

Earth's Future

RESEARCH ARTICLE

10.1029/2020EF001610

Key Points:

- Initial condition large ensembles of climate simulations have their scientific challenges, being expensive, possibly redundant, and biased
- Despite such challenges, large ensembles provide unique information to the scientific and policy communities
- Large ensembles are a valuable tool for robust decision-making, which is a strategy for making difficult decisions under deep uncertainty

Correspondence to:

J. S. Mankin,
mankin@dartmouth.edu

Citation:

Mankin, J. S., Lehner, F., Coats, S., & McKinnon, K. A. (2020). The value of initial condition large ensembles to robust adaptation decision-making. *Earth's Future*, 8, e2012EF001610. <https://doi.org/10.1029/2020EF001610>

Received 4 MAY 2020

Accepted 3 AUG 2020

Accepted article online 10 AUG 2020

Author Contributions:

Conceptualization: Justin S. Mankin

Data curation: Justin S. Mankin, Flavio Lehner, Sloan Coats, Karen A. McKinnon

Formal analysis: Justin S. Mankin, Flavio Lehner, Sloan Coats, Karen A. McKinnon

Funding acquisition: Justin S. Mankin

Investigation: Justin S. Mankin

Methodology: Justin S. Mankin, Flavio Lehner, Sloan Coats, Karen A. McKinnon

Project administration: Justin S. Mankin

(continued)

©2020. The Authors. Earth's Future published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

The Value of Initial Condition Large Ensembles to Robust Adaptation Decision-Making

Justin S. Mankin^{1,2} , Flavio Lehner^{3,4} , Sloan Coats^{5,6} , and Karen A. McKinnon⁷ 

¹Department of Geography, Dartmouth College, Hanover, NH, USA, ²Lamont-Doherty Earth Observatory, Columbia University, New York, NY, USA, ³Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland, ⁴National Center for Atmospheric Research, Boulder, CO, USA, ⁵Woods Hole Oceanographic Institution, Woods Hole, MA, USA, ⁶Department of Earth Sciences, University of Hawai'i at Mānoa, Honolulu, HI, USA, ⁷Department of Statistics and the Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA

Abstract The origins of uncertainty in climate projections have major consequences for the scientific and policy decisions made in response to climate change. Internal climate variability, for example, is an inherent uncertainty in the climate system that is undersampled by the multimodel ensembles used in most climate impacts research. Because of this, decision makers are left with the question of whether the range of climate projections across models is due to structural model choices, thus requiring more scientific investment to constrain, or instead is a set of equally plausible outcomes consistent with the same warming world. Similarly, many questions faced by scientists require a clear separation of model uncertainty and that arising from internal variability. With this as motivation and the renewed attention to large ensembles given planning for Phase 7 of the Coupled Model Intercomparison Project (CMIP7), we illustrate the scientific and policy value of the attribution and quantification of uncertainty from initial condition large ensembles, particularly when analyzed in conjunction with multimodel ensembles. We focus on how large ensembles can support regional-scale robust adaptation decision-making in ways multimodel ensembles alone cannot. We also acknowledge several recently identified problems associated with large ensembles, namely, that they are (1) resource intensive, (2) redundant, and (3) biased. Despite these challenges, we show, using examples from hydroclimate, how large ensembles provide unique information for the scientific and policy communities and can be analyzed appropriately for regional-scale climate impacts research to help inform risk management in a warming world.

Plain Language Summary Estimating uncertainties in projections of climate change poses challenges but is crucial to focusing scientific and policy efforts. Initial condition large ensembles (the same model run many times with the same set of assumptions) has revealed that irreducible uncertainty arising from natural variations in the climate system—called internal variability—can be larger and more persistent than expected when compared to the set of models typically used in climate impacts assessments. Because of this, some argue that the large magnitude of internal variability presents a challenge to effective adaptations in response to climate change. Here we show using examples from water management that characterizing internal variability, even if it is large and irreducible, is the means to more effective decision-making, pointing to the importance of initial condition large ensembles in this effort. We also discuss the criticisms of large ensembles: that they are costly, redundant, and biased. We show that despite these challenges, large ensembles provide unique information that is consistent with the insights from decision science about how to position effective decisions under conditions of deep uncertainty.

1. Introduction

Borges said that *the future is inevitable and precise, but it may not happen* (Borges, 1941). For the science of climate change projections, the implications of his insight are clear: Across multimodel projections of climate change (like those from Phase 6 of the Coupled Model Intercomparison Project, or CMIP6; Eyring et al., 2015), warming is inevitable. And while each model's simulation is a precise rendering of the future climate, no one model projection will happen.

The inevitability of warming creates a desire for scientists to provide accurate estimates of the future to best inform adaptation decisions. Given scant resources, where do decision makers put their efforts to greatest

Resources: Justin S. Mankin, Flavio Lehner, Sloan Coats, Karen A. McKinnon

Software: Justin S. Mankin

Supervision: Justin S. Mankin

Validation: Justin S. Mankin

Visualization: Justin S. Mankin, Flavio Lehner, Sloan Coats

Writing - original draft: Justin S. Mankin

Writing - review & editing: Justin S. Mankin, Flavio Lehner, Sloan Coats, Karen A. McKinnon

effect (Paté-Cornell, 1996)? The deep uncertainty (Lempert et al., 2003) in regional-scale projections of climate change naturally leads to questions about whether climate science can inform such decisions. The models give different answers to the same sets of assumptions, so which model do we choose to trust? And what is the basis for that trust? Because of these difficulties, decision makers are left with questions about whether the range of climate projections across models is due to scientific choices about how best to simulate processes like convection, clouds, or tree mortality, or whether it is instead due to natural internal variability of the climate system, which should be interpreted as a set of possible real-world outcomes consistent with the same global warming. Because of this, climate uncertainty, such as different answers across models, can be cast as a justification for indecision, suggesting that the science needs to mature before it can be helpful to decision-making (e.g., Koonin, 2014; Palutikof et al., 2019; Weaver et al., 2013). In a recent survey of local-scale adaptation to sea level rise in coastal Australia, for example, respondents found scientific uncertainty to be the second most important barrier to the development of effective responses, after lack of leadership (Palutikof et al., 2019).

Initial condition large ensembles of fully coupled Earth System Model (ESM) simulations, have, at first glance, further complicated this climate-science-for-decisions gap (Deser, Knutti, et al., 2012; Hawkins et al., 2016). Such ensembles are experimentally similar to those used in weather forecasts: Run one model with one set of boundary conditions many times with different initial conditions to generate a distribution of outcomes consistent with the same assumptions. The key differences from weather forecasting are threefold: (1) The spatiotemporal scale of the integration is larger in initial condition large ensembles—typically global in scale and centennial in length; (2) there are different methods for initializing the model, such as introducing round-off errors into atmospheric fields (e.g., Kay et al., 2015; Kirchmeier-Young et al., 2017; Selten et al., 2004; Sterl et al., 2008) versus sampling the ocean-atmosphere and/or land states (e.g., Maher et al., 2019; Rodgers et al., 2015), a distinction that is sometimes called microinitializations versus macroinitializations (Hawkins et al., 2016); and (3) that each model realization is fully coupled, simulating the interactions of the ocean, atmosphere, cryosphere, and land surface, often with active biogeochemistry under evolving greenhouse gas emissions. This experimental design has only recently become computationally feasible and is not one of the handful of common “Diagnostic, Evaluation and Characterization of Klima” (DECK) experiments mandated by CMIP6. Large ensembles have revealed an underappreciated source of uncertainty in regional climate projections—that of internal climate variability (Deser, Knutti, et al., 2012; Deser, Phillips, et al 2012; Deser et al., 2020; Kay et al., 2015). Therefore, not only do different models give different answers about future climate (as in the CMIP5), but so does the same model when run many times (as in a large ensemble). In the former situation, there remains hope that we can constrain uncertainty by improving models. In the latter situation, the uncertainty emerges within the *same* model, and so it cannot necessarily be tied back to how we build models. Instead, this uncertainty is irreducible as it is intrinsic to the climate system—both simulated and real.

The existence of irreducible uncertainty in model predictions has been known since Lorenz (1963). What large ensembles have illustrated, however, is how persistent and large irreducible uncertainty is in centennial-scale climate simulations from state-of-the-art ESMs. In fact, at a recent community workshop hosted by CLIVAR on large ensembles, (<https://usclivar.org/meetings/large-ensembles-workshop>) there was an explicit discussion among attendees about moving away from using the word “uncertainty” when talking about the range of outcomes from large ensembles for fear of reinforcing mistrust in model projections. This is because of the possibility that the large magnitude of “irreducible uncertainty” in climate projections can confound people’s expectations of what climate change looks like at regional scales (Deser, Knutti, et al., 2012; Deser et al., 2020), perhaps reinforcing the sentiment that climate science is not positioned to inform adaptation decisions.

Here we illustrate why this sentiment is not correct. Characterizing the magnitude and sources of uncertainty—even when it is large, irreducible, and poses communication challenges—is not a bulwark to effective decision-making, *it is the means to effective decision-making*. Risk assessment tools require the full range of probabilities of the most damaging risks; there are rightful calls for climate science to take risk assessment much more seriously (Sutton, 2019). As we discuss in the sections below, large ensembles are a crucial tool for climate change risk assessment and adaptation decision-making because they allow us to (1) better attribute uncertainty and estimate its irreducible magnitude, something that cannot be done well in multimodel ensembles like the CMIP, and (2) because the noise they generate allows

decision makers to consider the *full* range of outcomes consistent with same forcing to ensure they are robustly prepared for the future.

Several arguments about the importance of large ensembles to decision-making have already been made in the literature (e.g., Deser, Knutti, et al., 2012; Deser et al., 2020). The most common one made is that large ensembles help robustly identify the emergence of signals in climate projections, particularly noisy systems like hydroclimate (Deser, Phillips, et al., 2012). A second important argument emerging from large ensembles is about how uncertainty in climate projections should best be communicated to policymakers. Because the climate system is inherently noisy, there are appropriate arguments that scientists need to manage expectations about the limits of regional climate prediction (Deser, Knutti, et al., 2012; Deser et al., 2020). Here we extend these important points, arguing that the noise is not only a means to an end, such as signal identification or better climate communication. Instead, we point out that the *noise itself* provides information that *is as valuable to planning as the signal*. One may ask, if more than 40 ensemble members are needed to detect a significant change, is the change even relevant? Such a question comes from a signal-dominated perspective: The noise is not simply noise. They are outcomes—even if rare—entirely consistent with the same change in the mean state, and decision makers can only be prepared for those outcomes if they are communicated as possibilities. For example, even if all models agreed that the next two decades should on average see more rainfall because of climate change, it does not preclude the possibility to experience a drought. How could a water manager know to prepare for the possibility of drought if it is simply dismissed as noise (Lawrence et al., 2020)?

Nevertheless, it is important to highlight that large ensembles also have considerable drawbacks. Some common critiques are that they are (1) resource intensive, (2) redundant, and (3) biased, potentially misrepresenting the true magnitude of internal variability. Focusing on hydroclimate, we engage each of these major drawbacks in the sections below as well, illustrating how they inform responsible use of initial condition large ensembles to aid regional-scale adaptation decisions.

2. The Importance of Sourcing Uncertainty for Decision-Making

Part of the power of large ensembles lies in their ability to rigorously quantify and attribute uncertainty in climate projections in ways multimodel ensembles alone cannot. This insight has recently been concretized by comparing several large ensembles to the multimodel CMIP5 ensemble. Such a comparison shows that internal variability can be underestimated *and* overestimated at the same time, depending on the variable, region, or temporal frequency being analyzed (Deser et al., 2020; Lehner et al., 2020). This implies that one cannot distinguish true model-to-model differences from noise in a multimodel ensemble without large errors (Lehner et al., 2020). To understand why this is, it helps to recall the sources of climate uncertainty, why they are important to quantify, how they have been quantified to date, and why large ensembles better articulate the range of real-world outcomes a decision maker will be faced with.

Canonically, uncertainty in multimodel projections of climate change is partitioned across three sources (Hawkins & Sutton, 2009): (1) forcing (or scenario) uncertainty, which is a function of ambiguity about people's greenhouse gas emissions over the coming century. This is fundamentally a boundary condition problem (Kirtman et al., 2012; Meehl et al., 2009). (2) Structural (or model) uncertainty, which arises from models giving different answers in response to the same forcing and initialization. This uncertainty is a function of both computational limits and scientists' imperfect knowledge of the physics of the climate system, and how those factors are reflected in modeling choices. (3) Internal variability, which are climate variations that occur in the absence of any changes in boundary conditions. It suggests that the climate we experience is but one realization of many plausible futures, given the chaos innate to the climate system (Lorenz, 1963). This uncertainty in model projections is sometimes called irreducible (e.g., Deser, Knutti, et al., 2012; Hawkins et al., 2016) because this variability is intrinsic to model representations of climate. Because we cannot characterize the state of the climate everywhere simultaneously to perfectly initialize a model simulation (there will always be uncertainties in observations), estimates of internal variability are framed as an initial condition problem (Kirtman et al., 2012; Lorenz, 1963; Meehl et al., 2009).

Partitioning uncertainty in projections of climate change is not a pedantic exercise. The origins of such uncertainty have important consequences for the scientific and policy decisions made in response to climate change. For example, where do we focus our scientific efforts to best project climate change? And given these

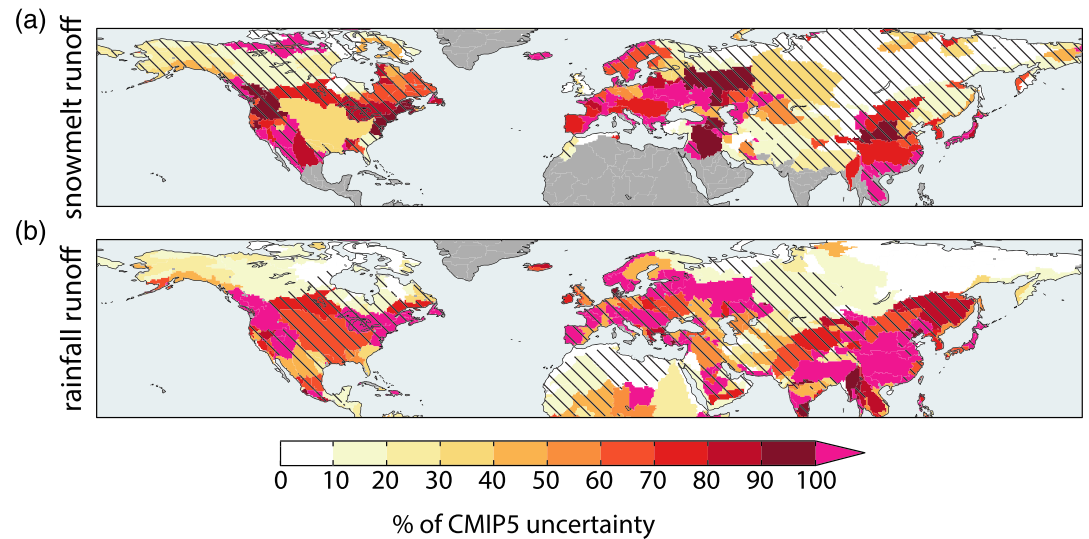


Figure 1. The percent of the total range of 75-yr linear relative trends (%/yr) estimated from 2006–2080 from CMIP5 (19 models with available data) that is spanned by the CESM1 large ensemble (CESM1-LE) in (a) runoff from snowmelt and (b) runoff from rainfall over the Northern Hemisphere at the basin scale following Mankin et al. (2015). The single-model large ensemble (which only includes uncertainty from internal variability) spans the majority or exceeds CMIP5 uncertainty (which includes uncertainty from internal variability and model uncertainty) in many midlatitude basins where there are competing demands on water resources and over time scales where one might expect model uncertainty to dominate. Hatching indicates regions that have both positive and negative trends in both ensembles. Gray color indicates no data.

ongoing efforts, should a policymaker wait for more information before making a decision? Of these three sources of uncertainty, only two exist in the real world: scenario uncertainty and internal variability. In contrast, model uncertainty is a function of both our lack of knowledge about some climate processes and the computational challenges inherent in modeling a multiscale, complex system. This uncertainty does not exist in the real world and, unlike the other two uncertainty sources, will not shape the real-world climate people experience. Therefore, model uncertainty provides a veil of plausible deniability for making decisions—it is the uncertainty that suggests that climate science needs to mature.

While the sourcing of uncertainty in projections of climate change is important to decision-making, it is not clear that multimodel ensembles like the CMIP5 provide a rigorous quantification of these uncertainties. Large ensembles have revealed that the CMIP5 conflates model uncertainty with that from internal variability (Kay et al., 2015; Lehner et al., 2020; Mankin et al., 2015, 2017). One can see this clearly by comparing the total uncertainty in a climate projection from a multimodel ensemble to that in a single-model large ensemble. We make this comparison in Figure 1, using the large ensemble from the National Center for Atmospheric Research's Community Earth System Model (NCAR CESM1 Large Ensemble, or CESM1-LE) and the CMIP5 multimodel ensemble both forced with a high-emissions pathway, called Representative Concentration Pathway 8.5, or RCP8.5 (Riahi et al., 2011). We show the relative magnitudes of uncertainty (measured as the full range of the CESM1-LE as a percentage of that from the CMIP5) focusing on two variables important to water management: basin-scale 75-yr trends in spring and summer runoff from snowmelt (Figure 1a) versus that from rainfall (Figure 1b), both estimated from 2006–2080. For many midlatitude basins, the uncertainty in CESM1-LE, which is the range of simulated outcomes from internal variability alone, spans the majority of (or exceeds) the uncertainty in the CMIP5, which includes both model and internal variability uncertainties. These basins have large populations and varied demands on water resources that require confident estimates of future climate. For regional hydroclimate, the expectation based on multimodel ensembles is that model uncertainty should dominate that from internal variability on multidecadal time scales (e.g., Giuntoli et al., 2018; Hawkins & Sutton, 2011; Zhang & Soden, 2019). But Figure 1 shows this is not always the case—for many basins, the CMIP5 ensemble is undersampling the uncertainty from internal variability—at least as estimated using the CESM1-LE—because there are too few simulations with each model to quantify their representation of internal climate variability.

Large ensembles, therefore, demonstrate that uncertainty quantification in multimodel ensembles can be misleading for some regions and variables. An implication of this for decision-making is that the fractional uncertainty from internal variability is larger and more persistent than assumed and, more importantly, that its true extent (the form of uncertainty that exists in the real world) may not be accurately reflected in projected climate impacts from the CMIP5 multimodel ensemble. From a decision-making perspective, this is crucial, as decision makers are left with the question of whether the range of climate projections across models is due to model uncertainty or instead is entirely consistent with the same warming world. Consider a water manager in the western United States deliberating long-range risk management options: Figure 1 shows (via hatching) that snowmelt and rainfall runoff have the possibility for 75-yr increases *and* decreases in *both ensembles* despite the fact that the ensembles represent different sources of uncertainty. Looking at such results from CMIP5 alone, that water manager might argue that the models do not agree on anything, leading to a “wait and see” approach in decision-making. That same manager, looking at the results from the CESM1-LE, may recognize that there are many basin-scale climate futures that are consistent with the same warming, and that we need to be prepared for all of them.

3. The Value of Large Ensembles to Robust Decision-Making

Adaptations to climate change will require regional responses encompassing institutional, social, and economic changes, as well as significant time horizons for planning. Since adaptations are best made proactively, their implementation should necessarily occur when internal climate variability still has the potential to amplify, mask, or reverse expected climate changes, even in the face of a clearly detected and attributed anthropogenic signal. Evaluating tradeoffs associated with adaptation options is difficult in a world where internal variability can create the appearance of maladaptation or be truly maladaptive over the short term, diverting resources to activities that may increase people's vulnerability to climate risks. Furthermore, adaptations require benchmarks for evaluation. A farmer may ask, for example, “Did changing my sowing date based on my expectations of climate change actually improve my yields?” Such evaluations are based on actual yields, meaning they are, in part, a function of the actual climate state and thus subject to irreducible uncertainty from internal variability. But humans, including water managers and farmers, have a long history of optimizing decisions under deep uncertainty, and scholars in decision sciences, engineering, and epidemiology have worked hard to develop tools to ensure that the decisions people make in response to things like climate change are robust despite such uncertainty.

Robust adaptation decision-making is the process of choosing a strategy or set of strategies that have the greatest benefit across the broadest range of potential real-world outcomes (Dessai & Hulme, 2004; Lempert & Groves, 2010; Lempert & Schlesinger, 2001; Lorenz et al., 2015), such as the outcomes consistent with internal variability. Consider the set of water management adaptations that would be evaluated under two different distributions of decadal precipitation changes over the western United States, inspired by Dessai and Hulme (2004) (Figure 2). The distribution in blue is an estimate of irreducible uncertainty following Hawkins & Sutton (2009) from 40 models in the CMIP5 multimodel ensemble, totaling 40 simulations. The distribution in red is the same but from seven different large ensembles, totaling 286 simulations. Both distributions are shifted to match the (potentially biased) CMIP5 ensemble mean, or “best estimate” of climate change. As such, it isolates only modeled internal variability. Across the bottom are possible adaptations in response to the different magnitudes of decadal-scale precipitation change, from sourcing new supply if the climate shifts to that exemplified by the decade centered on the 2012–2015 California drought to increasing dam heights for high precipitation decades like the decade centered on 1982–1983, when Lake Powell overtopped (Figure 2). The large ensembles, which have 33 fewer models than the CMIP5, encompass a larger range of variability in projected western U.S. rainfall. This larger variability occurs because the additional realizations in the large ensembles provide a more complete sampling of internal variability than one can get from the smaller number of simulations from each model in the CMIP5. This presents a challenge for robust decision-making.

Robust decision-making is a computational approach to stress-testing decisions for vulnerabilities against a set of possible scenarios (Lempert et al., 2013). To carry out a formal decision analysis, one needs a full characterization of the possible outcomes, including the range related to the sampling of internal variability. In the case of the western U.S. water manager presented in Figure 2, decisions around long-term,

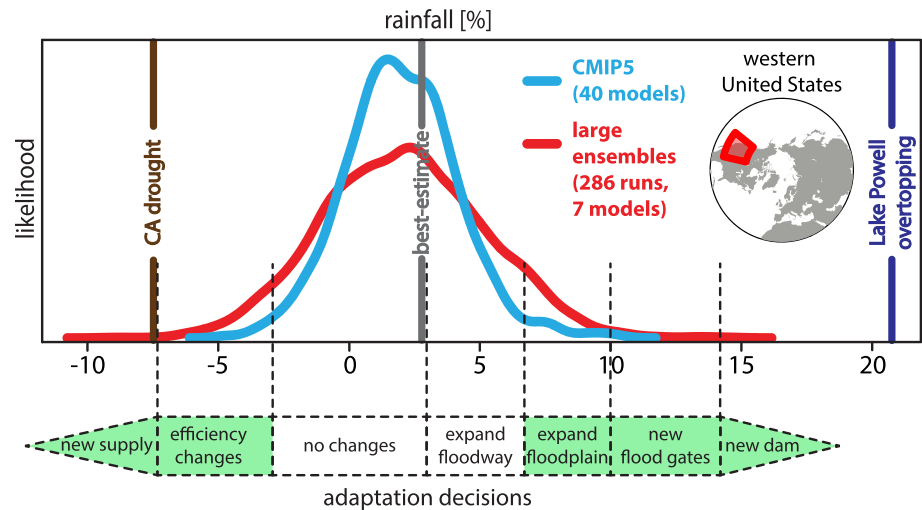


Figure 2. The influence of mischaracterized internal variability on adaptation decision-making for water management in the western United States (29.3–51.6°N; 230.6–255.9°E). The distribution in blue shows the change in decadal rainfall between 2061 and 2090 as a percent of the historical (1971–2000) mean in 40 models (40 simulations) from the CMIP5, where the range is that due to internal variability (model uncertainty is removed following Hawkins and Sutton (2009) by estimating the residuals off of a fourth-order polynomial fit to each model realization). The distribution in red is the same but from seven initial condition large ensembles archived by NCAR totaling 286 simulations. Both distributions are shifted to match a specified forced response, here taken as the CMIP5 ensemble mean for RCP8.5 (gray). To put these distributions in perspective, we use Global Precipitation Climatology Centre (GPCC) rainfall observations to show the western U.S. average decadal precipitation anomalies centered on the 1982–1983 Lake Powell overtopping (21.3%) and centered on the height of the 2012–2015 California drought (–7.5%). Under the figure are potential adaptation decisions that could be considered in response to particular magnitudes of change (dotted lines). The adaptations highlighted in green, for example, would not be considered without full estimates of internal variability; as such, water managers would not be able to “robustly adapt” to rainfall changes. Inspired by Dessai and Hulme (2004).

capital-intensive hydrologic infrastructure and planning can be evaluated for vulnerabilities against the set of real-world outcomes consistent with internal variability. From a decision standpoint, the potential that internal variability’s full extent is not being quantified or attributed in our multimodel projections is critical—because the set of strategies a decision maker adopts may be different if it is based on a truncated estimate of internal variability, as under the CMIP5 distribution, than if it is based on a fuller estimate of internal variability, as provided by the large ensembles. For example, if a water manager is only presented with the CMIP5 distribution, then the possibility of needing to seek new water supplies would not be part of the formal decision-making process. The consequences of that could include a misallocation of resources and a lack of preparedness for the water availability risks facing the region.

The impulse to give decision makers the *most likely outcome* as the basis for sound policy still pervades discussions on how climate science can best inform climate policy (Hausfather & Peters, 2020; Lawrence et al., 2020). But such a focus on the most likely set of outcomes (rather than all possible outcomes) can wrongly give decision makers a “false sense of certainty”, increasing vulnerability and exposure to climate damages and locking-in costly decisions that will need to be made later on (Lawrence et al., 2020). Decision makers can only make flexible and robust decisions if they evaluate the set of responses that are beneficial across the broadest range of outcomes; this inherently requires we scientists to deemphasize our focus on the most likely outcome (the signal) and consider those outcomes that are typically dismissed as noise. The variability from large ensembles in conjunction with the signal from multimodel ensembles—as we show above—could be a key tool for that work.

4. Addressing the Problems With Large Ensembles

While we have illustrated the value that large ensembles have to robust decision-making, they are not without their problems. Rightful and common critiques are (1) that they are expensive to run and archive, (2) that one can estimate the same range of irreducible uncertainty via other means, and (3) that they are biased,

improperly estimating real-world internal variability. In this section, we examine each critique and consider how that should inform our evaluation of their continued scientific and policy value.

4.1. They Are Expensive

The CESM1-LE cost NCAR approximately 17 million core hours on the Yellowstone supercomputer, producing over 200 TB of archived model output (Kay et al., 2015). This is a massive resource investment of computing, data storage, and labor. NCAR could have used these resources elsewhere, and thus it is appropriate to ask whether it makes sense for a modeling center to invest in generating a large ensemble, particularly given the constraints and production schedules associated with CMIP. Certainly, modeling centers have to reconcile the generation of a large ensemble against competing scientific priorities such as higher resolutions or additional physical processes, both of which also improve our understanding of the real world. High-resolution simulations are also valuable to adaptation decision-making because they can better capture fine-scale processes and extremes (e.g., Diffenbaugh et al., 2005; Hall, 2014) at the scales most useful for decision-making. So why invest resources in a large ensemble? On this question we note three points:

First, if one is in the model improvement business, as all modeling centers are, then knowing the model's irreducible uncertainty can be valuable as a benchmark. For instance, analysis of extreme climate outcomes in a large ensemble can help to elucidate the physical mechanisms that drive these outcomes, providing grounds to improve simulations, prediction, and model intercomparison. Consider the experience of model developers at NCAR, who discovered that their model, CESM2, had multiple equilibria in sea ice concentrations over the Labrador Sea that only emerged sporadically in their preindustrial control simulations, wreaking havoc on how best to initialize the model for other production runs associated with CMIP6 (Danabasoglu et al., 2020). While the underlying physics of overly extensive sea ice has not been identified, a solution to the challenge it presented to other production runs only emerged through the use of initial condition large ensembles of preindustrial simulations (Danabasoglu et al., 2020). Second, while higher resolutions are crucial to representing fine-scale processes that regulate the response of climate extremes to forcings (Diffenbaugh et al., 2005), there is mixed evidence as to whether higher resolutions necessarily increase a model's actual predictive skill (Scaife et al., 2019), which is what would be most valuable from a decision maker's perspective. Finally, there is the important question of the resolution-realization computational trade-off. Both higher resolutions and additional realizations are crucial, and each modeling center must determine its scientific priorities: The pursuit of higher-resolution simulations is motivated by different scientific aims than additional realizations, but given finite computing, they come to the sacrifice of one another. For example, if we take the simplistic question of computational resources for a standard CMIP-style component set (compset) of the CESM1, a factor N increase in resolution is far more expensive than N additional realizations (of order N versus N^4 , though factor increases could be as much as N^5 due to input-output of data). For example, assuming a base cost of a centennial-scale simulation of $\sim 300,000$ core hours, then 10 model realizations are approximately 1,000 times cheaper than a factor 10 increase in resolution. This kind of comparison is inherently reductionist, and as such it does not compare the actual payoffs of investing in resolution versus realizations. Resolution increases could pay real scientific and decision-making dividends that vastly outweigh their costs (e.g., Schneider et al., 2017). But additional realizations are, as we note above, also prized. How might we overcome this resolution-realization trade-off?

Scientists are undertaking this challenge, synergistically combining large ensembles with high-resolution regional climate simulations to efficiently provide information for decision-making. For instance, what if you want to inform a decision maker of the dynamics and impacts of a 1-in-1,000-yr event? A single centennial-scale realization at high resolution has a less than 10% chance of including that event as part of its simulation. A 40-member large ensemble of centennial-scale realizations, like the CESM1-LE, has over a 98% chance of including that event. A modeler can then use that large ensemble realization to drive a Regional Climate Model (RCM) to produce the fine-scale features and impacts associated with such an anomalous event (see, e.g., <https://usclivar.org/sites/default/files/meetings/2019/presentations/SWAIN-DANIEL-LE19.pdf>). This also points to the importance of archiving (at a minimum) the variables needed to force an RCM as part of any large ensemble project.

More widely, new methods are being developed to answer the crucial question of the “necessary” ensemble size (Beusch et al., 2020; Coats & Mankin, 2016; Link et al., 2019; Milinski et al., 2019), and there are active discussions in the large ensemble community (US CLIVAR, 2020) about how to best balance the length of

simulation, number of ensemble members, and the degree of model complexity. Ultimately, it is for each modeling center to decide on the worth of a particular resource investment, as the number of required simulations, as others have noted (Milinski et al., 2019), is a function of the question you are seeking to answer.

4.2. They Are Redundant

A second critique is that the range of irreducible uncertainty in future climate change can be estimated via other, less resource-intensive means. For example, recent research has shown that for certain variables one can recover nearly the same variability estimated from a large ensemble with a simple red noise model trained on a preindustrial control simulation or even observations (Thompson et al., 2015). Such an insight is powerful, but it does not dismiss the value of the large ensemble to science and decision-making for at least four reasons: First, statistical models do not allow for the same biogeophysical diagnosis and evaluation that process-based ones do. Indeed, a fundamental goal of process-based modeling is to produce a physically consistent system against which one can test hypotheses. While empirical models are useful for rigorous hypothesis testing, they are less useful for tracing the physical processes underpinning climate variability. Second, large ensembles are an ideal experimental testbed for statistical methods and approaches. Third, a large ensemble provides a robust estimate of the climate response to changes in boundary conditions like anthropogenic emissions (Deser et al. 2016; Selten et al., 2004). Finally, statistical modeling approaches tend to assume that internal variability is not itself altered by anthropogenic forcing (Coats & Mankin, 2016; Pendergrass et al., 2017). We examine these last three reasons more closely, below.

Large ensembles provide a perfect model framework for testing statistical and other ensemble analysis approaches that can then be applied to multimodel ensembles and/or the real-world climate system—even for those approaches that suggest they supersede a large ensemble's value in estimating internal variability (e.g., Thompson et al., 2015). Using models as a testbed for cheaper analytical approaches has a long history in the sciences when evaluating, for example, the skill of a coarse numerical implementation in capturing the “perfect model” provided by an analytical solution (Machete & Smith, 2016). It has also long been applied by the weather and climate communities for everything from evaluating filtering schemes for the Navier-Stokes equations (e.g., De Stefano & Vasilyev, 2002), to data assimilation (e.g., Anderson et al., 2009), to numerical weather prediction and operational forecasting (see, e.g., the Developmental Testbed Center's Ensemble Testbed (DET): <https://dtcenter.org/det/index.php>) (Anderson, 1996; Gallo et al., 2017; Schwartz et al., 2019). Such a philosophy has naturally extended to large ensembles of climate simulations as they have become more widely available. One example of this is the testing of statistical methods to decompose signal and noise in the climate system, such as dynamical adjustment procedures to isolate thermodynamic from dynamic responses (e.g., Lehner et al., 2017; Saffioti et al., 2015; Sippel et al., 2019; Wallace et al., 2012). In the absence of large ensembles, we would not be able to vet the robustness of such approaches for assessing the climate response to greenhouse gas forcing.

Estimates of the climate system response to greenhouse gas emissions is critical from an adaptation decision standpoint, and large ensembles allow for a robust estimate of each model's forced response. This is crucial because even if one can estimate the distribution of internal variability through statistical methods, one still needs an estimate of the forced response to shift the center of mass of that distribution to its new future position. Typically, to assess the forced response, one averages across ensemble members and uses the resulting ensemble mean or fits a trend to that ensemble mean. Importantly, the ensemble mean is itself a function of the ensemble size, n : A larger ensemble will produce a more stable estimate of the forced response. This begs a question that climate scientists have long wondered—how many ensemble members does one need to robustly estimate the forced response (e.g., Daron & Stainforth, 2013; Deser, Phillips, et al., 2012; Drótos et al., 2015; Hawkins & Sutton, 2012; Hawkins et al., 2016; Maher et al., 2019; Milinski et al., 2019)?

We can combine the perfect model framework philosophy discussed above with a large ensemble to address this question. Here the framework involves using standard techniques on randomly chosen subsets of a large ensemble (like calculating a mean centennial-scale change) to assess how well such subsets capture a model's “known” characteristics, where the “known” characteristics are those from the full ensemble. So, for example, we can assess the informational value of an additional ensemble member to improved estimation of a known quantity, like the full 40-member ensemble mean, which is often taken as the forced signal of climate change.

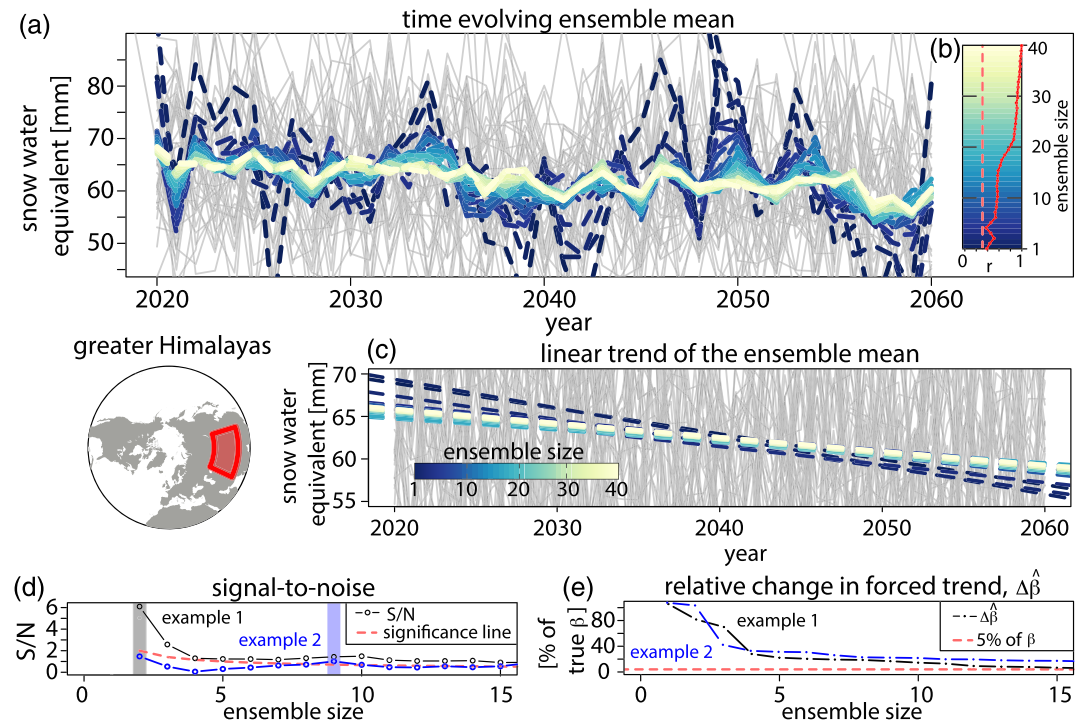


Figure 3. Uncertainty in the forced response as a function of the ensemble size for Himalayan snowpack. In (a), we show an example of the time-evolving ensemble mean of Himalayan (map inset) snow water equivalent (mm) between 2020 and 2060 as estimated from different numbers of ensemble members (colors, see color bar in (b)) from the 40-member CESM1-LE. Note that the order in which ensemble members are added changes the evolution of the ensemble mean because the average is taken over a different subset of ensemble members. Based on the example in (a), we show in (b) that it can take over 30 ensemble members to have a Pearson correlation coefficient (r) with the “true” 40-member ensemble mean that exceeds 0.9. In (c), we show that a large number of ensemble members is also needed to estimate the “true” (full ensemble) forced response in the case where it is estimated as a linear trend of the ensemble mean (c). A typical metric of the number of ensemble members needed is the signal-to-noise ratio (S/N) (d). We show two examples of how the S/N in snow water equivalent can vary as a function of ensemble size, from only needing 2 ensemble members (Example 1, black line, gray bar) to achieve statistical significance ($(S/N) \times \sqrt{n} > 2$), to 9 ensemble members (Example 2, blue line, blue bar). Yet in (e) it is clear that it can take more than 15 ensemble members in both cases to estimate the forced response ($\Delta\hat{\beta}$) to within 5% of the forced response from all 40 ensemble members (β).

We show such an analysis for a 40-member ensemble projection of snowpack (snow water equivalent, mm) just before springtime melt in the headwaters of the Ganges and Brahmaputra Rivers from the CESM-LE (Figure 3). If one estimates the forced response as the time-evolving mean (Figure 3a), we can see that there is a high degree of variation in that estimate as a function of the total number of ensemble members and the unique ensemble members used in that calculation. For example, in Figure 3a, one can see that it takes 10 or more ensemble members before there is reasonable convergence on the ensemble’s “true” forced response, which here is the mean across all 40 members. Taking the Pearson correlation between each estimate of the forced response and the “true” forced response, some 30 ensemble members from the example in Figure 3a are needed to bring the correlation above 0.9. (Figure 3b). If instead one estimates the forced response as the linear trend on the ensemble mean, we can see that measure too is dependent on the unique ensemble members and ensemble size used in its calculation (Figure 3c), though less so than for the time-evolving ensemble mean.

A common benchmark for assessing the ensemble size required to capture the forced response is the signal-to-noise ratio (S/N). Here we estimate the S/N in Himalayan snowpack as the ratio of the magnitude of the ensemble mean linear trend to the standard deviation across the ensemble (Figure 3d). We perform the exercise of estimating the S/N as a function of ensemble size two times (Examples 1 and 2), illustrating that the S/N , particularly for small ensemble sizes (>10), can vary drastically. Even if the S/N is valuable for detection and attribution studies despite these issues, it is less valuable from a decision-making standpoint. A decision

maker in this snow-dependent region is likely to be less concerned with precisely when a significant signal in snow declines emerges, than the actual range of outcomes for which to be prepared.

We illustrate this point most clearly in Figure 3e, where we show how the estimate of the forced response $\hat{\beta}$, converges on the “true” forced response, β . In Example 1, it can take up to 15 or more ensemble members to get $\hat{\beta}$ to fall within 5% of the linear trend calculated on the full 40-member ensemble mean (Figure 3e). From a robust decision-making standpoint, where the distribution of future outcomes is centered matters as it helps determine the magnitude of changes in the extremes. In nonlinear systems like water management, this can matter greatly—modest linear increases in rainfall extremes, for example, could be sufficient to overtop dams or levees, causing nonlinear impacts, like floods.

While there are clearly diminishing marginal returns with each additional ensemble member (Figure 3e), they still contain important information about the forced response. This is also true of more sophisticated statistical approaches that could theoretically yield a faster convergence on the true forced response (such as nonlinear fits, time series decomposition, or dynamical adjustment). But it is rarely clear a priori how many ensemble members are needed to achieve convergence on the true forced response—one inevitably needs large ensembles to develop and test such approaches. Again, such exercises are not pedantic, because the information about the true forced response is essential from an impacts perspective: The full distribution of outcomes that must be adapted to is shaped not simply by its irreducible uncertainty *but also by where you place its center of mass* (i.e., the forced response). Such a lesson holds for the real world as well: We do not know the true forced response to climate change, which necessarily increases our reliance on robust model estimates. In particular, one's estimate of the forced response has a crucial influence over what constitutes changes in the most damaging tail risks, as we illustrate below.

Finally, estimating irreducible uncertainty from a control simulation of a model or from observations assumes that anthropogenic forcing does not project onto internal variability to change its characteristics. As models become more complex (Marvel et al., 2015), stationarity assumptions, particularly for hydroclimate, begin to fall apart (Pendergrass et al., 2019), implying that there is little basis for assuming the irreducible uncertainty of the future will look like the irreducible uncertainty of the past (Lehner et al., 2020). Even the CMIP5 archive (which as we discuss undersamples internal variability) projects that the dominant mode of ocean-atmosphere variability, El Niño–Southern Oscillation (ENSO), may change in the future, bringing both more El Niños and more La Niñas (Cai et al., 2018). This assumption of stationary variability also breaks down in CESM1-LE, as we show at the grid point scale for multidecadal variability in persistent (35-yr) aridity using a common aridity metric called the Palmer Drought Severity Index, or PDSI (Figure 4) (Coats & Mankin, 2016; Cook et al., 2018). Over at least 12% of land area, it is nearly guaranteed that this assumption is untrue in the CESM1-LE (red colors). This can immensely impact our characterization of tail risks, such as for multidecadal droughts, which are of utmost importance for robust decision-making (Figure 2) (Sutton, 2019).

4.3. They Are Biased

A final critique points to the fact that the irreducible uncertainty from a large ensemble is itself a function of the model and, as such, is shaped by model uncertainty—that is, any estimate of internal variability from a model is biased. This is no doubt true as these models are imperfect reductions of the real world. For some regions and quantities, climate models tend to overestimate variability, thereby leading to both an underestimate of the S/N (Scaife & Smith, 2018) and the predictability of the real world (Eade et al., 2014). At the same time, climate models can underestimate variability on multidecadal and longer time scales (Laepple & Huybers, 2014). Large ensembles do not solve the problem of model biases. And yet, despite this issue, large ensembles are important in part because of what they are telling us about sources of real-world uncertainty. We do not have a perfect sense of the dividing line between model uncertainty and that from internal variability: Some of a large ensemble's internal variability is polluted by model uncertainty, making it overdispersive or underdispersive. But so too in the CMIP5, some of what is being called model uncertainty is actually attributable to internal variability. Without large ensembles, we would not know this. Large ensembles can help diagnose these biases and help scientists draw the decision-critical dividing line between model and irreducible uncertainty. It also enables searching for robust emergent constraints that allow scientists to reduce the spread in future projections (Hall et al., 2019; Lehner et al., 2019; Tokarska et al., 2020). This is

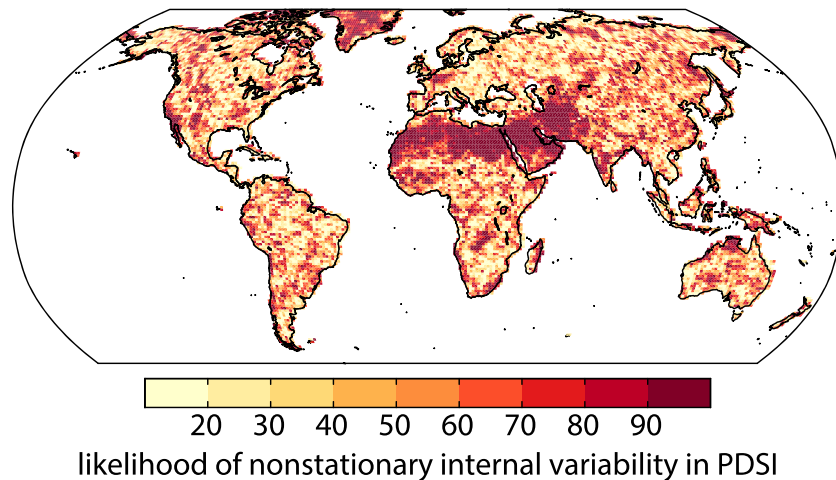


Figure 4. Characterizing drought risks in a nonstationary climate. We show the likelihood that internal variability in the 35-yr mean of the Palmer Drought Severity Index (PDSI) is nonstationary with anthropogenic forcing. We perform a bootstrapped Kolmogorov-Smirnov (K-S) test comparing all 35-yr hydroclimate means from the preindustrial control to all 35-yr means of the CESM1-LE (2006–2040), showing the percent of K-S tests where the null that the distributions are the same can be rejected. Following methods from Coats and Mankin (2016).

critical when it comes to making costly and long-term adaptation decisions, like hydrologic infrastructure, which can be too costly to build for all trajectories emerging from climate models (Vano et al., 2018).

5. Moving Forward

How do we, therefore, as Bertrand Russell noted, “live without certainty, and yet without being paralyzed by hesitation?” (Russell B: *History of Western Philosophy*. 2nd ed. London: Allen & Unwin, 1961, p. 14). While the observation of model bias projecting onto its representation of internal variability poses a new challenge, it is a challenge for which large ensembles are likely needed. Tools under active development like the Observational Large Ensemble (McKinnon et al., 2017; McKinnon & Deser, 2018), enable scientists to assess a large ensemble’s internal variability vis-à-vis a comparable estimate of observed variability. While an observational large ensemble does not position an evaluation of a large ensemble’s future variability, it could, in combination with estimates of nonstationary variability from a large ensemble itself (e.g., Poppick et al., 2016). Increasingly, climate scientists are combining analyses of large and multimodel ensembles in an effort to better attribute uncertainty in impacts analyses (see, e.g., <https://usclivar.org/working-groups/large-ensemble-working-group>). As such, large ensembles alone are not a perfect remedy unto themselves but instead are part of a wider constellation of tools to best position decision-making. In places where they show skill, initialized ensembles for decadal prediction will likely be part of such a toolbox. They enable reductions in the otherwise irreducible uncertainty from internal variability for lead times of several years (e.g., Hawkins & Sutton, 2009; Simpson et al., 2019; Yeager et al., 2018) and can provide value for decision makers. Existing uninitialized ensembles can in turn be used to search for analogs that complement initialized prediction efforts (Ding et al., 2018), ultimately leading to a more effective use of limited computing resources. A philosophy of “model for purpose” has prevailed in climate model experimental designs; perhaps this maxim should be extended to include “ensemble for purpose”.

What is clear is that given society’s commitment to warming, it is crucial to inform decision makers regarding the irreducible uncertainty in the climate system (Pielke, 2003), with an emphasis on how this can shape effective and sustainable climate risk management (Kunreuther et al., 2013; Stern et al., 2013). People consistently make decisions under conditions of deep uncertainty, and climate adaptation should be no different. In fact, decision makers increasingly have tools to optimize decisions under uncertainty, such as robust decision-making (Lempert & Collins, 2007; Lempert & Groves, 2010; Lempert et al., 2013). There is a danger in claiming certainty about inherently uncertain things, as it removes human agency to mitigate risks.

Uncertainty, therefore, is not a dirty word—it simply means that many outcomes are consistent with expectations (Dessai et al., 2009). Climate science works to update our expectations about the varied forms regional climates can take, all of which are consistent with the same warming world. Until recently, however, it was hard to appreciate the potential magnitude of this uncertainty. This is in part because the different answers provided by multimodel ensembles could be dismissed as due to model choices made by scientists, rather than the irreducible uncertainty that exists in the real world. But the scientific community, largely through computational advances, is now able to better characterize real-world uncertainty in regional climate change through experiments like initial condition large ensembles. From a decision standpoint, we know it is a problem if science is not getting the full range of outcomes correct. Initial condition large ensembles are thus a crucial tool to inform necessary decisions as the world warms.

Data Availability Statement

All data used in this manuscript are freely available. The Multi-model Large Ensemble (MMLE) project, provided by U.S. CLIVAR, is hosted at the Earth System Grid supported by NCAR (https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm4.CLIVAR_LE.html). The CESM-Large Ensemble is accessible at ESG at NCAR (at https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm4.CESM_CAM5_BGC_LE.html). The CMIP5 data are available at the Earth System Grid Federation (<https://esgf-node.llnl.gov/search/cmip5/>).

Acknowledgments

We thank two anonymous reviewers, Noah Diffenbaugh, and Fran Moore for helpful discussions, as well as the participants of the U.S. CLIVAR Large Ensemble Workshop held at NCAR in July 2019. We also thank the Earth System Grid Federation (ESGF) and their archiving of the Coupled Model Intercomparison Project (Phase 5) data, the US CLIVAR Working Group on Large Ensembles and the National Center for Atmospheric Research's Climate Data Gateway (NCAR ESG) for archiving the large ensemble data and the Cheyenne cluster, Arnold Song and Dartmouth's Discovery Computing Cluster, and Naomi Henderson and Haibo Liu for the data serving and computing support in the Division of Ocean and Climate Physics at Lamont-Doherty Earth Observatory of Columbia University. J. S. M. also thanks The Nelson A. Rockefeller Center at Dartmouth College. F. L. has been supported by the Swiss NSF (grant no. PZ00P2_174128), the NSF Division of Atmospheric and Geospace Sciences (grant no. AGS-0856145, Amendment 87), and the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) via NSF IA 1844590. This is SOEST publication no. 11115.

References

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., & Avellano, A. (2009). The data assimilation research testbed: A community facility. *Bulletin of the American Meteorological Society*, 90, 1283–1296. <https://doi.org/10.1175/2009BAMS2618.1>

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9, 1518–1530.

Beusch, L., Gudmundsson, L., & Seneviratne, S. I. (2020). Emulating Earth system model temperatures with MESMER: From global mean temperature trajectories to grid-point-level realizations on land. *Earth System Dynamics*, 11, 139–159. <https://doi.org/10.5194/esd-11-139-2020>

Borges, J. L. (1941). La creación y P.H. Gosse. *Sur*, Buenos Aires, no. 81, 70 pp.

Cai, W., Wang, G., Dewitte, B., Wu, L., Santoso, A., Takahashi, K., et al. (2018). Increased variability of eastern Pacific El Niño under greenhouse warming. *Nature*, 564(7735), 201–206. <https://doi.org/10.1038/s41586-018-0776-9>

Coats, S., & Mankin, J. S. (2016). The challenge of accurately quantifying future megadrought risk in the American Southwest. *Geophysical Research Letters*, 43, 1–9. <https://doi.org/10.1002/2016GL070445>

Cook, B. I., Mankin, J. S., & Anchukaitis, K. J. (2018). Climate change and drought: From past to future. *Current Climate Change Report*, 4, 164–179. <https://doi.org/10.1007/s40641-018-0093-2>

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001916. <https://doi.org/10.1029/2019MS001916>

Daron, J. D., & Stainforth, D. A. (2013). On predicting climate under climate change. *Environmental Research Letters*, 8. <https://doi.org/10.1088/1748-9326/8/3/034021>

De Stefano, G., & Vasilyev, O. V. (2002). Sharp cutoff versus smooth filtering in large eddy simulation. *Physics of Fluids*, 14, 362–369. <https://doi.org/10.1063/1.1421368>

Deser, C., Phillips, A., Bourdette, V., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, 38, 527–546. <https://doi.org/10.1007/s00382-010-0977-x>

Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, 2(11), 775–779. <https://doi.org/10.1038/NCLIMATE1562>

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., et al. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, 10(4), 277–286. <https://doi.org/10.1038/s41558-020-0731-2>

Deser, C., Terray, L., & Phillips, A. S. (2016). Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *Journal of Climate*, 29, 2237–2258. <https://doi.org/10.1175/JCLI-D-15-0304.1>

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

Dessai, S., Lempert, R. J., & Pielke, R. A. Jr. (2009). Climate prediction: A limit to adaptation? In N. Adger, I. Lorenzoni, & K. O'Brien (Eds.), *Adapting to climate change: Thresholds, values, governance* (pp. 64–78). Cambridge: Cambridge University Press.

Diffenbaugh, N. S., Pal, J. S., Trapp, R. J., & Giorgi, F. (2005). Fine-scale processes regulate the response of extreme events to global climate change. *Proceedings of the National Academy of Sciences*, 102, 15,774–15,778. <https://doi.org/10.1073/pnas.0506042102>

Ding, H., Newman, M., Alexander, M. A., & Wittenberg, A. T. (2018). Skillful climate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *Journal of Climate*, 31, 5437–5459. <https://doi.org/10.1175/JCLI-D-17-0661.1>

Drótos, G., Bódai, T., & Tél, T. (2015). Probabilistic concepts in a changing climate: A snapshot attractor picture. *Journal of Climate*, 28, 3275–3288. <https://doi.org/10.1175/JCLI-D-14-00459.1>

Eade, R., Smith, D. M., Scaife, A. A., Wallace, E., Dunstone, N. J., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, 41, 5620–5628. <https://doi.org/10.1002/2014GL061146>

Eyring, V., Bony, S., Meehl, G. A., Senior, C., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2015). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation. *Geoscientific Model Development Discussion*, 8, 10,539–10,583. <https://doi.org/10.5194/gmdd-8-10539-2015>

- Gallo, B. T., Clark, A. J., Jirak, I., Kain, J. S., Weiss, S. J., Coniglio, M., et al. (2017). Breaking new ground in severe weather prediction: The 2015 NOAA/hazardous weather testbed spring forecasting experiment. *Weather and Forecasting*, *32*(4), 1541–1568. <https://doi.org/10.1175/WAF-D-16-0178.1>
- Giuntoli, I., Villarini, G., Prudhomme, C., & Hannah, D. M. (2018). Uncertainties in projected runoff over the conterminous United States. *Climatic Change*, *150*, 149–162. <https://doi.org/10.1007/s10584-018-2280-5>
- Hall, A. (2014). Projecting regional change. *Science*, *346*(6216), 1461–1462. <https://doi.org/10.1126/science.aaa0629>
- Hall, P., Cox, C. H., & Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, *9*, 269–278. <https://doi.org/10.1038/s41558-019-0436-6>
- Hausfather, Z., & Peters, G. P. (2020). Emissions—The 'business as usual' story is misleading. *Nature*, *577*, 618–620.
- Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*, 1095–1107. <https://doi.org/10.1175/2009BAMS2607.1>
- Hawkins, E., & Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, *37*, 407–418. <https://doi.org/10.1007/s00382-010-0810-6>
- Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*, L01702. <https://doi.org/10.1029/2011GL050087>
- Hawkins, R. S., Smith, J. M. G., & Stainforth, D. A. (2016). Irreducible uncertainty in near-term climate projections. *Climate Dynamics*, *46*, 3807–3819. <https://doi.org/10.1007/s00382-015-2806-8>
- Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349. <https://doi.org/10.1175/BAMS-D-13-00255.1>
- Kirchmeier-Young, M. C., Zwiers, F. W., & Gillett, N. P. (2017). Attribution of extreme events in Arctic sea ice extent. *Journal of Climate*, *30*, 553–571. <https://doi.org/10.1175/JCLI-D-16-0412.1>
- Kirtman, B., Adedoyin, A., & Bindoff, N. (2012). Near-term climate change: Projections and predictability. *Notes*. <https://doi.org/10.1017/CBO9781107415324.023>
- Koonin, S. (2014). Climate science is not settled. *The Wall Street Journal*.
- Kunreuther, H., Heal, G., Allen, M. R., Edenhofer, O., Field, C. B., & Yohe, G. (2013). Risk management and climate change. *Nature Climate Change*, *3*, 447–450. <https://doi.org/10.1038/nclimate1740>
- Laepple, T., & Huybers, P. (2014). Ocean surface temperature variability: Large model-data differences at decadal and longer periods. *Proceedings of the National Academy of Sciences*, *111*, 16,682–16,687. <https://doi.org/10.1073/pnas.1412077111>
- Lawrence, J., Haasnoot, M., & Lempert, R. (2020). Climate change: Making decisions in the face of deep uncertainty. *Nature*, *580*, 456.
- Lehner, A. W., Wood, J. A., Vano, D. M., Lawrence, M. P. C., & Mankin, J. S. (2019). The potential to reduce uncertainty in regional runoff projections from climate models. *Nature Climate Change*, *9*, 926–933. <https://doi.org/10.1038/s41558-019-0639-x>
- Lehner, C., Deser, N., Maher, J., Marotzke, E., Fischer, L., Brunner, R. K., & Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics Discussions*, 1–28. <https://doi.org/10.5194/esd-2019-93>
- Lehner, F., Deser, C., & Terray, L. (2017). Towards a new estimate of “time of emergence” of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *Journal of Climate*, *JCLI-D-16-0792.1*. <https://doi.org/10.1175/JCLI-D-16-0792.1>
- Lempert, D., Groves, G., & Fischbach, J. R. (2013). Is it ethical to use a single probability density function? WR-992-NSF. https://www.rand.org/content/dam/rand/pubs/working_papers/WR900/WR992/RAND_WR992.pdf
- Lempert, R., & Schlesinger, M. E. (2001). Climate-change strategy needs to be robust. *Nature*, *412*, 375–375. <https://doi.org/10.1038/35086615>
- Lempert, R. J., & Collins, M. T. (2007). Managing the risk of uncertain threshold responses: Comparison of robust, optimum, and precautionary approaches. *Risk Analyses Off. Publ. Soc. Risk Anal*, *27*, 1009–1026. <https://doi.org/10.1111/j.1539-6924.2007.00940.x>
- Lempert, R. J., & Groves, D. G. (2010). Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the American west. *Technological Forecasting and Social Change*, *77*, 960–974. <https://doi.org/10.1016/j.techfore.2010.04.007>
- Lempert, J. L., Popper, S. W., & Bankes, S. C. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis. *Rand*, 187.
- Link, R., Snyder, A., Lynch, C., Hartin, C., Kravitz, B., & Bond-Lamberty, B. (2019). Fldgen v1.0: An emulator with internal variability and space-time correlation for Earth system models. *Geoscientific Model Development*, *12*, 1477–1489. <https://doi.org/10.5194/gmd-12-1477-2019>
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141. [https://doi.org/10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)
- Lorenz, S., Dessai, S., Paavola, J., & Forster, P. M. (2015). The communication of physical science uncertainty in European National Adaptation Strategies. *Climatic Change*, *132*, 143–155. <https://doi.org/10.1007/s10584-013-0809-1>
- Machete, R. L., & Smith, L. A. (2016). Demonstrating the value of larger ensembles in forecasting physical systems. *Tellus Series. Dynamics Meteorological Oceanography*, *68*. <https://doi.org/10.3402/tellusa.v68.28393>
- Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornbluh, L., et al. (2019). The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, *11*, 2050–2069. <https://doi.org/10.1029/2019MS001639>
- Mankin, J. S., Mekonnen, M. M., Hoekstra, A. Y., Horton, R. M., Smerdon, J. E., & Diffenbaugh, N. S. (2017). Influence of internal variability on population exposure to hydroclimatic changes. *Environmental Research Letters*, *12*. <https://doi.org/10.1088/1748-9326/aa5efc>
- Mankin, J. S., Viviroli, D., Singh, D., Hoekstra, A. Y., & Diffenbaugh, N. S. (2015). The potential for snow to supply human water demand in the present and future. *Environmental Research Letters*, *10*, 114016. <https://doi.org/10.1088/1748-9326/10/11/114016>
- Marvel, K., Schmidt, G. A., Shindell, D., Bonfils, C., LeGrande, A. N., Nazarenko, L., & Tsigaridis, K. (2015). Do responses to different anthropogenic forcings add linearly in climate models? *Environmental Research Letters*, *10*, 104010. <https://doi.org/10.1088/1748-9326/10/10/104010>
- McKinnon, A., Poppick, Dunn-Sigouin, E., & Deser, C. (2017). An 'observational large ensemble' to compare observed and modeled temperature trend uncertainty due to internal variability. *Journal of Climate*, <https://doi.org/10.1175/JCLI-D-16-0905.1>
- McKinnon, K. A., & Deser, C. (2018). Internal variability and regional climate trends in an observational large ensemble. *Journal of Climate*, *31*, 6783–6802. <https://doi.org/10.1175/JCLI-D-17-0901.1>
- Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., et al. (2009). Decadal prediction. *Bulletin of the American Meteorological Society*, *90*(10), 1467–1486. <https://doi.org/10.1175/2009BAMS2778.1>

- Milinski, S., Maher, N., & Olonscheck, D. (2019). How large does a large ensemble need to be? *Earth System. Dynamics Discussions*, 1–19. <https://doi.org/10.5194/esd-2019-70>
- Palutikof, J. P., Leitch, A. M., Rissik, D., Boulter, S. L., Campbell, M. J., Perez Vidaurre, A. C., et al. (2019). Overcoming knowledge barriers to adaptation using a decision support framework. *Climatic Change*, 153, 607–624. <https://doi.org/10.1007/s10584-018-2177-3>
- Paté-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering and System Safety*, 54, 95–111. [https://doi.org/10.1016/S0951-8320\(96\)00067-1](https://doi.org/10.1016/S0951-8320(96)00067-1)
- Pendergrass, A. G., Coleman, D., Deser, C., Lehner, F., Rosenbloom, N. A., & Simpson, I. R. (2019). Nonlinear response of extreme precipitation to warming in CESM1. *Geophysical Research Letters*, 46, 10551–10560. <https://doi.org/10.1029/2019GL084826>
- Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., & Sanderson, B. M. (2017). Precipitation variability increases in a warmer climate. *Scientific Reports*, 7, 17966. <https://doi.org/10.1038/s41598-017-17966-y>
- Pielke, R. A. Jr. (2003). The role of models in prediction for decision. In C. Canham, J. Cole & W. Lauenroth (Eds.), *Models in ecosystem science* (pp. 111–135). Princeton, NJ: Princeton University Press.
- Poppick, A., McInerney, D. J., Moyer, E. J., & Stein, M. L. (2016). Temperatures in transient climates: Improved methods for simulations with evolving temporal covariances. *The Annals of Applied Statistics*, 10, 477–505. <https://doi.org/10.1214/16-AOAS903>
- Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., et al. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, 109(1–2), 33–57. <https://doi.org/10.1007/s10584-011-0149-y>
- Rodgers, K. B., Lin, J., & Frölicher, T. L. (2015). Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences*, 12, 3301–3320. <https://doi.org/10.5194/bg-12-3301-2015>
- Saffioti, C., Fischer, E. M., & Knutti, R. (2015). The contribution of atmospheric circulation and coverage bias to the warming hiatus. *Geophysical Research Letters*, 17, 2385–2391. <https://doi.org/10.1002/2015GL063091>
- Scaife, A. A., Camp, J., Comer, R., Davis, P., Dunstone, N., Gordon, M., et al. (2019). Does increased atmospheric resolution improve seasonal climate predictions? *Atmospheric Science Letters*, 20(8), 1–10. <https://doi.org/10.1002/asl.922>
- Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *Npj Climate Atmospheric Science*, 1. <https://doi.org/10.1038/s41612-018-0038-4>
- Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schär, C., & Siebesma, A. P. (2017). Climate goals and computing the future of clouds. *Nature Climate Change*, 7, 3.
- Schwartz, C. S., Romine, G. S., Sobash, R. A., Fossell, K. R., & Weisman, M. L. (2019). NCAR's real-time convection-allowing ensemble project. *Bulletin of the American Meteorological Society*, 100, 321–343. <https://doi.org/10.1175/BAMS-D-17-0297.1>
- Selten, F. M., Branstator, G. W., Dijkstra, H. A., & Kliphuis, M. (2004). Tropical origins for recent and future Northern Hemisphere climate change. *Geophysical Research Letters*, 31, L21205. <https://doi.org/10.1029/2004GL020739>
- Simpson, I. R., Yeager, S. G., McKinnon, K. A., & Deser, C. (2019). Decadal predictability of late winter precipitation in western Europe through an ocean–jet stream connection. *Nature Geoscience*, 12, 613–619. <https://doi.org/10.1038/s41561-019-0391-x>
- Sippel, S., Meinshausen, N., Merrifield, A., Lehner, F., Pendergrass, A. G., Fischer, E., & Knutti, R. (2019). Uncovering the forced climate response from a single ensemble member using statistical learning. *Journal of Climate*, 5677–5699. <https://doi.org/10.1175/jcli-d-18-0882.1>
- Sterl, A., Severijns, C., Dijkstra, H., Hazeleger, W., Jan van Oldenborgh, G., van den Broeke, M., et al. (2008). When can we expect extremely high surface temperatures? *Geophysical Research Letters*, 35, L14703. <https://doi.org/10.1029/2008GL034071>
- Stern, P. C., Ebi, K. L., Leichenko, R., Olson, R. S., Steinbruner, J. D., & Lempert, R. J. (2013). Managing risk with climate vulnerability science. *Nature Climate Change*, 3, 607–609. <https://doi.org/10.1038/nclimate1929>
- Sutton, R. T. (2019). Climate science needs to take risk assessment much more seriously. *Bulletin of the American Meteorological Society*, 100, 1637–1642. <https://doi.org/10.1175/BAMS-D-18-0280.1>
- Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E., & Phillips, A. S. (2015). Quantifying the role of internal climate variability in future climate trends. *Journal of Climate*, 150521130832005. <https://doi.org/10.1175/JCLI-D-14-00830.1>
- Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, 6, 1–14. <https://doi.org/10.1126/sciadv.aaz9549>
- US CLIVAR. (2020). A summary of the US CLIVAR large ensembles workshop, July 24–26, 2019. 6 pp.
- Vano, J. A., Arnold, J. R., Nijssen, B., Clark, M. P., Wood, A. W., Gutmann, E. D., et al. (2018). DOs and DON'Ts for using climate change information for water resource planning and management: Guidelines for study design. *Climate Services*, 12, 1–13. <https://doi.org/10.1016/j.cliser.2018.07.002>
- Wallace, J. M., Fu, Q., Smoliak, B. V., Lin, P., & Johanson, C. M. (2012). Simulated versus observed patterns of warming over the extra-tropical Northern Hemisphere continents during the cold season. *Proceedings of the National Academy of Sciences*, 109, 14,337–14,342. <https://doi.org/10.1073/pnas.1204875109>
- Weaver, C. P., Lempert, R. J., Brown, C., Hall, J. A., Revell, D., & Sarewitz, D. (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, 39–60. <https://doi.org/10.1002/wcc.202>
- Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., et al. (2018). Predicting near-term changes in the Earth system: A large ensemble of initialized decadal prediction simulations using the Community Earth System Model. *Bulletin of the American Meteorological Society*, 99(9), 1867–1886. <https://doi.org/10.1175/BAMS-D-17-0098.1>
- Zhang, B., & Soden, B. J. (2019). Constraining climate model projections of regional precipitation change. *Geophysical Research Letters*, 46, 1–10. <https://doi.org/10.1029/2019GL083926>