UC Berkeley

UC Berkeley Previously Published Works

Title

Probabilistic Storylines: Characterizing the Likelihood of Impactful Events in an Uncertain World

Permalink

https://escholarship.org/uc/item/80h8k3x2

Journal

Bulletin of the American Meteorological Society, 107(10)

ISSN

0003-0007

Authors

Jones, Andrew D

Longmate, Julia M

Buddhavarapu, Smitha

et al.

Publication Date

2026-10-01

DOI

10.1175/bams-d-24-0301.1

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

BAMS Article

Probabilistic Storylines: Characterizing the Likelihood of Impactful Events in an Uncertain World

Andrew D. Jones,[a,b] Julia M. Longmate,[a,b] Smitha Buddhavarapu,[a,c] William Gutowski,[d] Antonia Hadjimichael,[e,f] Kripa Jagannathan,[a] Flavio Lehner,[g,h] Erwan Monier,[i] Joshua North,[a] Kevin A. Reed,[j] Alan Rhoades,[a] Mark Risser,[a] Andrew Schwarz,[k] Claudia Tebaldi,[l,m] Paul Ullrich,[i,n] and Colin Zarzycki[o]

KEYWORDS:
Extreme events;
Statistics;
Uncertainty;
Ensembles;
Decision making;
Risk assessment

ABSTRACT: Storylines are a useful approach for understanding and preparing for impactful events. For instance, resource managers often use "design droughts" or "floods of record" to stress-test system vulnerabilities and consider alternative responses. While plausible storylines are sufficient for many applications, there is a growing interest in rigorously evaluating storyline probabilities to inform risk-based adaptation investments, especially in cases where preparing for storylines is costly or where storyline probabilities are changing over time. However, assigning actionable probabilities to storylines presents conceptual, methodological, computational, and data challenges. One key challenge is how to reconcile deeply uncertain aspects of the future arising from different scenarios and model structural biases with those aspects that are more amenable to statistical characterization, such as stochastic variability. Another challenge is translating the idiosyncratic features of specific storylines into decision-relevant metrics that can be understood statistically, ideally delivering probability distributions of salient events. Informed by our experience coproducing storyline simulations with resource managers and Earth system scientists, and by a critical literature review, we propose a conceptual framework for rigorously incorporating probabilities into storyline-based analysis and planning. The framework isolates components of uncertainty addressable by extreme value analysis and encourages context-specific methods for evaluating variation across scenarios and models including methods from the field of decision-making under deep uncertainty. We present use cases for the framework that demonstrate how it might be applied for practical management purposes and highlight promising approaches from the literature that can overcome methodological, computational, and data challenges to implementation.

SIGNIFICANCE STATEMENT: Resource management practitioners often use impactful event sequences, or storylines, for stress-testing system resilience and considering alternative options for addressing vulnerabilities. Assigning probabilities to such storylines would increase their management value, especially if preparing for the storylines is costly. However, assigning probabilities to storylines presents significant conceptual, methodological, computational, and data challenges.

DOI: 10.1175/BAMS-D-24-0301.1
Corresponding author: Andrew D. Jones, adjones@lbl.gov
Manuscript received 7 November 2024, in final form 13 July 2026, accepted 21 August 2026
© 2026 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Here, we present a framework for understanding the different kinds of uncertainty that contribute to storyline probabilities and for using appropriate methods to address each type of uncertainty as part of a comprehensive approach. We also highlight emerging opportunities for improving the methods and data available for calculating storyline probabilities and their uncertainties.

AFFILIATIONS: [a] Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California; [b] Energy and Resources Group, University of California, Berkeley, Berkeley, California; [c] College of Environmental Design, University of California, Berkeley, Berkeley, California; [d] Department of the Earth, Atmosphere, and Climate, Iowa State University, Ames, Iowa; [e] Department of Geosciences, The Pennsylvania State University, University Park, Pennsylvania; [f] Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, Pennsylvania; [g] Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York; [h] Polar Bears International, Bozeman, Montana; [i] Department of Land, Air and Water Resources, University of California, Davis, Davis, California; [j] School of Marine and Atmospheric Sciences, Stony Brook University, State University of New York, Stony Brook, New York; [k] State Water Project, California Department of Water Resources, Sacramento, California; [l] Pacific Northwest National Laboratory, Richland, Washington; [m] Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland; [n] Division of Physical and Life Sciences, Lawrence Livermore National Laboratory, Livermore, California; [o] Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania

1. Introduction

In the context of climate vulnerability and risk assessment, a physical climate storyline (hereafter "storyline") can be defined as a self-consistent and plausible trajectory of the climate system or of a particular weather event (Shepherd et al. 2018). For example, a storyline might describe a slow-moving hurricane making landfall along a populated coastline, producing extreme rainfall, storm surge, and prolonged inland flooding. From a practitioner's perspective, storylines are useful for understanding and preparing for potentially high-impact events or climate trajectories. Practitioners often use storylines to stress test their operations, which provides insight into system vulnerabilities and the efficacy of management strategies (California Department of Water Resources 2023, 2025; Hydros Consulting Inc. 2018; Lehner 2024; Mankin et al. 2020; Smith et al. 2022). Storylines have also proven useful for performing controlled "what-if" experiments grounded in relationship to past experience, for instance by examining in detail how past events could play out under different climate conditions (Chan et al. 2023; Huang and Swain 2022; Huprikar et al. 2024; Reed and Wehner 2023; Rhoades et al. 2023, 2024; Swain et al. 2018; Ullrich et al. 2018). Storylines can also be effective in capturing the imagination of the public, planners, and scientists (Kahneman 2011; Roush 2024).

One advantage of storylines is their ability to sidestep difficult issues of characterizing uncertainty. To provide useful insights into system vulnerabilities and responses, a storyline must only be deemed "plausible" in the process of storyline construction or in assessing the usefulness of the storyline (Shepherd et al. 2018). Earth system models (ESMs) can be used to test the plausibility of the storyline (Yiou et al. 2024), rather than as sources of probabilistic prediction, thereby avoiding the need to address deeply uncertain factors such as model and scenario uncertainty and avoiding sensitivity to upstream uncertainties. For many

applications, simply knowing that a storyline is plausible is sufficient to raise awareness and develop action plans for responding if the storyline occurs. Despite these strengths, the absence of probability characterization from storylines also limits their usefulness in certain decision contexts, particularly when the cost of preparing for those storylines is substantial, for example, when major investments in infrastructure are required (Hinkel et al. 2019). Indeed, the origins of this paper can be traced to systematic feedback that we received from climate data users and practitioners during coproduction engagements that storylines alone, while valuable, do not present sufficient information for decision-making without associated probabilities. Several studies have elucidated that the storyline and the probabilistic approaches represent two ends of a spectrum in terms of risk preferences (Pulkkinen et al. 2022; Winsberg et al. 2020). While storylines excel in exploring the risks of falsely underestimating the likelihood of severe events, probabilistic approaches are useful in exploring the other end of the risk spectrum, i.e., the risks of falsely overestimating the likelihood of events, preparing for which might be costly (Parker and Lusk 2019). Identifying which of these two errors is most consequential depends on specific user contexts and their use cases (Jagannathan et al. 2023; Winsberg et al. 2020), but for the latter error, plausibility alone is an insufficient basis for decision-making.

Though desirable, attaching probabilities to storylines—especially across alternate climates with different degrees of global warming—is challenging for several reasons. First, storylines are idiosyncratic. Storylines describe unique trajectories and events arising from complex systems and processes; every trajectory or event, exactly as it unfolded, does not have a well-defined probability of occurrence. Metrics are needed to abstract unique storylines into a discrete set of salient features, defining classes of events that can possess probabilistic properties. However, the salient features of a storyline can vary substantially across decision contexts, and not all metrics of interest can be credibly evaluated with currently available datasets and models. Therefore, collaboration between practitioners and technical experts may be required to appropriately define storylines in a manner that is both decision-relevant and technically tractable.

Second, storyline uncertainties arise from multiple sources that are best understood through different methodological and potentially different philosophical approaches. Some uncertainties, such as stochastic weather variability, so-called "internal variability" of the climate system (Deser 2020; Lehner and Deser 2023), can be well characterized probabilistically through methods such as extreme value analysis (EVA). Other uncertainties, such as those arising from alternate future climate forcing scenarios and modeled responses to those forcings, are deeply uncertain—meaning their probability distributions are unknown, or experts do not agree on their distribution nor on a single model that can describe their distribution (Marchau et al. 2019; Walker et al. 2013). Focusing only on readily characterizable components of uncertainty (e.g., by analyzing a single model or emissions scenario) or treating deeply uncertain components inappropriately (e.g., by relying on narrow assumptions about the range of plausible emission scenarios) can result in spuriously precise and overly confident probability assessments (Sanderson et al. 2017). On the other hand, interpreting the presence of deep uncertainty as a complete lack of statistical constraints on likelihood or assuming that all models are equally likely means ignoring potentially useful information that could prevent overpreparation for storylines that are indeed plausible yet extremely rare, even when considering deeply uncertain components.

A number of methods have been employed in the literature to qualitatively constrain or quantitatively assign weights to deeply uncertain factors such as scenarios and models based on their presumed likelihood or credibility (Gillett 2024; Lehner et al. 2023; Sanderson and Wehner 2017). Meanwhile, the field of decision-making under deep uncertainty (DMDU) has developed a range of methods for informing robust and adaptable decisions, even when

uncertainty cannot be fully characterized (Marchau et al. 2019; Moallemi et al. 2020, 2023). Understanding when and how these different philosophical and methodological approaches can be combined with rigorous statistical analysis of the well-characterizable components of uncertainty is thus a key challenge for understanding the probabilities of storylines. "Rigorous" here means following a principled (i.e., grounded in statistical theory) and transparent methodology, which requires acknowledging and accounting for conditionality and uncertainty.

Last, evaluating storyline probabilities is a computationally and data-intensive process that is limited by currently available projection datasets. Many promising methodologies are available to overcome these challenges but have not yet been combined in a comprehensive manner to generate more ideal datasets that can be leveraged to systematically evaluate storyline probabilities and their uncertainties. These include methods for constructing storyline metrics, for correlating salient fine-scale outcomes with predictive quantities well represented in models, for efficiently generating large ensembles of climate data, and for designing climate model simulation protocols that systematically and consistently probe various dimensions of uncertainty.

In this paper, we propose a framework, informed by coproduction and a critical literature review, for characterizing storyline probabilities across a range of uncertainties. We propose partitioning model, scenario, and stochastic uncertainties into well-characterizable and deeply uncertain components and employing different methods for understanding these components separately. Within this framework, the concept of global warming levels provides a useful tool for convolving deep uncertainty arising from emission scenarios and global climate sensitivity into a single factor. For any given climate model, warming level, and metric, it is possible to rigorously evaluate storyline probability distributions, though the choice of metric can have a large effect. Various approaches can then be used to compare those probabilities and their decision relevance across warming levels and models. We conclude by highlighting promising methodologies and challenges that must be overcome to more systematically and effectively implement this framework. We note that our proposed framework shares some conceptual similarities and methodologies with probabilistic event attribution (PEA), which seeks to understand how present-day climate change has altered the frequency and characteristics of extreme events (Philip et al. 2020). The largest difference is that we focus on future changes in storyline probabilities, although our framework is also applicable for historical change. Considering future change requires that our framework address the conditionality of probabilistic change as a function of deeply uncertain scenario and model factors. We also place a strong emphasis on metric selection for practical salience.

2. Conceptual framework

a. Collaborative problem and metric definition. Assigning probabilities to storylines is motivated in large part by practitioner needs. Therefore, collaborative problem definition and solution framing are critical to ensure that the approach is salient and credible to assist in decision-making (Parker and Lusk 2019). Embarking on this novel integration of concepts will necessarily involve trade-offs between scientists' and users' values and priorities (i.e., scientific accuracy vs practical needs) (Undorf et al. 2022). Iterative engagements between the two groups can help to ensure that there is transparent understanding regarding what types of risk preferences, errors, and scientific approaches are most suitable in different decision contexts (Helgeson et al. 2022; Parker and Lusk 2019).

From an operational perspective, four key aspects of developing and evaluating storylines can benefit from iterative coassessment between scientists and users: identifying key events of interests (e.g., most extreme vs a range of intermediate events), identifying the specific characteristics of these events that are most impactful for management (e.g., timing, seasonality,

duration, spatial extent, compounding), identifying specific metrics and thresholds of importance (e.g., Hadjimichael et al. 2020, 2024; Jagannathan et al. 2021; Reed et al. 2022), and discussing the types of errors or uncertainties that are most relevant to the context (Parker and Lusk 2019; Winsberg et al. 2020).

Collaborative problem definition is especially important for identifying appropriate metrics for characterizing storylines. Because different metrics emphasize different features, the same storyline described by different metrics could acquire different probabilities (Jagannathan et al. 2021; Reed et al. 2022). Storm severity, for instance, could be characterized by cumulative precipitation, by maximum wind speed, or by other metrics, each of which would result in a different statistical distribution of storm severities and therefore a different likelihood assigned to any particular storm. Studies on Hurricane Harvey illustrate this, with varying event definitions yielding different return periods, e.g., 1-in-3000 years for a 7-day precipitation total (Risser and Wehner 2017), 1-in-9000 years for a 3-day precipitation total (van Oldenborgh et al. 2017), or 1-in-2000 years for "storm totals" of varying storm durations (Emanuel 2017). To be decision relevant, it is critical that metrics reflect the features of storylines that matter most to decision-makers.

It is also important to consider metric choice within a broader problem definition and decision framing. For instance, flood elevation is a decision-relevant metric that is affected by both physical hazard characteristics (e.g., precipitation severity) and human infrastructure and adaptive responses. If the goal is to understand return frequency of physical hazards to inform adaptive decisions, then it may be preferable to choose physical hazard metrics, like storm total precipitation or maximum runoff, to characterize event probabilities and then evaluate implications for downstream flood impacts across a range of potential responses. If the goal is to evaluate how adaptive responses change the probability of flood impacts, then it may be appropriate to use a more downstream metric, like flood elevation, that is influenced by both physical and human processes. In other contexts, decision-makers may need to consider the interplay of multiple risks with distinct metrics (e.g., Jézéquel et al. 2024). One could use high-level metrics like temperature and precipitation to characterize future trajectories, for example, as broadly "hot and wet," or use multiple more specific hazard-oriented metrics to characterize how the frequencies of droughts, floods, and other impactful events change within those trajectories. Compound hazards and cascading impacts across multiple systems may require an examination of joint metrics, such as the intersection of droughts with heat waves (Rastogi et al. 2023), or the co-occurrence of both human health and electricity grid vulnerabilities (Casey et al. 2020).

From a practical perspective, though, not all metrics can be credibly modeled by current global and regional climate models. Spatial or temporal aggregation, or identification of proxy metrics, may be required to evaluate storyline characteristics within model outputs (Fig. 1). For example, in Fig. 1, streamflow rates are not directly represented in ESMs, but aggregate runoff is. Some quantities, like peak snowpack accumulation and snowmelt rates, are represented in models but have significant biases (McCrary et al. 2017; Rhoades et al. 2018; Xu et al. 2019) that could affect storyline characterization. Many promising methodological innovations are becoming available for characterizing event metrics and identifying large-scale features that are predictive of fine-scale decision-relevant features. These are discussed in section 4 below.

Because metric definition requires a careful balance of decision-relevant salience and technical credibility and tractability (Vincent et al. 2020), collaborative approaches such as coproduction offer a promising set of methods and insights (Bremer and Meisch 2017; Jagannathan et al. 2021). Examples from across the world have shown that iterative interactions between climate scientists and practitioners have helped to identify metrics that are both tractable for modelers and also actionable for users (Jacob et al. 2025; Jagannathan et al. 2021; Vincent et al. 2020). However, extracting nuanced details of metrics through coproduction

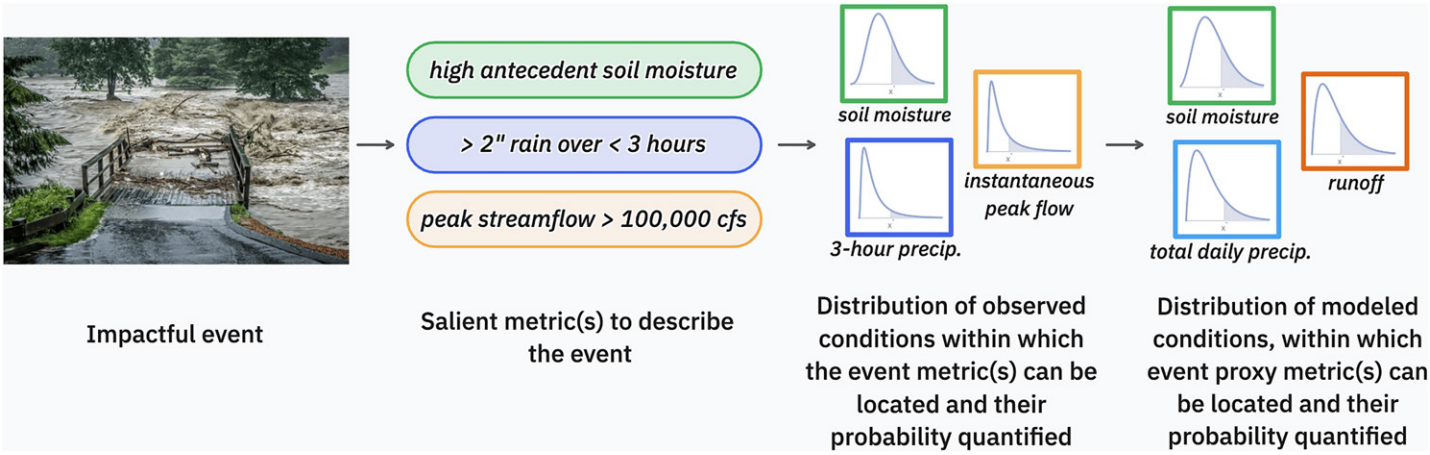


FIG. 1. **Translating storyline features into salient and credible probability distributions: To be statistically represented within a distribution of possible storylines in a climate model, the multidimensional features of a storyline must be abstracted using metrics. Multiple alternative metrics can represent different salient features of the storyline that are relevant for decision-makers, each of which would have its own statistical distribution. Some metrics can be observed directly but are not represented credibly within climate models. Therefore, aggregation or identification of proxy metrics may be required to further translate salient features of the storyline into quantities whose modeled distributions can support probabilistic characterization. This balance of decision salience and technical credibility can be mediated through collaborative approaches such as coproduction.**

is not an easy process and requires careful choice and implementation of the appropriate engagement methods and strategies (Jagannathan et al. 2021; Vincent et al. 2020). Skilled intermediaries or boundary spanners are often needed to effectively execute engagements that translate among the different actors and across epistemic divides (Goodrich et al. 2020; Jagannathan et al. 2021; Safford et al. 2017). Although there is growing evidence of the effectiveness and success of coproduction, it can be a time- and resource-intensive process, and strategies to cost-effectively scale-up coproduction are only starting to emerge (Lemos et al. 2025). In cases where intensive coproduction is infeasible, leveraging peer-reviewed and gray literature on decision-relevant metrics or storylines (Jagannathan et al. 2021; Vincent et al. 2020; Pacchetti et al. 2023), and reviewing existing coproduction frameworks to identify feasible low-resource engagement options (Selby et al. 2024; Bamzai-Dodson et al. 2021), could offer tangible starting points for including practitioner perspectives. Gathering feedback on problem framing and metrics from agencies that work at the boundary between climate science and decision-making (such as NOAA's Regional Integrated Science and Assessments, USGS's Climate Adaptation Science Centers, or USDA's Climate Hubs) could also provide an accessible entry point into making storyline research more actionable.

b. Tailoring of analytic methods to different types of uncertainty. As demonstrated by Hawkins and Sutton (2009), uncertainty in climate projections can be decomposed into contributions from scenario uncertainty, model uncertainty, and internal variability. In their work, this decomposition is for the purpose of quantifying how each uncertainty source contributes to overall variability and how those relative contributions change across time, space, and different variables or metrics. The decomposition has been extended to include additional uncertainty sources, such as downscaling uncertainty (Lafferty and Sriver 2023; Risser et al. 2024). While Hawkins and Sutton use this decomposition to quantify how each uncertainty source contributes to overall variability in an additive manner, our proposed framework uses a similar taxonomy of uncertainty sources to investigate how the structure of internal variability changes as a function of more deeply uncertain factors arising from scenario and model uncertainty, proposing a hierarchical framework germane to a conditional probability approach. The rationale for this approach follows from the notion that different sources of uncertainty are best understood and reckoned with using different analytic

Table 1. Partitioning of storyline uncertainties by type and scale with different methods that are appropriate for quantifying or informing decision-making under each type of uncertainty.

Uncertainty partitioning	Scale	Further uncertainty partitioning	Type of uncertainty	Methods to address types of uncertainty
Scenario uncertainty	Global		Constrained deep uncertainty	DMDU methods Scenario elimination based on expert judgment Bayesian weights based on observed socioeconomic, technology, and emissions trends
Model uncertainty	Global	Global mean response	Constrained deep uncertainty	DMDU methods Climate sensitivity constraints based on multiple lines of evidence
	Regional	Regional processes within global models plus downscaling uncertainty	Constrained deep uncertainty	DMDU methods Model culling based on credibility assessment Bayesian weighting based on credibility assessment including emergent constraints
Internal variability	Multiscale		Well-characterizable uncertainty	Repeated sampling of outcomes using initial condition ensembles, ensemble boosting, long time widows (exchanging time for replicates), and emulators. Data-driven and/or Bayesian probabilistic characterization including EVA, driven by data.

methods and potentially different philosophical orientations (Table 1; Kandlikar et al. 2005; Walker et al. 2003).

Focusing on how the structure of internal variability changes conditionally on other uncertainty sources has two key advantages for understanding storyline probabilities. First, when internal variability is isolated from external forcing, it is well suited to data-driven probabilistic methods. Internal variability is the stochastic noise of the climate system. In models, one can repeatedly sample internal variability using initial condition ensembles or long simulations that allow the system to chaotically evolve through its many possible states. Scenarios and models can be held fixed in such a way that repeated samples can be considered as drawn from stationary distributions of possible states. More samples lead to better statistical characterization of these distributions, and probabilistic methods like EVA can be applied to estimate storyline likelihoods based on these distributions. There is no upper limit to how many samples can be taken. Indeed, the increasing availability of single-model initial-condition large ensembles (SMILEs) (Deser et al. 2020; Maher et al. 2021) and advances in the development of emulators (Tebaldi et al. 2025) allow for an unprecedented exploration and understanding of internal variability, especially at the tails of distributions where extreme and often impactful events occur (Mahesh et al. 2025).

The second advantage of this approach is that it encourages and enables the storyline analyst to choose a context-appropriate method for dealing with the remaining scenario and model uncertainties. Unlike internal variability, scenarios and models are not randomly drawn from a stable distribution of all possible scenarios and models that can be further resolved by repeated sampling. Historically, scenario developers have characterized uncertainty as a set of plausible alternatives spanning a wide range of futures rather than a statistical distribution (O'Neill et al. 2016; Van Vuuren et al. 2026). Moreover, as deliberative tools for considering policy options, there may be philosophical reasons to treat emission scenarios as choices or "levers" rather than exogenous uncertainties to be characterized (Lempert et al. 2003). As for climate models, it has been long argued that they cannot be considered independent of each other, as they share ancestry and components, share common biases, and vary in resolution, process representation, and treatment of individual forcings, confounding statistical treatment (Gateño et al. 2024; Knutti et al. 2013), which has led to the development of various

approaches for climate model selection by independence, performance, and spread of future climate projections for decision-making purposes (Merrifield et al. 2023).

Various methods have been applied to characterize and constrain both scenario and model uncertainties, based on observational evidence and expert judgment. These include ruling out particular scenarios or models as implausible or insufficiently accurate (Gillett 2024; Goldenson et al. 2023; Hausfather and Peters 2020; Sanderson et al. 2017; Sobolowski et al. 2025) and Bayesian-based weights (Liu and Raftery 2021; Raftery et al. 2017; Sanderson and Wehner 2017). Meanwhile, DMDU is an area of decision science used to inform planning and management decisions in complex systems where experts do not know or cannot agree on probabilistic descriptions of potential outcomes (Marchau et al. 2019). Outside academic research, DMDU concepts and methods have been used in a variety of complex planning contexts, such as in the Thames Estuary 2100 project (Reeder and Ranger 2011), in the Dutch Delta Program (Bloemen et al. 2019), and in Colorado River basin planning efforts (Groves et al. 2013). DMDU offers several frameworks and analytical tools for evaluating system vulnerabilities and identifying robust and adaptive solutions that do not require formal quantitative evaluation of probabilities (Marchau et al. 2019; Moallemi et al. 2020). How best to combine these approaches with statistical evaluation of internal variability can vary with the goals of the analysis, preferences of decision-makers, and practical tractability.

Among these approaches, DMDU requires the least restrictions or weights on scenarios and models. DMDU accepts that some uncertainties are fundamentally unresolvable. Various methods are used to evaluate decision options across a range of plausible future states of the world to identify robust or adaptive solutions. DMDU approaches can be used to flip uncertainty analysis around and ask "which uncertainties are consequential to the decision at hand?" rather than "how do all the uncertainties propagate and interact to influence decision outcomes?" For instance, if the differences among alternative scenarios or models are not substantial enough to imply a different decision strategy, then it is not necessary to invest resources into determining precisely which scenarios or models are more or less likely. However, because DMDU seeks to find strategies that work well across a wide range of plausible future states, they might include and potentially overprepare for storylines that are extremely unlikely. Further, strategies could be vulnerable to different sets of scenario or model assumptions, making the value of understanding their relative likelihood higher, especially when alternative strategies have very different implementation costs. Combining DMDU for scenario and model uncertainty with probabilistic methods for internal variability offers a promising approach that could rule out storylines that are deemed to be sufficiently rare (in a formalized statistical sense) across a range of plausible scenarios and models.

To the extent that additional constraints can be defensibly applied to rule out or assess the relative likelihood of scenarios and models, this would further reduce the risk of overpreparation for unlikely storylines. Studies have argued for the elimination of certain scenarios based on expert judgment of their plausibility (Hausfather and Peters 2020) or culling of models based on an assessment of their credibility for representing key physical processes (California Department of Water Resources 2015). In general, it is easier to eliminate scenarios or models as implausible than to declare them to be likely with some quantitative certainty. For circumstances that call for such quantitative precision, Bayesian statistical methods have been used to develop likelihood weights across both scenarios and models. Bayesian methods can be helpful in a context where the uncertain quantity of interest cannot be easily seen as the realization of a random process. Rather than taking repeated samples to gain a more precise understanding of a stable distribution, Bayesian methods combine prior knowledge, sometimes in the form of expert judgment, to construct an initial probabilistic representation. The prior is updated using data-driven likelihood, possibly incorporating multiple lines of evidence, to arrive at a posterior distribution that reflects our best understanding of currently

available information. This approach is better suited to the nature of scenario and model uncertainty, which cannot be refined through repeated random sampling but can be updated as new evidence and understanding evolves (Gillett 2024; Lehner et al. 2023). It should be noted that Bayesian methods can only reflect current knowledge and assumptions, which may change over time. For instance, it was difficult to anticipate 20 years ago how artificial intelligence technologies would increase electricity demand globally, so Bayesian scenario weights based on historical population, GDP, and emissions data from this period (Raftery et al. 2005) would not reflect the role of this emerging technology on energy demand and associated greenhouse gas emissions. As these technologies proliferate, our Bayesian estimation of future scenario likelihoods can be updated, though.

Combining data-driven probabilistic methods for internal variability contingent on some combination of Bayesian and DMDU methods for scenario and model uncertainty is at the heart of our proposed framework. How these methods are best combined will depend on the specific circumstances. Regardless, it is valuable to be transparent and intentional about assumptions and thoughtful about the consequences of combining expert judgment, subjectivity, and data-driven methods with different weights to the various components of the problem. For instance, in a stochastic optimization context (Palmintier and Webster 2016; Tong et al. 2022), it would be ideal to develop Bayesian weights for all scenario and model uncertainties and convolve these with internal variability to produce a single best estimate of the probability distribution. A DMDU framework that treats some range of scenarios and models as plausible but uses data-driven methods for evaluation of storyline likelihoods within scenario-model combinations would expose the deeper uncertainties to deliberation and could be used to identify which ones matter most to the decision outcomes. Once the most consequential dimensions of uncertainty are identified, the notion of "sign posting" from the DMDU literature (Fletcher et al. 2019) calls for monitoring of those uncertainties over time, which could feed into a more formalized Bayesian approach that develops and updates distributions of those quantities over time.

c. Isolating self-consistent worlds with global warming levels. To probe the structure of internal variability contingent on scenarios and models, it is essential to hold scenario and model factors fixed in such a way that repeated samples of internal variability are drawn from a consistent and stable distribution. This is especially challenging for scenarios, which by their nature lead to time-varying changes in the distribution of climate and weather phenomena. Thus, each time slice of a scenario is associated with a different distribution of internal variability, and we would ideally have a large number of samples for each time, scenario, and model to fully understand how the structure of internal variability and, therefore, the probability of storylines changes as a function of its various uncertainty drivers.

The concept of global warming levels (GWLs) holds promise for simplifying this multidimensional problem by combining emissions scenario, time frame, and global mean model response into a single factor (Tebaldi et al. 2021). GWLs are annual global average temperature anomalies relative to preindustrial mean global temperature, evaluated over a given time frame, typically a climate normal on the order of 20–30 years (Arguez and Vose 2011; Seneviratne et al. 2021, cross-chapter box in 11.1 in chapter 11). GWLs are physically meaningful because global mean temperature governs many fundamental climate processes. The atmosphere's capacity to hold water via the Clausius–Clapeyron relationship, vertical temperature gradients, jet stream position, elevation at which snowmelt occurs, and most atmospheric variables having a short memory (i.e., not path dependent) have been found to behave as a direct function of global mean temperature. This motivates common language framing possible futures as, e.g., "a 2° world," and has become a central feature of international climate agreements.

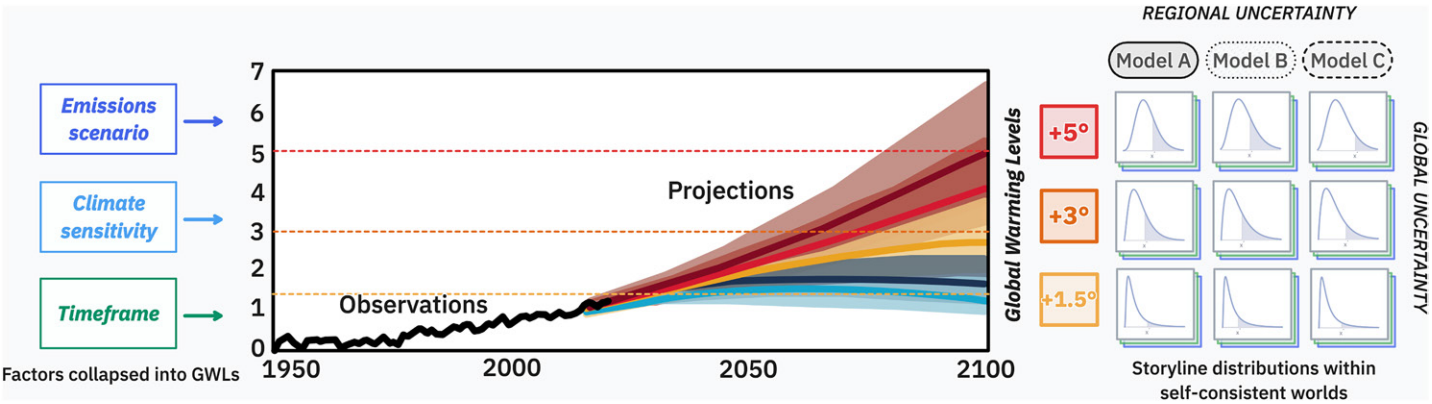


FIG. 2. GWLs (dashed lines) simplify time-varying differences across projections by combining emissions scenario, time frame, and global mean model response into a single factor. GWLs are calculated from the global surface temperature (°C relative to 1850–1900) of projections across each emissions scenario, model realization, and time frame. Evaluating the regional response including hazard characteristics, at a given GWL introduces additional model uncertainty. One model and one GWL comprise a world with internally consistent forcing and responses that can be repeatedly sampled to develop a statistical understanding of internal variability and therefore storyline probabilities. The likelihood of that world is then contingent on two factors; model uncertainty and the likelihood of that GWL. The central figure is adapted from Fig. TS.4 in the IPCC Sixth Assessment Report, Working Group I: The Physical Science Basis (2021).

For any given GWL, different models may represent regional processes differently, and so the regional structure of internal variability across models (potentially including additional downscaling models) remains as an additional axis of uncertainty (see Fig. 2). Each combination of one GWL and one model comprises a "world" with internally consistent forcing that can be repeatedly sampled with initial condition ensembles and/or sufficiently wide windows of time to develop a statistical understanding of internal variability and therefore storyline probabilities. We underline here that given the assumption of path independence, the sampling can be carried out across scenarios, for the same GWL, for most quantities of interest. As discussed in the previous section, scenario and model uncertainty, now expressed as GWLs and additional regional model uncertainty, can be treated using a variety of context-appropriate methods ranging from DMDU approaches to Bayesian weighting schemes.

The GWL approach effectively separates model uncertainty into its global and regional components and then convolves the global component with scenario and time frame considerations. This is advantageous for several reasons. First, there is a robust literature on both equilibrium and transient climate sensitivity that provides physical constraints on global mean response to climate forcing that can be used to constrain or apply Bayesian weights to the global mean model response (Hausfather 2025; Sherwood et al. 2020). These physical constraints on climate sensitivity have been combined with observational constraints on socioeconomic, emissions, and policy factors to produce Bayesian likelihood estimates for GWLs that are independent of the available ensemble of ESMs (Lehner et al. 2023). Second, this approach allows one to use model runs consistent with a given GWL irrespective of which particular scenario, time frame, and climate sensitivity produced that GWL. This is especially useful for so-called "hot models" (Hausfather et al. 2022), whose climate sensitivity exceeds our best estimate of the physically constrained range. Even though these models might overestimate the global mean response, their representation of regional processes such as jet streams and monsoon dynamics has been shown to be credible, so these models can contribute to our estimation of plausible regional outcomes, conditionally on a given GWL (Goldenson et al. 2018, 2023).

It is important to note that several factors could complicate the use of GWLs. For instance, due to varying time scales associated with different feedback processes, a transient 2° world reached relatively quickly (as in a high-emissions scenario) may differ from a

transient 2° world reached relatively slowly (as in a moderate emissions scenario), or a 2° world that is reached following an overshoot (Schleussner et al. 2024). Likewise, a 2° world reached through a different mix of climate forcings with different regional impacts such as land use (Jones et al. 2013) and aerosol emissions (Shiogama et al. 2023) can also differ. More work is needed to determine when these differences among GWL worlds are sufficient to impact regional storyline characteristics and probabilities.

3. Motivating examples

The practical needs motivating our framework for probabilistic storylines were elicited as part of scientist–practitioner coproduction efforts (Jagannathan et al. 2021, 2023). Initial problem identification and ideation on necessary elements of the framework originated at a scientist–practitioner workshop in 2023 that included several land, water, and energy sector resource managers. The framework was refined through a series of focus groups and 1:1 meetings with key climate data user agencies to detail examples of the use of probabilistic futures in practice. Additional experts, for example statisticians, were invited to join the manuscript to fill expertise gaps identified through coproduction.

Here, we provide an example based on the experience of the California Department of Water Resources (CDWR) in developing quasi-probabilistic future climate trajectories. This effort predates the development of the framework described in this paper and has served as a useful case study for motivating practical needs and identifying conceptual and methodological challenges encountered by practitioners.

CDWR operates California's State Water Project (SWP), delivering water to 27 million Californians and 750 000 acres of farmland across the 29 public water agencies it serves. The public water agencies have diverse risk tolerances and vulnerabilities that include both underpreparing for climate stress in a manner that results in water shortages and overinvesting in solutions that excessively raise water use rates. The CDWR provides estimates of delivery reliability under current and future (20 years out) conditions. Past CDWR reports provided multiple climate futures but with no associated probabilities. This lead to discounting of more extreme scenarios due to an inability to gauge how plausible those scenarios were.

In 2024, CDWR began providing quasi-probabilistic future climate trajectories across a range of plausible futures (California Department of Water Resources 2025). Using hydrologic and water systems models and a decision scaling approach (Brown et al. 2012), CDWR developed a stress response mapping to evaluate how changes in physical climate variables (temperature, precipitation, and sea level rise) impact key management–relevant metrics, including SWP deliveries. An ensemble of downscaled GCM projections was then used to calculate a bivariate probability density of climate variable change at future time periods. Combining the stress response mapping with the distribution of climate outcomes allowed CDWR to develop a continuous relationship that quantifies the relative model-informed likelihood of any level of system performance. From this relationship, specific system risk–informed climate scenarios with assigned stress severity percentiles or "levels of concern" were developed. Through system manager and water user collaboration, a limited ensemble was selected for full planning scenario development. These include scenarios representing the median, 75th percentile, and 95th percentile system performance outcomes, which were identified as an appropriate range over which water system planners could gauge both expected future conditions and more extreme (i.e., lower system performance) conditions that can be used for risk management and adaptation decision-making.

From a user perspective, having a principled and transparent method for evaluating how likely various outcomes are within the ensemble of available projections represents a significant improvement over simply treating all projections as plausible. Users understand that the

95th percentile level of concern is not a "worst-case scenario" but is a severe outcome that exceeds most of the other trajectories in expected impact. However, this approach was limited by available projections and could be improved in light of the framework presented in this paper. One of the specific challenges noted by CDWR was distinguishing internal variability from model and emissions scenario uncertainty. Models with multiple initial condition runs (3 or greater) proved more valuable in this respect, and CDWR found that large ensembles of initial conditions with the same model-downscaling-emissions combinations would be most valuable. Because CDWR focused on near-term outcomes (in year 2043), and because GWLs do not diverge substantially across emissions scenarios until after 2050, it was justifiable to treat projections from multiple emissions scenarios as equally likely. However, an extension to later time periods would require more nuanced approaches to scenario or GWL uncertainty. Additionally, the approach focused on a single integrated metric of system stress that is correlated with several important performance metrics, but the need to serve multiple decision contexts points to the value of evaluating multiple outcome metrics with different likelihoods.

Beyond decadal climate trajectories, a number of shorter-duration event storylines have been developed for understanding water resource management challenges in California (Rhoades et al. 2023, 2024; Ullrich et al. 2018). One of the most prominent is the ARkStorm storyline, which is based on a megastorm series of severe atmospheric rivers that caused widespread flooding across California in 1861–62. The ARkStorm is a good demonstration of the capacity of storylines to capture the public imagination and drive media and conversations. Importantly, simulations of ARkStorm or ARkStorm2 (Huang and Swain 2022; Swain et al. 2018) calculated likelihoods using a large ensemble from a GCM, executing part of the methodology proposed in this concept paper and providing valuable risk information beyond plausibility. The single-model approach leaves open the question of model uncertainty, however, and evaluation across additional GWLs would enable practitioners to better understand how risk changes across emissions scenarios and time.

4. Promising methodological approaches and challenges

Recent innovations in event-based metric analysis, simulation protocols, model emulators, and use of multimodel ensembles offer promising methodologies for implementing aspects of our proposed framework. Here, we highlight some of those emerging opportunities and discuss how they could be leveraged for probabilistic storylines, as well as key gaps and challenges that must be overcome.

a. Metric construction. The most salient features of storylines, especially those corresponding to fine-scale phenomena requiring high-resolution data, may not be represented well in ESMs, or it may not be feasible to produce sufficient data for exploring their probabilities due to the computational expense and complexity associated with high-resolution ensembles of ESM simulations. One solution is to tie fine-scale extremes to large-scale predictors—fields that are well captured by grid spacings used to generate large climate ensembles. Through this bridge, changes in large-scale predictor characteristics, strengths, and frequencies can be mapped onto probabilities of changes in fine-scale phenomena. Large-scale meteorological patterns (LSMPs) are atmospheric circulation features, larger than mesoscale systems but smaller than global climate modes, that accompany short-duration extreme events (Grotjahn et al. 2016). The circulation patterns associated with extremes (including compound and sequential events) can therefore be used to explore the spatial distributions and duration of these events (Barlow et al. 2019).

LSMPs can be identified using a range of statistical and pattern-recognition approaches. Common methods include composites, which average maps for preidentified extreme-event dates (e.g., Gutowski et al. 2008; Gao et al. 2014; Schoof et al. 2019); regression techniques,

which relate predictor fields to extremes at each grid point (e.g., Lau and Nath 2012); and empirical orthogonal functions (EOFs), which extract dominant spatial modes from event-date data (e.g., Sukhdeo et al. 2022). More advanced and increasingly automated machine learning approaches, such as self-organizing maps (Dong et al. 2021; Glisan et al. 2016; Gore et al. 2023; Song et al. 2019), clustering algorithms (Agel et al. 2018; Peters and Schumacher 2014; Srivastava et al. 2025), and emerging deep learning methods (e.g., Chattopadhyay et al. 2020), objectively group or classify patterns to reveal recurrent large-scale structures linked to extremes. Such strategies are intimately linked to the "analog technique," which permits the exploration of potential changes in weather conditions similar to those for a storyline event of interest (Faranda et al. 2020; Terray 2021; Yiou 2014; Gao et al. 2017). This approach has been used to find that European summertime heat waves and wintertime windstorms are becoming more frequent by investigating analogs of the large-scale atmospheric patterns that promote these events (Faranda et al. 2023). A related, but distinct, strategy to produce unresolved probabilistic hazard information from large-scale environments is the use of statistical or statistical–dynamical methods, where synoptic-scale predictor environments from climate models are used in conjunction with simplified process models or other stochastic techniques to derive computationally efficient statistics of large numbers of synthetic extreme weather events. These techniques have traditionally been popular for assessing regional tropical cyclone risk (e.g., Bloemendaal et al. 2020; Emanuel 2006; Lee et al. 2018; Lin et al. 2023; Xu et al. 2024) but have recently been applied to study other hazards such as midlatitude storms (Hall and Booth 2017; Li et al. 2026).

Another methodological challenge for storyline metric construction relates to the identification and characterization of impactful events within gridded climate model datasets. Many hazards can be understood as having a spatial extent and time duration over which some critical thresholds are crossed. Salient features might include the duration and extent of the event or some other metrics calculated over the period of time and space affected (e.g., cumulative storm precipitation or minimum nighttime temperature during a heat wave). A number of analytic methods for identification of such distinct events or "objects" within gridded datasets are becoming available, often by identifying a region in a space–time domain that exceeds a threshold, such as for heavy precipitation or temperatures yielding heat stress (Fisel et al. 2023, 2024; Ullrich et al. 2021; Stone et al. 2025). Such objects can provide event characteristics, including area, duration, intensity, frequency, and locations across a region. These impactful events can be the foundation for storylines, and the statistical distributions can become the basis for probabilistic storylines. Extending this approach to multiple event types can allow analysis of compounding behavior by exploring when the object types overlap, at least partially, so that they can be viewed as simultaneous events (Fisel et al. 2024; Rastogi et al. 2023). Comparing event characteristics with and without compounding would indicate the probability that they intensify each other. Observing the time evolution of each type of event leading to compounding can identify which tends to lead the other, which again can have a probabilistic foundation from the events' statistics.

b. Coordinated simulation protocols for bridging scales and uncertainties. To provide a complete understanding of storylines and their probabilities, ESM simulation protocols can be designed to recreate specific events or trajectories, to understand how key processes and metrics respond to forcing, and to generate datasets that provide insight into different dimensions of uncertainty. Due to the computational cost and specialized expertise required to implement these simulations, storyline analysts are often faced with a less than ideal "ensemble of opportunity" (Tebaldi and Knutti 2007). Here, we highlight three simulation types that are essential for understanding probabilistic storylines. Conducting all three consistently with a single ESM would provide a comprehensive understanding of storyline

probabilities within that ESM, and doing so across multiple ESMs would provide an ideal foundation for characterizing how model uncertainty shapes storyline probabilities.

1) Decadal-to-centennial-scale reanalysis: Reanalysis can be used to demonstrate that models are able to reproduce storyline-defining conditions and to reconcile differences in observed vs modeled storyline-defining metrics. Storyline work in advance of estimating storyline probabilities can explore the relevance of different storyline metrics and their causal roles as storyline drivers. To construct a reanalysis simulation, a model is configured to simulate a historical climatology that ideally overlaps with the storyline event(s) of choice (Zarzycki et al. 2024), although some studies have still leveraged reanalysis to provide insight into out-of-sample storylines (Risser and Wehner 2017). Over the entirety of the historical climatology, the model's prognostic variables (e.g., wind and moisture fields) are periodically nudged to conditions provided by an existing reanalysis dataset. The reanalysis simulation enables statistical characterization of the variables that shaped the storyline event by constraining historically observed atmospheric conditions within the model used, while simultaneously preserving structural uncertainty (e.g., subgrid-scale parameterizations and their influence on a storyline metric). By preserving structural uncertainties, the reanalysis simulation provides a means to identify storyline metrics in a historically constrained manner and allows those storyline metrics to be directly comparable with additional simulations discussed below.
2) Decadal-to-centennial-scale large ensembles: Large ensemble simulations consist of multiple realizations of a single-model experiment (Deser et al. 2012; Monier et al. 2015; Fasullo et al. 2024; Stevenson et al. 2023). Large ensembles can be used to identify how the statistical properties of storyline metrics might change under different climate scenarios (including past, future, or preindustrial climates). Most commonly, large ensemble simulations are designed according to specific scenarios. To circumvent issues of scenario uncertainty and "hot models," storyline metrics can be sampled within a large ensemble conditioned on GWLs, and changes in their return frequencies estimated accordingly. A methodological issue with constructing GWL-specific large ensembles is how to select and align ensemble members. For example, in a given member, low-frequency modes of variability can render a decade warmer or colder than a specific GWL, yet this realization might be consistent with the GWL within the range of internal variability. This could be solved by defining a fixed time period from a model's ensemble mean corresponding to the GWL. Additionally, these fixed-time segments from a large ensemble could be used to create a sea surface temperature climatology to then drive an even larger atmosphere-only model ensemble to more thoroughly sample extreme events conditional on a specific GWL's boundary conditions (Mitchell et al. 2017).
3) Hindcasts and futurecasts: The last simulation configuration uses the model in a hindcast and futurecast mode (Huprikar et al. 2024; Pettett and Zarzycki 2023; Rhoades et al. 2023, 2024) to first recreate the land–atmosphere conditions of the storyline and then reforecast that same storyline under different GWLs. Different from the reanalysis simulations, which need to match the resolution of the large ensemble simulations, hindcasts and futurecasts would identify necessary and sufficient resolutions and forecast lead times to skillfully recreate the event and validate if the configuration is fit for purpose for storyline analysis. The hindcast simulation is important in identifying structural uncertainty and understanding credibility of the model in representing the hazards that shaped the storyline event. Although finer horizontal resolution in models is not always a panacea for more skillful simulations, these simulations would likely require fine-resolution grid spacing, particularly in representing metrics that are either extreme or in regions of heterogeneity (e.g., mountains). The hindcast simulations also enable more direct

comparisons of the model-simulated variables of concern to the best available in situ observations and/or observationally constrained gridded products to contextualize model skill. Futurecast simulations can then be compared with hindcast simulations to identify how storyline metrics would respond to different GWLs and better understand physical processes that shape these responses.

c. Emulators to generate data for probabilistic analysis. With large ensembles, statistical characterization of internal variability has become more robust. Large ensembles, however, are tied to specific ESMs and scenario experiments and have high computational cost. Recent advances in the development of emulators of climate model output can enable the exploration of internal variability and therefore its statistical characterization by computationally cheaper means (Mahesh et al. 2025; Tebaldi et al. 2025). After being trained on available ESM experiments, emulators can generate novel realizations of internal variability according to a user-specified forcing profile. In fact, often emulators are driven by global average temperature, which makes their use particularly well suited in our GWL framing. Thus, emulators can either boost an existing ensemble or generate novel ones according to different forcings. Given the ability of emulators to generate tens of thousands of realizations, their output not only supports robust statistical characterization through methods like EVA but also can directly allow for the quantification of event probabilities through empirical frequency calculations (i.e., counting how many of the realizations show the extreme behavior of interest out of all the realizations), without having to rely on the asymptotic properties of EVA. The field of climate model emulation is extremely dynamic and fast evolving. Machine learning emulators aimed at replacing climate models and able to generate a large number of atmospheric variables at daily or even subdaily frequency are already available (Watt-Meyer et al. 2025), but also more targeted emulators able to generate a subset of them jointly have been developed, especially for the joint emulation of temperature and precipitation (Estrada et al. 2022; Mathison et al. 2025; Osborn et al. 2016). Not as developed are emulators for variables other than atmospheric, like ocean, land surface, or cryospheric variables. Given the pace of emulators' development, however, we expect these limits to be overcome in the near future. Given the importance of appropriate metric selection discussed above, it will be valuable to evaluate emulator performance for impact-relevant quantities.

d. Leveraging multimodel ensembles. In our framework, we have highlighted how data-driven statistical analysis (e.g., EVA) is most straightforwardly applied to the realizations of a metric of interest within one world, which we define by a GWL/model combination. The idiosyncratic nature of multimodel ensembles (Tebaldi and Knutti 2007) has defied a robust statistical treatment of their joint behavior, despite many published approaches. Redundancies and dependencies, common deficiencies, and spurious degrees of complexity are all found in the ensembles of ESMs gathered by projects like CMIP (Sanderson et al. 2015). Models share ancestry and components, common biases remain, and they vary widely in resolution, process representation, and treatment of individual forcings.

Despite all this, there is value in considering ensembles of different models. In fact, according to standard metric packages, like ESMValTool (Eyring et al. 2016), iLAMB (Collier et al. 2018), or CMAT (Fasullo 2020), for many diagnostic metrics, the multimodel mean outperforms individual models due to the cancellation of errors when output from different models is averaged. However, many of these metrics focus on mean state rather than model sensitivity, i.e., a model's ability to not only simulate a realistic present-day climate but also simulate a realistic response to external forcing changes. The lengthening observational record increasingly allows us to validate models also for long-term trends and identify model–observation discrepancies that cannot be explained by internal variability (Simpson et al. 2025; Shaw

et al. 2024). Large ensembles are particularly useful here, enabling a probabilistic model validation (e.g., Phillips et al. 2020; Zscheischler and Lehner 2022; Lehner 2024). Such diagnostic metrics represent and alert us to model structural uncertainties, which, depending on the metric, spatial scale, time horizon, and scenario of interest, may have a large role in the overall uncertainty characterization of climate projections (Hawkins and Sutton 2009; Lehner et al. 2020). However, it is notable that limited observational records, particularly for very rare extreme events, can complicate efforts to adequately characterize historical internal variability (Philip et al. 2022).

Some informed approaches to dealing with intermodel variation have sought to cull models because they do not perform acceptably over the historical period, as a necessary but not sufficient condition (e.g., Goldenson et al. 2023; Sobolowski et al. 2025). As for the unverifiable part of the models' output, future projections, researchers have explored ways to maximize the coverage, by a limited ensemble, of the dimensions of uncertainty deemed critical (Snyder et al. 2024). Or they have explored emergent constraints, where current behavior is related to future projections and constrained by observables, such that a subset of models can be identified as consistent with the latter (Hall et al. 2019; Lehner et al. 2019). Once that subset is identified, the resulting projections are still in most cases treated as a range, with little formal statistical modeling applied. But the maximization of the uncertainty representation by a—hopefully constrained—multimodel ensemble is critical input to any downstream analysis. It is possible that an emergent constraint analysis finds all models biased in the same way, which would require a bias correction rather than model weighting approach, which presents its own philosophical and methodological challenges (Spuler et al. 2026), making it difficult to actually constrain the range of future projections (e.g., Armour et al. 2024). Ideally, however, if the multimodel ensemble identified as appropriate for a given application could produce a set of individual large ensembles—for example, by deploying emulators trained on the individual ESMs (Tebaldi et al. 2025) or by boosting existing ensembles (Bloin-Wibe et al. 2025; Fischer et al. 2023)—the analysis chain would likely cover the main sources of variability and uncertainty.

5. Conclusions

The decision relevance of storylines in climate risk assessment would be greatly enhanced by incorporating probabilities but only if probabilities can be attached to storylines defensibly. Here, we propose a framework whereby different components of storyline uncertainty can be addressed separately using the methods best suited to each. When isolated from deeper model and emissions scenario uncertainty, internal variability can be well characterized using data-driven methods. For global modelers, this points to a need for more large ensembles of climate simulations constructed to systematically probe how the structure of internal variability changes across different models and levels of warming. Grappling with deeper uncertainties necessarily involves incomplete knowledge that cannot be resolved simply through additional data generation. For analysts ranging from statisticians to climate services professionals to planners, this calls for transparency regarding assumptions about conditionality and experimentation with DMDU methods that are designed to expose such uncertainties to deliberation or with Bayesian statistical approaches that allow for incorporation of subjective expert judgment. Identifying the right combination of analytic methods and metrics for defining storylines requires collaboration across professional and disciplinary boundaries. This points broadly to the value of coproduction and the critical role of boundary spanners for further refining the framework presented here and adapting it to specific storylines and decision contexts.

Acknowledgments. This study was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy Regional and Global Model Analysis and Multisector Dynamics Program Areas. Authors Jones, Longmate, Buddhavarapu, Gutowski, Jagannathan, Reed, Rhoades, Ullrich, Monier, and Zarzycki were funded by the "An Integrated Evaluation of the Simulated Hydroclimate System of the Continental U.S." (HyperFACETS) project (Award DE-SC0016605). Authors North, Rhoades, and Risser were funded by the "Calibrated and Systematic Characterization, Attribution and Detection of Extremes (CASCADE)" Science Focus Area (Award DE-AC02-05CH11231). Author Hadjimichael was funded by the "Integrated Multisector Multiscale Modeling (IM3)" Scientific Focus Area (Award DE-AC05-76RL01830). Author Tebaldi was funded by the "Water cycle: Modeling of Circulation, Convection, and Earth system Mechanisms (WACCEM)" Science Focus Area (Award DE-AC05-76RL01830). Author Lehner was funded by DoD ESTCP Award W912HQ23C0002 (Project NH22-7259). Work by P. A. Ullrich was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. We thank Cora Kingdon for helpful review of the manuscript during preparation.

Data availability statement. No datasets were generated or analyzed during the current study.

References

Agel, L., M. Barlow, S. B. Feldstein, and W. J. Gutowski Jr., 2018: Identification of large-scale meteorological patterns associated with extreme precipitation in the US northeast. *Climate Dyn.*, **50**, 1819–1839, https://doi.org/10.1007/s00382-017-3724-8.

Arguez, A., and R. S. Vose, 2011: The definition of the standard WMO climate normal: The key to deriving alternative climate normals. *Bull. Amer. Meteor. Soc.*, **92**, 699–704, https://doi.org/10.1175/2010BAMS2955.1.

Armour, K. C., and Coauthors, 2024: Sea-surface temperature pattern effects have slowed global warming and biased warming-based constraints on climate sensitivity. *Proc. Natl. Acad. Sci. USA*, **121**, e2312093121, https://doi.org/10.1073/pnas.2312093121.

Bamzai-Dodson, A., A. E. Cravens, A. A. Wade, and R. A. McPherson, 2021: Engaging with stakeholders to produce actionable science: A framework and guidance. *Wea. Climate Soc.*, **13**, 1027–1041, https://doi.org/10.1175/WCAS-D-21-0046.1.

Barlow, M., and Coauthors, 2019: North American extreme precipitation events and related large-scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **53**, 6835–6875, https://doi.org/10.1007/s00382-019-04958-z.

Bloemen, P. J. T. M., F. Hammer, M. J. van der Vlist, P. Grinwis, and J. van Alphen, 2019: DMDU into practice: Adaptive Delta Management in The Netherlands. *Decision Making under Deep Uncertainty: From Theory to Practice*, V. A. W. J. Marchau, Walker et al., Eds., Springer, 321–351, https://doi.org/10.1007/978-3-030-05252-2_14.

Bloemendaal, N., I. D. Haigh, H. de Moel, S. Muis, R. J. Haarsma, and J. C. J. H. Aerts, 2020: Generation of a global synthetic tropical cyclone hazard dataset using STORM. *Sci. Data*, **7**, 40, https://doi.org/10.1038/s41597-020-0381-2.

Bloin-Wibe, L., R. Noyelle, V. Humphrey, U. Beyerle, R. Knutti, and E. Fischer, 2025: Estimating return periods for extreme events in climate models through Ensemble Boosting. *Wea. Climate Dyn.*, **6**, 1147–1177, https://doi.org/10.5194/wcd-6-1147-2025.

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

Brown, C., Y. Ghile, M. Laverty, and K. Li, 2012: Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. *Water Resour. Res.*, **48**, W09537, https://doi.org/10.1029/2011WR011212.

California Department of Water Resources, 2015: Perspectives and guidance for climate change analysis. California Department of Water Resources Tech. Rep., 142 pp., https://water.ca.gov/-/media/DWR-Website/Web-Pages/Programs/All-Programs/Climate-Change-Program/Climate-Program-Activities/Files/Reports/Perspectives-Guidance-Climate-Change-Analysis.pdf.

——, 2023: Risk-informed future climate scenario development for the state water project delivery capability report. California Natural Resources Agency, 128 pp., https://data.cnra.ca.gov/dataset/finaldcr2023/resource/e41f531d-dace-4d37-b52e-35a6ddd2224e.

——, 2025: State Water Project Adaptation Strategy (DWR Climate Action Plan, Phase III). California Natural Resources Agency, 88 pp., https://mavensnotebook.com/wp-content/uploads/2025/08/SWP-AdaptationStrategy_Final.pdf.

Casey, J. A., M. Fukurai, D. Hernández, S. Balsari, and M. V. Kiang, 2020: Power outages and community health: A narrative review. *Curr. Environ. Health Rep.*, **7**, 371–383, https://doi.org/10.1007/s40572-020-00295-0.

Chan, W. C. H., N. W. Arnell, G. Darch, K. Facer-Childs, T. G. Shepherd, M. Tanguy, and K. Van Der Wiel, 2023: Current and future risk of unprecedented hydrological droughts in Great Britain. *J. Hydrol.*, **625**, 130074, https://doi.org/10.1016/j.jhydrol.2023.130074.

Chattopadhyay, A., E. Nabizadeh, and P. Hassanzadeh, 2020: Analog forecasting of extreme-causing weather patterns using deep learning. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001958, https://doi.org/10.1029/2019MS001958.

Collier, N., F. M. Hoffman, D. M. Lawrence, G. Keppel-Aleks, C. D. Koven, W. J. Riley, M. Mu, and J. T. Randerson, 2018: The International Land Model Benchmarking (ILAMB) system: Design, theory, and implementation. *J. Adv. Model. Earth Syst.*, **10**, 2731–2754, https://doi.org/10.1029/2018MS001354.

Deser, C., 2020: Certain uncertainty: The role of internal climate variability in projections of regional climate change and risk management. *Earth's Future*, **8**, e2020EF001854, https://doi.org/10.1029/2020EF001854.

——, A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, https://doi.org/10.1007/s00382-010-0977-x.

——, and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Climate Change*, **10**, 277–286, https://doi.org/10.1038/s41558-020-0731-2.

Dong, L., L. R. Leung, Y. Qian, Y. Zou, F. Song, and X. Chen, 2021: Meteorological environments associated with California wildfires and their potential roles in wildfire changes during 1984–2017. *J. Geophys. Res. Atmos.*, **126**, e2020JD033180, https://doi.org/10.1029/2020JD033180.

Emanuel, K., 2006: Climate and tropical cyclone activity: A new model downscaling approach. *J. Climate*, **19**, 4797–4802, https://doi.org/10.1175/JCLI3908.1.

——, 2017: Assessing the present and future probability of Hurricane Harvey's rainfall. *Proc. Natl. Acad. Sci. USA*, **114**, 12 681–12 684, https://doi.org/10.1073/pnas.1716222114.

Estrada, F., O. Calderón-Bustamante, W. Botzen, J. A. Velasco, and R. S. J. Tol, 2022: AIRCC-Clim: A user-friendly tool for generating regional probabilistic climate change scenarios and risk measures. *Environ. Modell. Software*, **157**, 105528, https://doi.org/10.1016/j.envsoft.2022.105528.

Eyring, V., and Coauthors, 2016: ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci. Model Dev.*, **9**, 1747–1802, https://doi.org/10.5194/gmd-9-1747-2016.

Faranda, D., G. Messori, and P. Yiou, 2020: Diagnosing concurrent drivers of weather extremes: Application to warm and cold days in North America. *Climate Dyn.*, **54**, 2187–2201, https://doi.org/10.1007/s00382-019-05106-3.

——, ——, A. Jezequel, M. Vrac, and P. Yiou, 2023: Atmospheric circulation compounds anthropogenic warming and impacts of climate extremes in Europe. *Proc. Natl. Acad. Sci. USA*, **120**, e2214525120, https://doi.org/10.1073/pnas.2214525120.

Fasullo, J. T., 2020: Evaluating simulated climate patterns from the CMIP archives using satellite and reanalysis datasets using the Climate Model Assessment Tool (CMATv1). *Geosci. Model Dev.*, **13**, 3627–3642, https://doi.org/10.5194/gmd-13-3627-2020.

——, and Coauthors, 2024: An overview of the E3SM version 2 large ensemble and comparison to other E3SM and CESM large ensembles. *Earth Syst. Dyn.*, **15**, 367–386, https://doi.org/10.5194/esd-15-367-2024.

Fischer, E. M., and Coauthors, 2023: Storylines for unprecedented heatwaves based on ensemble boosting. *Nat. Commun.*, **14**, 4643, https://doi.org/10.1038/s41467-023-40112-4.

Fisel, B. J., and Coauthors, 2023: Evaluating contemporary and future-scenario substantial-precipitation events in the Missouri River Basin using object-oriented analysis. *Climate*, **11**, 112, https://doi.org/10.3390/cli11050112.

——, S. J. Garbers, D. Haar, M. M. Zoerner, and W. J. Gutowski Jr., 2024: Object-oriented analysis as a foundation for building climate storylines of compounding short-term drought and crop heat stress. *Front. Climate*, **6**, 1357391, https://doi.org/10.3389/fclim.2024.1357391.

Fletcher, S., M. Lickley, and K. Strzepek, 2019: Learning about climate change uncertainty enables flexible water infrastructure planning. *Nat. Commun.*, **10**, 1782, https://doi.org/10.1038/s41467-019-09677-x.

Gao, X., C. A. Schlosser, P. Xie, E. Monier, and D. Entekhabi, 2014: An analogue approach to identify heavy precipitation events: Evaluation and application to CMIP5 climate models in the United States. *J. Climate*, **27**, 5941–5963, https://doi.org/10.1175/JCLI-D-13-00598.1.

——, ——, P. A. O'Gorman, E. Monier, and D. Entekhabi, 2017: Twenty-first-century changes in U.S. regional heavy precipitation frequency based on

resolved atmospheric patterns. *J. Climate*, **30**, 2501–2521, https://doi.org/10.1175/JCLI-D-16-0544.1.

Gateño, F., and Coauthors, 2024: Screening CMIP6 models for Chile based on past performance and code genealogy. *Climatic Change*, **177**, 87, https://doi.org/10.1007/s10584-024-03742-1.

Gillett, N. P., 2024: Halving of the uncertainty in projected warming over the past decade. *npj Climate Atmos. Sci.*, **7**, 146, https://doi.org/10.1038/s41612-024-00693-3.

Glisan, J. M., W. J. Gutowski Jr., J. J. Cassano, E. N. Cassano, and M. W. Seefeldt, 2016: Analysis of WRF extreme daily precipitation over Alaska using self-organizing maps. *J. Geophys. Res. Atmos.*, **121**, 7746–7761, https://doi.org/10.1002/2016JD024822.

Goldenson, N., G. Mauger, L. R. Leung, C. M. Bitz, and A. Rhines, 2018: Effects of ensemble configuration on estimates of regional climate uncertainties. *Geophys. Res. Lett.*, **45**, 926–934, https://doi.org/10.1002/2017GL076297.

——, and Coauthors, 2023: Use-inspired, process-oriented GCM selection: Prioritizing models for regional dynamical downscaling. *Bull. Amer. Meteor. Soc.*, **104**, E1619–E1629, https://doi.org/10.1175/BAMS-D-23-0100.1.

Goodrich, K. A., K. D. Sjostrom, C. Vaughan, L. Nichols, A. Bednarek, and M. C. Lemos, 2020: Who are boundary spanners and how can we support them in making knowledge more actionable in sustainability fields? *Curr. Opin. Environ. Sustainability*, **42**, 45–51, https://doi.org/10.1016/j.cosust.2020.01.001.

Gore, M. J., C. M. Zarzycki, and M. M. Gervais, 2023: Connecting large-scale meteorological patterns to extratropical cyclones in CMIP6 climate models using self-organizing maps. *Earth's Future*, **11**, e2022EF003211, https://doi.org/10.1029/2022EF003211.

Grotjahn, R., and Coauthors, 2016: North American extreme temperature events and related large scale meteorological patterns: A review of statistical methods, dynamics, modeling, and trends. *Climate Dyn.*, **46**, 1151–1184, https://doi.org/10.1007/s00382-015-2638-6.

Groves, D. G., J. R. Fischbach, E. Bloom, D. Knopman, and R. Keefe, 2013: *Adapting to a Changing Colorado River: Making Future Water Deliveries More Reliable Through Robust Management Strategies.* RAND Corporation, 70 pp.

Gutowski, W. J., Jr., S. S. Willis, J. C. Patton, B. R. J. Schwedler, R. W. Arritt, and E. S. Takle, 2008: Changes in extreme, cold-season synoptic precipitation events under global warming. *Geophys. Res. Lett.*, **35**, L20710, https://doi.org/10.1029/2008GL035516.

Hadjimichael, A., J. Quinn, E. Wilson, P. Reed, L. Basdekas, D. Yates, and M. Garrison, 2020: Defining robustness, vulnerabilities, and consequential scenarios for diverse stakeholder interests in institutionally complex river basins. *Earth's Future*, **8**, e2020EF001503, https://doi.org/10.1029/2020EF001503.

——, P. M. Reed, J. D. Quinn, C. R. Vernon, and T. Thurber, 2024: Scenario storyline discovery for planning in multi-actor human-natural systems confronting change. *Earth's Future*, **12**, e2023EF004252, https://doi.org/10.1029/2023EF004252.

Hall, A., P. Cox, C. Huntingford, and S. Klein, 2019: Progressing emergent constraints on future climate change. *Nat. Climate Change*, **9**, 269–278, https://doi.org/10.1038/s41558-019-0436-6.

Hall, T., and J. F. Booth, 2017: SynthETC: A statistical model for severe winter storm hazard on eastern North America. *J. Climate*, **30**, 5329–5343, https://doi.org/10.1175/JCLI-D-16-0711.1.

Hausfather, Z., 2025: An assessment of current policy scenarios over the 21st century and the reduced plausibility of high-emissions pathways. *Dialogues Climate Change*, **2**, 26–32, https://doi.org/10.1177/29768659241304854.

——, and G. P. Peters, 2020: Emissions—The 'business as usual' story is misleading. *Nature*, **577**, 618–620, https://doi.org/10.1038/d41586-020-00177-3.

——, K. Marvel, G. A. Schmidt, J. W. Nielsen-Gammon, and M. Zelinka, 2022: Climate simulations: Recognize the 'hot model' problem. *Nature*, **605**, 26–29, https://doi.org/10.1038/d41586-022-01192-2.

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

Helgeson, C., R. E. Nicholas, K. Keller, C. E. Forest, and N. Tuana, 2022: Attention to values helps shape convergence research. *Climatic Change*, **170**, 17, https://doi.org/10.1007/s10584-021-03274-y.

Hinkel, J., and Coauthors, 2019: Meeting user needs for sea level rise information: A decision analysis perspective. *Earth's Future*, **7**, 320–337, https://doi.org/10.1029/2018EF001071.

Huang, X., and D. L. Swain, 2022: Climate change is increasing the risk of a California megaflood. *Sci. Adv.*, **8**, eabq0995, https://doi.org/10.1126/sciadv.abq0995.

Huprikar, A. S., A. M. Stansfield, and K. A. Reed, 2024: A storyline analysis of Hurricane Irma's precipitation under various levels of climate warming. *Environ. Res. Lett.*, **19**, 014004, https://doi.org/10.1088/1748-9326/ad0c89.

Hydros Consulting Inc., 2018: Colorado River Risk Study: Phase I summary report. Colorado River District and Project Participants, 14 pp., https://www.waterinfo.org/files/039586998/West-Slope-BRT-Risk-Study-Phase-I-Report-FINAL-8-1-18.pdf.

Jacob, D., and Coauthors, 2025: Co-production of climate services: Challenges and enablers. *Front. Climate*, **7**, 1507759, https://doi.org/10.3389/fclim.2025.1507759.

Jagannathan, K., A. D. Jones, and I. Ray, 2021: The making of a metric: Co-producing decision-relevant climate science. *Bull. Amer. Meteor. Soc.*, **102**, E1579–E1590, https://doi.org/10.1175/BAMS-D-19-0296.1.

——, S. Buddhavarapu, P. A. Ullrich, A. D. Jones, and the HyperFACETS Project Team, 2023: Typologies of actionable climate information and its use. *Global Environ. Change*, **82**, 102732, https://doi.org/10.1016/j.gloenvcha.2023.102732.

Jézéquel, A., and Coauthors, 2024: Broadening the scope of anthropogenic influence in extreme event attribution. *Environ. Res.*, **3**, 042003, https://doi.org/10.1088/2752-5295/ad7527.

Jones, A. D., W. D. Collins, and M. S. Torn, 2013: On the additivity of radiative forcing between land use change and greenhouse gases. *Geophys. Res. Lett.*, **40**, 4036–4041, https://doi.org/10.1002/grl.50754.

Kahneman, D., 2011: *Thinking, Fast and Slow.* Farrar, Straus and Giroux, 499 pp.

Kandlikar, M., J. Risbey, and S. Dessai, 2005: Representing and communicating deep uncertainty in climate-change assessments. *C. R. Géosci.*, **337**, 443–455, https://doi.org/10.1016/j.crte.2004.10.010.

Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, **40**, 1194–1199, https://doi.org/10.1002/grl.50256.

Lafferty, D. C., and R. L. Sriver, 2023: Downscaling and bias-correction contribute considerable uncertainty to local climate projections in CMIP6. *npj Climate Atmos. Sci.*, **6**, 158, https://doi.org/10.1038/s41612-023-00486-0.

Lau, N.-C., and M. J. Nath, 2012: A model study of heat waves over North America: Meteorological aspects and projections for the twenty-first century. *J. Climate*, **25**, 4761–4784, https://doi.org/10.1175/JCLI-D-11-00575.1.

Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2018: An environmentally forced tropical cyclone hazard model. *J. Adv. Model. Earth Syst.*, **10**, 223–241, https://doi.org/10.1002/2017MS001186.

Lehner, F., 2024: Climate model large ensembles as test beds for applied compound event research. *iScience*, **27**, 111113, https://doi.org/10.1016/j.isci.2024.111113.

——, and C. Deser, 2023: Origin, importance, and predictive limits of internal climate variability. *Environ. Res.*, **2**, 023001, https://doi.org/10.1088/2752-5295/accf30.

——, A. W. Wood, J. A. Vano, D. M. Lawrence, M. P. Clark, and J. S. Mankin, 2019: The potential to reduce uncertainty in regional runoff projections from climate models. *Nat. Climate Change*, **9**, 926–933, https://doi.org/10.1038/s41558-019-0639-x.

——, C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins, 2020: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.*, **11**, 491–508, https://doi.org/10.5194/esd-11-491-2020.

——, E. Hawkins, R. Sutton, A. G. Pendergrass, and F. C. Moore, 2023: New potential to reduce uncertainty in regional climate projections by combining

physical and socio-economic constraints. *AGU Adv.*, **4**, e2023AV000887, https://doi.org/10.1029/2023AV000887.

Lemos, M. C., and Coauthors, 2025: Scaling up actionable climate knowledge. *Proc. Natl. Acad. Sci. USA*, **122**, e2515771122, https://doi.org/10.1073/pnas.2515771122.

Lempert, R. J., S. W. Popper, and S. C. Bankes, 2003: *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis.* RAND Corporation, 187 pp., https://www.rand.org/pubs/monograph_reports/MR1626.html.

Li, D.-Y., K. A. Reed, S. J. Camargo, C.-Y. Lee, A. H. Sobel, C. M. Zarzycki, and M. Bieli, 2026: A statistical model for post-tropical cyclone hazard assessment. *J. Geophys. Res. Atmos.*, **131**, e2025JD044936, https://doi.org/10.1029/2025JD044936.

Lin, J., R. Rousseau-Rizzi, C.-Y. Lee, and A. Sobel, 2023: An open-source, physics-based, tropical cyclone downscaling model with intensity-dependent steering. *J. Adv. Model. Earth Syst.*, **15**, e2023MS003686, https://doi.org/10.1029/2023MS003686.

Liu, P. R., and A. E. Raftery, 2021: Country-based rate of emissions reductions should increase by 80% beyond nationally determined contributions to meet the 2 °C target. *Commun. Earth Environ.*, **2**, 29, https://doi.org/10.1038/s43247-021-00097-8.

Maher, N., S. Milinski, and R. Ludwig, 2021: Large ensemble climate model simulations: Introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth Syst. Dyn.*, **12**, 401–418, https://doi.org/10.5194/esd-12-401-2021.

Mahesh, A., and Coauthors, 2025: Huge ensembles—Part 1: Design of ensemble weather forecasts using spherical Fourier neural operators. *Geosci. Model Dev.*, **18**, 5575–5603, https://doi.org/10.5194/gmd-18-5575-2025.

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

Marchau, V. A. W. J., W. E. Walker, P. J. T. M. Bloemen, and S. W. Popper, Eds., 2019: *Decision Making under Deep Uncertainty: From Theory to Practice.* Springer, 21 pp., https://doi.org/10.1007/978-3-030-05252-2.

Mathison, C., and Coauthors, 2025: A rapid-application emissions-to-impacts tool for scenario assessment: Probabilistic Regional Impacts from Model Patterns and Emissions (PRIME). *Geosci. Model Dev.*, **18**, 1785–1808, https://doi.org/10.5194/gmd-18-1785-2025.

McCrary, R. R., S. McGinnis, and L. O. Mearns, 2017: Evaluation of snow water equivalent in NARCCAP simulations, including measures of observational uncertainty. *J. Hydrometeor.*, **18**, 2425–2452, https://doi.org/10.1175/JHM-D-16-0264.1.

Merrifield, A. L., L. Brunner, R. Lorenz, V. Humphrey, and R. Knutti, 2023: Climate model Selection by Independence, Performance, and Spread (ClimSIPS v1.0.1) for regional applications. *Geosci. Model Dev.*, **16**, 4715–4747, https://doi.org/10.5194/gmd-16-4715-2023.

Mitchell, D., and Coauthors, 2017: Half a degree additional warming, prognosis and projected impacts (HAPPI): Background and experimental design. *Geosci. Model Dev.*, **10**, 571–583, https://doi.org/10.5194/gmd-10-571-2017.

Moallemi, E. A., J. Kwakkel, F. J. De Haan, and B. A. Bryan, 2020: Exploratory modeling for analyzing coupled human-natural systems under uncertainty. *Global Environ. Change*, **65**, 102186, https://doi.org/10.1016/j.gloenvcha.2020.102186.

——, and Coauthors, 2023: Knowledge co-production for decision-making in human-natural systems under uncertainty. *Global Environ. Change*, **82**, 102727, https://doi.org/10.1016/j.gloenvcha.2023.102727.

Monier, E., X. Gao, J. R. Scott, A. P. Sokolov, and C. A. Schlosser, 2015: A framework for modeling uncertainty in regional climate change. *Climatic Change*, **131**, 51–66, https://doi.org/10.1007/s10584-014-1112-5.

O'Neill, B. C., and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9**, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016.

Osborn, T. J., C. J. Wallace, I. C. Harris, and T. M. Melvin, 2016: Pattern scaling using ClimGen: Monthly-resolution future climate scenarios including changes in the variability of precipitation. *Climatic Change*, **134**, 353–369, https://doi.org/10.1007/s10584-015-1509-9.

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

Palmintier, B. S., and M. D. Webster, 2016: Impact of operational flexibility on electricity generation planning with renewable and carbon targets. *IEEE Trans. Sustainable Energy*, **7**, 672–684, https://doi.org/10.1109/TSTE.2015.2498640.

Parker, W. S., and G. Lusk, 2019: Incorporating user values into climate services. *Bull. Amer. Meteor. Soc.*, **100**, 1643–1650, https://doi.org/10.1175/BAMS-D-17-0325.1.

Peters, J. M., and R. S. Schumacher, 2014: Objective categorization of heavy-rain-producing MCS synoptic types by rotated principal component analysis. *Mon. Wea. Rev.*, **142**, 1716–1737, https://doi.org/10.1175/MWR-D-13-00295.1.

Pettett, A., and C. M. Zarzycki, 2023: The 1996 mid-Atlantic winter flood: Exploring climate risk through a storyline approach. *J. Hydrometeor.*, **24**, 2259–2280, https://doi.org/10.1175/JHM-D-22-0146.1.

Philip, S., and Coauthors, 2020: A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Climate Meteor. Oceanogr.*, **6**, 177–203, https://doi.org/10.5194/ascmo-6-177-2020.

Philip, S. Y., and Coauthors, 2022: Rapid attribution analysis of the extraordinary heat wave on the Pacific coast of the US and Canada in June 2021. *Earth Syst. Dyn.*, **13**, 1689–1713, https://doi.org/10.5194/esd-13-1689-2022.

Phillips, A. S., C. Deser, J. Fasullo, D. P. Schneider, and I. R. Simpson, 2020: Assessing climate variability and change in model large ensembles: A user's guide to the "climate variability diagnostics package for large ensembles". Climate Analysis Section, NCAR, 53 pp., https://opensky.ucar.edu/islandora/object/manuscripts:1001.

Pulkkinen, K., and Coauthors, 2022: The value of values in climate science. *Nat. Climate Change*, **12**, 4–6, https://doi.org/10.1038/s41558-021-01238-9.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, https://doi.org/10.1175/MWR2906.1.

——, A. Zimmer, D. M. W. Frierson, R. Startz, and P. Liu, 2017: Less than 2 °C warming by 2100 unlikely. *Nat. Climate Change*, **7**, 637–641, https://doi.org/10.1038/nclimate3352.

Rastogi, D., J. Trok, N. Depsky, E. Monier, and A. Jones, 2023: Historical evaluation and future projections of compound heatwave and drought extremes over the conterminous United States in CMIP6. *Environ. Res. Lett.*, **19**, 014039, https://doi.org/10.1088/1748-9326/ad0efe.

Reed, K. A., and M. F. Wehner, 2023: Real-time attribution of the influence of climate change on extreme weather events: A storyline case study of Hurricane Ian rainfall. *Environ. Res.*, **2**, 043001, https://doi.org/10.1088/2752-5295/acfd4e.

——, and Coauthors, 2022: Metrics as tools for bridging climate science and applications. *Wiley. Interdiscip. Rev.: Climate Change*, **13**, e799, https://doi.org/10.1002/wcc.799.

Reeder, T., and N. Ranger, 2011: How do you adapt in an uncertain world? Lessons from the Thames Estuary 2100 project. World Resources Institute Rep., 16 pp., http://wriorg.s3.amazonaws.com/s3fs-public/uploads/wrr_reeder_and_ranger_uncertainty.pdf.

Rhoades, A. M., A. D. Jones, and P. A. Ullrich, 2018: Assessing mountains as natural reservoirs with a multimetric framework. *Earth's Future*, **6**, 1221–1241, https://doi.org/10.1002/2017EF000789.

——, and Coauthors, 2023: Recreating the California New Year's flood event of 1997 in a regionally refined Earth system model. *J. Adv. Model. Earth Syst.*, **15**, e2023MS003793, https://doi.org/10.1029/2023MS003793.

——, and Coauthors, 2024: Anticipating how rain-on-snow events will change through the 21st century: Lessons from the 1997 New Year's flood event. *Climate Dyn.*, **62**, 8615–8637, https://doi.org/10.1007/s00382-024-07351-7.

Risser, M. D., and M. F. Wehner, 2017: Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. *Geophys. Res. Lett.*, **44**, 12 457–12 464, https://doi.org/10.1002/2017GL075888.

——, S. Rahimi, N. Goldenson, A. Hall, Z. J. Lebo, and D. R. Feldman, 2024: Is bias correction in dynamical downscaling defensible? *Geophys. Res. Lett.*, **51**, e2023GL105979, https://doi.org/10.1029/2023GL105979.

Roush, T., 2024: What is an ARkStorm? Here's why California officials have prepared for a 'megastorm' scenario. *Forbes*, 2 February, https://www.forbes.com/sites/tylerroush/2024/02/02/what-is-an-arkstorm-heres-why-california-officials-have-prepared-for-a-megastorm-scenario/.

Safford, H. D., S. C. Sawyer, S. D. Kocher, J. K. Hiers, and M. Cross, 2017: Linking knowledge to action: The role of boundary spanners in translating ecology. *Front. Ecol. Environ.*, **15**, 560–568, https://doi.org/10.1002/fee.1731.

Sanderson, B. M., and M. F. Wehner, 2017: Model weighting strategy. *Climate Science Special Report: Fourth National Climate Assessment*, U.S. Global Change Research Program, 436–442.

——, R. Knutti, and P. Caldwell, 2015: A representative democracy to reduce interdependency in a multimodel ensemble. *J. Climate*, **28**, 5171–5194, https://doi.org/10.1175/JCLI-D-14-00362.1.

——, M. Wehner, and R. Knutti, 2017: Skill and independence weighting for multi-model assessments. *Geosci. Model Dev.*, **10**, 2379–2395, https://doi.org/10.5194/gmd-10-2379-2017.

Schleussner, C.-F., and Coauthors, 2024: Overconfidence in climate overshoot. *Nature*, **634**, 366–373, https://doi.org/10.1038/s41586-024-08020-9.

Schoof, J. T., S. C. Pryor, and T. W. Ford, 2019: Projected changes in United States regional extreme heat days derived from bivariate quantile mapping of CMIP5 simulations. *J. Geophys. Res. Atmos.*, **124**, 5214–5232, https://doi.org/10.1029/2018JD029599.

Selby, L. B., and Coauthors, 2024: A toolkit for coproducing actionable science to support public land management. U.S. Department of the Interior, Bureau of Land Management, https://www.blm.gov/noc/report/toolkit-coproducing-actionable-science-support-public-land-management.

Seneviratne, S. I., and Coauthors, 2021: Weather and climate extreme events in a changing climate. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 1513–1766, https://doi.org/10.1017/9781009157896.013.

Shaw, T. A., and Coauthors, 2024: Regional climate change: Consensus, discrepancies, and ways forward. *Front. Climate*, **6**, 1391634, https://doi.org/10.3389/fclim.2024.1391634.

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

Sherwood, S. C., and Coauthors, 2020: An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.*, **58**, e2019RG000678, https://doi.org/10.1029/2019RG000678.

Shiogama, H., and Coauthors, 2023: Important distinctiveness of SSP3–7.0 for use in impact assessments. *Nat. Climate Change*, **13**, 1276–1278, https://doi.org/10.1038/s41558-023-01883-2.

Simpson, I. R., and Coauthors, 2025: Confronting Earth System Model trends with observations. *Sci. Adv.*, **11**, eadt8035, https://doi.org/10.1126/sciadv.adt8035.

Smith, R., E. Zagona, J. Kasprzyk, N. Bonham, E. Alexander, A. Butler, J. Prairie, and C. Jerla, 2022: Decision science can help address the challenges of long-term planning in the Colorado River Basin. *J. Amer. Water Resour. Assoc.*, **58**, 735–745, https://doi.org/10.1111/1752-1688.12985.

Snyder, A., N. Prime, C. Tebaldi, and K. Dorheim, 2024: Uncertainty-informed selection of CMIP6 Earth system model subsets for use in multisectoral and impact models. *Earth Syst. Dyn.*, **15**, 1301–1318, https://doi.org/10.5194/esd-15-1301-2024.

Sobolowski, S., and Coauthors, 2025: GCM selection and ensemble design: Best practices and recommendations from the EURO-CORDEX community. *Bull. Amer. Meteor. Soc.*, **106**, E1834–E1850, https://doi.org/10.1175/BAMS-D-23-0189.1.

Song, F., Z. Feng, L. R. Leung, R. A. Houze Jr., J. Wang, J. Hardin, and C. R. Homeyer, 2019: Contrasting spring and summer large-scale environments associated with mesoscale convective systems over the U.S. Great Plains. *J. Climate*, **32**, 6749–6767, https://doi.org/10.1175/JCLI-D-18-0839.1.

Spuler, F. R., J. B. Wessel, J. Jebeile, J. Zscheischler, and T. G. Shepherd, 2026: Bias adjustment and the question of usable climate information: Methodological assumptions and value judgments. *Bull. Amer. Meteor. Soc.*, **107**, E79–E102, https://doi.org/10.1175/BAMS-D-25-0022.1.

Srivastava, A. K., R. Grotjahn, A. M. Rhoades, and P. A. Ullrich, 2025: Large-scale statistically meaningful patterns (LSMPs) associated with precipitation extremes over Northern California. *J. Geophys. Res. Atmos.*, **130**, e2024JD041307, https://doi.org/10.1029/2024JD041307.

Stevenson, S., X. Huang, Y. Zhao, E. Di Lorenzo, M. Newman, L. van Roekel, T. Xu, and A. Capotondi, 2023: Ensemble spread behavior in coupled climate models: Insights from the Energy Exascale Earth System Model version 1 large ensemble. *J. Adv. Model. Earth Syst.*, **15**, e2023MS003653, https://doi.org/10.1029/2023MS003653.

Stone, J., M. Gervais, K. A. Bowley, and C. Zarzycki, 2025: Identifying, tracking, and evaluating mechanisms of North American cold air outbreaks (CAOs) using a feature tracking approach. *Mon. Wea. Rev.*, **153**, 153–168, https://doi.org/10.1175/MWR-D-23-0265.1.

Sukhdeo, R., P. A. Ullrich, and R. Grotjahn, 2022: Assessing the large-scale drivers of precipitation in the northeastern United States via linear orthogonal decomposition. *Climate Dyn.*, **59**, 3657–3681, https://doi.org/10.1007/s00382-022-06289-y.

Swain, D. L., B. Langenbrunner, J. D. Neelin, and A. Hall, 2018: Increasing precipitation volatility in twenty-first-century California. *Nat. Climate Change*, **8**, 427–433, https://doi.org/10.1038/s41558-018-0140-y.

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

——, and Coauthors, 2021: Extreme sea levels at different global warming levels. *Nat. Climate Change*, **11**, 746–751, https://doi.org/10.1038/s41558-021-01127-1.

——, N. E. Selin, R. Ferrari, and G. Flierl, 2025: Emulators of climate model output. *Annu. Rev. Environ. Resour.*, **50**, 709–737, https://doi.org/10.1146/annurev-environ-012125-085838.

Terray, L., 2021: A dynamical adjustment perspective on extreme event attribution. *Wea. Climate Dyn.*, **2**, 971–989, https://doi.org/10.5194/wcd-2-971-2021.

Tong, S., A. Subramanyam, and V. Rao, 2022: Optimization under rare chance constraints. *SIAM J. Optim.*, **32**, 930–958, https://doi.org/10.1137/20M1382490.

Ullrich, P. A., Z. Xu, A. M. Rhoades, M. D. Dettinger, J. F. Mount, A. D. Jones, and P. Vahmani, 2018: California's drought of the future: A midcentury recreation of the exceptional conditions of 2012–2017. *Earth's Future*, **6**, 1568–1587, https://doi.org/10.1029/2018EF001007.

——, C. M. Zarzycki, E. E. McClenny, M. C. Pinheiro, A. M. Stansfield, and K. A. Reed, 2021: TempestExtremes v2.1: A community framework for feature detection, tracking, and analysis in large datasets. *Geosci. Model Dev.*, **14**, 5023–5048, https://doi.org/10.5194/gmd-14-5023-2021.

Undorf, S., K. Pulkkinen, P. Wikman-Svahn, and F. A.-M. Bender, 2022: How do value-judgements enter model-based assessments of climate sensitivity? *Climatic Change*, **174**, 19, https://doi.org/10.1007/s10584-022-03435-7.

van Oldenborgh, G. J., and Coauthors, 2017: Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environ. Res. Lett.*, **12**, 124009, https://doi.org/10.1088/1748-9326/aa9ef2.

Van Vuuren, D., and Coauthors, 2026: The Scenario Model Intercomparison Project for CMIP7 (ScenarioMIP-CMIP7). *Geosci. Model Dev.*, **19**, 2627–2656, https://doi.org/10.5194/gmd-19-2627-2026.

Vincent, K., E. Archer, R. Henriksson, J. Pardoe, and N. Mittal, 2020: Reflections on a key component of co-producing climate services: Defining climate metrics from user needs. *Climate Serv.*, **20**, 100204, https://doi.org/10.1016/j.cliser.2020.100204.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. Krayer von Krauss, 2003: Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integr. Assess.*, **4**, 5–17, https://doi.org/10.1076/iaij.4.1.5.16466.

——, V. A. W. J. Marchau, and J. H. Kwakkel, 2013: Uncertainty in the framework of policy analysis. *Public Policy Analysis: New Developments*, W. A. H. Thissen and W. E. Walker, Eds., Springer, 215–261, https://doi.org/10.1007/978-1-4614-4602-6_9.

Watt-Meyer, O., and Coauthors, 2025: ACE2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *npj Climate Atmos. Sci.*, **8**, 205, https://doi.org/10.1038/s41612-025-01090-0.

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

Xu, W., K. Balaguru, D. R. Judi, J. Rice, L. R. Leung, and S. Lipari, 2024: A North Atlantic synthetic tropical cyclone track, intensity, and rainfall dataset. *Sci. Data*, **11**, 130, https://doi.org/10.1038/s41597-024-02952-7.

Xu, Y., A. Jones, and A. Rhoades, 2019: A quantitative method to decompose SWE differences between regional climate models and reanalysis datasets. *Sci. Rep.*, **9**, 16520, https://doi.org/10.1038/s41598-019-52880-5.

Yiou, P., 2014: AnaWEGE: A weather generator based on analogues of atmospheric circulation. *Geosci. Model Dev.*, **7**, 531–543, https://doi.org/10.5194/gmd-7-531-2014.

——, R. Vautard, Y. Robin, N. de Noblet-Ducoudré, F. d'Andrea, and R. Noyelle, 2024: How could 50 °C be reached in Paris: Analyzing the CMIP6 ensemble to design storylines for adaptation. *Climate Serv.*, **36**, 100518, https://doi.org/10.1016/j.cliser.2024.100518.

Zarzycki, C. M., B. D. Ascher, A. M. Rhoades, and R. R. McCrary, 2024: Algorithmically detected rain-on-snow flood events in different climate datasets: A case study of the Susquehanna River Basin. *Nat. Hazards Earth Syst. Sci.*, **24**, 3315–3335, https://doi.org/10.5194/nhess-24-3315-2024.

Zscheischler, J., and F. Lehner, 2022: Attributing compound events to anthropogenic climate change. *Bull. Amer. Meteor. Soc.*, **103**, E936–E953, https://doi.org/10.1175/BAMS-D-21-0116.1.