Why Simpler Computer Simulation Models Can Be Epistemically Better for Informing Decisions

Casey Helgeson, Vivek Srikrishnan, Klaus Keller, and Nancy Tuana*†

For computer simulation models to usefully inform climate risk management, uncertainties in model projections must be explored and characterized. Because doing so requires running the model many times over, and because computing resources are finite, uncertainty assessment is more feasible using models that demand less computer processor time. Such models are generally simpler in the sense of being more idealized, or less realistic. So modelers face a trade-off between realism and uncertainty quantification. Seeing this trade-off for the important epistemic issue that it is requires a shift in perspective from the established simplicity literature in philosophy of science.

1. Introduction. Computer simulation models are now essential tools in many scientific fields, and a rapidly expanding philosophical literature examines a host of accompanying methodological and epistemological questions about their roles and uses (e.g., Frigg and Reiss 2009; Grüne-Yanoff and Weirich 2010; Winsberg 2010, 2018; Weisberg 2013; Parke 2014; Jebeile

Received July 2019; revised January 2020.

*To contact the authors, please write to: Casey Helgeson, Earth and Environmental Systems Institute, Pennsylvania State University; e-mail: helgeson@psu.edu.

†The authors thank Hayley Clatterbuck, Kevin Elliott, Ben Seiyon Lee, and Elliott Sober for comments on a draft and Robert Fuller, Kelsey Ruckert, and Tony Wong for sharing expertise on the models discussed in the article. Three anonymous referees helped to improve the article. This work was cosupported by the National Science Foundation (NSF) through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the Penn State Center for Climate Risk Management. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding entities.

Philosophy of Science, 88 (April 2021) pp. 213–233. 0031-8248/2021/8802-0002$10.00
Copyright 2021 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/711501 Published online by Cambridge University Press

2017; Beisbart and Saam 2019). Climate science is one such field (Edwards 2001), and questions about the interpretation and reliability of the simulation models used to understand, attribute, and predict climate change have received considerable attention (e.g., Lloyd 2010, 2015; Oreskes, Stainforth, and Smith 2010; Parker 2011, 2013; Petersen 2012; Frigg, Smith, and Stainforth 2013, 2015; Steele and Werndl 2016; Thompson, Frigg, and Helgeson 2016; Vezér 2016; Lloyd and Winsberg 2018).

One conspicuous feature of scientific discourse about the simulation models used in climate science, and in environmental modeling more broadly, is the attention given to where a model lies on a spectrum from simple to complex (e.g., McGuffie and Henderson-Sellers 2001; Jakeman, Letcher, and Norton 2006; Smith et al. 2014). While this attention to model complexity has informed some of the philosophical discourse on simulation modeling (e.g., Parker 2010), its relevance for the literature on simplicity in science remains largely unexplored.

This literature on simplicity addresses whether and why simpler theories (or hypotheses, models, etc.) might be—other things being equal—better than complex ones. Different ways of unpacking "simpler" and "better" yield a diversity of specific theses, with correspondingly different justifications (see Sober [2015] and Baker [2016], for details and history). A number of distinctly modern variants rest on mathematical theorems tying well-defined notions of simplicity to benefits such as predictive accuracy (Akaike 1973; Forster and Sober 1994), reliability (Vapnik 1998; Harman and Kulkarni 2007), and efficient inquiry (Kelly 2004, 2007). Arguably more domain-specific appeals to parsimony include instances in phylogenetics (see Sober 1988, 2015) and animal cognition (Sober 2009, 2015; Clatterbuck 2015).

Here we discuss a notion of simplicity drawn from scientific discourse on environmental simulation modeling and expound its importance in the context of climate risk management. The new idea that we bring to the simplicity literature is that simplicity benefits the assessment of uncertainty in the model's predictions. The short explanation for this is that quantifying uncertainty in the predictions of computer simulation models requires running the model many times over using different inputs, and simpler models enable this because they use less computer processor time. (The quantification of uncertainty in light of present knowledge and available data should be clearly distinguished from the reduction of uncertainty that may occur as knowledge and data accumulate over time. We address the former, not the latter.)

While complexity obstructs uncertainty quantification, complex models may behave more like the real-world system, especially when pushed into Anthropocene conditions. So there is a trade-off between a model's capacity to realistically represent the system and its capacity to tell us how confident it is in its predictions. Both are desirable from a purely scientific or epistemic perspective as well as for their contributions to the model's utility in climate

https://doi.org/10.1086/711501 Published online by Cambridge University Press

risk management. Whether simpler is better in any given case depends on details that go beyond the scope of this article, but the critical importance of uncertainty assessment for addressing climate risks (e.g., Reilly et al. 2001; Smith and Stern 2011) is why simpler models can be epistemically better for informing decisions.

In what follows, we introduce the relevant notion of simplicity and a way to measure it (sec. 2). We then explain the link from simplicity to uncertainty quantification (sec. 3), arguing that through this link, simplicity becomes epistemically relevant to model choice and model development (sec. 4). We next briefly discuss the resulting trade-off, highlighting the roles of nonepistemic values and high-impact, low-probability outcomes in mediating the importance of uncertainty assessment for climate risk management (sec. 5).

2. Simplicity and Run Time. Environmental simulation models populate a spectrum from simple to complex, and attention to a model's position on this spectrum is a pervasive feature of both published research and everyday scientific discourse. All computational models idealize their target systems by neglecting less important processes and by discretizing space and time. What makes a model comparatively complex is the explicit representation of more processes and feedback thought to operate in the real-world system or greater resolution in the discretization (i.e., smaller grid size or a shorter time step). Greater complexity can allow for a more realistic depiction of the target system, while simpler models must work with a more idealized picture.

Realistic depictions can provide benefits but come at a cost: complex models demand more computer processor time. Here we use the length of time needed to run the simulation model on a computer, or the model's *run time*, as a proxy for model complexity. (To begin, suppose the model runs on a single processor; we discuss parallel computing further below.) Run time is, of course, a processor-dependent measure: a faster computer processor can run the same program in less time. But this will not hamper our discussion since we are ultimately concerned with between-model comparisons that can be relativized to fixed hardware without loss of import.[1] Moreover, differences in processor speed are in practice relatively small (in the neighborhood of times two for processors in service at any given time) when compared with differences in run time across models (factors of tens to trillions). Run time also depends on the time span simulated, but this, too, is something we can hold constant across models in order to compare apples to apples.

1. An alternative, processor-independent way to quantify computational expense might be to count the number of floating-point operations (FLOPs) required by the model's compiled program. But this is not a common metric; modelers typically know their model's run time but not its FLOP count, and the processor-hour is the usual unit in which scientific computing resources are allocated and purchased. So we continue with run time.

https://doi.org/10.1086/711501 Published online by Cambridge University Press

Another feature of run time as a measure of simplicity is that it applies to models understood in the most concrete sense. Run time is not a feature of calculations understood abstractly but of a specific piece of computer code written in a specific programming language (and run on a specific machine). A consequence of this is that two pieces of code with meaningfully different run times can instantiate what is, in some sense, the same model. What that means for our discussion is that the trade-off we examine applies most unyieldingly to computationally efficient programming; inefficiently coded models can, to a point, be sped up without sacrificing realism.

Run time can be contrasted with other concepts already implicated in discussions of simplicity's role in science, for example, the number of adjustable parameters in a hypothesis. Run time quantifies the amount of calculating needed to use the model, while the number of parameters concerns the model's plasticity in the face of observations. Simulation models contain adjustable parameters, but their number is often poorly defined, since quantities appearing in the computer code might be fixed in advance for one application but allowed to vary in the next.[2]

To make the discussion more concrete (and point readers to further details), we introduce a small collection of models that together illustrate the simplicity-complexity spectrum in environmental simulation modeling. We choose from models that have been used to investigate the contribution to sea level rise from the Antarctic ice sheet (AIS), a key source of uncertainty about future sea level rise on time scales of decades to centuries (DeConto and Pollard 2016; Bakker, Louchard, and Keller 2017; Bakker, Wong, et al. 2017).

The Danish Antarctic Ice Sheet model (DAIS; Shaffer 2014; Ruckert et al. 2017) is the simplest of four models we consider here. It represents the AIS as a perfect half spheroid resting on a shallow cone of land—a highly aggregate representation in the sense that just a few numbers summarize a vast and varied landscape (the AIS is larger than the United States). DAIS represents several key processes governing ice mass balance, including snow accumulation and melting at contact surfaces with air, water, and land.

The Building Blocks for Relevant Ice and Climate Knowledge model (BRICK; Bakker, Wong, et al. 2017; Wong, Bakker, Ruckert, et al. 2017) couples a slightly expanded version of DAIS with similarly aggregate models of global atmosphere and ocean temperature, thermal expansion of ocean water,

2. Strictly speaking, a model does not have a single run time but a range of run times, one for each parameter choice. Still, we can continue to speak sensibly of run time as a property of the model, since different parameter choices typically result in similar run times, so long as time step and resolution are not treated as parameters. Because these two features contribute to our notion of simplicity, and because they are generally held fixed during calibration and projection, we see them as part of what defines the model.

https://doi.org/10.1086/711501 Published online by Cambridge University Press

and contributions to sea level from other land ice (glaciers and the Greenland ice sheet). Compared to DAIS, BRICK represents an additional AIS process (marine ice sheet instability), as well as a number of interactions with other elements of the global climate system, including feedback between sea level change and AIS behavior.

The Pennsylvania State University 3-D ice sheet-shelf model (PSU3D; Pollard and DeConto 2012; DeConto and Pollard 2016) includes fewer global-scale interconnections than BRICK but a much richer representation of both AIS processes and local ocean and atmosphere interactions. PSU3D is a spatially resolved model of AIS in the sense that spatial variation in ice thickness and underlying topography are explicitly represented and incorporated into AIS dynamics. PSU3D represents many additional AIS processes (beyond DAIS), including ice flow through deformation and sliding, marine ice cliff instability, and ice shelf calving.

The last and most complex model we use to illustrate the simplicity spectrum is the Community Earth System Model (CESM; Hurrell et al. 2013; Lipscomb and Sacks 2013; Lenaerts et al. 2016), incorporating spatially resolved atmosphere, ocean, land surface, and sea ice components and allowing global ocean and atmosphere circulation to interact with the AIS. While dynamic (two-way) coupling of CESM with a full AIS model is still under development (Lipscomb 2017, 2018), recent work (Lenaerts et al. 2016) uses a static ice sheet surface topography to investigate one aspect of future AIS dynamics, the *surface mass balance* (net change in ice mass due to precipitation, sublimation, and surface melt).

The resolution and run time of the models described above are provided in table 1. Figure 1 summarizes and illustrates the models with special attention to key differences in complexity that account for their different run times.

3. Epistemic Relevance for Model Choice. Having introduced a notion of simplicity and a way to measure it, we now turn to the benefits enabled by this kind of simplicity. The proximate consequence of using a simpler model is that a shorter run time allows for more model runs. That, in turn, has consequences for what one can learn from the model. But we begin with the proximate step.

Given a computing budget of some number of processor-hours, a simple calculation of computing budget divided by run time yields a theoretical maximum for the number of times the model can be run. Figure 2*a* displays this reciprocal relationship for two example computing budgets. Each point along such a curve corresponds to a different model choice: moving from left to right, one trades away model complexity (run time) in exchange for more runs. (Because such plots become hard to read with larger numbers, it will be helpful to use a logarithmic scale on the axes; we introduce this visualization in fig. 2*b*.)

https://doi.org/10.1086/711501 Published online by Cambridge University Press

TABLE 1. Four Simulation Models

Model	Resolution, AIS (km)	Resolution, Atmosphere (km)	Approximate Run Time (min)	Reference
DAIS	NA	NA	.001	Ruckert et al. (2017)
BRICK	NA	NA	.1	Wong, Bakker, Ruckert, et al. (2017)
PSU3D	10	40 (regional)	25,000	DeConto and Pollard (2016)
CESM	110*	110 (global)	2,000,000,000†	Lenaerts et al. (2016)

Sources.—Bakker, Applegate, and Keller (2016), DeConto and Pollard (2016), Lenaerts et al. (2016), UCAR (2016), Ruckert et al. (2017), Wong, Bakker, Ruckert, et al. (2017), and personal communication with Kelsey Ruckert, Tony Wong, and Robert Fuller.

Note.—Resolution and approximate run time of four simulation models. Run times are for 240,000-year hindcasts, which enable model calibration to incorporate key paleoclimate data. Model configurations are as per the reference.

* Resolution of CESM's land surface component used to calculate surface mass balance.

† The hindcast length used for this comparison is impractical for models as complex as CESM. Yet because of parallel computing (see sec. 3.1) this number is not quite as outlandish as it may seem.

As per figure 2, run time and a computing budget determine how many model runs can be carried out. This run limit in turn constrains what methods can be employed at key stages of a modeling study, including calibration and projection. Model *calibration* is a process of tuning, weighting, or otherwise constraining the values of adjustable parameters in order to make the model the best representation that it can be of the system under study. Subsequent *projection* involves running the calibrated model into the future to see what it foretells, conditional on assumptions about how required exogenous (supplied from outside the model) inputs will play out over the time frame in question.[3] We discuss calibration and projection in turn, in each case detailing how the feasible number of model runs constrains the approach taken to these modeling tasks and what those constraints mean for the quantification of uncertainty.

3.1. Calibration. In the geosciences, model calibration typically aims to exploit both fit with data and prior knowledge about parameters (which often have a physical interpretation; see, e.g., Pollard and DeConto 2012; Shaffer 2014). How this is done varies, and some methods require more model runs than others. To illustrate the dependence between model simplicity and

3. Although we used the word "prediction" in our broad-brush introduction, strictly speaking, we are concerned with projections. While projection is a subspecies of prediction construed broadly, the two are mutually exclusive on some finer-grained nomenclatures (MacCracken 2001; Bray and von Storch 2009). We use "projection" to emphasize the hypothetical nature of some assumptions—especially exogenous inputs such as future greenhouse gas emissions. Another way to put it is that by "projection" we mean a prediction conditional on certain inputs (like future emissions) about which the modeler explicitly makes no probability judgments.

https://doi.org/10.1086/711501 Published online by Cambridge University Press

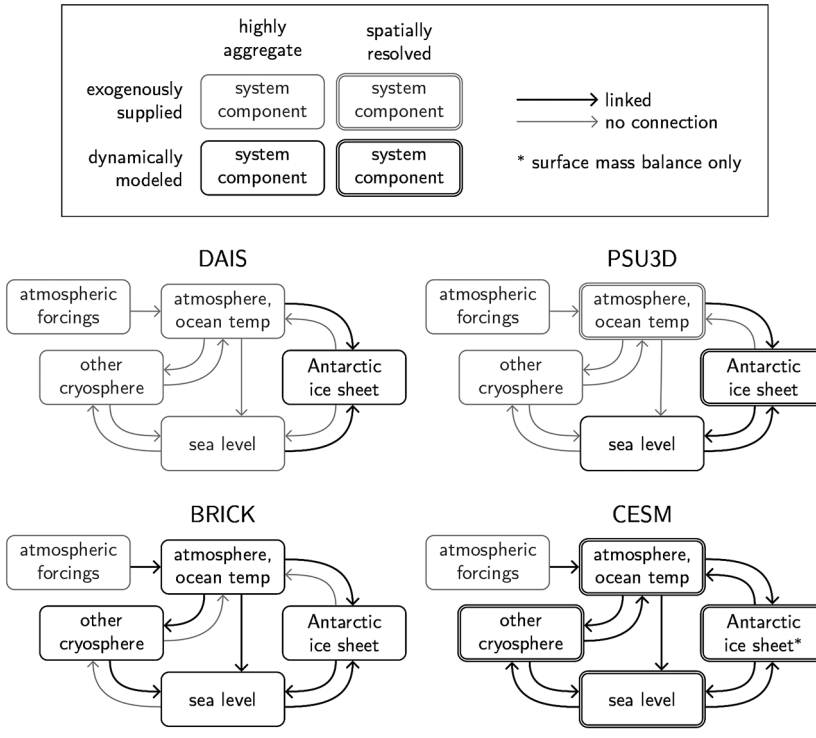


Figure 1. Four environmental simulation models discussed in the text. Based on configurations of DAIS, BRICK, PSU3D, and CESM in Ruckert et al. (2017), Wong, Bakker, Ruckert, et al. (2017), DeConto and Pollard (2016) (Pliocene simulation), and Lenaerts et al. (2016), respectively. Different configurations of the same model may correspond to somewhat different depictions within the visual schema used here. CESM includes many additional system components not pictured.

calibration methods, we contrast three methods that together span the runs-required gamut. At one extreme lies Markov Chain Monte Carlo (MCMC), the gold standard in Bayesian model calibration (Bayesian inference is a natural fit for the task of integrating observations with prior knowledge). Near the middle is an approach called (somewhat confusingly) *precalibration*.[4] At the other extreme lies *hand tuning*. We describe each below.

For the following discussion, assume a model that is deterministic and at each time step calculates the next system state as a function of the current

4. The approach can be applied as a preliminary step before further calibration (Edwards, Cameron, and Rougier 2011, sec. 7), but here we discuss its use as a standalone method of calibration (as per Sriver et al. 2012; Ruckert et al. 2017).

https://doi.org/10.1086/711501 Published online by Cambridge University Press

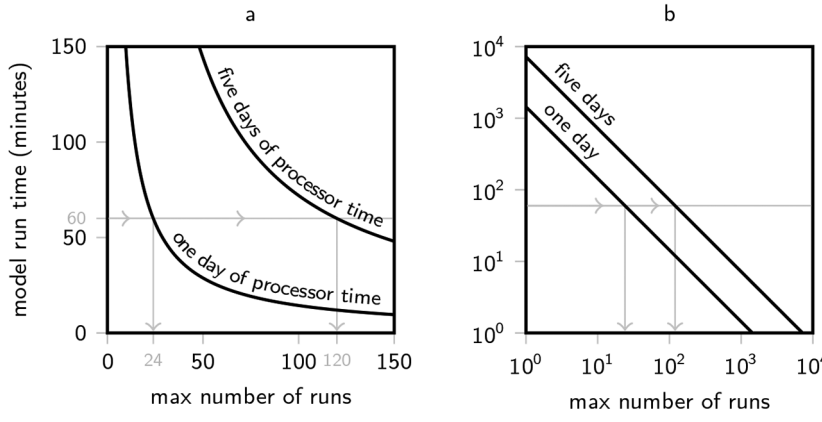


Figure 2. *a*, Two computing budgets, illustrating the reciprocal relationship between run time and number of runs. Arrows illustrate how to read the figure: a model that runs in 1 hour (e.g.) can be run 24 times on a 1-day budget and 120 times on a 5-day budget. *b*, Same figure, now plotted on logarithmic axes for a wider view (we use this format to accommodate very large and very small numbers in one plot).

state plus exogenous inputs (also called *forcings*) impinging on the system. Assume a set of historical observations, both of the forcings and of quantities corresponding to the model's state variables. Finally, assume a measure of fit between the time series of observations and the corresponding series of values produced by the model when driven by historical forcings (a *hindcast*).

Bayesian calibration begins with a prior probability distribution over all parameter combinations (the model's *parameter space*) and updates that prior in light of observations to arrive at a posterior distribution. In principle, the updating takes into account fit between those observations and every possible version of the model (every combination of parameter values). The posterior can therefore be calculated exactly (analytically) only where a suitably tractable mathematical formula maps parameter choices to hindcast-observation fit (the likelihood function). This is not the case for computer simulation models, where fit can be assessed only by running the simulation and comparing the resulting hindcast with the observations. In this case, the posterior can still be approximated numerically, but this requires evaluating fit with observations (running the model each time) for a very large number of parameter choices. MCMC is a standard approach to numerically approximating Bayesian posterior distributions and requires tens of thousands to millions of model runs to implement (Metropolis et al. 1953; Kennedy and O'Hagan 2001; van Ravenzwaaij, Cassey, and Brown 2018; for application to ice sheet modeling, see Bakker, Applegate, and Keller [2016] and Ruckert et al. [2017]).

https://doi.org/10.1086/711501 Published online by Cambridge University Press

In contrast to MCMC's thorough survey of parameter space, precalibration involves running the model at a smaller number of strategically sampled parameter-value combinations (e.g., 250, 1,000, and 2,000, in Sriver et al. [2012], Ruckert et al. [2017], and Edwards et al. [2011], respectively). Each of the resulting hindcasts is compared against observations using a streamlined, binary standard of fit—sorting the hindcasts into two classes: those *reasonably* similar to the observations and those not. The result is a dichotomous characterization of parameter choices as *plausible* or *implausible*, the latter unsuitable for use in projection.

The third method we highlight, hand tuning, encompasses a diverse set of practices that share some common features including varying parameters one at a time rather than jointly, using different approaches to model-observation fit (or different observations altogether) for different parameters, calibrating submodel components separately, a greater emphasis on expert assessment of parameter values, and a goal of identifying a single best parameter choice (for at least a majority of the parameters, sometimes for all). Hand tuning may not be clearly separated from model development, tends to be less transparent than the previous two approaches, and examines overall model behavior (and model-observation fit) for a relatively small number of parameter choices (very roughly, less than 100). Examples include Scheller et al. (2007) and Pollard and DeConto (2012); calibration of general circulation models (Meehl et al. 2007) and earth system models such as CESM also generally falls into this category (Hourdin et al. 2017).

To summarize key points for current purposes, hand tuning samples the smallest number of possible parameter choices and aims to identify the best among them. Precalibration examines (on the order of) 100 times more parameter choices and issues a dichotomous division of those into the plausible and implausible. MCMC examines (roughly) 1,000 times more than that and exhaustively quantifies the relative plausibility of every possible parameter choice in the form of a probability distribution. Methods that demand more model runs do more to characterize uncertainty about parameter choices, both by testing more possible values for the parameters and by furnishing a richer characterization of uncertainty about those values.

By limiting the feasible number of runs, model complexity undermines the characterization of parameter uncertainty. Figure 3 illustrates this point using the models from section 2 and two example computing budgets. The lower diagonal line shows a 10-day budget (240 processor-hours). For researchers working on a single processor, this is a plausible limit on the computing time that can realistically be devoted to model calibration (in part since the calibration procedure might be repeated three to five times in the course of troubleshooting and replicating results). Points on or below this line are feasible on a 240 processor-hour budget. The figure shows that on

https://doi.org/10.1086/711501 Published online by Cambridge University Press

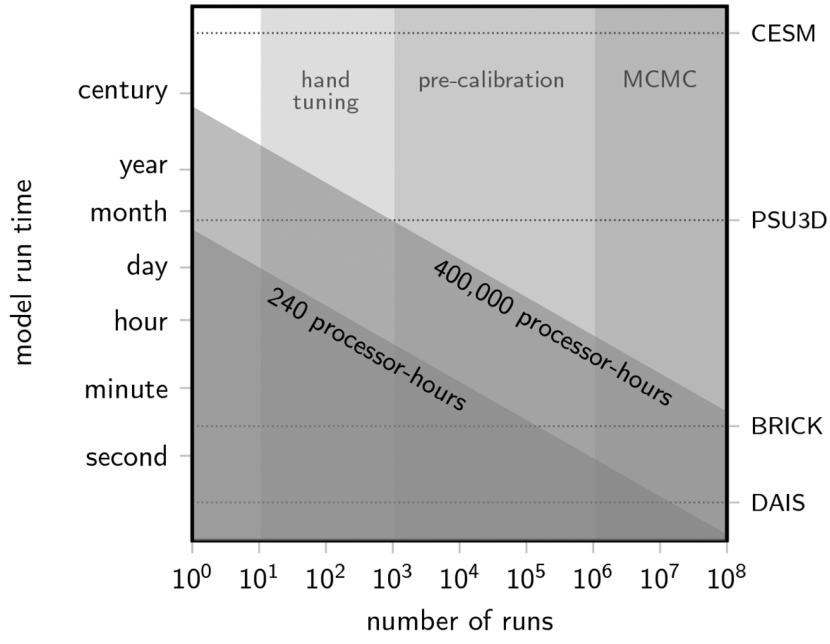


Figure 3. Example computing budgets compared with approximate run times for four simulation models (see table 1). Left boundaries of the shaded columns show approximate minimum run requirements for calibration methods discussed in the text (supposing ~10 parameters are calibrated). The region below a computing-budget diagonal shows which calibration methods are feasible for each model on that budget. Expands on Bakker, Applegate, and Keller (2016). Color version available as an online enhancement.

this budget, DAIS can be calibrated by MCMC and BRICK by precalibration; PSU3D and CESM cannot be calibrated by any means.

With access to a high-performance computing cluster, much larger computing budgets can be contemplated: 400,000 processor-hours is a routine high-performance computing allocation in 2019 for research supported by awards from the US National Science Foundation (UCAR 2019). Since 400,000 hours would occupy a single processor for 46 years, such a budget can be properly exploited only where the computing workload can be *parallelized* (split between multiple processors that run in parallel). Spread over 1,000 processors, 400,000 hours lasts a little over 2 weeks. While precalibration is easily parallelized, the most widely used algorithm for implementing MCMC (Metropolis et al. 1953; Robert and Casella 1999) requires that model runs be executed serially. There are, however, a number of kindred

https://doi.org/10.1086/711501 Published online by Cambridge University Press

approaches to numerically approximating a Bayesian posterior, some of which can be substantially parallelized (Lee et al. 2020, and references therein).

The upper diagonal in figure 3 shows a 400,000-hour budget. With those resources, BRICK can easily be calibrated by numerical Bayesian methods, and PSU3D moves to the edge of precalibration territory. A single hindcast using CESM is still well out of reach. Parallelization and large computing clusters substantially shift the goalposts, but they cannot dissolve the fundamental trade-off between model complexity and uncertainty quantification.

3.2. Projection. To the degree that parameter uncertainty has been characterized during calibration, it can then be propagated into projections. The most frugal approach to projection would be a single model run looking into the future. This can provide a best guess about future system behavior but does not offer any characterization of uncertainty around that guess. To do that requires additional runs using alternate, also-plausible parameter choices to generate correspondingly also-plausible projections. A collection of different parameter choices leads to an *ensemble* of projections that can collectively characterize how uncertainty in parameter values translates to uncertainty in future system behavior.

The characterization of parameter uncertainty provided by MCMC (or other Bayesian numerical methods) allows for projection ensembles that are interpretable probabilistically (Ruckert et al. 2017; Wong, Bakker, Ruckert, et al. 2017; Lee et al. 2020). Precalibration allows for a dichotomous grading of plausibility in projected futures. Hand tuning offers little information about parameter uncertainty that could be propagated into a projection ensemble.

The size of the ensemble (i.e., number of projection runs) needed for high-fidelity propagation of characterized parameter uncertainty varies depending on many particulars, including the number of parameters calibrated (ensemble sizes in the thousands are typical; e.g., Ruckert et al. 2017; Wong, Bakker, Ruckert, et al. 2017). By limiting the feasible number of runs, complexity can preclude projection ensembles of sufficient size. In this way, model complexity constrains not only the characterization of parameter uncertainty but also its propagation into projections.

Moreover, parameter values are not the only uncertain inputs into model projections. Incorporating additional sources of uncertainty requires expanding the projection ensemble. Where uncertainty about initial conditions makes a meaningful difference to model projections, these conditions can be varied across ensemble members (e.g., Daron and Stainforth 2013; Deser et al. 2014; Sriver, Forest, and Keller 2015). *Exogenous forcings* are often deeply uncertain and treated using a *scenarios* approach (Schwartz 1996; Carlsen et al. 2016), multiplying the projection ensemble by the number of scenarios used (three to five is typical). The *model structure* (assumptions built into the model regardless of parameter choice) can also be questioned. Expanding the model

https://doi.org/10.1086/711501 Published online by Cambridge University Press

structure (Draper 1995) or explicitly characterizing model discrepancy (Brynjarsdóttir and O'Hagan 2014) adds new parameters that, in turn, raise the run demands of both calibration and projection. Alternatively, repeating the entire work flow (calibration and projection) with several different models multiplies the required runs by the number of different models used.

The overall message on model runs and projection is that the more thoroughly one wishes to characterize uncertainty in projections, the larger the required ensemble. Specifically, more thorough characterization of uncertainty means that more of the assumptions built into the modeling have been questioned, with the consequences of questioning (varying) those assumptions having been propagated into projected system behavior. Jointly addressing multiple sources of uncertainty can lead to very large projection ensembles (e.g., 10 million model runs; Wong and Keller 2017).[5]

We have discussed two key phases in simulation modeling studies: calibration and projection. In each phase, the approach taken and the results obtainable are strongly constrained by a model's run time.[6] The absolute numbers of runs needed for thorough characterization of known uncertainties can be very large, easily outstripping realistic computing budgets for more complex models. A fixed budget can therefore enforce a harsh trade-off between model complexity (and the realism it enables) and characterization of uncertainty in model projections. Another way to put it is that complexity limits what can be learned from the model. For this reason, simplicity is epistemically relevant to model choice and model development.

4. Epistemic and Nonepistemic Benefits. We have argued that simplicity, measured via run time, is epistemically relevant to model choice. But some readers may be drawn to another framing of the issue, on which the value of this kind of simplicity is in fact not epistemic but merely practical (and therefore outside the traditional focus of the simplicity literature). After all, it would seem that however complex the model, projection uncertainties could be thoroughly characterized with enough processor time. The benefit of

5. The computing demands of projection relative to calibration vary from one modeling exercise to the next, and they depend on not only the number of runs used at each stage but also the length of those runs. Regarding the AIS, sparse data and processes operating on geological time scales motivate calibration hindcasts of at least hundreds of millennia. The length of projections, however, often reflects policy planning horizons and may extend only a century or two into the future. Since a model's run time is generally proportional to the number of simulated years, in this case one calibration run (hindcast) needs a thousand times more processor time than one projection run.

6. Another important modeling activity (that we lack space to discuss but where a similar lesson applies) is *sensitivity testing* (Bankes 1993; Sobol 2001; Wong and Keller 2017); also see the notion of *relevant dominant uncertainty* (Smith and Petersen 2014).

https://doi.org/10.1086/711501 Published online by Cambridge University Press

simplicity, then, is that it reduces the processor time needed to complete the research—which may sound like a practical matter rather than an epistemic one.

When comparing the consequences of different model choices, one must hold something else fixed in order to structure the comparison. The reasoning sketched above implicitly holds fixed the desired approach to calibration and projection (keeping the number and length of model runs the same). But you can make a different sort of comparison by holding something else fixed. We compare the consequences of model choice by holding the computing budget fixed, in which case simpler models enable different approaches to calibration and projection, resulting in better uncertainty quantification—a recognizably epistemic upshot. Which comparison gives the "right" answer? The two results are complementary, not contradictory. Neither comparison tells the whole story; each isolates and illuminates one aspect of a bigger-picture bundle of trade-offs.[7]

Analogous contrasting perspectives can be applied to the issue of cognitive benefits (such as ease of use), which are routinely dismissed as nonepistemic advantages of simplicity (e.g., Kelly and Mayo-Wilson 2010; Sober 2015; Baker 2016). One way to reach this dismissive conclusion is to assume a fixed research plan detailing the concrete steps to be taken within a research project (analogous to fixing the desired approach to calibration and projection). On this sort of comparison, the benefit of employing a simple, easy-to-use theory rather than a complex and burdensome one appears to be getting the proposed work done faster or with less effort (a seemingly nonepistemic benefit). Yet, a different sort of comparison can be made by supposing a fixed cognitive-effort "budget," in which case easier use translates to more research completed and, as a result, more knowledge (or greater fulfillment of some epistemic value or other).[8]

For comparison, it is worth noting that the benefits of other, well-discussed notions of simplicity also admit of multiple framings, where one perspective highlights an epistemic upshot and another highlights a practical one. Notions of simplicity that concern the flexibility of a model or hypothesis get their epistemic relevance as a result of viewing the choice between simple

7. Acknowledging different ways of making such comparisons (by holding different things fixed) may help clarify a disagreement on whether using streamlined, less reliable scientific methods in resource-constrained regulatory contexts illustrates nonepistemic factors intruding on method choice (Elliott and McKaughan 2014; Steel 2016a). If the resource budget is framed as a part of the give and take, then yes, epistemic considerations have been traded against nonepistemic. But if that budget is viewed as a fixed, external constraint, then the methods trade-off pits quantity against quality of scientific results, both of which are recognizably epistemic.

8. For more discussion connecting simplicity with ease of use, see Frigg and Hoefer (2015) on a notion of "derivation-simplicity" and Douglas (2013, 800) on simpler claims being "easier to follow through to their implications."

https://doi.org/10.1086/711501 Published online by Cambridge University Press

and complex models while holding fixed the quantity of data available. Akaike information criterion (AIC) scores (see Forster and Sober 1994), for example, give advice about which statistical model will yield more accurate out-of-sample predictions after fitting to data. AIC does this by rewarding fit with data while penalizing flexibility (number of parameters). But the influence of the parameter penalty decreases as the number of data increase, so the more data one has, the less simplicity matters. This means that if we instead hold fixed the goal of some desired degree of predictive accuracy, the benefit of simplicity will now show up in the quantity of data needed to achieve that goal—or more to the point, the time and expense of obtaining those data. As before, making a different sort of comparison pivots an epistemic consideration into one more naturally viewed as nonepistemic.

The fixed-data perspective is often salient because obtaining more data can be costly, slow, or otherwise impractical (and because the division of scientific labor often divorces statistical analyses from data gathering). But developments in the nature of scientific research have made our fixed-budget perspective equally salient. The growth of scientific computing has shifted work from brains to computer processors where it is more easily quantified and tracked. The complexity of computational simulation models has shadowed the exponential growth of computing power, massively increasing the calculating required to answer even simple questions using a model. At the same time, run-hungry statistical computing methods for calibration and projection of these simulations multiply the "cognitive effort" advantage of simpler models thousands to millions of times over, all within the scope of a single study or publishable unit of research. As a result, the trade-offs illuminated by contrasting modeling options on a fixed computing budget are now critical to a full understanding of the epistemology of simulation modeling.

5. Purpose and Values. Simplicity facilitates uncertainty quantification, but complex models can be more realistic and may behave more like the real-world system. How much complexity is the right amount? The question raises challenging scientific and technical issues requiring deep, case-by-case integration of geoscience, statistics, computing, and numerical approximation (issues that go far beyond the scope of this article). But equally important is the general qualitative point that, like other aspects of model evaluation (Parker 2009; Haasnoot et al. 2014; Addor and Melsen 2019), much depends on the purpose of the modeling exercise. Simulation modeling to improve scientific understanding, for example, may demand realism and benefit little from uncertainty quantification. Informing decisions, however, often demands attention to uncertainty (Smith and Stern 2011; Keller and Nicholas 2015; Rougier and Crucifix 2018).

Broadly speaking, risk assessment involves contemplating what outcomes might occur, how likely each is, and how bad each would be (Kaplan and

https://doi.org/10.1086/711501 Published online by Cambridge University Press

Garrick 1981). These components jointly characterize the risk associated with a given course of action. Thinking in terms of probability and cost, for example, risk might be expressed as *expected cost*. This is not to say that risk management requires probabilities (Dessai and Hulme 2004; Lempert et al. 2013; Weaver et al. 2013), only that some sense of the plausibility of different outcomes is needed to assess and manage risk (and that probability estimates are a common medium for this). Because simplicity enables the required uncertainty quantification while complexity impedes it, the simplicity-complexity dimension of model choice strongly influences a model's adequacy for the purpose of supporting decisions.

In climate risk management specifically, the importance of simplicity is magnified by the role of high-impact, low-probability outcomes. The limits that complexity places on uncertainty quantification are particularly unfavorable to estimating the chances of extreme possibilities, or what are referred to (in probability terms) as the *tails* of a distribution (e.g., Sriver et al. 2012; Wong and Keller 2017; Lee et al. 2020). But since these extreme outcomes (e.g., large or rapid sea level rise) are also the most dangerous and costly, estimating their probability can be central to managing risks, and relatively small changes to their estimated probability can have an outsize impact on risk calculations and management strategies.

A study by Wong, Bakker, and Keller (2017) serves to illustrate these points. The authors use a relatively simple model of the AIS and other contributors to sea level rise (BRICK; sec. 2), allowing for rigorous quantification of parameter uncertainty via MCMC, followed by multiple large ensembles to propagate that uncertainty into local sea level rise projections for the city of New Orleans under each of several forcing (greenhouse gas concentration) scenarios. The resulting projections (plus other inputs and assumptions) allow for estimation of a site-specific, economically optimal levee height (such that building any higher costs more than the flood damage it would be expected to prevent) for each concentration scenario.

The simplicity of BRICK also enabled Wong et al. to characterize some model uncertainty by repeating the entire workflow for two different model configurations: one with and one without an additional (poorly understood but potentially important) mechanism of ice sheet behavior labeled *fast dynamics* (Pollard, DeConto, and Alley 2015; DeConto and Pollard 2016). Focusing on just one of the city's five levee rings and assuming a business-as-usual greenhouse gas scenario (RCP8.5; van Vuuren et al. 2011), the authors quantify the impact of this model uncertainty by confronting the base-case model's economically optimal levee with projections from the fast-dynamics model configuration. The result is an increase in estimated annual chance of flooding (seawater overtopping the levee) of one-half of one-tenth of a percent (from eight in 10,000 to 13 in 10,000). This seemingly small change adds $175 million in expected flood damage between now and 2100. A levee

https://doi.org/10.1086/711501 Published online by Cambridge University Press

25 centimeters higher could prevent much of that damage, with estimated net savings of $53 million.

To underline the key points of the illustration: simplicity can contribute to a model's adequacy for purpose by enabling quantification and propagation of parameter uncertainty into projections; estimation of probabilities for high-impact, low-probability outcomes; and characterization of deeper uncertainties (e.g., model structure, forcing scenario) by spelling out how alternative assumptions affect management strategies. Where complexity undermines such modeling activities, the model's adequacy for purpose suffers.

The broad-brush purpose of supporting climate risk management can be analyzed further in any particular instance to reveal specific nonepistemic concerns such as protecting livelihoods, preserving culture, and saving money and lives (Bessette et al. 2017; CPRAL 2017). By judging models in light of purpose while also viewing these motivating values as a part of that purpose, the simplicity-complexity dimension of model choice can be seen as a coupled ethical-epistemic problem (Tuana 2013, 2017; Vezér et al. 2018) in which motivations and trade-offs encompass both epistemic and ethical values.

The prospect of ethical values motivating model choice may raise concerns about such values overstepping their proper role in science, and at this point our discussion links up with a large literature on ethical (or more broadly, nonepistemic) values in science (Douglas 2009; Elliott 2017), a portion of which addresses climate science specifically (e.g., Steele 2012; Winsberg 2012; Betz 2013; Parker 2014; Intemann 2015; Steel 2016b). Here we can only note this connection, leaving further exploration of the topic for future work.

6. Conclusion. Discussions of simplicity's role in scientific method and reasoning have often recognized a loose notion of cognitive benefit—or benefit in terms of cognitive effort—associated with simple theories or models. Yet this aspect of simplicity has received relatively little attention in philosophy of science, either because the advantage is seen as self-evident and trivial or because the upshot is judged a matter of convenience, not epistemology.

This convenience-not-epistemology verdict is a natural consequence of the practice (common in much traditional philosophy of science) of attending to formal relationships between theory and data while idealizing away the messy human elements of science. But for today's computer simulation models, the "effort" required to operate the model—now understood in terms of computing resources, not cognitive burden—is too consequential to neglect. Computing demands sharply constrain how a model can be used and what can be learned from it.

We have used the run time of a simulation model as a measure of the model's complexity: simple models run faster, and complex models run more slowly. The importance of run time to the epistemology of computer simulation can be seen clearly by adopting what we have called a fixed-budget

https://doi.org/10.1086/711501 Published online by Cambridge University Press

perspective: compare what can be achieved with a simpler model to what can be achieved with a more complex one on the same computing budget. On such a comparison, simplicity facilitates quantification of parameter uncertainty and propagation of this and other sources of uncertainty into model projections, including estimates of chances for low-probability, high-impact outcomes.

How much one values these benefits is a further question that is tied up with the purpose of a modeling activity. One purpose for which uncertainty assessment can be critical is informing climate risk management. One specific example is managing flood risk in costal communities facing sea level rise, but there are, of course, many others (e.g., Butler et al. 2014; Keller and Nicholas 2015; Hoegh-Guldberg et al. 2018).

None of this takes away from the important purposes served by very complex—and maximally realistic—environmental simulation models, including advancing understanding of processes and their interactions across multiple scales and expanding the range of model structures that can be explored by the research community as a whole. Our discussion highlights the high stakes and harsh trade-offs inherent in model choice and model development—and the central role of simplicity in prioritizing the various scientific and social benefits gleaned from environmental simulation modeling.

REFERENCES

Addor, N., and L. Melsen. 2019. "Legacy, Rather than Adequacy, Drives the Selection of Hydrological Models." *Water Resources Research* 55 (1): 378–90.

Akaike, H. 1973. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Proceedings of the 2nd International Symposium on Information Theory*, ed. B. N. Petrov and F. Csaki, 267–81. Budapest: Akademiai Kiado.

Baker, A. 2016. "Simplicity." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. https://plato.stanford.edu/archives/win2016/entries/simplicity/.

Bakker, A. M., P. J. Applegate, and K. Keller. 2016. "A Simple, Physically Motivated Model of Sea-Level Contributions from the Greenland Ice Sheet in Response to Temperature Changes." *Environmental Modelling and Software* 83:27–35.

Bakker, A. M., D. Louchard, and K. Keller. 2017. "Sources and Implications of Deep Uncertainties Surrounding Sea-Level Projections." *Climatic Change* 140 (3–4): 339–47.

Bakker, A. M., T. E. Wong, K. L. Ruckert, and K. Keller. 2017. "Sea-Level Projections Representing the Deeply Uncertain Contribution of the West Antarctic Ice Sheet." *Scientific Reports* 7 (1): 3880.

Bankes, S. 1993. "Exploratory Modeling for Policy Analysis." *Operations Research* 41 (3): 435–49.

Beisbart, C., and J. J. Saam, eds. 2019. *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*. Dordrecht: Springer.

Bessette, D. L., L. A. Mayer, B. Cwik, M. Vezér, K. Keller, R. J. Lempert, and N. Tuana. 2017. "Building a Values-Informed Mental Model for New Orleans Climate Risk Management." *Risk Analysis* 37 (10): 1993–2004.

Betz, G. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3 (2): 207–20.

Bray, D., and H. von Storch. 2009. "'Prediction' or 'Projection'? The Nomenclature of Climate Science." *Science Communication* 30 (4): 534–43.

https://doi.org/10.1086/711501 Published online by Cambridge University Press

Brynjarsdóttir, J., and A. O'Hagan. 2014. "Learning about Physical Parameters: The Importance of Model Discrepancy." *Inverse Problems* 30 (11): 114007.
Butler, M. P., P. M. Reed, K. Fisher-Vanden, K. Keller, and T. Wagener. 2014. "Inaction and Climate Stabilization Uncertainties Lead to Severe Economic Risks." *Climatic Change* 127 (3–4): 463–74.
Carlsen, H., R. Lempert, P. Wikman-Svahn, and V. Schweizer. 2016. "Choosing Small Sets of Policy-Relevant Scenarios by Combining Vulnerability and Diversity Approaches." *Environmental Modelling and Software* 84:155–64.
Clatterbuck, H. 2015. "Chimpanzee Mindreading and the Value of Parsimonious Mental Models." *Mind and Language* 30 (4): 414–36.
CPRAL (Coastal Protection and Restoration Authority of Louisiana). 2017. "Louisiana's Comprehensive Master Plan for a Sustainable Coast." Technical report, CPRAL, Baton Rouge.
Daron, J. D., and D. A. Stainforth. 2013. "On Predicting Climate under Climate Change." *Environmental Research Letters* 8 (3): 034021.
DeConto, R. M., and D. Pollard. 2016. "Contribution of Antarctica to Past and Future Sea-Level Rise." *Nature* 531 (7596): 591–97.
Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak. 2014. "Projecting North American Climate over the Next 50 Years: Uncertainty due to Internal Variability." *Journal of Climate* 27 (6): 2271–96.
Dessai, S., and M. Hulme. 2004. "Does Climate Adaptation Policy Need Probabilities?" *Climate Policy* 4 (2): 107–28.
Douglas, H. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
———. 2013. "The Value of Cognitive Values." *Philosophy of Science* 80 (5): 796–806.
Draper, D. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society* B 57 (1): 45–70.
Edwards, N. R., D. Cameron, and J. Rougier. 2011. "Precalibrating an Intermediate Complexity Climate Model." *Climate Dynamics* 37 (7): 1469–82.
Edwards, P. N. 2001. "Representing the Global Atmosphere: Computer Models, Data, and Knowledge about Climate Change." In *Changing the Atmosphere: Expert Knowledge and Environmental Governance*, ed. C. A. Miller and P. N. Edwards. Cambridge, MA: MIT Press.
Elliott, K. C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford: Oxford University Press.
Elliott, K. C., and D. J. McKaughan. 2014. "Nonepistemic Values and the Multiple Goals of Science." *Philosophy of Science* 81 (1): 1–21.
Forster, M., and E. Sober. 1994. "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions." *British Journal for the Philosophy of Science* 45 (1): 1–35.
Frigg, R., and C. Hoefer. 2015. "The Best Humean System for Statistical Mechanics." *Erkenntnis* 80 (3): 551–74.
Frigg, R., and J. Reiss. 2009. "The Philosophy of Simulation: Hot New Issues or Same Old Stew?" *Synthese* 169 (3): 593–613.
Frigg, R., L. A. Smith, and D. A. Stainforth. 2013. "The Myopia of Imperfect Climate Models: The Case of UKCP09." *Philosophy of Science* 80 (5): 886–97.
———. 2015. "An Assessment of the Foundational Assumptions in High-Resolution Climate Projections: The Case of UKCP09." *Synthese* 192 (12): 3979–4008.
Grüne-Yanoff, T., and P. Weirich. 2010. "The Philosophy and Epistemology of Simulation: A Review." *Simulation and Gaming* 41 (1): 20–50.
Haasnoot, M., W. Van Deursen, J. H. Guillaume, J. H. Kwakkel, E. van Beek, and H. Middelkoop. 2014. "Fit for Purpose? Building and Evaluating a Fast, Integrated Model for Exploring Water Policy Pathways." *Environmental Modelling and Software* 60:99–120.
Harman, G., and S. Kulkarni. 2007. *Reliable Reasoning: Induction and Statistical Learning Theory*. Cambridge, MA: MIT Press.
Hoegh-Guldberg, O., et al. 2018. "Impacts of 1.5°C Global Warming on Natural and Human Systems." In *Global Warming of 1.5°C: An IPCC Special Report on the Impacts of Global Warming of 1.5°C Above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate*

https://doi.org/10.1086/711501 Published online by Cambridge University Press

*Change, Sustainable Development, and Efforts to Eradicate Poverty*j, ed. V. Masson-Delmotte, et al. Geneva: Intergovernmental Panel on Climate Change.
Hourdin, F., et al. 2017. "The Art and Science of Climate Model Tuning." *Bulletin of the American Meteorological Society* 98 (3): 589–602.
Hurrell, J. W., et al. 2013. "The Community Earth System Model: A Framework for Collaborative Research." *Bulletin of the American Meteorological Society* 94 (9): 1339–60.
Intemann, K. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32.
Jakeman, A. J., R. A. Letcher, and J. P. Norton. 2006. "Ten Iterative Steps in Development and Evaluation of Environmental Models." *Environmental Modelling and Software* 21 (5): 602–14.
Jebeile, J. 2017. "Computer Simulation, Experiment, and Novelty." *International Studies in the Philosophy of Science* 31 (4): 379–95.
Kaplan, S., and B. J. Garrick. 1981. "On the Quantitative Definition of Risk." *Risk Analysis* 1 (1): 11–27.
Keller, K., and R. Nicholas. 2015. "Improving Climate Projections to Better Inform Climate Risk Management." In *The Oxford Handbook of the Macroeconomics of Global Warming*, 9–18. New York: Oxford University Press.
Kelly, K. T. 2004. "Justification as Truth-Finding Efficiency: How Ockham's Razor Works." *Minds and Machines* 14 (4): 485–505.
———. 2007. "A New Solution to the Puzzle of Simplicity." *Philosophy of Science* 74 (5): 561–73.
Kelly, K. T., and C. Mayo-Wilson. 2010. "Ockham Efficiency Theorem for Stochastic Empirical Methods." *Journal of Philosophical Logic* 39 (6): 679–712.
Kennedy, M. C., and A. O'Hagan. 2001. "Bayesian Calibration of Computer Models." *Journal of the Royal Statistical Society* B 63 (3): 425–64.
Lee, B. S., M. Huran, R. Fuller, D. Pollard, and K. Keller. 2020. "A Fast Particle-Based Approach for Calibrating a 3-D Model of the Antarctic Ice Sheet." *Annals of Applied Statistics* 14 (2): 605–34.
Lempert, R. J., et al. 2013. "Making Good Decisions without Predictions." Technical report, Rand, Santa Monica, CA.
Lenaerts, J. T., M. Vizcaino, J. Fyke, L. Van Kampenhout, and M. R. van den Broeke. 2016. "Present-Day and Future Antarctic Ice Sheet Climate and Surface Mass Balance in the Community Earth System Model." *Climate Dynamics* 47 (5–6): 1367–81.
Lipscomb, W. 2017. "Steps toward Modeling Marine Ice Sheets in the Community Earth System Model." Technical Report LA-UR-17-21665, Los Alamos National Laboratory.
———. 2018. "Ice Sheet Modeling and Sea Level Rise." Lecture, NCAR CESM Sea Level Session, January 10.
Lipscomb, W., and W. Sacks. 2013. "The CESM Land Ice Model Documentation and User's Guide." Technical report, National Center for Atmospheric Research.
Lloyd, E. A. 2010. "Confirmation and Robustness of Climate Models." *Philosophy of Science* 77 (5): 971–84.
———. 2015. "Model Robustness as a Confirmatory Virtue: The Case of Climate Science." *Studies in History and Philosophy of Science* A 49:58–68.
Lloyd, E. A., and E. Winsberg, eds. 2018. *Climate Modelling: Philosophical and Conceptual Issues*. Dordrecht: Springer.
MacCracken, M. 2001. "Prediction versus Projection-Forecast versus Possibility." *WeatherZine*, no. 26, February. https://sciencepolicy.colorado.edu/zine/archives/1-29/26/guest.html.
McGuffie, K., and A. Henderson-Sellers. 2001. "Forty Years of Numerical Climate Modelling." *International Journal of Climatology* 21 (9): 1067–109.
Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. Mitchell, R. J. Stouffer, and K. E. Taylor. 2007. "The WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research." *Bulletin of the American Meteorological Society* 88 (9): 1383–94.
Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *Journal of Chemical Physics* 21 (6): 1087–92.
Oreskes, N., D. A. Stainforth, and L. A. Smith. 2010. "Adaptation to Global Warming: Do Climate Models Tell Us What We Need to Know?" *Philosophy of Science* 77 (5): 1012–28.

https://doi.org/10.1086/711501 Published online by Cambridge University Press

Parke, E. C. 2014. "Experiments, Simulations, and Epistemic Privilege." *Philosophy of Science* 81 (4): 516–36.
Parker, W. 2014. "Values and Uncertainties in Climate Prediction, Revisited." *Studies in History and Philosophy of Science* A 46:24–30.
Parker, W. S. 2009. "Confirmation and Adequacy-for-Purpose in Climate Modelling." *Proceedings of the Aristotelian Society Supplementary Volume* 83 (1): 233–49.
———. 2010. "Predicting Weather and Climate: Uncertainty, Ensembles and Probability." *Studies in History and Philosophy of Science* B 41 (3): 263–72.
———. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78 (4): 579–600.
———. 2013. "Ensemble Modeling, Uncertainty and Robust Predictions." *Wiley Interdisciplinary Reviews: Climate Change* 4 (3): 213–23.
Petersen, A. C. 2012. *Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice*. Boca Raton, FL: CRC.
Pollard, D., and R. DeConto. 2012. "Description of a Hybrid Ice Sheet-Shelf Model, and Application to Antarctica." *Geoscientific Model Development* 5 (5): 1273–95.
Pollard, D., R. M. DeConto, and R. B. Alley. 2015. "Potential Antarctic Ice Sheet Retreat Driven by Hydrofracturing and Ice Cliff Failure." *Earth and Planetary Science Letters* 412:112–21.
Reilly, J., P. H. Stone, C. E. Forest, M. D. Webster, H. D. Jacoby, and R. G. Prinn. 2001. "Uncertainty and Climate Change Assessments." *Science* 293 (5529): 430–33.
Robert, C. P., and G. Casella. 1999. *Monte Carlo Statistical Methods*, chap. 7. Dordrecht: Springer.
Rougier, J., and M. Crucifix. 2018. "Uncertainty in Climate Science and Climate Policy." In *Climate Modelling: Philosophical and Conceptual Issues*, ed. E. A. Lloyd and E. Winsberg, 361–80. Dordrecht: Springer.
Ruckert, K. L., G. Shaffer, D. Pollard, Y. Guan, T. E. Wong, C. E. Forest, and K. Keller. 2017. "Assessing the Impact of Retreat Mechanisms in a Simple Antarctic Ice Sheet Model Using Bayesian Calibration." *PLoS ONE* 12 (1): e0170052.
Scheller, R. M., J. B. Domingo, B. R. Sturtevant, J. S. Williams, A. Rudy, E. J. Gustafson, and D. J. Mladenoff. 2007. "Design, Development, and Application of LANDIS-II, a Spatial Landscape Simulation Model with Flexible Temporal and Spatial Resolution." *Ecological Modelling* 201 (3–4): 409–19.
Schwartz, P. 1996. *The Art of the Long View: Planning in an Uncertain World*. New York: Currency-Doubleday.
Shaffer, G. 2014. "Formulation, Calibration and Validation of the DAIS Model (version 1): A Simple Antarctic Ice Sheet Model Sensitive to Variations of Sea Level and Ocean Subsurface Temperature." *Geoscientific Model Development* 7 (4): 1803–18.
Smith, L. A., and A. C. Petersen. 2014. "Variations on Reliability: Connecting Climate Predictions to Climate Policy." In *Error and Uncertainty in Scientific Practice*, ed. M. Boumans, G. Hon, and A. C. Petersen, 137–56. London: Pickering & Chatto.
Smith, L. A., and N. Stern. 2011. "Uncertainty in Science and Its Role in Climate Policy." *Philosophical Transactions of the Royal Society* A 369 (1956): 4818–41.
Smith, M. J., P. I. Palmer, D. W. Purves, M. C. Vanderwel, V. Lyutsarev, B. Calderhead, L. N. Joppa, C. M. Bishop, and S. Emmott. 2014. "Changing How Earth System Modeling Is Done to Provide More Useful Information for Decision Making, Science, and Society." *Bulletin of the American Meteorological Society* 95 (9): 1453–64.
Sober, E. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.
———. 2009. "Parsimony and Models of Animal Minds." In *The Philosophy of Animal Minds*, ed. R. W. Lurz, 237–57. Cambridge: Cambridge University Press.
———. 2015. *Ockham's Razors*. Cambridge: Cambridge University Press.
Sobol, I. M. 2001. "Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates." *Mathematics and Computers in Simulation* 55 (1–3): 271–80.
Sriver, R. L., C. E. Forest, and K. Keller. 2015. "Effects of Initial Conditions Uncertainty on Regional Climate Variability: An Analysis Using a Low-Resolution CESM Ensemble." *Geophysical Research Letters* 42 (13): 5468–76.
Sriver, R. L., N. M. Urban, R. Olson, and K. Keller. 2012. "Toward a Physically Plausible Upper Bound of Sea-Level Rise Projections." *Climatic Change* 115 (3–4): 893–902.

https://doi.org/10.1086/711501 Published online by Cambridge University Press

Steel, D. 2016a. "Accepting an Epistemically Inferior Alternative? A Comment on Elliott and McKaughan." *Philosophy of Science* 83 (4): 606–12.
———. 2016b. "Climate Change and Second-Order Uncertainty: Defending a Generalized, Normative, and Structural Argument from Inductive Risk." *Perspectives on Science* 24 (6): 696–721.
Steele, K. 2012. "The Scientist qua Policy Advisor Makes Value Judgments." *Philosophy of Science* 79 (5): 893–904.
Steele, K., and C. Werndl. 2016. "The Diversity of Model Tuning Practices in Climate Science." *Philosophy of Science* 83 (5): 1133–44.
Thompson, E., R. Frigg, and C. Helgeson. 2016. "Expert Judgement for Climate Change Adaptation." *Philosophy of Science* 83 (5): 1110–21.
Tuana, N. 2013. "Embedding Philosophers in the Practices of Science: Bringing Humanities to the Sciences." *Synthese* 190 (11): 1955–73.
———. 2017. "Understanding Coupled Ethical-Epistemic Issues Relevant to Climate Modeling and Decision Support Science." In *Scientific Integrity and Ethics in the Geosciences*, ed. L. C. Gundersen, 157–73. Hoboken, NJ: Wiley.
UCAR (University Corporation for Atmospheric Research). 2016. "CESM 1.2 Timing Table." Technical report, UCAR. http://www.cesm.ucar.edu/models/cesm1.2/timing/.
———. 2019. "University Allocations." Technical report, UCAR. https://www2.cisl.ucar.edu/user-support/allocations/university-allocations.
van Ravenzwaaij, D., P. Cassey, and S. D. Brown. 2018. "A Simple Introduction to Markov Chain Monte-Carlo Sampling." *Psychonomic Bulletin and Review* 25 (1): 143–54.
van Vuuren, D. P., et al. 2011. "The Representative Concentration Pathways: An Overview." *Climatic Change* 109 (1–2): 5.
Vapnik, V. 1998. *Statistical Learning Theory*. New York: Wiley.
Vezér, M., A. Bakker, K. Keller, and N. Tuana. 2018. "Epistemic and Ethical Trade-Offs in Decision Analytical Modelling." *Climatic Change* 147 (1): 1–10.
Vezér, M. A. 2016. "Computer Models and the Evidence of Anthropogenic Climate Change: An Epistemology of Variety-of-Evidence Inferences and Robustness Analysis." *Studies in History and Philosophy of Science* A 56:95–102.
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 Interdisciplinary Reviews: Climate Change* 4 (1): 39–60.
Weisberg, M. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.
Winsberg, E. 2010. *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press.
———. 2012. "Values and Uncertainties in the Predictions of Global Climate Models." *Kennedy Institute of Ethics Journal* 22 (2): 111–37.
———. 2018. "Computer Simulations in Science." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. https://plato.stanford.edu/archives/sum2018/entries/simulations-science/.
Wong, T. E., A. M. Bakker, and K. Keller. 2017. "Impacts of Antarctic Fast Dynamics on Sea-Level Projections and Coastal Flood Defense." *Climatic Change* 144 (2): 347–64.
Wong, T. E., A. M. Bakker, K. Ruckert, P. Applegate, A. Slangen, and K. Keller. 2017. "BRICK v0.2, a Simple, Accessible, and Transparent Model Framework for Climate and Regional Sea-Level Projections." *Geoscientific Model Development* 10 (7): 2741–60.
Wong, T. E., and K. Keller. 2017. "Deep Uncertainty Surrounding Coastal Flood Risk Projections: A Case Study for New Orleans." *Earth's Future* 5 (10): 1015–26.

https://doi.org/10.1086/711501 Published online by Cambridge University Press