Thompson, J. S. Risbey, D. Stainforth, S. Bradley, and M. Frisch, 2021: On the appropriate and inappropriate uses of probability distributions in climate projections, and some alternatives. *Climatic Change*, **169**, 15, https://doi.org/10.1007/s10584-021-03267-x.

Kim, S. K., H. J. Kim, H. A. Dijkstra, and S.-I. An, 2022: Slow and soft passage through tipping point of the Atlantic meridional overturning circulation in a changing climate. *npj Climate Atmos. Sci.*, **5**, 13, https://doi.org/10.1038/s41612-022-00236-8.

Kirchhoff, C. J., M. C. Lemos, and S. Dessai, 2013: Actionable knowledge for environmental decision making: Broadening the usability of climate science. *Annu. Rev. Environ. Resour.*, **38**, 393–414, https://doi.org/10.1146/annurev-environ-022112-112828.

Knüsel, B., and C. Baumberger, 2020: Understanding climate phenomena with data-driven models. *Stud. Hist. Philos. Sci.*, **84**, 46–56, https://doi.org/10.1016/j.shpsa.2020.08.003.

Lam, R., and Coauthors, 2023: Learning skillful medium-range global weather forecasting. *Science*, **382**, 1416–1421, https://doi.org/10.1126/science.adi2336.

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

Lloyd, E. A., and T. G. Shepherd, 2020: Environmental catastrophes, climate change, and attribution. *Ann. N. Y. Acad. Sci.*, **1469**, 105–124, https://doi.org/10.1111/nyas.14308.

Longino, H. E., 1996: Cognitive and non-cognitive values in science: Rethinking the dichotomy. *Feminism, Science, and the Philosophy of Science*, L. H. Nelson and J. Nelson, Eds., Kluwer, 39–58.

Met Office, 2021: The Unified Model. Accessed 9 July 2021, https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model.

Mishra, S. K., P. Upadhyaya, J. T. Fasullo, N. P. Keshri, P. Salunke, A. Ghosh, A. B. Sainudeen, and I.-S. Kang, 2023: A need for actionable climate projections across the Global South. *Nat. Climate Change*, **13**, 883–886, https://doi.org/10.1038/s41558-023-01778-2.

Nordhaus, W. D., and J. Boyer, 2003: *Warming the World: Economic Models of Global Warming.* The MIT Press, 232 pp.

Overland, I., and B. K. Sovacool, 2020: The misallocation of climate research funding. *Energy Res. Soc. Sci.*, **62**, 101349, https://doi.org/10.1016/j.erss.2019.101349.

Palmer, T., 2011: A CERN for climate change. *Phys. World*, **24**, 14–15, https://doi.org/10.1088/2058-7058/24/03/24.

——, 2014: Climate forecasting: Build high-resolution global climate models. *Nature*, **515**, 338–339, https://doi.org/10.1038/515338a.

——, and B. Stevens, 2019: The scientific challenge of understanding and estimating climate change. *Proc. Natl. Acad. Sci. USA*, **116**, 24 390–24 395, https://doi.org/10.1073/pnas.1906691116.

Porter, J. J., and S. Dessai, 2017: Mini-me: Why do climate scientists' misunderstand users and their needs? *Environ. Sci. Policy*, **77**, 9–14, https://doi.org/10.1016/j.envsci.2017.07.004.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204, https://doi.org/10.1038/s41586-019-0912-1.

Risbey, J. S., P. J. Lamb, R. L. Miller, M. C. Morgan, and G. H. Roe, 2002: Exploring the structure of regional climate scenarios by combining synoptic and dynamic guidance and GCM output. *J. Climate*, **15**, 1036–1050, https://doi.org/10.1175/1520-0442(2002)015<1036:ETSORC>2.0.CO;2.

Rodrigues, R. R., and T. Shepherd, 2022: Small is beautiful: Climate-change science as if people mattered. *PNAS Nexus*, **1**, pgac009, https://doi.org/10.1093/pnasnexus/pgac009.

Rosenzweig, C., and Coauthors, 2017: Assessing inter-sectoral climate change risks: The role of ISIMIP. *Environ. Res. Lett.*, **12**, 010301, https://doi.org/10.1088/1748-9326/12/1/010301.

Rounsevell, M. D. A., and M. J. Metzger, 2010: Developing qualitative scenario storylines for environmental change assessment. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 606–619, https://doi.org/10.1002/wcc.63.

——, and Coauthors, 2021: Identifying uncertainties in scenarios and models of socio-ecological systems in support of decision-making. *One Earth*, **4**, 967–985, https://doi.org/10.1016/j.oneear.2021.06.003.

Schmidt, G. A., and Coauthors, 2017: Practice and philosophy of climate model tuning across six US modeling centers. *Geosci. Model Dev.*, **10**, 3207–3223, https://doi.org/10.5194/gmd-10-3207-2017.

Schultz, M. G., C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, and S. Stadtler, 2021: Can deep learning beat numerical weather prediction? *Philos. Trans. Roy. Soc.*, **A379**, 20200097, https://doi.org/10.1098/rsta.2020.0097.

Shackley, S., and B. Wynne, 1995: Integrating knowledges for climate change: Pyramids, nets and uncertainties. *Global Environ. Change*, **5**, 113–126, https://doi.org/10.1016/0959-3780(95)00017-I.

——, P. Young, S. Parkinson, and B. Wynne, 1998: Uncertainty, complexity and concepts of good science in climate change modelling: Are GCMs the best tools? *Climatic Change*, **38**, 159–205, https://doi.org/10.1023/A:1005310109968.

Shepherd, T. G., 2021: Bringing physical reasoning into statistical practice in climate-change science. *Climatic Change*, **169**, 2, https://doi.org/10.1007/s10584-021-03226-6.

——, and Coauthors, 2018: Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, **151**, 555–571, https://doi.org/10.1007/s10584-018-2317-9.

Shukla, J., R. Hagedorn, M. Miller, T. N. Palmer, B. Hoskins, J. Kinter, J. Marotzke, and J. Slingo, 2009: Strategies: Revolution in climate prediction is both necessary and possible: A declaration at the world modelling summit for climate prediction. *Bull. Amer. Meteor. Soc.*, **90**, 175–178, https://doi.org/10.1175/2008BAMS2759.1.

Sillmann, J., T. G. Shepherd, B. van den Hurk, W. Hazeleger, O. Martius, J. Slingo, and J. Zscheischler, 2021: Event-based storylines to address climate risk. *Earth's Future*, **9**, e2020EF001783, https://doi.org/10.1029/2020EF001783.

Slingo, J., and Coauthors, 2022: Ambitious partnership needed for reliable climate prediction. *Nat. Climate Change*, **12**, 499–503, https://doi.org/10.1038/s41558-022-01384-8.

Stainforth, D. A., and R. Calel, 2020: New priorities for climate science and climate economics in the 2020s. *Nat. Commun.*, **11**, 3864, https://doi.org/10.1038/s41467-020-16624-8.

Steffen, W., K. Richardson, J. Rockström, H. J. Schellnhuber, O. P. Dube, S. Dutreuil, T. M. Lenton, and J. Lubchenco, 2020: The emergence and evolution of Earth system science. *Nat. Rev. Earth Environ.*, **1**, 54–63, https://doi.org/10.1038/s43017-019-0005-6.

Stevens, B., and Coauthors, 2024: Earth Virtualization Engines (EVE). *Earth Syst. Sci. Data*, **16**, 2113–2122, https://doi.org/10.5194/essd-16-2113-2024.

Trenberth, K. E., J. T. Fasullo, and T. G. Shepherd, 2015: Attribution of climate extreme events. *Nat. Climate Change*, **5**, 725–730, https://doi.org/10.1038/nclimate2657.

Uhrqvist, O., 2015: One model to fit all? The pursuit of integrated Earth system models in GAIM and AIMES. *Hist. Soc. Res.*, **40**, 271–297, https://doi.org/10.12759/hsr.40.2015.2.271-297.

van den Hurk, B. J. J. M., and Coauthors, 2023: Climate impact storylines for assessing socio-economic responses to remote events. *Climate Risk Manage.*, **40**, 100500, https://doi.org/10.1016/j.crm.2023.100500.

Walmsley, L. D., 2020: The strategy of model building in climate science. *Synthese*, **199**, 745–765, https://doi.org/10.1007/s11229-020-02707-y.

Wilby, R. L., and S. Dessai, 2010: Robust adaptation to climate change. *Weather*, **65**, 180–185, https://doi.org/10.1002/wea.543.

Winsberg, E., N. Oreskes, and E. Lloyd, 2020: Severe weather event attribution: Why values won't go away. *Stud. Hist. Philos. Sci.*, **84**, 142–149, https://doi.org/10.1016/j.shpsa.2020.09.003.