Uncertainty analysis in multi-sector systems: Considerations for risk analysis, projection, and planning for complex systems

Vivek Srikrishnan[1], David C. Lafferty[2], Tony E. Wong[3], Jonathan R. Lamontagne[4], Julianne D. Quinn[5], Sanjib Sharma[6], Nusrat J. Molla[7], Jonathan D. Herman[7], Ryan L. Sriver[2], Jennifer F. Morris[9], Ben Seiyon Lee[8]

[1]Department of Biological & Environmental Engineering, Cornell University, Ithaca, NY, USA
[2]Department of Atmospheric Sciences, University of Illinois, Urbana-Champaign, IL, USA
[3]School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY, USA
[4]Department of Civil & Environmental Engineering, Tufts University, Medford, MA, USA
[5]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA
[7]Department of Civil & Environmental Engineering, University of California, Davis, CA, USA
[6]Earth & Environmental Systems Institute, Pennsylvania State University, University Park, PA, USA
[8]MIT Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, MA, USA
[9]Department of Statistics, George Mason University, Fairfax, VA, USA

Key Points:

- Uncertainty is an inherent part of multi-sector systems analysis;
- Approaches to addressing uncertainty involve deliberate tradeoffs;
- Best practices involve standardizing communication and improving transparency

Corresponding author: Vivek Srikrishnan, `viveks@cornell.edu`

Abstract

Simulation models of multi-sector systems are increasingly used to understand societal resilience to climate and economic shocks and change. However, multi-sector systems are also subject to numerous uncertainties that prevent the direct application of simulation models for prediction and planning, particularly when extrapolating past behavior to a nonstationary future. Recent studies have developed a combination of methods to characterize, attribute, and quantify these uncertainties for both single- and multi-sector systems. Here we review challenges and complications to the idealized goal of fully quantifying all uncertainties in a multi-sector model and their interactions with policy design as they emerge at different stages of analysis: (1) inference and model calibration; (2) projecting future outcomes; and (3) scenario discovery and identification of risk regimes. We also identify potential methods and research opportunities to help navigate the tradeoffs inherent in uncertainty analyses for complex systems. During this discussion, we provide a classification of uncertainty types and discuss model coupling frameworks to support interdisciplinary collaboration on multi-sector dynamics (MSD) research. Finally, we conclude with recommendations for best practices to ensure that MSD research can be properly contextualized with respect to the underlying uncertainties.

1 Introduction

Simulation models of multi-sector systems are increasingly used to understand societal resilience to climate and economic shocks and long-term change. To faithfully represent societal systems across spatiotemporal scales, such multi-sector system representations need to account for dynamic and endogenous interactions between sectors, rather than treating other sectors as exogenous boundary conditions and forcings. This approach is at the heart of the emerging field of MultiSector Dynamics (MSD). However, this growing complexity increases the number and types of uncertainties that affect both the inverse problem (calibration and inference) as well as the forward projection of system dynamics and resilience into the future, which is critical for decision support. This paper identifies and reviews the key challenges involved in uncertainty analysis for MSD. We discuss why they arise (or are made more acute) in the multi-sectoral modeling context, the current state of the art, and what research opportunities may help address them going forward.

Our focus is on quantitative aspects of multi-sectoral modeling. However it is important to note that there are also many semi- and non-quantitative aspects of multi-sector modeling and risk analysis. These considerations, which are critical in the development of the conceptual model of the system (S. Robinson et al., 2015) and the use of uncertainty analysis to inform policy and governance of complex, multi-sector systems in the face of systemic risk (Renn et al., 2020; Hochrainer-Stigler et al., 2020).

We begin with definitions of several key terms:

- *Sector*: a complex system-of-systems that delivers services, amenities, and products critical to a subdivision of society. Components of sectors may include infrastructure, environmental systems, governing institutions (public and private), labor force capacity, finance, and a range of actors (*e.g.*, firms, regulatory agencies, investors, consumers) involved in producing and consuming services and products (Reed et al., 2022);
- *Multi-sector system*: a set of interacting sectors that yield emergent dynamics beyond that which could be predicted from each sector alone (Reed et al., 2022);
- *Uncertainty*: "a departure from the (unachievable) ideal of complete determinism" (Walker et al., 2003) in any aspect of the system.

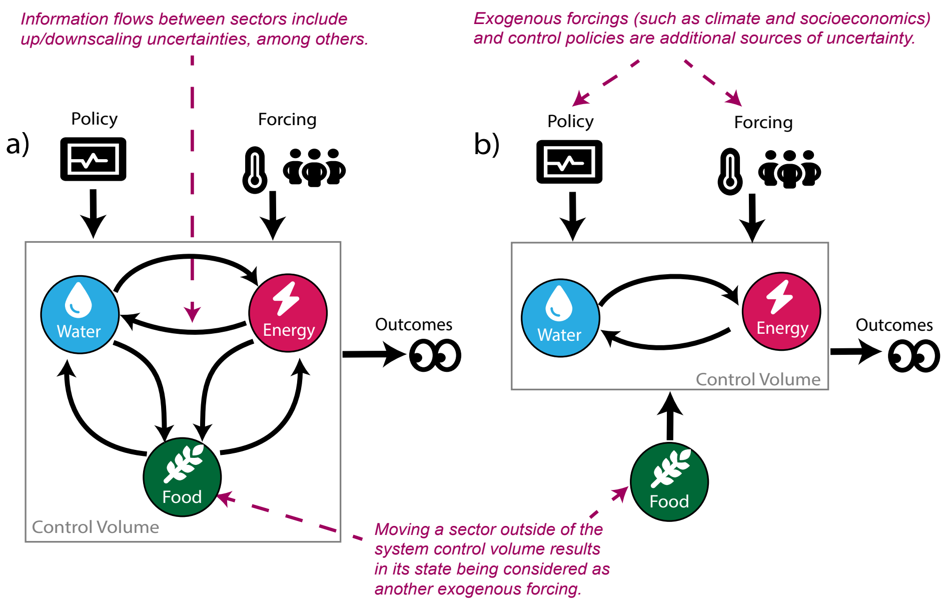

Figure 1. Schematic of a multi-sector system model. Two conceptual examples of how a food-energy-water system can be represented by coupled models of each sector. In panel a), the control volume includes all three sectors, allowing feedbacks between the food system and the water and energy system(s) that are not possible when the food system is outside of the control volume (panel b)) and is therefore treated as exogenous. The endogenous dynamics within the control volume can be further influenced by exogenous forcings, such as socioeconomic and climate inputs, and policies, which determine how sectors respond to changes in the internal system state and external forcings.

These definitions highlight the fact that each sector alone is a complex system of agents, institutions, and infrastructure interacting with the natural environment, and each other. A useful notion is the idea of the *control volume* of an analysis, which is a concept borrowed from thermodynamics. We use "control volume" to refer to the portion of the analyzed system(s) whose dynamics are modeled endogenously, as contrasted with any exogenous inputs and the model outputs. The shift to studying a multi-sector system-of-systems adds complexity by expanding the control volume under analysis to encompass feedbacks between systems, potentially across different characteristic spatial and temporal scales, and across different resolutions of the system (*e.g.*, individual agents vs. aggregations). These dynamics are represented in Figure 1, which is a schematic of a coupled multi-sector system-of-systems. Many of the challenges that we review in this paper are present in the single-sector case, but are amplified in the multi-sector setting.

One of the main strategic goals of MSD research is the identification and analysis of key uncertainties influencing the evolution of a particular system-of-systems. These analyses are often conducted using simulation models, which are a set of coupled numerical equations and/or agent-based rules describing the time evolution of the system state(s), given inputs of forcing variables that are external to the system. In general, multi-sector system models are subject to several sources of uncertainty, as illustrated in Figure 1.

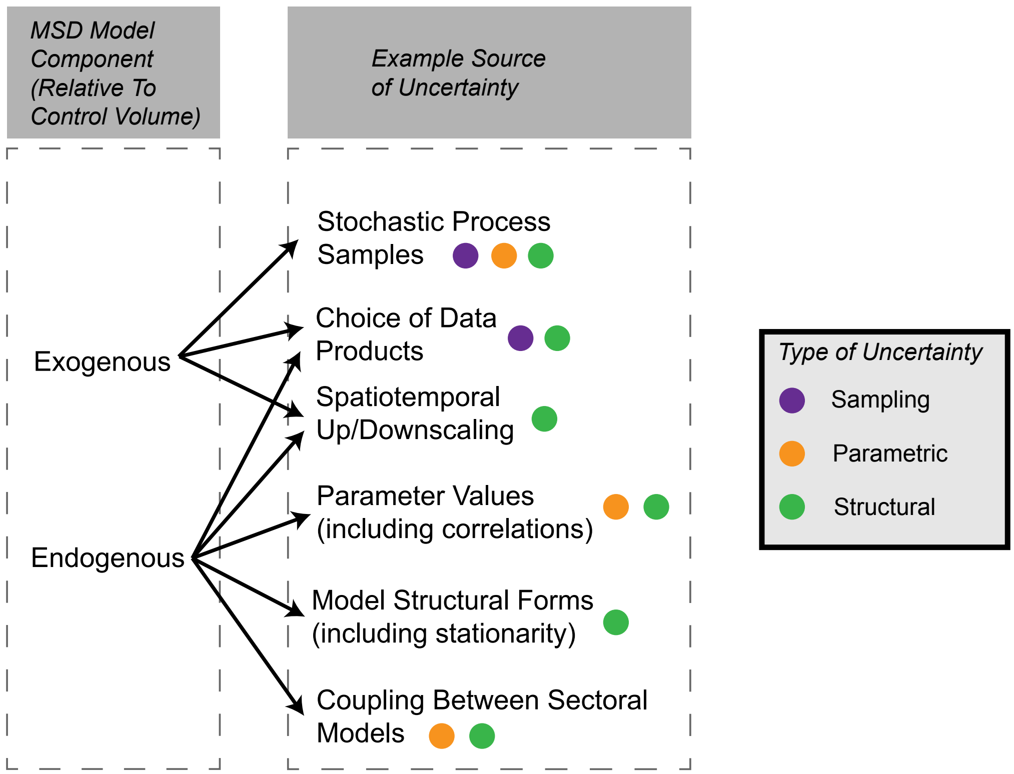

Figure 2. Example sources of uncertainty that are relevant to model components that are exogenous and/or endogenous to the control volume under analysis. Sampling, parametric, and structural uncertainties can enter the system through both exogenous and endogenous components of the system. This demonstrates some of the many ways in which MSD researchers can make choices which affect how uncertainty influences their analyses. Many of these specific examples are further discussed in the subsequent sections of this paper. The colored circles relate the uncertainty sources to the uncertainty classification in Table 1.

These can stem from exogenous or endogenous model components, as shown in Figure 2.

Exogenous model components can be classified as either forcings or policy inputs. By forcings, we refer to inputs which serve as boundary conditions, representing characteristics of the external environment that are relevant to one sector and/or the coupled system within the control volume. Many variables could either be an exogenous forcing or an endogenous component of the modeled system, depending on the boundaries of the control volume. For example, global temperature would be considered a forcing if it is input into the modeled system as an exogenous factor. However, if instead temperature is generated within the control volume, it would be considered an endogenous component of the multi-sector coupling.

By policies, we mean rules which dictate actions taken by humans or institutions. Such policies influence how the sector or coupled system responds to changes in the internal state or external environment. Analogously to forcings, while policy rules can change endogenously in response to system dynamics, we focus here on policies (or meta-policies) which are supplied exogenously. Figure 2 illustrates some of the uncertainties which can influence these exogenous components.

Uncertainties related to endogenous system dynamics include model structure and parameters describing each sector as well as those describing interactions between sectors. In combination. All of these uncertainties interact and propagate to influence the modeled outcomes of interest, which could include error metrics (if observations are available for calibration) and/or performance metrics in the case of planning for future scenarios.

Accurate representation of state changes and variability within multi-sector systems requires careful consideration of interactions and feedbacks within the coupled systems. Coupling multiple sectors in a unified modeling framework creates two broad challenges that will be recurring themes throughout this paper: 1) scaling, and 2) the complexity-uncertainty tradeoff. First, the relevant scales at which each sector is modeled may not align with each other, or with influential climate and weather conditions. This creates a need for upscaling or downscaling to adequately model responses and the feedbacks between sectors, which introduces additional uncertainty beyond the dynamics alone.

As shown in Figure 1, the choice of control volume for an analysis is critical for establishing what sectors will be treated endogenously, and therefore what feedbacks, influences, and interactions are possible. We note that a broader control volume (with more endogenous sectors) is not necessarily "better", as the introduction of additional linkages and dynamics may make it correspondingly more difficult to analyze and trace the uncertainties which are most relevant to the research question.

Second, with a fixed computational budget, there is a tradeoff between the computational complexity of a model and the number of feasible model evaluations (Helgeson et al., 2021). Accounting for endogenous interactions within and between multiple sectors adds computational and parametric complexity. This can result in a more accurate representation of observed dynamics when appropriately calibrated, but can also result in unrealistic behavior when extrapolating beyond the data used for calibration due to overfitting. Added complexity only improves the representation of uncertainties if the primary contributors to those uncertainties were missing mechanisms in the original model (Figure 3).

When newly added model mechanisms include missing components which help explain variability in outcomes, added model complexity can decrease uncertainty despite the addition of new parameters and equations (the blue scenario in Figure 3). For example, the addition of equations allowing the Antarctic Ice Sheet to rapidly disintegrate in response to increased warming reduces uncertainty in ice sheet volume hindcasts (Wong et al., 2017). This effect is the result of other unrelated parameters no longer compensating for the missing structural dynamics. However, the inclusion of additional model complexity can increase uncertainty if additional parameters which were not related to the underlying sources of variability need to be calibrated (the green scenario in Figure 3). If these two outcomes are mixed, so that some missing mechanisms are included, but the net effect on uncertainty is dampened by additional calibration needs, the result will be a more moderate reduction or increase in total uncertainty (the orange scenario in Figure 3).

If this is not the case — for example, if future forcing scenarios dominate the total uncertainty in the outcomes — increased complexity in model representation may be a detriment to understanding the range of potential system dynamics, as the computational cost will limit the ability to evaluate an ensemble of scenarios. In other words, finer scales and/or increased complexity do not improve model performance if there are key processes missing that control variability within the coupled system. Increasing model complexity may also result in negative learning or poor inferences if inadequacies in model structure persist or are poorly constrained by observations (Draper, 1995; Oppenheimer et al., 2008; Small & Fischbeck, 1999). Therefore, it is critical to analyze the sources of

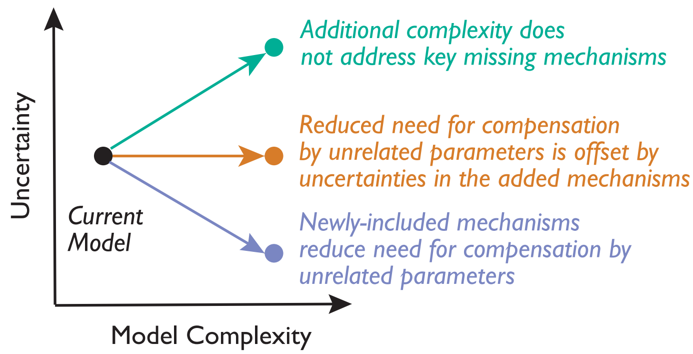

Figure 3. How adding model complexity can change model uncertainty. If the additional complexity causes key missing mechanisms to be included (blue), overall uncertainty can be reduced, as parameter distributions have less need to compensate for the missing components. If the additional complexity does not include representations of the missing mechanisms which were related to system uncertainties, overall uncertainty can increase due to the inclusion of parameters and interactions which need to be calibrated (green). In some cases, these two effects may partially cancel each other out (orange), leading to more moderate or no net changes to the total uncertainty.

uncertainty in multi-sector models to ensure that any additional complexity is appropriately targeted.

As a consequence of these challenges, studies of single- and multi-sector systems have developed several approaches to analyzing and representing uncertainty:

- *Uncertainty Characterization (UC)*: Mapping how alternative representations of the stressors and form and function of modeled systems influence outcomes of interest (Moallemi, Kwakkel, de Haan, & Bryan, 2020; Walker et al., 2003);
- *Uncertainty Quantification (UQ)*: "the full specification of likelihoods as well as distributional forms necessary to infer the joint probabilistic response across all modeled factors of interest" (Cooke, 1991);
- *Sensitivity Analysis (SA)*: The study of how uncertainty in the output of a model (numerical or otherwise) is influenced by different sources of uncertainty in the model input (adapted from Saltelli et al. (2004)).

The goals of these methods are multifaceted: (a) to improve the accuracy of the models by identifying missing components; (b) to improve understanding of system dynamics, risks, and vulnerabilities; and (c) to design policies or infrastructure. These approaches to uncertainty analyses are not mutually exclusive, and are often combined. For example, initial studies of uncertainty characterization and system sensitivity may conclude with a formal quantification of uncertainties related to a specific decision problem (*e.g.* Shortridge & Zaitchik, 2018; Taner et al., 2019). In general, sensitivity analysis may be employed for either UC or UQ, depending on the mathematical description of the inputs and outputs. Uncertainty characterization approaches such as exploratory modeling may be more appropriate than UQ in situations where well-defined probability distributions over the sets of possible outcomes do not exist or cannot be agreed upon, a situation known as deep or Knightian uncertainty (Knight, 1921; Langlois & Cosgel, 1993;

Lempert, 2002). These steps may also be iterative and not always sequential. Specific methods for SA and UC are reviewed in detail by Pianosi et al. (2016) and Moallemi, Kwakkel, de Haan, and Bryan (2020), respectively.

Given the breadth of applications of uncertainty analysis in multi-sector systems modeling, we focus this review on key challenges related to the chain of uncertainty propagation throughout a multi-sector system. In Section 2, we discuss how choices made in the MSD modeling process exchange model and computational complexity for the ability to capture feedbacks and other dynamics, with implications for uncertainty representations. In Section 3, we discuss uncertainties in inference and calibration of multi-sector models, which can be both structural and parametric in nature. In Section 4, we discuss uncertainty in forward projections of multi-sector dynamics. In Section 5, we discuss how the increase in parametric and structural complexity associated with multi-sector analyses can result in high-dimensional outcomes that are difficult to attribute to particular sources of uncertainty, complicating the identification of scenarios of interest for further analysis or communication. Finally, we conclude by identifying some recommended best practices and cross-cutting research targets of opportunity which can help navigate some of these analytic trade-offs and complexities.

2 Types of Uncertainty and Model Coupling Regimes

In discussing the three key challenges we review in this paper, it is important to define the lexicon we will be using. MSD research is inherently interdisciplinary, and different communities of researchers focusing on different sectors often have different vocabularies, which is a fundamental challenge for interdisciplinary research teams (Bracken & Oughton, 2006; Henson et al., 2020; J. J. Cohen et al., 2021). MSD research, however, necessarily involves coupling and integration of simulation models and research outputs that may reflect differing disciplinary norms about the treatment of uncertainties. In this section, we classify key types of uncertainty and model coupling structures to help interdisciplinary teams communicate their research plans and outcomes.

2.1 Overview of MSD-Relevant Uncertainties

Simulation models are subject to several different types of uncertainty. From the perspective of multi-sector system analyses, we classify these uncertainties into three categories:

- *Structural uncertainty*: uncertainty in the mathematical and/or rule-based representation of processes within a simulation model;
- *Parametric uncertainty*: uncertainty in the numerical values of internal parameters representing endogenous model processes, given a fixed model structure; and
- *Sampling uncertainty*: uncertainty arising from the finite sampling of a stochastic process (including coverage of an output space).

Table 1 provides a brief overview of these types of uncertainty, along with examples. Parametric and structural uncertainties can be aleatory (stemming from irreducible randomness) or epistemic (stemming from a lack of knowledge about the "truth"), while sampling uncertainty typically reflects aleatory uncertainty (O'Hagan, 2004). One way to distinguish sampling uncertainty from parametric and structural uncertainty is that while sampling uncertainty relates to sampling from a stochastic process (represented exogenously or endogenously), parametric and structural uncertainties refer to uncertainty in how a simulation model responds to changes in external inputs, policies, and boundary conditions. For example, one might consider uncertainties related to model-data residuals (Brynjarsdóttir & O'Hagan, 2014) to be structural when those discrepancies are the result of choices or ignorance related to the representation of system com-

Uncertainty Type	**Associated Uncertainties**	**Examples**	**Sample Method of Exploration**
Structural	Model inadequacy, (epistemic) residual uncertainty	Choices of which physical processes to include and the equations used to represent them	Multi-model ensembles, multi-physics ensembles
Parametric	Parameter uncertainty	Choice of parameter vector between alternatives producing similar results, strength of coupling between models	Perturbed-physics ensembles, posterior predictive samples
Sampling	Natural variability, (aleatory) residual uncertainty, observation error	Sample realizations from a fixed stochastic process, internal variability, uncertain boundary conditions or forcings	Initial conditions ensembles, forcing scenarios

Table 1. Categories of uncertainty relevant for multi-sector models, including associated uncertainties from the taxonomy in Kennedy and O'Hagan (2001) and examples.

ponents (in this case, they would represent epistemic uncertainties). Alternatively, these model-data residual uncertainties could be considered sampling uncertainty when they represent particular realizations of "true" underlying stochastic processes (hence they would represent aleatory uncertainties).

Structural uncertainty can be defined as the consideration or inclusion/exclusion of one or more relevant structural variants. This could include different sectoral model representations, different policy or decision rules, or different choices of data products and statistical representations for exogenous forcings or model calibration (Bojke et al., 2006). Another consideration is the alignment (or lack thereof) of modeling paradigms, or formalisms, across sub-components of the system (Davis & Tolk, 2007). Decisions about how to couple models with different formalisms (*e.g.*, co-simulation, translation into a common formalism, or construction of a super-formalism) adds another level of structural uncertainty (Vangheluwe et al., 2002).

The line between structural and parametric uncertainties can be blurry. For example, a parametrized regression model with at least one zero coefficient is the same as a simpler regression model with the relevant variable omitted. Whether this should be classified as a case of structural or parametric uncertainty is highly contextual, and dependent on the broader analysis, *i.e.* is there a more formal variable selection procedure, or is zero included as one element of the possible coefficient values?

In many cases, the same conceptual uncertainties can be classified differently according to this taxonomy depending on the control volume of a particular analysis. Figure 2 shows example uncertainties which are relevant for exogenous and/or endogenous system components and how they might be classified. While there is a large amount of overlap, the nature of how uncertainties are represented can differ. For example, the Representative Concentration Pathways-Shared Socioeconomic Pathways (RCP-SSP) scenarios of future global change (O'Neill et al., 2016; Gidden et al., 2019) can be treated as a representation of sampling uncertainty when used as exogenous inputs or boundary conditions for a model of future climate or socioeconomic change. However, these same scenarios also reflect parametric and structural differences that may be relevant for a model or model component with feedbacks to global emissions or economic growth. Thus, it is important for MSD analyses to be transparent about not only which uncertainties they are treating, but how those uncertainties are represented in the context of the analytic control volume.

2.2 Coupling Frameworks and Control Volumes

Structural uncertainty is an essential feature of any modeling exercise, as all modelers necessarily make choices about what system dynamics will be modeled endogenously and at what resolution(s). Multi-sector modeling activities also necessarily involve coupling representations of multiple systems together, as in Figure 1. Model coupling can take a number of forms, even for a fixed system-of-systems, as illustrated in Figure 4. These choices have impacts on uncertainty propagation and analysis. We provide a brief overview of the types of coupling regimes and their implications for the resulting analyses.

An essential modeling decision is the selection of the control volume through the choice of endogenously- and exogenously-represented system components. Model structures with a greater share of exogenous components are typically less computationally expensive than those that feature more endogenous dynamics (assuming similar spatiotemporal resolutions). However, this comes at the expense of being able to analyze the feedbacks and interdependencies between subsystems, such as uncertainties and hypotheses related to the strength and patterns of influence of one sector on another. Whether this is acceptable depends on the research question and control volume. For example, many climate impact studies consist of one or more sectoral models forced by a climate model ensemble to produce a set of outcomes of interest (Grogan et al., 2020; Piontek et al., 2014; van Vliet et al., 2016). This choice might be reasonable if there is no clear pathway for the system contained within the control volume to dynamically influence greenhouse gas emissions trajectories.

Another critical structural distinction involving coupled models is whether a given coupling is *unidirectional* or *multidirectional*. Unidirectional coupling involves chaining models together in series, with no feedbacks between the modeled subsystems. The resulting wiring diagram (the directed model graph) is acyclic. Conversely, multidirectional coupling allows two model components to interact with each other, creating the possibility for feedbacks. Models involving multidirectionally-coupled components can have richer dynamics, but have an increased number of uncertain parameters due to the additional couplings. The potentially nonlinear dynamics introduced by the multidirectional couplings can also complicate analyses of uncertainty propagation. To date, most examples of coupled multidirectional frameworks come from the multi-sector Integrated Assessment Models (IAMs), rather than from the coupling of independently-developed sectoral models. Examples of coupled multidirectional modeling frameworks include Yoon et al. (2021), Mosnier et al. (2014), and Walsh et al. (2019).

For a concrete example of the implications of the choice of model coupling regime and control volume design, consider the coupled agricultural-hydrological system depicted in Figure 4. In the unidirectional example (Figure 4a), local hydrology, crop production, and crop prices are modeled endogenously, allowing farmer income to reflect the coupled hydro-agricultural-economic dynamics (some analyses that use similar modeling frameworks include Davies et al. (2013), Ma et al. (2016), and Stevanović et al. (2016)). There are already large uncertainties in this unidirectional case: for example, how to construct the impact of production shocks on prices (Nelson et al., 2014). However, additional in-

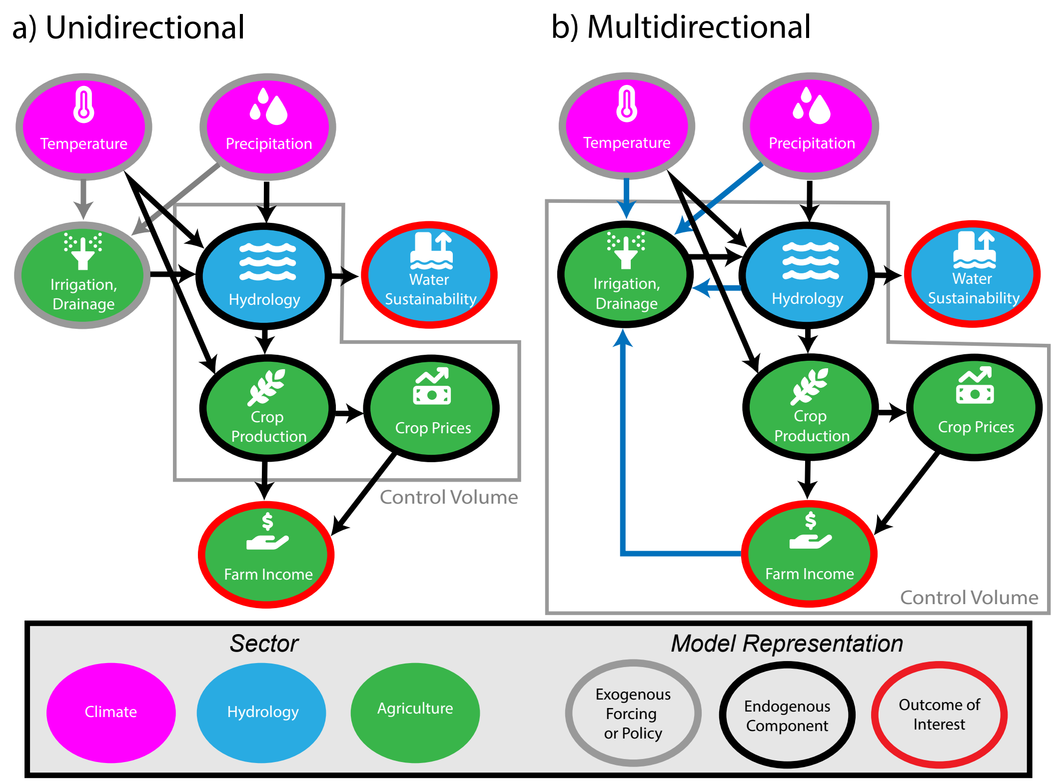

Figure 4. Simplified examples of possible model coupling configurations for a linked agriculture-water system. Panel a) represents a unidirectional coupling scheme between local hydrology, crop production, and crop prices, prohibiting the presence of feedback loops. Decisions about irrigation are supplied exogenously, either as forcings or policy rules (and may or may not be correlated or influenced by the climate forcings; those optional connections are shown with gray arrows). In panel b), the control volume has been expanded to include irrigation decisions, which allows for a multi-directional coupling scheme and feedbacks between the hydrological and economic systems and irrigation choices. Additional model couplings and dependencies in the multi-directional case represented in panel b) are represented by blue arrows. While the switch from unidirectional to multi-directional coupling makes it possible to represent richer and more realistic dynamics, the additional complexity may create computational and/or conceptual challenges for uncertainty analysis.

come from the joint agricultural-economic system is not allowed to directly feed back and induce changes in irrigation and drainage infrastructure. As a result, these influences on the local hydrology must be treated exogenously. In Figure 4b, which features multidirectional feedbacks through the introduction of a cycle, farm income is allowed to be invested into expanded irrigation and drainage, allowing farmers to alter the local hydrology to their benefit (with potential consequences for the broader hydrological system). This allows the analysis to more accurately capture the influence of agricultural decision-making and economic dynamics on the hydrological system and future production, but at the expense of additional data requirements and model complexity, since the relationships between farm incomes, investment decisions, irrigation operations, and local hydrology needs to be parameterized and (ideally) calibrated (Holtz & Pahl-Wostl, 2012).

Although Figure 4 portrays only a simple and stylized example, it nonetheless illustrates many of the important implications of the (linked) choices regarding control volume design and coupling regime for model complexity, the associated data and computational requirements, and how the results of the analysis can be interpreted with respect to relevant uncertainties. MSD investigators should hence make these choices as transparent as possible when reporting results, including by presenting a wiring diagram illustrating the coupled model structure.

One further consideration when coupling models of different sectors is that their characteristic scales may differ with respect to space and/or time. This can require up- and/or downscaling model structures and forcings to adequately model the dynamics within and across sectors. Coupling models with different spatiotemporal scales introduces new uncertainties in how the output of one model is translated to another, which should be accounted for in model calibration. We discuss implications of scales as they relate to forcings in Section 4.2, as this is a key issue when making forward projections, though some of these considerations may also be relevant for calibration.

3 Uncertainty in Model Calibration and Inference

The first step in uncertainty analysis is to determine the space over which the analysis will be conducted (including input and subsystem model structures and/or parameter values), as well as ranges or distributions for the parameters which are treated as uncertain. We refer to the selection of model parameters and structures to maximize the fidelity of the system model to observational data given model and computational constraints as *calibration* (Oreskes et al., 1994). Model calibration methods can span a range of techniques from hand-tuning model parameters until the output looks "right" to fully probabilistic approaches (Helgeson et al., 2021). With sufficient data, the uncertainty in these inputs can be estimated through statistical calibration. When calibration is conducted using statistical methods, it can be considered a backward estimation of uncertainty (Kennedy & O'Hagan, 2001). While calibration aims to approximate observations of the modeled system with model output, statistical inference focuses on obtaining estimates, probabilistic or summary, of the system parameters to learn about their values. Statistical calibration and inference are closely related, but have different (if complementary) goals.

Not all MSD analyses will require model calibration. For example, certain UC and SA studies may focus on understanding how a particular model structure responds to varying inputs over ranges or samples, rather than trying to select among model structures or infer probabilities. However, whether we are engaged in UC, UQ, or SA, we necessarily make some assumptions about parameter ranges and distributional forms (particularly in the case of UQ). These assumptions have implications for which variables we find to be most influential on the outputs and which decision alternatives we find to be most robust to that uncertainty (Quinn et al., 2020; McPhail et al., 2020; Reis & Shortridge, 2022). Moreover, a model calibrated to match observations with respect to one output may not sufficiently capture the dynamics of another (Efstratiadis & Koutsoyiannis, 2010). This is unsurprising given the choices made in the modeling process, but highlights the fact that "model calibration" is not a single method: different calibrations and calibration approaches are needed for different research questions.

As such, many questions surround how to best infer uncertainties through calibration, even in single-sector systems. These choices, whether they involve the selection of input data, the choice of model structures, or whether to calibrate system components independently or jointly, must be made with the goals of the research in mind, as they involve tradeoffs from the perspective of uncertainty analysis. We briefly discuss these challenges to uncertainty analysis here, consider how they are magnified in multi-sector systems, and discuss open research questions for how they should best be addressed. Answering these questions will be a critical first step before estimating how these uncertainties propagate forward to influence outcomes in multi-sector systems.

3.1 Exogenous Uncertainties

Model-based projections of outcomes in multi-sector systems require forcing multi-sector models with exogenous variables. These are often climate variables, such as precipitation and temperature, but may represent the output of other linked processes and systems, depending on the specified control volume of the analysis. How these inputs are modeled has implications for the resulting projections and output analysis. Ignoring uncertainty in the marginal and joint distributions of these forcing variables can bias projected system outcomes. This raises questions about 1) how to identify the structure and parameters defining the joint distribution of system inputs given limited data and 2) whether data from the past that must be used for this estimation will be representative of the future. In this section, we discuss how backwards uncertainty analyses can help address these questions and how choices in data sets and modeling can influence subsequent results.

3.1.1 Observational Data

Observational climate data plays an important role in model calibration. Several model parameters typically need to be calibrated by relying on historical data of climatological variables, which may take the form of (interpolated) station data or reanalysis products (Auffhammer et al., 2020), or streamflow observations (Kiang et al., 2018). There are observational uncertainties associated with the measurements underlying each of these, as well as parametric and structural uncertainties in any data assimilation procedure that might be used (Zumwald et al., 2020). In some cases, different choices of observational datasets can lead to significantly different estimates of endogenous model parameters (Parkes et al., 2019), although such uncertainties are typically neglected during model construction and parameter calibration. It may be difficult to know *a priori* whether observational uncertainties are important relative to endogenous and/or forcing uncertainties, and solutions such as explicitly modeling measurement errors (Schennach, 2016) or using a "dataset ensemble" (Zumwald et al., 2020)] may be computationally expensive.

Observations of socioeconomic data are subject to uncertainties which are unique to the specific product, and these observational uncertainties could be accounted for in the calibration process via the probability model (see the discussion of likelihood function specification in Section 3.2.1).

3.1.2 Statistical Modeling of Correlated Events

Within the risk analysis literature for individual sectors, challenges in answering these questions have been acknowledged and researched (Stedinger et al., 1993). How-

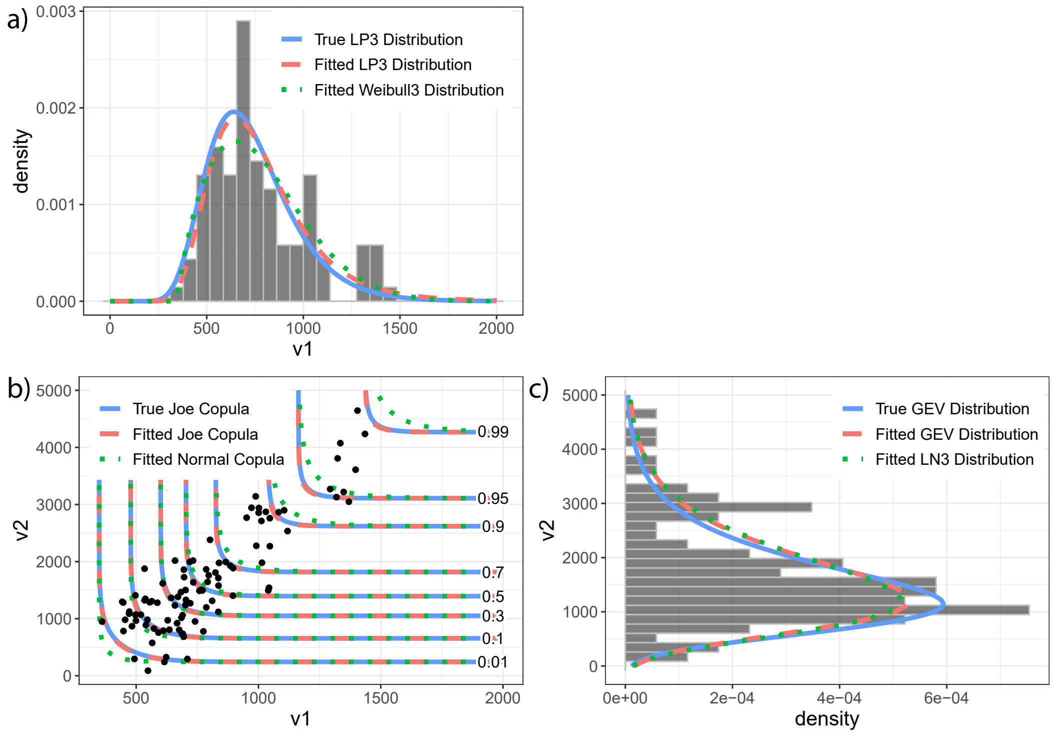

Figure 5. Example of parametric and structural uncertainty in estimating the joint distribution of exogenous variables. (a) Uncertainty in estimating the marginal distribution of variable $v1$. (b) Uncertainty in estimating the copula describing the joint distribution of variables $v1$ and $v2$. (c) Uncertainty in estimating the marginal distribution of variable $v2$.

ever, explicitly modeling the linked dynamics of multi-scale, multi-sector systems may reveal additional vulnerabilities due to the interactions between sectors and correlations across spatiotemporal domains (Su et al., 2020; Dolan et al., 2021). This emphasizes the importance of accounting for joint extremes and compound events in multi-sector risk analyses.

Figure 5 shows a stylized example of this challenge: estimating the structure and parameters defining the joint distribution of two exogenous variables. In this example, synthetic observations of these two variables were generated from the joint (Figure 5b) and marginal (Figure 5a,c) distributions shown by solid blue lines in the figure. In the real world, we do not know these underlying distributions but have to estimate them from observed or modeled data. If we correctly assume the structure of these distributions and simply estimate their parameters through statistical approaches such as maximum likelihood estimation, we might estimate that the data came from the dashed pink distributions. If we incorrectly assume the structure, we might estimate they were generated from the dotted green distributions.

All of these fits rely on point estimates of the parameters of each distribution. The implications of errors in these point estimates are most prominent in the tails of the distribution, where impacts are generally greatest, and data is most limited, resulting in the greatest uncertainty in estimation. In the example in Figure 5, both fitted distributions have fatter upper tails than the true distribution, which could lead to overestimation of the frequency of extreme events, and hence alter the resulting risk analysis. For example, if these variables represented rainfall volumes and peak storm surge or drought intensity and duration, we might overestimate the impacts of floods on coastal infrastructure or droughts on agricultural production. These errors could influence decision-making processes, resulting in overinvestment in stormwater infrastructure or irrigation reservoirs. Underinvestment is similarly likely if we underestimate the occurrence of these joint extremes. Alternative parameter estimators may result in a higher, equal, or lower probability of underdesign than overdesign. If one is more risk averse, a Bayesian prior can initiate the parameter estimates such that the probability of underdesign is less likely (Stedinger, 1983).

Risks of under-design can be compounded when considering joint drivers. The most common approach to fitting joint distributions of stochastic variables is through copulas (Nelsen, 2007), which model the dependence between variables in quantile-space. First, marginal distributions are fit to the individual variables and then the observations are transformed into quantiles of these distributions through inversion, where their dependency is modeled. There are many families of copulas that can capture this dependency, some of which exhibit tail dependency, meaning the variables are more highly correlated in the tails (upper, lower or both) than in the middle of the distribution (Schmidt, 2005). Fitting a copula that does not exhibit tail dependency when the observations do can lead to underestimation of the probability of joint extremes (Poulin et al., 2007). This occurs in Figure 5 when assuming the two variables come from a normal copula, which does not exhibit tail dependence, as opposed to the true Joe copula (Joe, 1993), which exhibits upper tail dependence, meaning high values of $v1$ are more highly correlated with high values of $v2$ than in the middle of the distribution.

The consequences of errors in marginal distribution estimation have been well-documented in the literature on single-sector systems, most predominantly with respect to floods (Wong et al., 2018). The negative consequences of incorrectly estimating the joint distribution of exogenous variables, particularly in the tails, or worse, assuming independence, have recently been raised in the literature with respect to coastal flooding (Moftakhari et al., 2017), agricultural production (Haqiqi et al., 2021), and wildfires (Brown et al., 2021), among others.

These consequences can be mitigated by not only only using point estimates of the most likely distribution parameters, but accounting for parametric uncertainty, such as through sampling from frequentist confidence intervals or Bayesian credible intervals (Sadegh et al., 2017, 2018). Bayesian approaches have the advantage of explicitly encoding prior knowledge about parameter values as prior distributions, which can be updated using Bayes' Theorem with information from data to obtain posterior distributions (P. M. Lee, 1989). This allows researchers to be transparent about these assumptions, which facilitates exploration of alternative hypotheses and sensitivities. Generating realizations from the distributions parameterized by multiple posterior samples results in draws from the posterior predictive distribution, which combines parametric and sampling uncertainty.

Bayesian estimation approaches can be applied to capture structural uncertainty as well through Bayesian model averaging (Madigan et al., 1996; Hoeting et al., 1999). However, depending on the complexity of the statistical and process models, propagating samples of exogenous variables through a multi-sector model to quantify output uncertainty can become computationally challenging or intractable if those samples are generated from the posterior distributions of multiple model structures. Another option is the use of principled model selection techniques, which we discuss further in Section 3.2 — the key point is that each approach to model selection reflects different modeling and epistemic goals, and care should be taken to align the selection criteria with the goals of the analysis.

3.1.3 Nonstationarity in Exogenous Processes

The example illustrated in Figure 5 assumes the stochastic process being estimated is stationary, meaning its distribution does not change over time (Koutsoyiannis & Montanari, 2015). For many exogenous variables, this may not be true, particularly in the context of climate change. For example, we are confident that increasing global carbon emissions have resulted in nonstationary temperature time series, but are more uncertain on how this has impacted precipitation and other climate variables (Arias et al., 2021). Assuming these other climate variables are stationary when they are not could exacerbate over or under-estimation errors, particularly in the tails (Milly et al., 2008; Wong et al., 2018). However, modeling them as nonstationary introduces greater uncertainty in the structure of that nonstationarity, as well as uncertainty in the parameters of that structure. For example, a modeler must determine which variables are non-stationary, what covariates influence those non-stationary variables, and the form of that dependency, *e.g.*, linear, log-linear, quadratic, or some other functional form (Grinsted et al., 2013; Wong et al., 2018; Wong, 2018). Time is a common choice of covariate, but loses ties to physical processes (Koutsoyiannis & Montanari, 2015), however conditioning on other covariates requires projecting how that variable will change in the future as well. With limited data to constrain the additional parameter estimates required to model these dependencies, particularly in the tails of concern, uncertainty can balloon to levels uninformative for decision-making (Serinaldi & Kilsby, 2015). Thus, modeling these processes as stationary vs. nonstationary is often a tradeoff between bias and variance (Ceres et al., 2017), and the decision about which to favor should depend on the consequences of each type of error (Rosner et al., 2014), which may differ across sectors.

Another issue is that the models used for projections may operate on scales that are misaligned with decision processes. Returning to the temperature and precipitation example, flood managers and urban planners are often concerned with daily, local-scale projections which climate models are not designed to generate. Statistical bias correcting and downscaling based on historical observations generally ignores the physical process reasons why projections misrepresent history, and so may propagate unjustifiable physical distortion into the future (Steinschneider et al., 2015). An alternative is stochastic weather generation (Steinschneider et al., 2019), wherein small scale weather realizations are simulated through a stochastic model that ties weather conditions to observable weather regimes (Robertson et al., 2015) that are better represented by climate models (Johnson & Sharma, 2009; Farnham et al., 2018). Thus temperature and precipitation realizations can be obtained at decision-relevant scales, leveraging climate models' strengths, conditional on deeply uncertain emissions trajectories. The advantage of such an approach is the ability to produce large samples of future climate or weather conditions. Indeed, such exploratory methods can be useful for multi-sector planning studies in order to identify critical uncertainties and design adaptive monitoring systems (Quinn et al., 2020). In the broader MSD context, analogous approaches hold promise where model and decision scales are misaligned.

3.2 Endogenous Uncertainties

In addition to quantifying uncertainty in the exogenous forcing to our models, it is crucial to consider uncertainty in the relationships between model components themselves, both within sectors and between sectors. While individual systems, considered in isolation, may primarily face risk from extreme, tail-area events, the nonlinear dynamics associated with coupled systems-of-systems could result in more moderate stressors simultaneously affecting multiple parts of the system. An illustrative example is the impact of Winter Storm Uri on the Texas infrastructure system in February 2021. While the severity of the triggering cold snap had precedent (Doss-Gollin et al., 2021), its impact on the natural gas and electric power systems was disproportionate due to the tight coupling between these systems and socioeconomic stresses such as increased heating demand (Busby et al., 2021).

3.2.1 Addressing Uncertainty in Model Parameters and Structures

Uncertainty in endogenous model components can be both parametric and structural. Conceptually, it is not always easy to untangle these two different types of uncertainties. Within single-sector models, it is well-known that multiple combinations of parameters and structures can produce dynamics similar to observations. From a Bayesian perspective, this reflects a posterior distribution over the space of joint structural and parametric combinations which does not have a unique maximum. In the hydrological literature, this non-uniqueness is typically referred to as equifinality (Beven, 2006). In such cases, Bayesian methods that explicitly estimate the posterior probability of different parameter combinations are recommended over single-objective calibration approaches that provide parameter point estimates that minimize an objective function, such as the sum of squared errors between observed and modeled output variables (Vrugt et al., 2008). Uncertainty estimates from bootstrap replications (Efron & Tibshirani, 1986; Efron, 2014) are a reasonable alternative to Bayesian methods, though care should be taken to account for dependence and potential non-stationarities.

Additional uncertainties come from the choice of model structures under consideration, as all models are necessarily just approximations to the "truth" (Oreskes et al., 1994) (or, in the common phrasing, "all models are wrong" (Box, 1979)). In general, a preferred structure is as parsimonious as possible while accurately reproducing held-out observations. There are a number of important considerations when deciding on a model selection or averaging approach, with different choices being more or less appropriate for different modeling goals (Höge et al., 2019; Bojke et al., 2006). Computational constraints may also play a role in whether a single model is selected (as opposed to averaging an ensemble of model structures), but care should be taken to acknowledge the ambient structural uncertainty in the interpretation of results.

Potential nonstationarity in endogenous dynamics further complicates model selection. Model selection and averaging techniques based on optimizing out-of-sample predictive performance (Gelman et al., 2014; Vehtari et al., 2017; Yao et al., 2018) may help, but still require the model structures under consideration to be capable of capturing appropriate changes to dynamics. Bottom-up modeling methods, such as those from the generative social sciences, can be used to explore the impacts of structural and parametric uncertainties related to alternative theories of human and institutional behavior, including potential nonstationarity (Epstein, 1999). For example, several different agent-based models of flood risk have explored different theories of human behavior within a consistent modeling framework (Haer et al., 2017; de Koning et al., 2017; Magliocca & Walls, 2018). These bottom-up methods can also be used to identify the emergence of new regimes of behavior (see Section 5.2.3 for discussion of these methods). Additionally, The critical transitions literature provides tools for modeling and empirically detecting shifts in endogenous dynamics (Lade et al., 2013; Scheffer et al., 2009). Models incorporating human and institutional decisions may also be able to incorporate data-driven generation of model structure (Ekblad & Herman, 2021) coupled with dimension reduction to support feature engineering for dynamic multisector datasets (Cominola et al., 2019; Giuliani & Herman, 2018) to generate structural and parametric variants which are consistent with past observations.

It is unclear whether multi-sector models mitigate or exacerbate this challenge. On one hand, the models become more complex: the more complex the model, the greater the number of parameters that need to be calibrated and the more challenging this estimation problem becomes, as more data is needed to constrain the likely parameter space (Srikrishnan & Keller, 2021). On the other hand, data from another sector might help constrain the likely parameter set. For example, a set of soil parameters that perform well in simulating hydrologic behavior, may not simulate crop yields well, and that might only be discovered through a coupled agro-hydrological model. This is an example of how adding model complexity could result in less uncertainty, as depicted in Figure 3.

Another challenge is the specification of a likelihood function. Calibration that does not properly account for the statistical structure of model-data discrepancies can result in biased inferences and hence projections (Brynjarsdóttir & O'Hagan, 2014). This likelihood function should ideally include different sources of uncertainties, such as both model-data discrepancy and observational errors. When these can both be modeled as independent errors with no correlation, they can be combined into a single error term. Srikrishnan et al. (2022) and Ruckert et al. (2017) provide examples of likelihood specifications which mix autocorrelated model-data discrepancies and independent observation errors.

For particularly complex models, the likelihood function may be mathematically or computationally intractable. Likelihood-free methods, such as precalibration (Edwards et al., 2011), Generalized Likelihood Uncertainty Estimation (GLUE) (Beven & Binley, 1992), and approximate Bayesian computation (ABC) (Sisson et al., 2018) can be used in these settings to obtain a representation of "behavioral" parameter sets. However, care should be taken when interpreting these results: Stedinger et al. (2008) notes that precalibration and GLUE parameterizations should not be treated probabilistically, and ABC results can show strong sensitivity to the choice of summary statistics and distance thresholds.

3.2.2 Addressing Computational Expense

Even if multi-sector models can constrain the domain of likely parameter sets and structures, calibration problems could still be more challenging computationally, both because the greater number of parameters increases the dimension of the search, requiring more model simulations to fully characterize the posterior distribution, and because the multi-sector model itself takes longer to run. Additionally, some model components may be more trusted than others, either in terms of model fidelity or quality of calibration data, and there might be concerns about "contaminating" the calibration of one module through these interactions. One approach to this problem is to calibrate the single sector models separately. However, combining the parameter sets from separate calibrations could yield unrealistic multi-sector dynamics by neglecting correlations. Alternatively, one could calibrate the multi-sector model for performance in a single sector first and then fix those parameters for a second calibration of parameters controlling another sector. This approach is common in the hydrological literature, *e.g.*, calibrating for streamflow and then nutrients (Arnold et al., 2012), but it is still likely to neglect correlations and may underestimate multi-modality. Jacob et al. (2017) provides some guidance on navigating this problem, but the implications of these choices for MSD calibration are not well understood in general.

Bayesian (or approximately Bayesian) calibration methods such as Markov chain Monte Carlo can require many thousands to millions of model evaluations, potentially making them computationally prohibitive for models that are too expensive for a sufficient number of runs on a given computational budget. There exist a suite of methods for speeding up Bayesian inference (Robert et al., 2018), but these may not be generally applicable to MSD calibration exercises. For example, Hamiltonian Monte Carlo methods (Betancourt, 2018), which are implemented in the Stan probabilistic programming language (Stan Development Team, 2019) and language-specific packages such as Julia's Turing.jl (Ge et al., 2018) and Python's pyMC3 (Salvatier et al., 2016), are extremely efficient, but require information about the gradient of the posterior, which can be difficult to obtain from simulation models that are not written to be parsed by an automatic differentiation package. Another approach can be to exploit parallelization in a high-performance computing environment, which is taken by sequential Monte Carlo-like algorithms like FAMOUS (B. S. Lee et al., 2020).

One approach to managing computational expense is reducing the number of parameters which need to be calibrated through *factor fixing*. In factor fixing, sensitivity analysis is used to identify groups of parameters or model components which are not influential and might be fixed without substantially impacting the analysis (Saltelli et al., 2008). This allows the analyst to focus their computational resources on simulating from the distributions of influential factors by justifying the deterministic treatment of non-influential factors. Different sensitivity analyses can be used for factor fixing. An important consideration is that a factor may not be influential when varied individually, but may exhibit significant influence through interactions (*e.g.*, the sensitivity analysis in Srikrishnan et al. (2022)). Consequently, the Method of Morris is commonly used for factor fixing (Cariboni et al., 2007) because it efficiently provides estimates of total order sensitivities that include individual and interactive effects (M. D. Morris, 1991). Other methods can be used for factor fixing (for instance elementary effects), but the key feature of any approach is that it should approximate total sensitivity (i.e. individual and interactive effects (Campolongo et al., 2007)), and be computationally efficient.

When the original model does not need to be used directly, surrogate models (or emulators) can be employed to reduce computational and parametric complexity. A number of different surrogate model structures can be used, including Gaussian processes (Kennedy et al., 2006), support vector machines (Bouboulis et al., 2015), and artificial neural networks (Eason & Cremaschi, 2014). These methods have different pros and cons; for example, Gaussian processes can only handle a limited parameter space, which can have implications for resulting risk analyses (B. S. Lee et al., 2020), while the machine-learning methods may be easy to overfit to data if not tuned carefully and may limit learning about system dynamics due to their black-box nature if not accompanied by careful diagnostics and sensitivity analyses. In many cases, a primary limitation in training good surrogate models is the number of available model evaluations (due to computational constraints), particularly as MSD outcomes of interest are likely to emerge from the interactions of a relatively large number of parameters and exogenous forcings. More sophisticated sampling strategies, such as adaptive designs of experiment (Burnaev & Panov, 2015; Gramacy & Lee, 2009; Chang et al., 2016) may be useful to maximize computing budgets, allowing surrogates to be trained on a larger subset of the parameter space. Evolutionary approaches to co-tune and select surrogate models have been proposed (Gorissen et al., 2009), which may be useful if building the surrogate model itself requires a large number of model runs to capture the dynamics of the model response surface, so surrogate modeling alone does not fully solve the problem of computational expense.

Another approach is the use of simple models to act as emulators of more complex models. This results in emulators which are mechanistically-motivated and can provide more direct insight into system dynamics and parameter values, but which may be less flexible in fitting the original model's response surface. For example, reduced-complexity climate models have been calibrated and used instead of more computationally-expensive models (Dorheim et al., 2020; Nicholls et al., 2020). While these simple models may lack the full richness and mechanistic detail of the complex models they're emulating, their increased ability to capture uncertainties may make their use more appropriate for certain research questions than the original models would have been (Helgeson et al., 2021).

However, there may be cases when emulation is insufficient due to the large number of parameters which need to be considered or the complexity of the system response surface, and full model evaluations are required for projections and scenario discovery. In this case, advances in efficient model calibration are necessary to facilitate uncertainty quantification and propagation. For example, B. S. Lee et al. (2020) demonstrate how a parallelized sequential Monte Carlo algorithm can treat a relatively large number of parameters of a complex Antarctic ice sheet model as uncertain, resulting in higher potential contributions to future sea levels.

An interesting approach is the application of machine learning methods for uncertainty quantification. Klotz et al. (2021) demonstrate how deep neural networks, typically thought of as black-box models, can be used to estimate uncertainties for a hydrological system, while also showing an example of how to obtain some measure of interpretability with a *post hoc* interrogation of fitted machine learning models. The power of careful implementations of machine learning methods, which embed mechanistic insights into the model structure, as an alternative for learning and uncertainty quantification for complex systems, rather than explicitly process-based modeling, is starting to be explored in the hydrological literature (Kratzert, Klotz, Herrnegger, et al., 2019; Kratzert, Klotz, Shalev, et al., 2019). These approaches may be a promising alternative to the use of computationally-expensive, mechanistic models for broader multi-sector analyses when large training data sets are available.

4 Uncertainty in Forward Projections

After calibrating a multi-sector model, we can use that model to project future outcomes. Analyses projecting outcomes for MSD systems involve uncertainty in two separate but overlapping ways: a) accounting for uncertainty in exogenous forcings and b) understanding the relative influence of various sampling, parametric, and structural uncertainties on model projections. Due to the number of relevant uncertainties, several of them deep, forward projection exercises in MSD are typically exploratory in nature (Bankes, 1993; Moallemi, Kwakkel, de Haan, & Bryan, 2020), which is why we use the term *projection* rather than *prediction* (MacCracken, 2001; Bray & von Storch, 2009).

In this section, we focus primarily on the influence of the treatment of exogenous forcings and up- and downscaling on uncertainties in projections. This focus is informed by the existence of several comprehensive reviews on techniques for SA (see *e.g* Pianosi et al. (2016)). However, the role of computational expense, as discussed in Section 3.2.2, is a major consideration for developing projection ensembles and SA with MSD models, as it is for calibration. One additional challenge here for emulation is the presence of spatiotemporal teleconnections due to the complex dynamics of cross-sectoral and regional connections (Helbing, 2013; Dolan et al., 2021). Mismatches between the "true" and emulated response surfaces could result in very different dynamical patterns and bias estimates of sensitivity, risk, and policy effectiveness. A related challenge is the use of a resulting ensemble to understand how uncertainties propagate through and interact within the system; we discuss these issues in the context of scenario discovery in Sections 5.2.1 and 5.2.3.

4.1 Exogenous Forcings and Joint Extremes

As we discuss in Section 2.2, control volume design, including the decision of which components to treat endogenously, is centrally important to uncertainty analysis. Increasing the number of components that are treated endogenously can facilitate a more complete uncertainty analysis, since model structures, parameters, and dynamic interactions can be more systematically varied and tested. However, it is important to recognize that in practice, computational constraints and/or issues of scale and scope lead modelers to externalize much of the system dynamics into fixed, exogenous boundary conditions. For example, it may be computationally intractable to include the impact of MSD system evolution on emissions to endogenously represent changes to the climate system. These external forcings are often outputs from a separate set of models, for example one or more climatological variables simulated by an ensemble of climate models or a set of socioeconomic projections produced by an IAM. Uncertainties surrounding exogenous forcings can often exceed the uncertainty associated with endogenous dynamics. Several studies across hydrology (J. Chen et al., 2011; Chegwidden et al., 2019; Vetter et al., 2017), agriculture (Asseng et al., 2013; Rosenzweig et al., 2014), health (Sanderson et al., 2017), and energy (van Ruijven et al., 2019; Bloomfield et al., 2021; Deroubaix et al., 2021) find that uncertainty arising from climate models can represent a substantial fraction of the total. Similarly, many studies find large uncertainties surrounding socioeconomic inputs, including emissions scenarios (Paltsev et al., 2015), population growth (Veldkamp et al., 2016), energy costs and demand (Lamontagne et al., 2018; Su et al., 2020), economic growth (Gillingham et al., 2018), and parameterization of damages (Errickson et al., 2021).

Biased or low-coverage realizations of these uncertainties could interact with errors in the emulated response surface to compound failures to identify potential teleconnections. This section hence discusses uncertainties associated with exogenous forcing. We distinguish between climate forcing (Section 4.1.1) and socioeconomic forcing (Section 4.1.2) with further breakdowns given in each section. We provide a brief overview of how each type of forcing is typically employed in single sector models and discuss the challenges and opportunities of moving to the multi-sector case.

4.1.1 Climate Forcing

Perhaps the most common type of climate forcing data takes the form of gridded simulation outputs of meteorological variables from global climate models (GCMs). GCMs are subject to the same types of uncertainties outlined previously (structural, parametric, and sampling) and the climate modeling community typically probes each of these through ensemble frameworks. As different ensemble outputs address uncertainty differently, the choice of climate product influences how climate uncertainty is treated in the resulting MSD analysis.

Multi-Model Ensembles (MMEs), such as the Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016; Taylor et al., 2012), are the most commonly used framework. MMEs do not represent a systematic sampling of any one type of uncertainty but instead represent an "ensemble of opportunity" (Tebaldi & Knutti, 2007). That is, they are collections of models from various institutions that often share code and expertise (Abramowitz et al., 2019), with parameters tuned in complex ways (Mauritsen et al., 2012) and simulations reported without an estimate of internal variability (Maher, Power, & Marotzke, 2021). MMEs thus combine all three sources of uncertainty into one ensemble (which may or may not be desirable depending on the specific research question), but are typically framed as focusing on structural uncertainty.

In contrast to MMEs, Single Model Initial condition Large Ensembles (SMILEs) are designed specifically to estimate the effects of internal variability, which here we classify under sampling uncertainty. SMILEs are constructed by perturbing the initial conditions of a single GCM to produce varying climate and weather trajectories (Hawkins et al., 2016). The number of publicly available SMILEs (Deser et al., 2020) and the number of studies employing SMILEs (Maher, Milinski, & Ludwig, 2021) have increased considerably in recent years. One advantage of SMILEs is an improved sampling of extreme events (Wiel et al., 2019; Haugen et al., 2018) relative to MMEs.

Finally, single-model Perturbed Physics Ensembles (PPEs) are designed to sample parametric uncertainty (Murphy et al., 2004). In PPEs, the parameters or configurations of each ensemble member are systematically varied while keeping other factors fixed (Sexton et al., 2019). This framework isolates the impact of parametric uncertainties, which are typically neglected in the other frameworks, on model projections (L. A. Lee et al., 2011). PPEs may also be used to produce probabilistic projections (conditioned on model structure) if employed in a Bayesian framework (Sexton et al., 2012).

Each of the above ensemble frameworks exhibits distinct advantages and disadvantages for sectoral modeling. The different representations of uncertainty in each framework may render some ensembles particularly useful for a given research question. For example, the interpretation of ensemble spread in SMILEs as arising from irreducible or aleatory uncertainty (and therefore as a representation of sampling uncertainty) makes them uniquely well-suited as decision-making tools; each ensemble member represents a plausible real-world outcome that could be included in a robust risk management strategy (Mankin et al., 2020). However, any single GCM used to produce a SMILE is still subject to structural and parametric uncertainties which may bias its representation of internal variability. Multi-model large ensembles have been proposed as one method to address this limitation (Deser et al., 2020). Utilizing both SMILEs and MMEs concurrently can help quantify what fraction of uncertainty is irreducible (Lehner et al., 2020), a metric with important policy implications (Palutikof et al., 2019). Additional considerations include ensemble configuration and data access. Given the large number of simulation members in a typical SMILE (on the order of 20 to 100), their use may exacerbate challenges related to computational tractability of MSD uncertainty analysis.

The main disadvantage of global, gridded, process-based Earth system models is their high computational cost. In contrast, simple climate models (SCMs) are generally much faster to run and thus might be preferable in a variety of modeling setups, particularly for uncertainty analyses. SCMs, which for our purposes include all climate models other than full-scale Earth system models, span a large range of structures and complexities, from one- or few-line models that aim to emulate global responses of selected outcomes (for example, global mean surface temperature or sea-level rise), to intermediate complexity Earth system models that might be spatially resolved but with very coarse resolutions and highly parameterized representations of physical dynamics (Weber, 2010). Examples of prominent SCMs include MAGICC (Meinshausen et al., 2011), FAIR (Leach et al., 2021), and Hector (Hartin et al., 2015).

The reduced computational burden of SCMs allows a better sampling of uncertainty, including the ability to produce probabilistic simulations. SCMs can also be tailored to specific, possibly novel research questions more easily than gridded climate products from GCMs (Forster et al., 2020). As noted, these advantages typically come at the expense of spatial resolution and the variety of available output variables. Given their increased reliance on parameterized processes, care must also be taken to avoid overfitting the model to calibration data; the main value of SCMs is their ability to give reliable out-of-sample estimates.

4.1.2 Socioeconomic Forcing

Sectoral and multi-sectoral analyses typically require exogenous assumptions about broader socioeconomic dynamics. Key socioeconomic variables generally revolve around demographics, economics, land-use, and emissions, but certain sectoral modeling efforts also require relatively more obscure quantities such as price trajectories of specific technologies (Auping et al., 2016), or local government structures (Andrijevic et al., 2020).

Projections of the future of the global economy and its associated socio-political dynamics are characterized by deep and dynamic uncertainties. As such, the global change research community typically relies on carefully crafted sets of plausible alternative futures known as scenarios, the canonical example being the Shared Socioeconomic Pathways (SSPs) (Riahi et al., 2017). Here, we briefly discuss the design and usage of the SSPs as well as their characterization of associated uncertainties. Our discussion can be generalized to other unrelated but similarly constructed scenario-based approaches (for example, as in Gurgel et al. (2021) and citeAwildImplicationsGlobalChange2021).

SSPs provide global trajectories of socioeconomic factors such as demographics, health, education, urbanization, economic growth and inequality, governance, technology, and policy. There are five SSPs, each reflecting qualitative global narratives that represent equally plausible future socioeconomic and geopolitical trends along axes of high or low challenges to climate change mitigation and adaptation (O'Neill et al., 2017). These trajectories are passed to IAMs that generate quantitative projections of energy use (Bauer et al., 2017), land use (Popp et al., 2017), and associated emissions, among other outcomes (Riahi et al., 2017). These projections may represent a "baseline" scenario without climate policy or under various Shared climate Policy Assumptions (SPAs) that represent different sets of climate policy attributes (Kriegler et al., 2014). By design, the SSPs are parsimonious representations of future socioeconomic conditions at the global scale (Kriegler et al., 2012). As such, they often need to be supplemented with sector-specific (*e.g.*, Rao et al. (2017); Graham et al. (2018)) and/or localized scenarios (*e.g.*, Kok et al. (2019)).

There are large uncertainties both within and among the SSPs, many of which apply to scenario-based approaches more broadly. First, the key socioeconomic drivers of a given outcome often do not obey consistent narratives, but instead arise from a mixture of components from the narrative-driven scenarios (Lamontagne et al., 2018; Dolan et al., 2021). This highlights the difficult but important task of designing suitably encompassing scenarios from which such hybrids can be drawn. Multi-model comparisons often find large structural (Duan et al., 2019) and parametric (Krey et al., 2019) differences across IAMs that propagate into simulated outcomes (von Lampe et al., 2014; Harmsen et al., 2021). Behind any given quantitative projection in the SSP framework is an assumption that the underlying IAM has produced a plausible real-world trajectory, but this has been increasingly challenged, particularly with respect to energy mixes (Ritchie & Dowlatabadi, 2017a, 2017b; Burgess et al., 2021; Hausfather & Peters, 2020). It remains challenging, in general, to evaluate the efficacy of IAMs across the wide range of research objectives for which they are employed (Wilson et al., 2021; Schwanitz, 2013). Some authors advocate for a more holistic approach with a diminished role for IAMs (Morgan & Keith, 2008). Some technical details of SSP design may also limit their utility for decision making. As the baseline SSPs do not include climate policy or climate impacts, there is no single scenario that incorporates the best estimates of impacts or the latest governmental mitigation targets (Grant et al., 2020). A related concern is that scenarios can become out of date, particularly for near-term projections, either as more recent data is made available or through improvements in scientific understanding and modeling capabilities (Hausfather & Peters, 2020; Burgess et al., 2021).

Scenarios such as the SSPs are also typically not accompanied by probabilistic information, which can make them difficult to integrate into risk assessments and may make their interpretation more susceptible to typical cognitive biases (Tversky & Kahneman, 1974; Morgan et al., 1992; Webster et al., 2001). Alternatively, probabilistic approaches can be used to systematically explore the uncertainty space and provide insight into the likelihoods of both inputs and outcomes. While uncertainty quantification and the use of probabilities is not always appropriate, there are cases in which it is defensible and can provide useful information for risk-based decision-making, particularly when relevant assumptions about likelihoods, data sources, and distributional forms are made transparent (Morgan & Keith, 2008). Such an approach has particularly been used for key socioeconomic drivers such as population, GDP, and emissions (*e.g.*, Gillingham et al. (2018)). IAMs may also be employed in a probabilistic setting, sampling from distributions of key inputs to explore the uncertainty space (Webster et al., 2012; J. Morris et al., 2022), although this remains a rare approach. Methods employed to develop probability distributions for model inputs vary on a case-by-case basis and can involve time series forecasting (Keilman, 2020; Vollset et al., 2020), broader statistical approaches (Raftery et al., 2017; Liu & Raftery, 2021), and expert elicitation (Christensen et al., 2018), possibly used alongside process-based models (Güneralp & Seto, 2013; Seto et al., 2012; Srikrishnan et al., 2022). Employing a statistical model for exogenous forcings may offer some advantages, including the ability to validate out of sample and to more completely probe structural and parametric uncertainty owing to the reduced computational expense. However, the core difficulties lie in carefully quantifying "standard" uncertainties, including specifying the relevant, possibly multivariate, probability distributions, as well as properly characterizing deep uncertainties. It is also important to include the correlation structure of uncertainty across outputs, even for univariate distributions. For example, many probabilistic population forecasts include 95% confidence intervals for each country, without explicitly specifying the correlation among countries: is the 90th percentile for US population in 2070 coincident with the 90th percentile for Canada in 2070? Such correlational effects have important implications for sectoral dynamics across space and time.

In addition to socioeconomic forcings, another important sources of uncertainty in human system modeling are socioeconomic policies, or the rules by which the model reflects human responses to changes in the internal state or external environment. There are many theories from political science, sociology, psychology and other social science fields that are relevant to the modeling of human, firm, and government behavior and shifts. A growing and diverse literature draws on these theories to model dynamics such as the evolution or breakdown of cooperation (Stewart & Plotkin, 2014; Auer et al., 2015), the diffusion of opinions or innovation (Janssen & Jager, 2001), and the behavior of investors or consumers in markets (Bonabeau, 2002). Multi-formalism modeling or multi-paradigm modeling presents the potential for integrating such dynamics into multi-sectoral models (Vangheluwe et al., 2002).

4.2 Changing Scales

Mismatches between the characteristic scales of forcing inputs and system models creates challenges and uncertainties that are somewhat distinct from those discussed thus far. In this section, we discuss the impacts of downscaling climate and socioeconomic data to match the spatiotemporal scales relevant for models.

4.2.1 Downscaling Climate Data

Downscaling, and the oftentimes related process of bias-correction, has received considerable attention in the hydrology and climate impacts communities. There are two broad categories: dynamical, which involves running a high-resolution regional climate model forced with boundary conditions provided by a GCM (Giorgi & Gutowski, 2015), and statistical, which involves modeling a statistical relationship between large-scale atmospheric predictors and local predictands (Hewitson et al., 2014). Both methods can involve some form of bias-correction, although typically more so for statistical approaches (Maraun, 2016). Known uncertainties, which apply equally to dynamic and statistical downscaling, include the validity of any stationarity assumptions, the physical plausibility of results across space and time (Maraun, 2016), and the resulting representation of (multivariate) extremes (Werner & Cannon, 2016; Zscheischler et al., 2019). An additional uncertainty that is relevant for bias-correction and statistical downscaling, and can be difficult to account for, is the choice of observational product (Lopez-Cantu et al., 2020).

When possible, careful consideration should be given to what information is most important for the relevant sectoral dynamics and/or decision problems — for example, methods that jointly process temperature and precipitation (*e.g.*, Abatzoglou and Brown (2012)) may be better suited for analyses where risks are driven by multivariate hazards, whereas methods that place a higher emphasis on capturing spatial structure (*e.g.*, Pierce et al. (2014)) might be preferred for sectors in which spatial heterogeneity is important. In any case, performing a hindcast test, where sectoral outcomes simulated by the original GCMs are compared to those simulated by downscaled outputs, can be useful to uncover biases directly relevant to sectoral dynamics that might otherwise go unnoticed (Lafferty et al., 2021). Practical considerations such as the spatiotemporal domain and resolution, as well as the number of variables included, are also likely to be important factors in determining which datasets are widely used. Ease of access is also crucial: products that abide by community standards and strive towards the FAIR principles (Wilkinson et al., 2016) will better facilitate inter-comparisons and research extensions.

4.2.2 Downscaling Socioeconomic Forcings

Downscaling is also increasingly relevant for socioeconomic projections, with important differences in understanding and application relative to climate simulations. Socioeconomic dynamics are inherently multi-scale in that different national or regional policies can interact with the same global drivers to produce a broad and possibly diverging set of outcomes. Downscaling in the socioeconomic context can mean the generation of additional regional/local scenarios that fit into a broader global context, or the more traditional exercise of interpolating gridded data to a higher resolution.

For the former case, there are a number of possible approaches to multi-scale scenario generation, each differing in the level of interconnectedness across scales (Biggs et al., 2007). Downscaling, in the sense of generating regional scenarios from a set of global or otherwise larger-scale contexts, should hence be understood as only one possible "top-down" option. Other participatory "bottom-up" (Kok et al., 2006) or hybrid (Nilsson et al., 2017) approaches may be more suitable in some situations. However, even within the downscaling paradigm there is a considerable degree of heterogeneity regarding, for example, the strictness of quantitative boundary conditions and the consistency of qualitative storylines (Zurek & Henrichs, 2007). Additionally, downscaling can follow a "one-to-one" approach where regional storylines follow as closely as possible the global narratives, or a "many-to-one" approach where regional storylines are perturbed around a broadly consistent larger context (Absar & Preston, 2015). The many-to-one method better represents the increasing uncertainty at local scales but may quickly become challenging to manage (Kriegler et al., 2014). It may also be necessary to generate quantitative trajectories of important quantities, either to reflect the results of the regional scenario generation process or to include new factors that were previously unavailable. To this end, many IAMs can be employed at regional or national scales (*e.g.*, Palazzo et al. (2017)).

In many cases, modelers require spatially-resolved information beyond the highly aggregated outputs of most IAMs. For example, the SSPs provide projections of key drivers such as population structure only at the national scale and land-use at the regional/continental scale. As such, a number of methods are used to downscale these scenarios into gridded products. Most follow a similar framework, where statistical or process-based models are calibrated on historical data and then applied to aggregated IAM outputs in the future period.

Statistical methods are typically employed to downscale population and other demographic factors. One rudimentary approach is to fix the spatial pattern at the current distribution and scale each grid point with national factors (Caminade et al., 2014). More sophisticated approaches include gravity models that assume areas with certain characteristics, such as higher populations, attract more people (Jones & O'Neill, 2016) and regression methods that make use of auxiliary variables likely to be important in determining future growth (Murakami & Yamagata, 2019). Several studies jointly downscale population and GDP (*e.g.*, Wear and Prestemon (2019)). Across all methods, parametric and structural uncertainties are rarely explicitly included or examined.

Land-use downscaling is typically more involved than population or GDP downscaling, reflecting large uncertainties in socioeconomic and biophysical conditions. Many models allocate land via profit maximization, which can be employed within a statistical framework (Meiyappan et al., 2014) or within IAMs reconfigured to produce gridded outputs (Fujimori et al., 2018). As with population and GDP downscaling, parametric and structural uncertainties are typically neglected. One notable exception is M. Chen et al. (2019), which examined parametric uncertainty in Demeter, a downscaling algo-

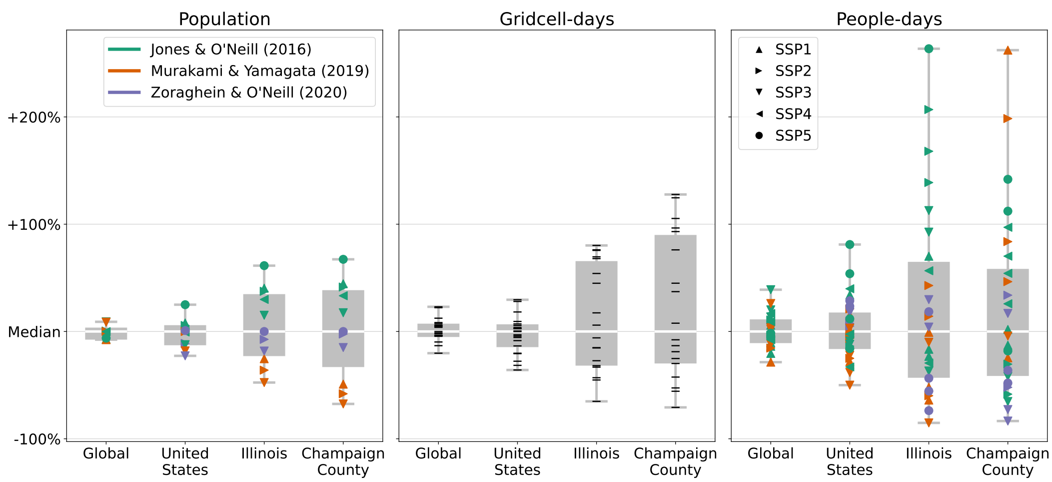

Figure 6. Exogenous uncertainty in compound extremes across scales. In each subplot, the boxplots show the distribution (relative to the median projection) of population (left), the number of annual gridcell-days above 35°C (center), and the number of annual people-days above 35°C (right), at different spatial scales (global, national, regional, and local). Boxplot whiskers extend over the full range of data and a sampling of individual points are shown by the markers, where different colors represent difference population downscaling methods and different symbols represent different SSP scenarios. Gridcell-days are calculated from the 21 models in the NEX-GDDP ensemble (Thrasher et al., 2012) as 2040-2060 averages and people-days are calculated by multiplying by the projected (2050) number of people in each gridcell; both metrics are then summed over the appropriate spatial region. Population distributions are taken from publicly available products that downscale the SSP population scenarios (Jones & O'Neill, 2016; Murakami & Yamagata, 2019; Zoraghein & O'Neill, 2020).

rithm that uses a rules-based approach to describe land conversion (M. Chen et al., 2020). M. Chen et al. (2019) finds a considerable propagation of uncertainty into future projections, with large effects on grasslands and cropland, but little influence on urban areas. Demeter is a relatively simple model with few parameters, but similar effects are likely to be found in more complex downscaling algorithms.

In general, the uncertainties of a scenario approach, used alone or in conjunction with climate projections, are amplified at smaller scales. This is demonstrated in Figure 6, which shows a simple socioeconomic metric (population), a simple climate metric (gridcell-days above 35°C), and a related joint metric (people-days above 35°C) at increasingly smaller spatial scales. In each case, relative uncertainty (as measured by the ensemble spread) increases at smaller scales. We also see that for population projections, the downscaling algorithm (delineated by different colors) becomes more important than the SSP scenario (delineated by different shapes) at smaller scales. At all scales, the joint metric is more uncertain than either single metric.

4.2.3 Temporal Scales

The appropriate temporal scale of forcing data should be dictated by the relevant sectoral dynamics and outcomes of interest. As such, choices should be carefully motivated by the relevant scales for system dynamics of primary interest: long-term trends or short-term stresses? Systems and outcomes that are more sensitive to transitory phenomena, including tail events that are limited in time, are likely to require forcing data generated by an alternative suite of models. For example, time series approaches can be used to model various economic indicators at different temporal resolutions (De Winne & Peersman, 2021; Koop & Korobilis, 2009). This contrasts with a SSP-like framework that aims to understand decadal-scale changes in socioeconomics and thus produce quantitative trajectories that are typically smoothly varying and with 5-year time steps.

The temporal scale of analysis may also influence other modeling decisions in non-trivial ways. For example, most population datasets project residential populations rather than ambient population, which accounts for daily population movements to and from work or school (McKee et al., 2015). Ambient population is likely a more useful metric for understanding exposure to short-lived climate or weather hazards.

5 Scenario Discovery and Characterizing Dynamics

5.1 The Role of Scenario Discovery in MSD

From its inception, MSD has aimed to be societally-relevant by improving our understanding of the dynamics of integrated human-Earth systems and impacts on complex societal changes (Reed et al., 2022). The identification and communication of key uncertainties to other researchers, decision-makers, and the public are therefore key components of MSD analyses. However, the high dimensionality, interconnectivity, and complexity of the uncertainty space discussed in the preceding sections presents a significant challenge to this goal. The response of the public and stakeholders to uncertainty is highly complex and dependent on a number of factors, including how that uncertainty is communicated to them (Ho & Budescu, 2019; Howe et al., 2019). As such, uncertainties often are not fully accounted for in planning processes (Carlsson Kanyama et al., 2019). To ensure stakeholders consider uncertainty in their decision-making, researchers must supply information that is relevant to addressing their concerns, but does not dangerously narrow the framing of the decision problem.

Our inability to predict *a priori* the leading sources of uncertainty and understand how they impact complex outcomes necessitates large ensemble simulations, with hundreds to millions of scenarios in order to capture tail risks, interactions, and key dynamics (Lamontagne et al., 2018). This requires a method to select a few key desirable or undesirable outcomes, ideally representative of a broader class of dynamics, from a large set of model runs. Scenario discovery (Bryant & Lempert, 2010) is one class of such methods, which has already seen wide adoption in MSD-related work (*e.g.*, Moallemi, Kwakkel, de Haan, and Bryan (2020); Lamontagne et al. (2018); Dolan et al. (2021); Jafino and Kwakkel (2021); Quinn et al. (2018); Guivarch et al. (2016); Halim et al. (2016); Wang et al. (2013)).

Scenario discovery is a computer-assisted approach to scenario development that identifies regions of the uncertainty space that are tied to outcomes of interest (Bryant & Lempert, 2010; Kwakkel, 2019). These methods begin by sampling possible values of uncertain factors, which are then simulated using one or more system models to generate a large ensemble of potential future system conditions. Typically, a binary classification is applied to designate scenarios of interest in which some notable outcome is observed (*e.g.*, a satisficing constraint for objective attainment) (Herman et al., 2015). Machine-learning classification methods are then applied to identify the leading predictors of a case of interest (Bryant & Lempert, 2010). The most commonly used methods are the Patient Rule Induction Method (PRIM, (Friedman & Fisher, 1999)) and Classification and Regression Trees (CART, (Breiman et al., 2017)), though other methods can be used, such as logistic regression (Quinn et al., 2018; Lamontagne et al., 2019). Once the leading predictors and conditions associated with the cases of interest are identified, they are ideally translated into qualitative, comprehensible narratives to facilitate communication and interpretability (Parker et al., 2015; Trutnevyte et al., 2016; Moallemi et al., 2017; Jafino & Kwakkel, 2021). As can be seen by its procedure, scenario discovery is primarily focused on parametric uncertainties, which are an accessible if incomplete way of defining the space of possible futures.

Scenario discovery is often referred to as a "bottom-up" or *a posteriori* approach because it defines key drivers and scenarios after generating and analyzing a large simulation ensemble. In contrast, "top down" or *a priori* approaches begin with expert assessment of key drivers and associated uncertainties to develop a small number of scenario narratives, which are in turn simulated with systems models (Bryant & Lempert, 2010; Kwakkel, 2019; Maier et al., 2016). The nature of multi-sector systems, which are characterized by a large number of uncertainties, emergent complexity, and correlated outcomes across sectors, severely limits the ability of any group of experts to anticipate key drivers and dynamics (Helbing, 2013; Marchau et al., 2019). In such cases, *a priori* approaches may suffer from narrow problem framing, inadequate coverage of surprising or paradoxical outcomes, and may be less conducive to participatory decision making with diverse stakeholders (Bryant & Lempert, 2010).

As an illustrative example, we once again turn to the impacts of Winter Storm Uri in Feburary 2021 (Busby et al., 2021). Despite recent precedent for similarly or more severe weather conditions (Doss-Gollin et al., 2021), energy and gas operators failed to winterize equipment in Texas. As a result, gas production and delivery were severely curtailed during the peak of the cold, disrupting electricity production from natural gas while smaller outages from wind, nuclear, and coal generating plants also occurred (Busby et al., 2021). At the same time, electricity demand for heating spiked, bringing the Texas power grid to within minutes of collapse, leading regulators to curtail electricity supply to millions of people. The days-long outage severely curtailed the delivery of basic services such as water and wastewater, internet, medical services, food, and heat (Busby et al., 2021; Watson et al., 2021). This is an example of a chain of events leading to the failure of a critical infrastructure system which, in retrospect, ought to have been foreseen, but which seems to have been missed in scenario planning, particularly as the disruption transcended traditional sectoral boundaries.

5.2 Challenges for Scenario Analysis in Multi-Sector Systems Modeling

The uncertainties that arise in multi-sector modeling often go beyond what has typically been explored with Scenario Discovery. In particular, MSD analyses present challenges to typical scenario discovery approaches for three reasons: (a) the high dimensionality of the uncertainty and outcome space, (b) the challenge of defining cases of interest across sectors, and (c) the difficulty of interpreting *a posteriori* scenarios. MSD researchers should be aware of these gaps, and potential alternative methods, as they identify scenarios of interest for further analysis and communication.

5.2.1 High Dimensional Uncertainty and Outcome Spaces

Any scenario analysis begins with a design of experiment, which is unavoidably an *a priori* narrowing of the uncertainty space to be explored. In the MSD setting this is increasingly difficult as complex systems interactions and teleconnections massively expand the space that needs to be considered, while simultaneously obscuring the key uncertainties and amplifying the consequences of an incomplete representation. This presents a major challenge to scenario discovery in MSD. Incomplete representations of uncertainty typically manifest in three ways: the selection of factors, the number of samples, and the range of samples.

In Section 3.2.2, we discussed factor-fixing through sensitivity analysis. These techniques may fail when confronted with path dependence, emergence, and multiple outputs of interest. Factor influence may evolve over time and depend on earlier systems evolution, and is unlikely to be the same across output metrics (Lamontagne et al., 2018). Often, a more informal factor-fixing ensues in MSD studies, driven by "lamp-post science," where factors are varied because existing databases, such as the RCPs or the SSPs, make them easy to include, while other factors are fixed simply because they are more difficult to sample or because existing products fail to account for their uncertainties (as discussed in Section 4). A common example is an under-representation of structural uncertainties in scenario discovery, such as model structural uncertainty or the decision problem framing (Quinn et al., 2017; Rozenberg et al., 2014). Such experimental designs are often necessary to limit computational expense, but the resulting consequences for projections and planning are difficult to quantify.

Sparse sampling of uncertainties is one way to limit the computational cost of generating ensembles of model runs, but this can severely limit our ability to identify leading drivers of outcomes. As an example, Lamontagne et al. (2018) considered more than 33,000 scenarios derived as hybrids of the SSPs: a marked increase over the 3-5 canonical SSPs considered in many analyses. This decoupling of the SSP dimensions highlighted plausible yet overlooked narratives with serious global consequences. However, the experimental design in Lamontagne et al. (2018) did not disentangle, for instance, the yield improvements for different crops in each of the 285 modeled land-use regions across the globe, nor were GDP or energy technology trajectories decoupled for individual countries, instead opting to vary "consistent" SSP narratives for different sectors. It is not clear that such consistency is epistemically valuable for scenario discovery. Sectoral studies in water resources suggest these choices may substantially bias robustness and scenario discovery assessments (Quinn et al., 2020; McPhail et al., 2020).

The high dimensionality of the uncertainty space often necessitates inadequate coverage of extreme cases that are likely to drive cases of interest. This is particularly acute in the case of deep uncertainty, where full UQ may be inappropriate. One approach is to use multiple models with varying structures instead of a single, more complex model; scenario discovery on the resulting multi-model ensemble can yield insights into different dynamical pathways leading to outcomes of interest under varying assumptions (Kwakkel et al., 2013; Auping et al., 2014).

5.2.2 Multiple Outcomes of Interest

Within a single-sector or regional analysis, defining cases of interest can be relatively straightforward (*e.g.*, when is a levee overtopped, or when is there a blackout?). Many sectoral studies have utilized satisficing criteria across several metrics, often utilizing the logical connection between those metrics to identify cases of interest, followed by binary classification on those scenarios (Herman et al., 2014). In MSD settings, this process is more complicated as the number of sectors and regions increase, with correspondingly more complex interactions and teleconnections. For these complex systems, it is not necessarily clear *a priori* which output(s) might be correlated and simultaneously achieve the satisficing criteria. For instance, Jafino and Kwakkel (2021) illustrate diverse inequality patterns in adaptive water-food management that defy binary classification. The dynamical nature of MSD systems also presents a challenge to traditional binary scenario discovery, as the timing of failure conditions can be an important consideration (Steinmann et al., 2020). Another complication is the presence of spatial and temporal teleconnections, which may mean that outcomes of interest in different sectors occur at different time steps or different spatial regions.

One potentially promising category of techniques is multinomial classification, wherein scenario discovery simultaneously identifies multiple different "cases of interest" (Gerst et al., 2013). Typically, this is performed in a sequential approach, where the output space is first partitioned into classes of interest, then classification tools are used to identify input factors that are most predictive of individual class membership (Jafino & Kwakkel, 2021). The partition of the outcome space could be manual (Lamontagne et al., 2018; Rozenberg et al., 2014), or utilize clustering algorithms (Gerst et al., 2013; Steinmann et al., 2020). Manual classification has the advantage of interpretability but suffers from the same weaknesses as *a priori* scenario development for high dimensional problems. On the other hand, while clustering with statistical algorithms is more scalable, the resulting classes can be difficult to interpret, and the results can be sensitive to a number of choices, such as the number of classes. Standard scenario discovery is then often implemented on each of the classes individually through a series of binary classification problems. One drawback of this is that the membership rules between classes might not be easily distinguishable (Kwakkel & Jaxa-Rozen, 2016), which may hinder stakeholder engagement (Jafino & Kwakkel, 2021). Because the classification is conducted independently, the relationship between classes may also be difficult to interpret. Alternatively, a concurrent multinomial scenario discovery approach has also been proposed (Jafino & Kwakkel, 2021), which simultaneously partitions the data and predicts class membership through the use of multivariate regression trees. This approach can reveal more detailed classes than the sequential approach, but this comes at the expense of interpretability and communicability.

The scale, diversity, and interconnectivity of the uncertainty space in MSD problems poses a significant challenge to traditional scenario discovery techniques. For example, how can we identify the potential for cross-sector interactions to lead to cascading failures? One route is through the application of methods from the complexity sciences to investigate nonlinear feedbacks, emergent behavior, and tipping points (Berkes, 2007). Similar to scenario discovery, these approaches aim to understand the space of possible trajectories of a system rather than prediction of the particular system state at a given point in time (Brugnach & Pahl-Wostl, 2008), in part reflecting the high levels of uncertainty in MSD systems (Vogel et al., 2015).

A complex systems approach to understanding parametric uncertainty can provide more information than typical sensitivity analyses about the importance of model parameters in determining qualitative behavior. Qualitative behavior of interest, for example, would involve regime shifts toward an unstable equilibrium consisting of a different set of feedbacks. Examples of regime shifts include a lake switching from being oligotrophic to eutrophic (Carpenter, 2005), as well as the collapse of communities that are economically dependent on local natural resources (Y. Chen et al., 2009). The possibility of this type of sudden, discontinuous change in equilibrium behavior does not necessarily exist in all systems, but becomes more likely in highly coupled systems (Lade et al., 2013).

One example of a dynamical systems tool with potential application to MSD analyses is topological data analysis (TDA) (Wasserman, 2018; Smith et al., 2021; Chazal & Michel, 2021) to understanding the network structure of coupled model output. Another example is generalized modeling, which is a form of dynamical systems analysis that does not require specifying functional forms. Instead, it allows the functional forms and magnitudes of relationships between variables to be treated as parameters (Gross & Feudel, 2006; Lade & Gross, 2012; Lade & Niiranen, 2017). Finally, structural uncertainty in agent-based modeling can be addressed using pattern-oriented modeling, a method that involves formulating alternative theories of agents' behavior and testing them by how well they reproduce characteristic patterns at multiple levels (Grimm et al., 2005).

5.2.3 Scenario Interpretability

A primary goal of decision support for MSD is to identify broadly plausible pathways by which good or bad outcomes might occur and be influenced by changes to exogenous forcings, system dynamics, and/or policy interventions. This requires identifying and articulating the patterns and mechanisms by which these changes propagate through the coupled system. However, a typical theme in statistical learning is the tension between classification ability and the interpretability of the resulting classes. In scenario discovery, the emphasis is to maximize interpretability, at the expense of "optimal" classification.

Interpretability is particularly difficult in MSD settings given the presence of teleconnections and emergent dynamics. While a powerful classifier may be able to identify the experimental factors related to scenarios of interest, the resulting scenarios may not be tied to a clear narrative explaining the circumstances and dynamics driving the outcomes. The interpretability-prediction tradeoff is not unique to scenario discovery or MSD, and there exists an opportunity to include emerging developments in machine learning and visual analytics with existing scenario discovery workflows to improve interpretability. One such direction is the use of machine learning methods to predict future vulnerable conditions based on observed system states and fluxes (B. Robinson et al., 2020), and to design dynamic adaptation policies to mitigate them (J. S. Cohen & Herman, 2021). Advances in interpretable machine learning (Rudin, 2014; Rudin et al., 2021; Murdoch et al., 2019; Molnar et al., 2020) also present opportunities to help navigate the tradeoff between interpretability and classification when analyzing model output ensembles. Interpretable approaches to machine learning also have the potential advantage of increased transparency, which might help expose systematic biases in MSD modeling which could be relevant to decision-making.

Despite challenges to interpretability, MSD model projections and analyses can be useful in informing policy under uncertainty. For example, lower-dimensional models have been used in social-ecological systems literature to provide broad insight into resource management problems relevant to MSD while remaining interpretable. The robust control framework has been used to identify fundamental tradeoffs in the robustness of different institutional arrangements, modeled as different controllers for the system, to parameter uncertainty (Anderies et al., 2007; Rodriguez et al., 2011). This approach has also shown how preparing for certain types of shocks may make a system more vulnerable to novel ones (Cifdaloz et al., 2010; Carlson & Doyle, 1999, 2000; Doyle & Carlson, 2000; Manning et al., 2005). This same modeling framework has also been used to explore how policy implementation issues that result from or exacerbate uncertainty, such as infrequent sampling or implementation delays, impact policy performance (Rodriguez et al., 2011), especially under the possibility of regime shifts (Polasky et al., 2011). Finally, MSD models have been used to identify safe operating spaces (Barfuss et al., 2018; Cooper & Dearing, 2019; Rockström et al., 2009) and identify threats to system resilience and the importance of cross-sectoral policies (Brunner & Grêt-Regamey, 2016).

Model development is a component of uncertainty characterization and can aid the process of communication, social learning, and exploration of scenarios and solutions among diverse stakeholders (Brugnach & Pahl-Wostl, 2008). Methods for exploring structural uncertainty, especially when paired with expert elicitation and participatory processes, help identify conflicts and agreements and make explicit different problem framings and mental models (Brugnach & Pahl-Wostl, 2008; Hare & Pahl-Wostl, 2002; Rouwette & Vennix, 2020). In addition to improving the model predictions, this process also increases the likelihood of stakeholders accepting model results (Pahl-Wostl, 2007; Giordano et al., 2020). For MSD systems, scaling these participatory modeling approaches to higher levels of governance with far more stakeholders remains a challenge, though there is an emerging environmental governance literature aimed at informing these higher level processes, particularly in the context of global climate change policy (Cloutier et al., 2015; Figueiredo & Perkins, 2013; Fröhlich & Knieling, 2013).

6 Conclusions & Best Practices

MSD is an emerging area of research focused on identifying and analyzing complex systems related to critical societal questions. Conclusions based on limited analysis (for example, analyses which only account for a handful of scenarios), could harm decision-making by anchoring stakeholders to a range of outcomes which might not be representative of true risks. As a result, all MSD analyses ought to explicitly discuss how the research methods treated uncertainty (or consciously chose not to, for example in a benchmarking activity).

It is not necessarily reasonable or even desirable for every MSD analysis to account for all types of uncertainties. For example, while we have focused on quantitative aspects of uncertainty analysis for MSD systems, there are a number of other considerations which might influence an MSD research design (Renn et al., 2020). For example, governance or stakeholder concerns might reduce the range of uncertainties, system configurations, or decision alternatives under consideration. The translation of systemic risk analyses into governance strategies is also critical, and requires an interdisciplinary, layered approach (Renn et al., 2020; Hochrainer-Stigler et al., 2020). Additionally, it can often be easier to focus on specific uncertainties or dynamics with reduced-form representations or samples of less-relevant model components.

Rather, best practices in MSD uncertainty analysis should facilitate communication across interdisciplinary teams of investigators and emphasize transparency, so that uncertainties that were not considered or fully treated in a given analysis can be examined in subsequent studies. One of the key points we have tried to emphasize is that many uncertainty-relevant research decisions should be made intentionally, to ensure that they are aligned with the research question, and that the resulting interpretation of results takes place within the context of the research design. To this end, we suggest that MSD research should include the following best practices and principles, though this list is by no means exhaustive and will likely evolve as practices and methods change over time.

1. *Develop consistent vocabulary*: Differing uses of terms such as "uncertainty characterization" can hinder the interdisciplinary collaboration which is intrinsically part of MSD. Standard definitions of approaches and a standard classification of uncertainty types can help clarify how uncertainties were and will be conceptualized and treated.

2. *Include wiring diagrams and graphical representations of modeling choices*: As discussed in Section 2.2, choices related to control volumes and coupling directionalities can limit how uncertainties can be represented and alter the resulting dynamics, such as introducing amplifying or dampening feedbacks. Contextualizing the results of an MSD analysis can be difficult without transparent communication of these choices. We prefer the inclusion of graphical representations of coupling frameworks, such as those seen in Figure 4, as they illustrate the control volume while making cycles and other connections clear.
3. *Deliberate selection of methods and data products for uncertainty analysis*: Almost every choice about the treatment of uncertainty, from calibration through scenario discovery, involves tradeoffs affecting the ability to address the driving research question. As a result, these choices should be justified based on the aims of the analysis. Documenting the motivation behind these choices, and their limitations, helps to contextualize the results and defines clear opportunities for future research.
4. *Test sensitivities to UQ assumptions about deep uncertainties*: In Section 3, we discussed the importance of the prior ranges and distributions used in an uncertainty analysis. When deep uncertainties are present and could influence calibration results through data or constraints, the use of a single input distribution to produce probabilistic projections could be misleading. When computationally tractable, one approach could be to re-calibrate the model under various realizations of deeply uncertain factors (*e.g.* , Srikrishnan et al. (2022)), but in general, a sensitivity analysis should be conducted to explore the dependence of the obtained projections on the choices made in quantifying inputs.
5. *Make model code and configurations open-source and open-access*: One category of uncertainties mentioned in Kennedy and O'Hagan (2001) that we do not explicitly account for in our taxonomy (though it is a subset of structural uncertainty in our framework) is "code uncertainty," as the specific implementation of model code can create uncertainty in outcomes. Well-documented and open-source code increases transparency around this class of uncertainties. Moreover, MSD modeling frameworks are complex, and potentially highly sensitive to specific choices of parameter values. Configuration files can be easily shared in public repositories along with the model code used for the analysis and documentation. Alignment with the FAIR principles (Wilkinson et al., 2016) for data and code sharing should also be encouraged.

Throughout our discussion, we have also identified several challenges and potential research opportunities, some of which cut across the different stages of MSD uncertainty analyses. One always-present challenge is created by the increased computational complexity of MSD models relative to single-sector models. Further advances in statistical computing via emulation or parallelized calibration methods can help navigate this tradeoff and leverage high-performance computing environments. Innovation applications of machine-learning methods could be particularly fruitful, either for use as emulators or as a direct replacement for mechanistically-motivated models (though this requires careful model construction and *post hoc* UC and SA exercises to avoid overfitting a black-box model). Advanced machine learning methods, particularly those that feature increased interpretability, could also be fruitful when applied to high-dimensional scenario classification and identification. Methods from closely-related disciplines, such as complexity science and network analysis, should also be tested for suitability in MSD applications, to further address these challenges.

Acknowledgments

The authors would like to thank Jared Smith, Mort Webster, Jim Yoon, and Klaus Keller for valuable inputs and helpful discussions. We would also like to thank Patrick Reed, Richard Moss, Erwan Monier, and Antonia Hadjimichael for their efforts on behalf of the MultiSector Dynamics Community of Practice. This work was partially supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics under Cooperative Agreements DE-SC0022141, DE-SC0016162, DE-AC05-76RL01830, DE-FG02-94ER61937, and the Global Change Intersectoral Modeling System Scientific Focus Area. The authors would like to thank Jared Smith, Mort Webster, Jim Yoon, and Klaus Keller for valuable inputs and helpful discussions. We would also like to thank Patrick Reed, Richard Moss, Erwan Monier, and Antonia Hadjimichael for their efforts on behalf of the MultiSector Dynamics Community of Practice. This work was partially supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics under Cooperative Agreements DE-SC0022141, DE-SC0016162, DE-AC05-76RL01830, DE-FG02-94ER61937, and the Global Change Intersectoral Modeling System Scientific Focus Area. Code to generate Figures 5 and 6 is available at `https://github.com/david0811/EarthsFuture_2022_MSD_Uncertainty`.

References

Abatzoglou, J. T., & Brown, T. J. (2012). A comparison of statistical downscaling methods suited for wildfire applications. *Int. J. Climatol.*, *32*(5), 772–780. doi: 10/dxdzv8

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., ... Schmidt, G. A. (2019). Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing. *Earth Syst. Dynam.*, *10*(1), 91–105. doi: 10/gnr5zw

Absar, S. M., & Preston, B. L. (2015). Extending the Shared Socioeconomic Pathways for sub-national impacts, adaptation, and vulnerability studies. *Global Environmental Change*, *33*, 83–96. doi: 10/f7j2kb

Anderies, J. M., Rodriguez, A. A., Janssen, M. A., & Cifdaloz, O. (2007). Panaceas, uncertainty, and the robust control framework in sustainability science. *Proceedings of the National Academy of Sciences*, *104*(39), 15194–15199. Retrieved 2021-12-19, from `https://www.pnas.org/content/104/39/15194` (ISBN: 9780702655104 Publisher: National Academy of Sciences Section: Research Articles) doi: 10/bf7gwm

Andrijevic, M., Crespo Cuaresma, J., Muttarak, R., & Schleussner, C.-F. (2020). Governance in socioeconomic pathways and its role for future adaptive capacity. *Nat Sustain*, *3*(1), 35–41. doi: 10/ggbxw8

Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., ... Zickfield, Kirsten (2021). Technical Summary. In Masson-Delmotte, V. et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., ... others (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1491–1508.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., ... Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Clim Change*, *3*(9), 827–832. doi: 10/f5f33r

Auer, S., Heitzig, J., Kornek, U., Schöll, E., & Kurths, J. (2015, August). The Dynamics of Coalition Formation on Complex Networks. *Scientific Reports*, *5*, 13386. doi: 10.1038/srep13386

Auffhammer, M., Hsiang, S. M., Schlenker, W., & Sobel, A. (2020). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*. doi: 10/gbn62k

Auping, W. L., Pruyt, E., de Jong, S., & Kwakkel, J. H. (2016). The geopolitical impact of the shale revolution: Exploring consequences on energy prices and rentier states. *Energy Policy*, *98*, 390–399. doi: 10/gf5bzb

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2014). Dealing with Multiple Models in System Dynamics: Perspectives on the Future of Copper. *International Journal of System Dynamics Applications (IJSDA)*, *3*(4), 17–35. doi: 10.4018/ijsda.2014100102

Bankes, S. (1993). Exploratory modeling for policy analysis. *Oper. Res.*, *41*(3), 435–449. doi: 10/c7rgcr

Barfuss, W., Donges, J. F., Lade, S. J., & Kurths, J. (2018). When optimization for governing human-environment tipping elements is neither sustainable nor safe. *Nature Communications*, *9*(1), 2354. Retrieved 2021-12-20, from `https://www.nature.com/articles/s41467-018-04738-z` (Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Applied mathematics;Sustainability Subject_term_id: applied-mathematics;sustainability) doi: 10.1038/s41467-018-04738-z

Bauer, N., Calvin, K., Emmerling, J., Fricko, O., Fujimori, S., Hilaire, J., ... van Vuuren, D. P. (2017). Shared socio-economic pathways of the energy sector – Quantifying the narratives. *Global Environmental Change*, *42*, 316–330. doi: 10/f9tspj

Berkes, F. (2007). Understanding uncertainty and reducing vulnerability: Lessons from resilience thinking. *Natural Hazards*, *41*(2), 283–295. Retrieved 2021-12-18, from `https://doi.org/10.1007/s11069-006-9036-7` doi: 10.1007/s11069-006-9036-7

Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [stat]*.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of hydrology*, *320*(1-2), 18–36. doi: 10/ccx2ks

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process.*, *6*(3), 279–298. doi: 10/bdrqd9

Biggs, R., Raudsepp-Hearne, C., Atkinson-Palombo, C., Bohensky, E., Boyd, E., Cundill, G., ... Zurek, M. (2007). Linking futures across scales: A dialog on multiscale scenarios. *Ecology and Society*, *12*(1). doi: 10/ggvvhm

Bloomfield, H. C., Brayshaw, D. J., Troccoli, A., Goodess, C. M., De Felice, M., Dubus, L., ... Saint-Drenan, Y. M. (2021). Quantifying the sensitivity of European power systems to energy scenarios and climate change projections. *Renewable Energy*, *164*, 1062–1075. doi: 10/ghfrq6

Bojke, L., Claxton, K., Palmer, S., & Sculpher, M. (2006). *Defining and Characterising Structural Uncertainty in Decision Analytic Models* (CHE Research Paper No. 9). York, UK: University of York Centre for Health Economics.

Bonabeau, E. (2002, May). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), 7280–7287. doi: 10.1073/pnas.082080899

Bouboulis, P., Theodoridis, S., Mavroforakis, C., & Evaggelatou-Dalla, L. (2015). Complex support vector machines for regression and quaternary classification. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(6), 1260–1274. doi: 10/f7cwbs

Box, G. (1979). Robustness in the strategy of scientific model building. In *Robustness in Statistics* (pp. 201–236). Elsevier. doi: 10.1016/B978-0-12-438150-6.50018-2

Bracken, L. J., & Oughton, E. A. (2006). 'What do you mean?' The importance of language in developing interdisciplinary research. *Transactions of the Institute of British Geographers*, *31*(3), 371–382. doi: 10/bf3p63

Bray, D., & von Storch, H. (2009). "Prediction" or "Projection"?: The nomenclature of climate science. *Science Communication*, *30*(4), 534–543. doi: 10/cgcqdn

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification And Regression Trees.* Boca Raton: Routledge. doi: 10.1201/9781315139470

Brown, E. K., Wang, J., & Feng, Y. (2021). US wildfire potential: A historical view and future projection using high-resolution climate data. *Environ. Res. Lett.*, *16*(3), 034060. doi: 10/gns4kh

Brugnach, M., & Pahl-Wostl, C. (2008). A broadened view on the role for models in natural resource management: Implications for model development. In C. Pahl-Wostl, P. Kabat, & J. Möltgen (Eds.), *Adaptive and Integrated Water Management: Coping with Complexity and Uncertainty* (pp. 187–203). Berlin, Heidelberg: Springer. Retrieved 2021-12-19, from https://doi.org/10.1007/978-3-540-75941-6_10 doi: 10.1007/978-3-540-75941-6_10

Brunner, S. H., & Grêt-Regamey, A. (2016). Policy strategies to foster the resilience of mountain social-ecological systems under uncertain global change. *Environmental Science & Policy*, *66*, 129–139. Retrieved 2021-12-20, from https://www.sciencedirect.com/science/article/pii/S1462901116301241 doi: 10.1016/j.envsci.2016.09.003

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, *77*(1), 34–49. Retrieved 2021-12-12, from https://www.sciencedirect.com/science/article/pii/S004016250900105X doi: 10/brmtgs

Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, *30*(11), 114007. doi: 10/gg7mwd

Burgess, M., Ritchie, J., Shapland, J., & Pielke, R. (2021). IPCC baseline scenarios have over-projected CO_2 emissions and economic growth. *Environmental Research Letters*, *16*(1). doi: 10/ghm2rr

Burnaev, E., & Panov, M. (2015). Adaptive design of experiments based on Gaussian processes. In A. Gammerman, V. Vovk, & H. Papadopoulos (Eds.), *Statistical Learning and Data Sciences* (pp. 116–125). Cham: Springer International Publishing. doi: 10/ghz2x2

Busby, J. W., Baker, K., Bazilian, M. D., Gilbert, A. Q., Grubert, E., Rai, V., ... Webber, M. E. (2021). Cascading risks: Understanding the 2021 winter blackout in texas. *Energy Research & Social Science*, *77*, 102106. doi: 10/gkcn8n

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A. M., Morse, A. P., Colón-González, F. J., ... Lloyd, S. J. (2014). Impact of climate change on global malaria distribution. *Proc Natl Acad Sci USA*, *111*(9), 3286–3291. doi: 10/f24tgt

Campolongo, F., Cariboni, J., & Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, *22*(10), 1509–1518. Retrieved 2021-12-21, from https://www.sciencedirect.com/science/article/pii/S1364815206002805 doi: 10.1016/j.envsoft.2006.10.004

Cariboni, J., Gatelli, D., Liska, R., & Saltelli, A. (2007). The role of sensitivity analysis in ecological modelling. *Ecological Modelling*, *203*(1), 167–182. Retrieved 2021-12-21, from https://www.sciencedirect.com/science/article/pii/S0304380006005734 doi: 10.1016/j.ecolmodel.2005.10.045

Carlson, J. M., & Doyle, J. (1999). Highly optimized tolerance: A mechanism for power laws in designed systems. *Phys. Rev. E*, *60*(2), 1412–1427. doi: 10/ctp7tq

Carlson, J. M., & Doyle, J. (2000). Highly optimized tolerance: Robustness and design in complex systems. *Phys. Rev. Lett.*, *84*(11), 2529–2532. doi: 10/fmbwmr

Carlsson Kanyama, A., Wikman-Svahn, P., & Mossberg Sonnek, K. (2019). "We want to know where the line is": Comparing current planning for future sea-level rise with three core principles of robust decision support approaches. *Journal of Environmental Planning and Management*, *62*(8), 1339–1358. doi: 10/gh9zbk

Carpenter, S. R. (2005). Eutrophication of aquatic ecosystems: Bistability and soil phosphorus. *Proceedings of the National Academy of Sciences*, *102*(29), 10002–10005. Retrieved 2021-12-20, from `https://www.pnas.org/content/102/29/10002` (Publisher: National Academy of Sciences Section: Biological Sciences) doi: 10.1073/pnas.0503959102

Ceres, R. L., Forest, C. E., & Keller, K. (2017). Understanding the detectability of potential changes to the 100-year peak storm surge. *Climatic Change*, *145*(1), 221–235. doi: 10/gckxr7

Chang, W., Haran, M., Applegate, P., & Pollard, D. (2016). Calibrating an ice sheet model using high-dimensional binary spatial data. *Journal of the American Statistical Association*, *111*(513), 57–72. doi: 10/f8mxq7

Chazal, F., & Michel, B. (2021). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *arXiv:1710.04019 [cs, math, stat]*.

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., ... Xiao, M. (2019). How do modeling decisions affect the spread among hydrologic climate change projections? Exploring a large ensemble of simulations across a diversity of hydroclimates. *Earth's Future*, *7*(6), 623–637. doi: 10/ghj99b

Chen, J., Brissette, F. P., & Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, *401*(3-4), 190–202. doi: 10/d6krzx

Chen, M., Vernon, C. R., Graham, N. T., Hejazi, M., Huang, M., Cheng, Y., & Calvin, K. (2020). Global land use for 2015–2100 at 0.05° resolution under diverse socioeconomic and climate scenarios. *Sci Data*, *7*(1), 320. doi: 10/gnr5zc

Chen, M., Vernon, C. R., Huang, M., Calvin, K. V., & Kraucunas, I. P. (2019). Calibration and analysis of the uncertainty in downscaling global land use and land cover projections from GCAM using Demeter (v1.0.0). *Geosci. Model Dev.*, *12*(5), 1753–1764. doi: 10/gnr5zd

Chen, Y., Irwin, E. G., & Jayaprakash, C. (2009). Dynamic modeling of environmental amenity-driven migration with ecological feedbacks. *Ecological Economics*, *68*(10), 2498–2510. Retrieved 2021-12-20, from `https://www.sciencedirect.com/science/article/pii/S0921800909000354` doi: 10.1016/j.ecolecon.2009.01.012

Christensen, P., Gillingham, K., & Nordhaus, W. (2018). Uncertainty in forecasts of long-run economic growth. *Proc Natl Acad Sci USA*, *115*(21), 5409–5414. doi: 10/gdmgt5

Cifdaloz, O., Regmi, A., Anderies, J. M., & Rodriguez, A. A. (2010). Robustness, vulnerability, and adaptive capacity in small-scale social-ecological systems: The Pumpa Irrigation System in Nepal. *Ecology and Society*, *15*(3). Retrieved 2021-12-19, from `https://www.jstor.org/stable/26268178` (Publisher: Resilience Alliance Inc.)

Cloutier, G., Joerin, F., Dubois, C., Labarthe, M., Legay, C., & Viens, D. (2015). Planning adaptation based on local actors' knowledge and participation: A climate governance experiment. *Climate Policy*, *15*(4), 458–474. Retrieved 2021-12-20, from `https://doi.org/10.1080/14693062.2014.937388` doi: 10.1080/14693062.2014.937388

Cohen, J. J., Azarova, V., Klöckner, C. A., Kollmann, A., Löfström, E., Pellegrini-Masini, G., ... Salt, D. (2021). Tackling the challenge of interdisciplinary energy research: A research toolkit. *Energy Research & Social Science*, *74*, 101966. doi: 10/gm4wh9

Cohen, J. S., & Herman, J. D. (2021). Dynamic adaptation of water resources systems under uncertainty by learning policy structure and indicators. *Water Resources Research*, *57*(11), e2021WR030433. Retrieved 2021-12-17, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2021WR030433 doi: 10.1029/2021WR030433

Cominola, A., Nguyen, K., Giuliani, M., Stewart, R. A., Maier, H. R., & Castelletti, A. (2019). Data mining to uncover heterogeneous water use behaviors from smart meter data. *Water Resources Research*, *55*(11), 9315–9333. doi: 10/gnsph2

Cooke, R. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. Oxford University Press.

Cooper, G. S., & Dearing, J. A. (2019). Modelling future safe and just operating spaces in regional social-ecological systems. *Science of The Total Environment*, *651*, 2105–2117. Retrieved 2021-12-20, from https://www.sciencedirect.com/science/article/pii/S0048969718339974 doi: 10.1016/j.scitotenv.2018.10.118

Davies, E. G. R., Kyle, P., & Edmonds, J. A. (2013). An integrated assessment of global and regional water demands for electricity generation to 2095. *Advances in Water Resources*, *52*, 296–313. doi: 10/f4nhqd

Davis, P. K., & Tolk, A. (2007). Observations on new developments in composability and multi-resolution modeling. In *2007 Winter Simulation Conference* (pp. 859–870). IEEE. doi: 10.1109/WSC.2007.4419682

de Koning, K., Filatova, T., & Bin, O. (2017). Bridging the Gap Between Revealed and Stated Preferences in Flood-prone Housing Markets. *Ecological Economics*, *136*, 1–13. doi: 10.1016/j.ecolecon.2017.01.022

Deroubaix, A., Labuhn, I., Camredon, M., Gaubert, B., Monerie, P.-A., Popp, M., ... Siour, G. (2021). Large uncertainties in trends of energy demand for heating and cooling under climate change. *Nat Commun*, *12*(1), 5197. doi: 10/gm4n3p

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., ... Ting, M. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Chang.*, *10*(4), 277–286. doi: 10/ggqkwn

De Winne, J., & Peersman, G. (2021). The adverse consequences of global harvest and weather disruptions on economic activity. *Nat. Clim. Chang.*, *11*(8), 665–672. doi: 10/gnspdb

Dolan, F., Lamontagne, J., Link, R., Hejazi, M., Reed, P., & Edmonds, J. (2021). Evaluating the economic impact of water scarcity in a changing world. *Nat Commun*, *12*(1), 1915. doi: 10/gjt94c

Dorheim, K., Link, R., Hartin, C., Kravitz, B., & Snyder, A. (2020). Calibrating Simple Climate Models to Individual Earth System Models: Lessons Learned From Calibrating Hector. *Earth and Space Science*, *7*(11), e2019EA000980. doi: 10/ghgb74

Doss-Gollin, J., Farnham, D. J., Lall, U., & Modi, V. (2021). How unprecedented was the February 2021 Texas cold snap? *Environ. Res. Lett.*, *16*(6), 064056. doi: 10/gnswvt

Doyle, J., & Carlson, J. M. (2000). Power laws, highly optimized tolerance, and generalized source coding. *Phys. Rev. Lett.*, *84*(24), 5656–5659. doi: 10/d8t9s8

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 45–97. doi: 10/gf9zvx

Duan, H., Zhang, G., Wang, S., & Fan, Y. (2019). Robust climate change research: A review on multi-model analysis. *Environ. Res. Lett.*, *14*(3), 033001. doi: 10/ggfzmk

Eason, J., & Cremaschi, S. (2014). Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers & Chemical Engineering*, *68*, 220–232. doi: 10/f6chqw

Edwards, N. R., Cameron, D., & Rougier, J. (2011). Precalibrating an intermediate complexity climate model. *Clim Dyn*, *37*(7-8), 1469–1482. doi: 10/c94d99

Efron, B. (2014). Estimation and accuracy after model selection. *Journal of the American Statistical Association*, *109*(507), 991–1007. doi: 10/gc3fv6

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.*, *1*(1), 54–75. doi: 10.1214/ss/1177013815

Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal–Journal Des Sciences Hydrologiques*, *55*(1), 58–78. doi: 10/fxhmd6

Ekblad, L., & Herman, J. D. (2021). Toward data-driven generation and evaluation of model structure for integrated representations of human behavior in water resources systems. *Water Resources Research*, *57*(2), e2020WR028148. doi: 10/gnsphz

Epstein, J. M. (1999). Agent-based computational models and generative social science. *Complexity*, *4*(5), 41–60. doi: 10.1002/(SICI)1099-0526(199905/06)4:5⟨41::AID-CPLX9⟩3.0.CO;2-F

Errickson, F. C., Keller, K., Collins, W. D., Srikrishnan, V., & Anthoff, D. (2021). Equity is more important for the social cost of methane than climate uncertainty. *Nature*, *592*(7855), 564–570. doi: 10.1038/s41586-021-03386-6

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, *9*(5), 1937–1958. doi: 10/gbpwpt

Farnham, D. J., Doss-Gollin, J., & Lall, U. (2018). Regional extreme precipitation events: Robust inference from credibly simulated GCM variables. *Water Resources Research*, *54*(6), 3809–3824. Retrieved 2021-12-19, from `https://onlinelibrary.wiley.com/doi/abs/10.1002/2017WR021318` doi: 10.1002/2017WR021318

Figueiredo, P., & Perkins, P. E. (2013). Women and water management in times of climate change: participatory and inclusive processes. *Journal of Cleaner Production*, *60*, 188–194. Retrieved 2021-12-20, from `https://www.sciencedirect.com/science/article/pii/S0959652612001011` doi: 10.1016/j.jclepro.2012.02.025

Forster, P. M., Forster, H. I., Evans, M. J., Gidden, M. J., Jones, C. D., Keller, C. A., ... Turnock, S. T. (2020). Current and future global climate impacts resulting from COVID-19. *Nat. Clim. Chang.*, *10*(10), 913–919. doi: 10/gg7s7w

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, *9*(2), 123–143. Retrieved 2021-12-17, from `https://doi.org/10.1023/A:1008894516817` doi: 10/c3cf4w

Fröhlich, J., & Knieling, J. (2013). Conceptualising climate change governance. In J. Knieling & W. Leal Filho (Eds.), *Climate Change Governance* (pp. 9–26). Berlin, Heidelberg: Springer. Retrieved 2021-12-20, from `https://doi.org/10.1007/978-3-642-29831-8_2` doi: 10.1007/978-3-642-29831-8_2

Fujimori, S., Hasegawa, T., Ito, A., Takahashi, K., & Masui, T. (2018). Gridded emissions and land-use data for 2005–2100 under diverse socioeconomic and climate mitigation scenarios. *Sci Data*, *5*(1), 180210. doi: 10/gh4339

Ge, H., Xu, K., & Ghahramani, Z. (2018). Turing: a language for flexible probabilistic inference. In *International conference on artificial intelligence and statistics, AISTATS 2018, 9-11 april 2018, playa blanca, lanzarote, canary islands, spain* (pp. 1682–1690). Retrieved from `http://proceedings.mlr.press/v84/ge18b.html`

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Stat Comput*, *24*(6), 997–1016. doi: 10/gc3f86

Gerst, M. D., Wang, P., & Borsuk, M. E. (2013). Discovering plausible energy and economic futures under global change using multidimensional scenario discovery. *Environmental Modelling & Software*, *44*, 76–86. Retrieved 2021-12-12, from https://www.sciencedirect.com/science/article/pii/S1364815212002320 doi: 10/f4x4zc

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., ... Takahashi, K. (2019). Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century. *Geosci. Model Dev.*, *12*(4), 1443–1475. doi: 10/gf8sgm

Gillingham, K., Nordhaus, W., Anthoff, D., Blanford, G., Bosetti, V., Christensen, P., ... Reilly, J. (2018). Modeling uncertainty in integrated assessment of climate change: A multimodel comparison. *Journal of the Association of Environmental and Resource Economists*, *5*(4), 791–826. doi: 10/gfknk4

Giordano, R., Pluchinotta, I., Pagano, A., Scrieciu, A., & Nanu, F. (2020). Enhancing nature-based solutions acceptance through stakeholders' engagement in co-benefits identification and trade-offs analysis. *Science of The Total Environment*, *713*, 136552. Retrieved 2021-12-20, from https://www.sciencedirect.com/science/article/pii/S0048969720300620 doi: 10.1016/j.scitotenv.2020.136552

Giorgi, F., & Gutowski, W. J. (2015). Regional dynamical downscaling and the CORDEX initiative. *Annual Review of Environment and Resources*, *40*(1), 467–490. doi: 10/gf4dvz

Giuliani, M., & Herman, J. D. (2018). Modeling the behavior of water reservoir operators via eigenbehavior analysis. *Advances in Water Resources*, *122*, 228–237. doi: 10/gfqkpf

Gorissen, D., Dhaene, T., & Turck, F. D. (2009). Evolutionary Model Type Selection for Global Surrogate Modeling. *Journal of Machine Learning Research*, *10*, 2039–2078. doi: 10.5555/1577069.1755854

Graham, N. T., Davies, E. G. R., Hejazi, M. I., Calvin, K., Kim, S. H., Helinski, L., ... Vernon, C. R. (2018). Water Sector Assumptions for the Shared Socioeconomic Pathways in an Integrated Modeling Framework. *Water Resour. Res.*, *54*(9), 6423–6440. doi: 10/gfk95h

Gramacy, R. B., & Lee, H. K. H. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics*, *51*(2), 130–145. doi: 10/c6dfdk

Grant, N., Hawkes, A., Napp, T., & Gambhir, A. (2020). The appropriate use of reference scenarios in mitigation analysis. *Nat. Clim. Chang.*, *10*(7), 605–610. doi: 10/gg3t37

Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W. M., Railsback, S. F., ... DeAngelis, D. L. (2005). Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, *310*(5750), 987–991. Retrieved 2021-12-20, from https://www.science.org/doi/10.1126/science.1116681 (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1116681

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2013). Projected atlantic hurricane surge threat from rising temperatures. *Proceedings of the National Academy of Sciences*, *110*(14), 5369–5373. doi: 10/kvj

Grogan, D. S., Burakowski, E. A., & Contosta, A. R. (2020). Snowmelt control on spring hydrology declines as the vernal window lengthens. *Environmental Research Letters*, *15*(11), 114040. Retrieved 2021-12-28, from https://iopscience.iop.org/article/10.1088/1748-9326/abbd00 doi: 10.1088/1748-9326/abbd00

Gross, T., & Feudel, U. (2006). Generalized models as a universal approach to the analysis of nonlinear dynamical systems. *Physical Review E*, *73*(1), 016205. Retrieved 2021-12-19, from https://link.aps.org/doi/10.1103/PhysRevE.73.016205 (Publisher: American Physical Society) doi: 10.1103/PhysRevE.73.016205

Guivarch, C., Rozenberg, J., & Schweizer, V. (2016). The diversity of socio-economic pathways and CO_2 emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling & Software*, *80*, 336–353. Retrieved 2021-12-17, from https://www.sciencedirect.com/science/article/pii/S1364815216300706 doi: 10/f8mqcc

Güneralp, B., & Seto, K. C. (2013). Futures of global urban expansion: Uncertainties and implications for biodiversity conservation. *Environ. Res. Lett.*, *8*(1), 014025. doi: 10/gf6379

Gurgel, A. C., Reilly, J., & Blanc, E. (2021). Agriculture and forest land use change in the continental United States: Are there tipping points? *iScience*, *24*(7), 102772. doi: 10/gmh87w

Haer, T., Botzen, W. J. W., de Moel, H., & Aerts, J. C. J. H. (2017). Integrating household risk mitigation behavior in flood risk analysis: An agent-based model approach: Integrating household risk mitigation. *Risk Analysis*, *37*(10), 1977–1992. doi: 10.1111/risa.12740

Halim, R. A., Kwakkel, J. H., & Tavasszy, L. A. (2016). A scenario discovery study of the impact of uncertainties in the global container transport system on European ports. *Futures*, *81*, 148–160. Retrieved 2021-12-12, from https://www.sciencedirect.com/science/article/pii/S0016328715001342 doi: 10/ghbjsv

Haqiqi, I., Grogan, D. S., Hertel, T. W., & Schlenker, W. (2021). Quantifying the impacts of compound extremes on agriculture. *Hydrology and Earth System Sciences*, *25*(2), 551–564. Retrieved 2021-12-28, from https://hess.copernicus.org/articles/25/551/2021/ doi: 10.5194/hess-25-551-2021

Hare, M., & Pahl-Wostl, C. (2002). Stakeholder categorisation in participatory integrated assessment processes. *Integrated Assessment*, *3*(1), 50–62. Retrieved 2021-12-20, from https://doi.org/10.1076/iaij.3.1.50.7408 doi: 10.1076/iaij.3.1.50.7408

Harmsen, M., Kriegler, E., van Vuuren, D. P., van der Wijst, K.-I., Luderer, G., Cui, R., ... Zakeri, B. (2021). Integrated assessment model diagnostics: Key indicators and model evolution. *Environ. Res. Lett.*, *16*(5), 054046. doi: 10/gkmz92

Hartin, C. A., Patel, P., Schwarber, A., Link, R. P., & Bond-Lamberty, B. P. (2015). A simple object-oriented and open-source model for scientific and policy analyses of the global climate system – Hector v1.0. *Geosci. Model Dev.*, *8*(4), 939–955. doi: 10/f7bz77

Haugen, M. A., Stein, M. L., Moyer, E. J., & Sriver, R. L. (2018). Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. *J. Climate*, *31*(20), 8573–8588. doi: 10/gfbghr

Hausfather, Z., & Peters, G. P. (2020). Emissions – The 'business as usual' story is misleading. *Nature*, *577*(7792), 618–620. doi: 10/ggjr4t

Hawkins, E., Smith, R. S., Gregory, J. M., & Stainforth, D. A. (2016). Irreducible uncertainty in near-term climate projections. *Clim Dyn*, *46*(11-12), 3807–3819. doi: 10/f8x83h

Helbing, D. (2013). Globally networked risks and how to respond. *Nature*, *497*(7447), 51–59. Retrieved 2021-11-04, from https://www.nature.com/articles/nature12047 (Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7447 Primary_atype: Reviews Publisher: Nature Publishing Group Subject_term: Complex networks;Phase transitions and critical phenomena;Socioeconomic scenarios;Sustainability Subject_term_id:

complex-networks;phase-transitions-and-critical-phenomena;socioeconomic-scenarios;sustainability) doi: 10.1038/nature12047

Helgeson, C., Srikrishnan, V., Keller, K., & Tuana, N. (2021). Why simpler computer simulation models can be epistemically better for informing decisions. *Philosophy of Science*, *88*(2), 213–233. doi: 10/gmv9w4

Henson, V. R., Cobourn, K. M., Weathers, K. C., Carey, C. C., Farrell, K. J., Klug, J. L., . . . Weng, W. (2020). A Practical Guide for Managing Interdisciplinary Teams: Lessons Learned from Coupled Natural and Human Systems Research. *Social Sciences*, *9*(7), 119. doi: 10/gnsd42

Herman, J. D., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2015). How should robustness be defined for water systems planning under change? *Journal of Water Resources Planning and Management*, *141*(10), 04015012. Retrieved 2021-12-17, from https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0000509 (Publisher: American Society of Civil Engineers) doi: 10.1061/(ASCE)WR.1943-5452.0000509

Herman, J. D., Zeff, H. B., Reed, P. M., & Characklis, G. W. (2014). Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, *50*(10), 7692–7713. Retrieved 2021-12-18, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015338 doi: 10.1002/2014WR015338

Hewitson, B. C., Daron, J., Crane, R. G., Zermoglio, M. F., & Jack, C. (2014). Interrogating empirical-statistical downscaling. *Climatic Change*, *122*(4), 539–554. doi: 10/f5sxtn

Ho, E. H., & Budescu, D. V. (2019). Climate uncertainty communication. *Nat. Clim. Chang.*, *9*(11), 802–803. doi: 10/ggbj4k

Hochrainer-Stigler, S., Colon, C., Boza, G., Poledna, S., Rovenskaya, E., & Dieckmann, U. (2020). Enhancing resilience of systems to individual and systemic risk: Steps toward an integrative framework. *International Journal of Disaster Risk Reduction*, *51*, 101868. doi: 10.1016/j.ijdrr.2020.101868

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.*, *14*(4), 382–417. doi: 10.1214/ss/1009212519

Höge, M., Guthke, A., & Nowak, W. (2019). The hydrologist's guide to Bayesian model selection, averaging and combination. *Journal of Hydrology*, *572*, 96–107. doi: 10/gfvz6r

Holtz, G., & Pahl-Wostl, C. (2012). An agent-based model of groundwater over-exploitation in the Upper Guadiana, Spain. *Reg Environ Change*, *12*(1), 95–121. doi: 10/bbjmwv

Howe, L. C., MacIperssnnis, B., Krosnick, J. A., Markowitz, E. M., & Socolow, R. (2019). Acknowledging uncertainty impacts public acceptance of climate scientists' predictions. *Nat. Clim. Chang.*, *9*(11), 863–867. doi: 10/ggbj4j

Jacob, P. E., Murray, L. M., Holmes, C. C., & Robert, C. P. (2017). Better together? Statistical learning in models made of modules. *arXiv:1708.08719 [stat]*.

Jafino, B. A., & Kwakkel, J. H. (2021). A novel concurrent approach for multiclass scenario discovery using Multivariate Regression Trees: Exploring spatial inequality patterns in the Vietnam Mekong Delta under uncertainty. *Environmental Modelling & Software*, *145*, 105177. Retrieved 2021-12-15, from https://www.sciencedirect.com/science/article/pii/S136481522100219X doi: 10/gnsdfb

Janssen, M. A., & Jager, W. (2001, December). Fashions, habits and changing preferences: Simulation of psychological factors affecting market dynamics. *Journal of Economic Psychology*, *22*(6), 745–772. doi: 10.1016/S0167-4870(01)00063-0

Joe, H. (1993). Parametric families of multivariate distributions with given margins.

Journal of multivariate analysis, *46*(2), 262–282. doi: 10/fjrb8b
Johnson, F., & Sharma, A. (2009). Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. *Journal of Climate*, *22*(16), 4373–4382. Retrieved 2021-12-19, from `https://journals.ametsoc.org/view/journals/clim/22/16/2009jcli2681.1.xml` (Publisher: American Meteorological Society Section: Journal of Climate) doi: 10.1175/2009JCLI2681.1
Jones, B., & O'Neill, B. C. (2016). Spatially explicit global population scenarios consistent with the Shared Socioeconomic Pathways. *Environ. Res. Lett.*, *11*(8), 084003. doi: 10/gbn6zr
Keilman, N. (2020). Uncertainty in population forecasts for the twenty-first century. *Annu. Rev. Resour. Econ.*, *12*(1), 449–470. doi: 10/gnr5zf
Kennedy, M. C., Anderson, C. W., Conti, S., & O'Hagan, A. (2006). Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, *91*(10), 1301–1309. doi: 10/b7v3mb
Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425–464. doi: 10/b3q4t2
Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., ... Mason, R. (2018). A comparison of methods for streamflow uncertainty estimation. *Water Resources Research*, *54*(10), 7149–7176. doi: 10/gfq7k4
Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., ... Nearing, G. (2021). Uncertainty estimation with deep learning for rainfall-runoff modelling. *Hydrology and Earth System Sciences Discussions*, 1–32. doi: 10/gns5kk
Knight, F. (1921). *Risk, uncertainty, and profit.* Boston, MA and New York, NY: Houghton Mifflin Company.
Kok, K., Pedde, S., Gramberger, M., Harrison, P. A., & Holman, I. P. (2019). New European socio-economic scenarios for climate change research: Operationalising concepts to extend the shared socio-economic pathways. *Reg Environ Change*, *19*(3), 643–654. doi: 10/gj264m
Kok, K., Rothman, D. S., & Patel, M. (2006). Multi-scale narratives from an IA perspective: Part I. European and Mediterranean scenario development. *Futures*, *38*(3), 261–284. doi: 10/b4n9n2
Koop, G., & Korobilis, D. (2009). Bayesian multivariate time series methods for empirical macroeconomics. *SSRN Journal*. doi: 10/fx96f2
Koutsoyiannis, D., & Montanari, A. (2015). Negligent killing of scientific concepts: The stationarity case. *Hydrological Sciences Journal*, *60*(7-8), 1174–1183. doi: 10/gmmdhj
Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*(12), 11344–11354. doi: 10/gg4ck8
Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. doi: 10/gghmz4
Krey, V., Guo, F., Kolp, P., Zhou, W., Schaeffer, R., Awasthy, A., ... van Vuuren, D. P. (2019). Looking under the hood: A comparison of techno-economic assumptions across national and global integrated assessment models. *Energy*, *172*, 1254–1267. doi: 10/gg3dgw
Kriegler, E., Edmonds, J., Hallegatte, S., Ebi, K. L., Kram, T., Riahi, K., ... van Vuuren, D. P. (2014). A new scenario framework for climate change research: The concept of shared climate policy assumptions. *Climatic Change*, *122*(3), 401–414. doi: 10/f5rp33

Kriegler, E., O'Neill, B. C., Hallegatte, S., Kram, T., Lempert, R. J., Moss, R. H., & Wilbanks, T. (2012). The need for and use of socio-economic scenarios for climate change analysis: A new approach based on shared socio-economic pathways. *Global Environmental Change*, *22*(4), 807–822. doi: 10/f4b9gk

Kwakkel, J. H. (2019). A generalized many-objective optimization approach for scenario discovery. *FUTURES & FORESIGHT SCIENCE*, *1*(2), e8. Retrieved 2021-12-12, from `https://onlinelibrary.wiley.com/doi/abs/10.1002/ffo2.8` doi: 10/ghkp9t

Kwakkel, J. H., Auping, W. L., & Pruyt, E. (2013). Dynamic scenario discovery under deep uncertainty: The future of copper. *Technological Forecasting and Social Change*, *80*(4), 789–800. Retrieved 2021-12-17, from `https://www.sciencedirect.com/science/article/pii/S0040162512002387` doi: 10.1016/j.techfore.2012.09.012

Kwakkel, J. H., & Jaxa-Rozen, M. (2016). Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes. *Environmental Modelling & Software*, *79*, 311–321. Retrieved 2021-12-12, from `https://www.sciencedirect.com/science/article/pii/S1364815215301092` doi: 10/gfvshb

Lade, S. J., & Gross, T. (2012). Early warning signals for critical transitions: A generalized modeling approach. *PLOS Computational Biology*, *8*(2), e1002360. Retrieved 2021-12-20, from `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002360` (Publisher: Public Library of Science) doi: 10.1371/journal.pcbi.1002360

Lade, S. J., & Niiranen, S. (2017). Generalized modeling of empirical social-ecological systems. *Natural Resource Modeling*, *30*(3), e12129. Retrieved 2021-12-20, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/nrm.12129` doi: 10.1111/nrm.12129

Lade, S. J., Tavoni, A., Levin, S. A., & Schlüter, M. (2013). Regime shifts in a social-ecological system. *Theoretical Ecology*, *6*(3), 359–372. Retrieved 2021-12-20, from `https://doi.org/10.1007/s12080-013-0187-3` doi: 10.1007/s12080-013-0187-3

Lafferty, D. C., Sriver, R. L., Haqiqi, I., Hertel, T. W., Keller, K., & Nicholas, R. E. (2021). Statistically bias-corrected and downscaled climate models underestimate the adverse effects of extreme heat on U.S. maize yields. *Commun Earth Environ*, *2*(1), 1–10. doi: 10/gnspgm

Lamontagne, J. R., Reed, P. M., Link, R., Calvin, K. V., Clarke, L. E., & Edmonds, J. A. (2018). Large ensemble analytic framework for consequence-driven discovery of climate change scenario. *Earth's Future*, *6*(3), 488–504. doi: 10/gdfdk4

Lamontagne, J. R., Reed, P. M., Marangoni, G., Keller, K., & Garner, G. G. (2019). Robust abatement pathways to tolerable climate futures require immediate global action. *Nature Climate Change*, *9*(4). Retrieved 2021-12-15, from `https://www.nature.com/articles/s41558-019-0426-8.` doi: https://doi.org/10.1038/s41558-019-0426-8

Langlois, R. N., & Cosgel, M. M. (1993). Frank Knight on risk, uncertainty, and the firm: A new interpretation. *Economic Inquiry*, *31*(3), 456–465. doi: 10/bjgdcm

Leach, N. J., Jenkins, S., Nicholls, Z., Smith, C. J., Lynch, J., Cain, M., ... Allen, M. R. (2021). FaIRv2.0.0: A generalized impulse response model for climate uncertainty and future scenario exploration. *Geosci. Model Dev.*, *14*(5), 3007–3036. doi: 10/gnr5zs

Lee, B. S., Haran, M., Fuller, R. W., Pollard, D., & Keller, K. (2020). A fast particle-based approach for calibrating a 3-D model of the Antarctic ice sheet. *The Annals of Applied Statistics*, *14*(2), 605–634. doi: 10/gns5km

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., & Spracklen, D. V. (2011).

Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmos. Chem. Phys.*, *11*(23), 12253–12273. doi: 10/cbqj49

Lee, P. M. (1989). *Bayesian statistics.* Oxford University Press London:.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., ... Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dynam.*, *11*(2), 491–508. doi: 10/gg33tn

Lempert, R. J. (2002). A new decision sciences for complex systems. *Proceedings of the National Academy of Sciences*, *99*(Supplement 3), 7309–7313. doi: 10/cs5mxr

Liu, P. R., & Raftery, A. E. (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*(1), 29. doi: 10/gnsn6q

Lopez-Cantu, T., Prein, A. F., & Samaras, C. (2020). Uncertainties in future us extreme precipitation from downscaled climate projections. *Geophysical Research Letters*, *47*(9), e2019GL086797.

Ma, D., Chen, W., Yin, X., & Wang, L. (2016). Quantifying the co-benefits of decarbonisation in China's steel sector: An integrated assessment approach. *Applied Energy*, *162*, 1225–1237. doi: 10/f75xpv

MacCracken, M. (2001). Prediction versus projection — Forecast versus possibility. *WeatherZine Guest Editorial*(26), 3–4.

Madigan, D., Raftery, A. E., Volinsky, C., & Hoeting, J. (1996). Bayesian model averaging. In *Proceedings of the AAAI workshop on integrating multiple learned models, portland, OR* (pp. 77–83).

Magliocca, N. R., & Walls, M. (2018). The role of subjective risk perceptions in shaping coastal development dynamics. *Computers, Environment and Urban Systems*, *71*, 1–13. doi: 10.1016/j.compenvurbsys.2018.03.009

Maher, N., Milinski, S., & Ludwig, R. (2021). Large ensemble climate model simulations: Introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth Syst. Dynam.*, *12*(2), 401–418. doi: 10/gmfhn5

Maher, N., Power, S. B., & Marotzke, J. (2021). More accurate quantification of model-to-model agreement in externally forced climatic responses over the coming century. *Nat Commun*, *12*(1), 788. doi: 10/gk45wf

Maier, H. R., Guillaume, J. H. A., van Delden, H., Riddell, G. A., Haasnoot, M., & Kwakkel, J. H. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? *Environmental Modelling & Software*, *81*, 154–164. Retrieved 2021-12-19, from `https://www.sciencedirect.com/science/article/pii/S1364815216300780` doi: 10.1016/j.envsoft.2016.03.014

Mankin, J. S., Lehner, F., Coats, S., & McKinnon, K. A. (2020). The value of initial condition large ensembles to robust adaptation decision-making. *Earth's Future*, *8*(10). doi: 10/gm8hvr

Manning, M., Carlson, J. M., & Doyle, J. (2005). Highly optimized tolerance and power laws in dense and sparse resource regimes. *Phys. Rev. E*, *72*(1), 016108. doi: 10/csp398

Maraun, D. (2016). Bias correcting climate change simulations - A critical review. *Curr Clim Change Rep*, *2*(4), 211–220. doi: 10/gfkbvj

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., & Popper, S. W. (2019). *Decision Making under Deep Uncertainty: From Theory to Practice.* Switzerland: Springer. Retrieved from `https://doi.org/10.1007/978-3-030-05252-2`

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., ... Tomassini, L. (2012). Tuning the climate of a global model. *J. Adv. Model. Earth Syst.*, *4*(3), n/a-n/a. doi: 10/gg2bwj

McKee, J. J., Rose, A. N., Bright, E. A., Huynh, T., & Bhaduri, B. L. (2015).

Locally adaptive, spatially explicit projection of US population for 2030 and 2050. *Proc Natl Acad Sci USA*, *112*(5), 1344–1349. doi: 10/gnr5x9

McPhail, C., Maier, H., Westra, S., Kwakkel, J., & Van Der Linden, L. (2020). Impact of scenario selection on robustness. *Water Resources Research*, *56*(9), e2019WR026515. doi: 10/ghd6bs

Meinshausen, M., Raper, S. C. B., & Wigley, T. M. L. (2011). Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration. *Atmos. Chem. Phys.*, *11*(4), 1417–1456. doi: 10/bgqxjm

Meiyappan, P., Dalton, M., O'Neill, B. C., & Jain, A. K. (2014). Spatial modeling of agricultural land use change at global scale. *Ecological Modelling*, *291*, 152–174. doi: 10/ggzmr7

Milly, P. C., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, *319*(5863), 573–574. doi: 10/bsrkbm

Moallemi, E. A., de Haan, F., Kwakkel, J., & Aye, L. (2017). Narrative-informed exploratory analysis of energy transition pathways: A case study of India's electricity sector. *Energy Policy*, *110*, 271–287. Retrieved 2021-12-17, from `https://www.sciencedirect.com/science/article/pii/S030142151730513X` doi: 10.1016/j.enpol.2017.08.019

Moallemi, E. A., Kwakkel, J., de Haan, F. J., & Bryan, B. A. (2020). Exploratory modeling for analyzing coupled human-natural systems under uncertainty. *Global Environmental Change*, *65*, 102186. doi: 10/ghkbnj

Moallemi, E. A., Kwakkel, J., de Haan, F. J., & Bryan, B. A. (2020). Exploratory modeling for analyzing coupled human-natural systems under uncertainty. *Global Environmental Change*, *65*, 102186. Retrieved 2021-06-14, from `https://linkinghub.elsevier.com/retrieve/pii/S095937802030769X` doi: 10/ghkbnj

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Compounding effects of sea level rise and fluvial flooding. *Proceedings of the National Academy of Sciences*, *114*(37), 9785–9790. doi: 10/gbxnqq

Molnar, C., Casalicchio, G., & Bischl, B. (2020). Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges. In I. Koprinska et al. (Eds.), *ECML PKDD 2020 Workshops* (pp. 417–431). Cham: Springer International Publishing. doi: 10/gjqs4k

Morgan, M. G., Henrion, M., & Small, M. (1992). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis* (Revised ed. edition ed.). Cambridge: Cambridge University Press.

Morgan, M. G., & Keith, D. W. (2008). Improving the way we think about projecting future energy use and emissions of carbon dioxide. *Climatic Change*, *90*(3), 189–215. doi: 10/dft6t7

Morris, J., Reilly, J., Paltsev, S., Sokolov, A., & Cox, K. (2022). Representing socio-economic uncertainty in human system models. *Earth's Future*, *10*(4). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021EF002239` doi: 10.1029/2021EF002239

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161–174. Retrieved 2021-12-21, from `https://www.jstor.org/stable/1269043` (Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality]) doi: 10.2307/1269043

Mosnier, A., Obersteiner, M., Havlík, P., Schmid, E., Khabarov, N., Westphal, M., ... Albrecht, F. (2014). Global food markets, trade and the cost of climate change adaptation. *Food Sec.*, *6*(1), 29–44. doi: 10/f5x596

Murakami, D., & Yamagata, Y. (2019). Estimation of gridded population and GDP scenarios with spatially explicit statistical downscaling. *Sustainability*, *11*(7), 2106. doi: 10/ghsqjx

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *PNAS*, *116*(44), 22071–22080. doi: 10/ggbhmq

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., & Stainforth, D. A. (2004). Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, *430*(7001), 768–772. doi: 10/dkhhgt

Nelsen, R. B. (2007). *An introduction to copulas.* Springer Science & Business Media.

Nelson, G. C., Valin, H., Sands, R. D., Havlík, P., Ahammad, H., Deryng, D., ... Willenbockel, D. (2014). Climate change effects on agriculture: Economic responses to biophysical shocks. *PNAS*, *111*(9), 3274–3279. doi: 10/f5wjg2

Nicholls, Z. R. J., Meinshausen, M., Lewis, J., Gieseke, R., Dommenget, D., Dorheim, K., ... Xie, Z. (2020). Reduced Complexity Model Intercomparison Project Phase 1: Introduction and evaluation of global-mean temperature response. *Geoscientific Model Development*, *13*(11), 5175–5190. doi: 10/ghq7rf

Nilsson, A. E., Bay-Larsen, I., Carlsen, H., van Oort, B., Bjørkan, M., Jylhä, K., ... van der Watt, L.-M. (2017). Towards extended shared socioeconomic pathways: A combined participatory bottom-up and top-down methodology with results from the Barents region. *Global Environmental Change*, *45*, 124–132. doi: 10/f99ptf

O'Hagan, T. (2004). Dicing with the unknown. *Significance*, *1*(3), 132–133. doi: 10/d7vn9v

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., ... Solecki, W. (2017). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, *42*, 169–180. doi: 10/f3pvqj

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, *9*(9), 3461–3482. doi: 10/f87n56

Oppenheimer, M., O'Neill, B. C., & Webster, M. (2008). Negative learning. *Climatic Change*, *89*(1-2), 155–172. doi: 10/bgbnq3

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the Earth sciences. *Science*, *263*(5147), 641–646. doi: 10/ct36kv

Pahl-Wostl, C. (2007). The implications of complexity for integrated resources management. *Environmental Modelling & Software*, *22*(5), 561–569. Retrieved 2021-12-20, from `https://www.sciencedirect.com/science/article/pii/S1364815206000417` doi: 10.1016/j.envsoft.2005.12.024

Palazzo, A., Vervoort, J. M., Mason-D'Croz, D., Rutting, L., Havlík, P., Islam, S., ... Zougmore, R. (2017). Linking regional stakeholder scenarios and shared socioeconomic pathways: Quantified West African food and climate futures in a global context. *Global Environmental Change*, *45*, 227–242. doi: 10/gb22ps

Paltsev, S., Monier, E., Scott, J., Sokolov, A., & Reilly, J. (2015). Integrated economic and climate projections for impact assessment. *Climatic Change*, *131*(1), 21–33. doi: 10/f7hzbj

Palutikof, J. P., Leitch, A. M., Rissik, D., Boulter, S. L., Campbell, M. J., Perez Vidaurre, A. C., ... Tonmoy, F. N. (2019). Overcoming knowledge barriers to adaptation using a decision support framework. *Climatic Change*, *153*(4), 607–624. doi: 10/gnr5zt

Parker, A. M., Srinivasan, S. V., Lempert, R. J., & Berry, S. H. (2015). Evaluating simulation-derived scenarios for effective decision support. *Technological Fore-*

casting and Social Change, *91*, 64–77. Retrieved 2021-12-16, from https://www.sciencedirect.com/science/article/pii/S0040162514000419 doi: 10/gc3f69

Parkes, B., Higginbottom, T. P., Hufkens, K., Ceballos, F., Kramer, B., & Foster, T. (2019). Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environ. Res. Lett.*, *14*(12), 124089. doi: 10/gns4jm

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. doi: 10/f8n6zw

Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical downscaling using localized constructed analogs (LOCA). *Journal of Hydrometeorology*, *15*(6), 2558–2585. doi: 10/f6rntd

Piontek, F., Müller, C., Pugh, T. A. M., Clark, D. B., Deryng, D., Elliott, J., … Schellnhuber, H. J. (2014). Multisectoral climate impact hotspots in a warming world. *PNAS*, *111*(9), 3233–3238. doi: 10/q4r

Polasky, S., de Zeeuw, A., & Wagener, F. (2011). Optimal management with potential regime shifts. *Journal of Environmental Economics and Management*, *62*(2), 229–240. Retrieved 2021-12-19, from https://www.sciencedirect.com/science/article/pii/S0095069611000556 doi: 10.1016/j.jeem.2010.09.004

Popp, A., Calvin, K., Fujimori, S., Havlik, P., Humpenöder, F., Stehfest, E., … van Vuuren, D. P. (2017). Land-use futures in the shared socio-economic pathways. *Global Environmental Change*, *42*, 331–345. doi: 10/f9tssx

Poulin, A., Huard, D., Favre, A.-C., & Pugin, S. (2007). Importance of tail dependence in bivariate frequency analysis. *Journal of Hydrologic Engineering*, *12*(4), 394–403. doi: 10/cjv2ph

Quinn, J. D., Hadjimichael, A., Reed, P. M., & Steinschneider, S. (2020). Can exploratory modeling of water scarcity vulnerabilities and robustness be scenario neutral? *Earth's Future*, *8*(11), e2020EF001650. doi: 10/ghkp45

Quinn, J. D., Reed, P. M., Giuliani, M., & Castelletti, A. (2017). Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems. *Water Resources Research*, *53*(8), 7208–7233. Retrieved 2021-12-17, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2017WR020524 doi: 10.1002/2017WR020524

Quinn, J. D., Reed, P. M., Giuliani, M., Castelletti, A., Oyler, J. W., & Nicholas, R. E. (2018). Exploring how changing monsoonal dynamics and human pressures challenge multireservoir management for flood protection, hydropower production, and agricultural water supply. *Water Resources Research*, *54*(7), 4638–4662. Retrieved 2021-12-16, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022743 doi: 10/gd6z59

Raftery, A. E., Zimmer, A., Frierson, D. M. W., Startz, R., & Liu, P. (2017). Less than 2°C warming by 2100 unlikely. *Nature Clim Change*, *7*(9), 637–641. doi: 10/gbwjsk

Rao, S., Klimont, Z., Smith, S. J., Van Dingenen, R., Dentener, F., Bouwman, L., … Tavoni, M. (2017). Future air pollution in the Shared Socio-economic Pathways. *Global Environmental Change*, *42*, 346–358. doi: 10/f9vds3

Reed, P. M., Hadjimichael, A., Moss, R. H., Brelsford, C., Burleyson, C. D., Cohen, S., … Yoon, J. (2022). Multisector Dynamics: Advancing the Science of Complex Adaptive Human-Earth Systems. *Earth's Future*, *10*(3), e2021EF002621. doi: 10.1029/2021EF002621

Reis, J., & Shortridge, J. (2022). Robust decision outcomes with induced correlations in climatic and economic parameters. *Mitigation and Adaptation Strate-*

gies for Global Change, *27*(1), 1–23.

Renn, O., Laubichler, M., Lucas, K., Kröger, W., Schanze, J., Scholz, R. W., & Schweizer, P.-J. (2020). Systemic Risks from Different Perspectives. *Risk Analysis*, *n/a*(n/a). doi: 10.1111/risa.13657

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., ... Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. doi: 10/brg6

Ritchie, J., & Dowlatabadi, H. (2017a). The 1000 GtC coal question: Are cases of vastly expanded future coal combustion still plausible? *Energy Economics*, *65*, 16–31. doi: 10/gbrx53

Ritchie, J., & Dowlatabadi, H. (2017b). Why do climate change scenarios return to coal? *Energy*, *140*, 1276–1291. doi: 10/gcmrdm

Robert, C. P., Elvira, V., Tawn, N., & Wu, C. (2018). Accelerating MCMC algorithms. *arXiv:1804.02719 [stat]*.

Robertson, A. W., Kushnir, Y., Lall, U., & Nakamura, J. (2015). Weather and climatic drivers of extreme flooding events over the Midwest of the United States. In *Extreme Events* (pp. 113–124). American Geophysical Union (AGU). Retrieved 2021-12-19, from https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119157052.ch9 doi: 10.1002/9781119157052.ch9

Robinson, B., Cohen, J. S., & Herman, J. D. (2020). Detecting early warning signals of long-term water supply vulnerability using machine learning. *Environmental Modelling & Software*, *131*, 104781. Retrieved 2021-12-15, from https://www.sciencedirect.com/science/article/pii/S1364815220303078 doi: 10/gg9fmt

Robinson, S., Arbez, G., Birta, L. G., Tolk, A., & Wagner, G. (2015). Conceptual modeling: Definition, purpose and benefits. In *2015 Winter Simulation Conference (WSC)* (pp. 2812–2826). doi: 10.1109/WSC.2015.7408386

Rockström, J., Steffen, W., Noone, K., Persson, r., Chapin, F. S., Lambin, E. F., ... Foley, J. A. (2009, September). A safe operating space for humanity. *Nature*, *461*(7263), 472–475. Retrieved 2021-12-20, from https://www.nature.com/articles/461472a (Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7263 Primary_atype: Special Features Publisher: Nature Publishing Group) doi: 10.1038/461472a

Rodriguez, A. A., Cifdaloz, O., Anderies, J. M., Janssen, M. A., & Dickeson, J. (2011). Confronting management challenges in highly uncertain natural resource systems: A robustness–vulnerability trade-off approach. *Environmental Modeling & Assessment*, *16*(1), 15–36. Retrieved 2021-12-19, from https://doi.org/10.1007/s10666-010-9229-z doi: 10.1007/s10666-010-9229-z

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., ... Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc Natl Acad Sci USA*, *111*(9), 3268–3273. doi: 10/f5wjnx

Rosner, A., Vogel, R. M., & Kirshen, P. H. (2014). A risk-based approach to flood management decisions in a nonstationary world. *Water Resources Research*, *50*(3), 1928–1942.

Rouwette, E. A. J. A., & Vennix, J. A. M. (2020). Group model building. In B. Dangerfield (Ed.), *System Dynamics: Theory and Applications* (pp. 91–107). New York, NY: Springer US. Retrieved 2021-12-20, from https://doi.org/10.1007/978-1-4939-8790-0_264 doi: 10.1007/978-1-4939-8790-0_264

Rozenberg, J., Guivarch, C., Lempert, R., & Hallegatte, S. (2014). Building SSPs for climate policy analysis: A scenario elicitation methodology to map the space of possible future challenges to mitigation and adaptation. *Climatic Change*, *122*(3), 509–522. Retrieved 2021-12-17, from https://doi.org/10.1007/s10584-013-0904-3 doi: 10/f5rpnz

Ruckert, K. L., Guan, Y., Bakker, A. M. R., Forest, C. E., & Keller, K. (2017). The effects of time-varying observation errors on semi-empirical sea-level projections. *Climatic Change*, *140*(3-4), 349–360. doi: 10/f9rc3z

Rudin, C. (2014). Algorithms for interpretable machine learning. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (p. 1519). New York, NY, USA: Association for Computing Machinery. doi: 10/gkrbnk

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2021). Interpretable Machine Learning: Fundamental Principles and 10 Grand Challenges. *arXiv:2103.11251 [cs, stat]*.

Sadegh, M., Moftakhari, H., Gupta, H. V., Ragno, E., Mazdiyasni, O., Sanders, B., . . . AghaKouchak, A. (2018). Multihazard scenarios for analysis of compound extreme events. *Geophysical Research Letters*, *45*(11), 5470–5480.

Sadegh, M., Ragno, E., & AghaKouchak, A. (2017). Multivariate c opula a nalysis t oolbox (mvcat): describing dependence and underlying uncertainty using a b ayesian framework. *Water Resources Research*, *53*(6), 5166–5183.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., . . . Tarantola, S. (2008). *Global sensitivity analysis: The primer.* John Wiley & Sons.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models.* Hoboken, NJ: Wiley.

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, *2*, e55. Retrieved from `https://doi.org/10.7717/peerj-cs.55` doi: 10.7717/peerj-cs.55

Sanderson, M., Arbuthnott, K., Kovats, S., Hajat, S., & Falloon, P. (2017). The use of climate information to estimate future mortality from high ambient temperature: A systematic literature review. *PLoS ONE*, *12*(7), e0180369. doi: 10/gbmsfr

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., . . . Sugihara, G. (2009, September). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53–59. doi: 10.1038/nature08227

Schennach, S. M. (2016). Recent advances in the measurement error literature. *Annual Review of Economics*, *8*(1), 341–377. doi: 10/gfghb8

Schmidt, R. (2005). Tail dependence. In *Statistical tools for finance and insurance* (pp. 65–91). Springer.

Schwanitz, V. J. (2013). Evaluating integrated assessment models of global climate change. *Environmental Modelling & Software*, *50*, 120–131. doi: 10/gnswvx

Serinaldi, F., & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, *77*, 17–36. doi: 10/gg3s55

Seto, K. C., Guneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, *109*(40), 16083–16088. doi: 10/f4bhs4

Sexton, D. M. H., Karmalkar, A. V., Murphy, J. M., Williams, K. D., Boutle, I. A., Morcrette, C. J., . . . Vosper, S. B. (2019, July). Finding plausible and diverse variants of a climate model. Part 1: Establishing the relationship between errors at weather and climate time scales. *Clim Dyn*, *53*(1-2), 989–1022. doi: 10/gnr5zv

Sexton, D. M. H., Murphy, J. M., Collins, M., & Webb, M. J. (2012). Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. *Clim Dyn*, *38*(11-12), 2513–2542. doi: 10/bt5s9g

Shortridge, J. E., & Zaitchik, B. F. (2018). Characterizing climate change risks by linking robust decision frameworks and uncertain probabilistic projections. *Climatic Change*, *151*(3-4), 525–539. doi: 10/gnr6gq

Sisson, S. A., Fan, Y., & Beaumont, M. A. (2018). Overview of Approximate

Bayesian Computation. In *Handbook of Approximate Bayesian Computation* (pp. 3–54). Chapman and Hall/CRC.

Small, M. J., & Fischbeck, P. S. (1999). False precision in Bayesian updating with incomplete models. *Human and Ecological Risk Assessment: An International Journal*, *5*(2), 291–304. doi: 10/dgjtd8

Smith, A. D., Dłotko, P., & Zavala, V. M. (2021). Topological data analysis: Concepts, computation, and applications in chemical engineering. *Computers & Chemical Engineering*, *146*, 107202. doi: 10/gk89ch

Srikrishnan, V., Guan, Y., Tol, R. S. J., & Keller, K. (2022). Probabilistic projections of baseline twenty-first century CO2 emissions using a simple calibrated integrated assessment model. *Climatic Change*, *170*(3), 37. doi: 10.1007/s10584-021-03279-7

Srikrishnan, V., & Keller, K. (2021). Small increases in agent-based model complexity can result in large increases in required calibration data. *Environmental Modelling & Software*, *138*, 104978. doi: 10/gnsph3

Stan Development Team. (2019). *Stan Modeling Language Users Guide and Reference Manual 2.28.*

Stedinger, J. R. (1983). Design events with specified flood risk. *Water Resources Research*, *19*(2), 511–522.

Stedinger, J. R., Vogel, R. M., Foufoula-Georgiou, E., & Maidment, D. R. (1993). Handbook of hydrology. *McGraw-Hill, Inc., New York, USA*.

Stedinger, J. R., Vogel, R. M., Lee, S. U., & Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (glue) method. *Water resources research*, *44*(12).

Steinmann, P., Auping, W. L., & Kwakkel, J. H. (2020). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change*, *156*, 120052. Retrieved 2021-12-12, from `https://www.sciencedirect.com/science/article/pii/S0040162519302380` doi: 10/gnrj6c

Steinschneider, S., McCrary, R., Wi, S., Mulligan, K., Mearns, L. O., & Brown, C. (2015). Expanded decision-scaling framework to select robust long-term water-system plans under hydroclimatic uncertainties. *Journal of Water Resources Planning and Management*, *141*(11), 04015023. Retrieved 2021-12-19, from `https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0000536` (Publisher: American Society of Civil Engineers) doi: 10.1061/(ASCE)WR.1943-5452.0000536

Steinschneider, S., Ray, P., Rahat, S. H., & Kucharski, J. (2019). A weather-regime-based stochastic weather generator for climate vulnerability assessments of water systems in the Western United States. *Water Resources Research*, *55*(8), 6923–6945. Retrieved 2021-12-19, from `https://onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024446` doi: 10.1029/2018WR024446

Stevanović, M., Popp, A., Lotze-Campen, H., Dietrich, J. P., Müller, C., Bonsch, M., . . . Weindl, I. (2016). The impact of high-end climate change on agricultural welfare. *Science Advances*, *2*(8), e1501452. doi: 10/gfxx5r

Stewart, A. J., & Plotkin, J. B. (2014, December). Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences*, *111*(49), 17558–17563. doi: 10.1073/pnas.1408618111

Su, Y., Kern, J. D., Reed, P. M., & Characklis, G. W. (2020). Compound hydrometeorological extremes across multiple timescales drive volatility in California electricity market prices and emissions. *Applied Energy*, *276*, 115541. doi: 10/gg53zp

Taner, M. Ü., Ray, P., & Brown, C. (2019). Incorporating multidimensional probabilistic information into robustness-based water systems planning. *Water Resources Research*, *55*(5), 3659–3679. doi: 10/gnr6gr

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. doi: 10/b7xgw2

Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A.*, *365*(1857), 2053–2075. doi: 10/dx4f9f

Thrasher, B., Maurer, E. P., McKellar, C., & Duffy, P. B. (2012). Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrology and Earth System Sciences*, *16*(9), 3309–3314. Retrieved 2021-12-28, from `https://hess.copernicus.org/articles/16/3309/2012/` doi: 10.5194/hess-16-3309-2012

Trutnevyte, E., Guivarch, C., Lempert, R., & Strachan, N. (2016). Reinvigorating the scenario technique to expand uncertainty consideration. *Climatic Change*, *135*(3), 373–379. Retrieved 2021-12-16, from `https://doi.org/10.1007/s10584-015-1585-x` doi: 10/f8jv5j

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. doi: 10/gwh

van Ruijven, B. J., De Cian, E., & Sue Wing, I. (2019). Amplification of future energy demand growth due to climate change. *Nat Commun*, *10*(1), 2762. doi: 10/ggf74m

van Vliet, M. T. H., Wiberg, D., Leduc, S., & Riahi, K. (2016). Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nature Clim Change*, *6*(4), 375–380. doi: 10/bbsp

Vangheluwe, H., de Lara, J., & Mosterman, P. J. (2002). An introduction to multi-paradigm modeling and simulation. In *Proceedings of the AIS'2002 conference (AI, Simulation and Planning in High Autonomy Systems)* (p. 13). Lisboa, Portugal.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*, *27*(5), 1413–1432. doi: 10/gdj2kz

Veldkamp, T. I. E., Wada, Y., Aerts, J. C. J. H., & Ward, P. J. (2016). Towards a global water scarcity risk assessment framework: Incorporation of probability distributions and hydro-climatic variability. *Environ. Res. Lett.*, *11*(2), 024006. doi: 10/gkq6hk

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., . . . Krysanova, V. (2017). Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Climatic Change*, *141*(3), 419–433. doi: 10/f928z4

Vogel, R. M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P. K., Hooper, R. P., & Matalas, N. C. (2015). Hydrology: The interdisciplinary science of water. *Water Resources Research*, *51*(6), 4409–4430. Retrieved 2021-12-19, from `https://onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017049` doi: 10.1002/2015WR017049

Vollset, S. E., Goren, E., Yuan, C.-W., Cao, J., Smith, A. E., Hsiao, T., . . . Murray, C. J. L. (2020). Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: A forecasting analysis for the Global Burden of Disease Study. *The Lancet*, *396*(10258), 1285–1306. doi: 10/gg53xk

von Lampe, M., Willenbockel, D., Ahammad, H., Blanc, E., Cai, Y., Calvin, K., . . . van Meijl, H. (2014). Why do global long-term scenarios for agriculture differ? An overview of the AgMIP global economic model intercomparison. *Agricultural Economics*, *45*(1), 3–20. doi: 10/f5qd49

Vrugt, J. A., Ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12). doi: 10/cj9k99

Walker, W., Harremoës, P., Rotmans, J., van der Sluijs, J., van Asselt, M., Janssen, P., & Krayer von Krauss, M. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, *4*(1), 5–17. doi: 10/fmwdhs

Walsh, C., Lazarou, N.-J., Traut, M., Price, J., Raucci, C., Sharmina, M., ... Smith, T. (2019). Trade and trade-offs: Shipping in changing climates. *Marine Policy*, *106*, 103537. doi: 10/ggzgfk

Wang, P., Gerst, M. D., & Borsuk, M. E. (2013). Exploring energy and economic futures using agent-based modeling and scenario discovery. In H. Qudrat-Ullah (Ed.), *Energy Policy Modeling in the 21st Century* (pp. 251–269). New York, NY: Springer. Retrieved 2021-12-12, from `https://doi.org/10.1007/978-1-4614-8606-0_13` doi: 10.1007/978-1-4614-8606-0_13

Wasserman, L. (2018). Topological data analysis. *Annu. Rev. Stat. Appl.*, *5*(1), 501–532. doi: 10/gc7rx3

Watson, K., Cross, R., Jones, M., Buttorff, G., Granato, J., Pinto, P., ... Vellejo, A. (2021). *The Winter Storm of 2021* (Tech. Rep.). Houston, TX: Hobby School of Public Affairs, University of Houston. Retrieved 2021-12-17, from `https://uh.edu/hobby/winter2021/storm.pdf`

Wear, D. N., & Prestemon, J. P. (2019). Spatiotemporal downscaling of global population and income scenarios for the United States. *PLoS ONE*, *14*(7), e0219242. doi: 10/gf5pd2

Weber, S. L. (2010). The utility of Earth system Models of Intermediate Complexity (EMICs). *WIREs Clim Change*, *1*(2), 243–252. doi: 10/c7px26

Webster, M., Babiker, M., Mayer, M., Reilly, J., Harnisch, J., Hyman, R., ... Wang, C. (2001). *Uncertainty in emissions projections for climate models* (Tech. Rep. No. 79). Cambridge, MA: MIT Joint Program.

Webster, M., Sokolov, A. P., Reilly, J. M., Forest, C. E., Paltsev, S., Schlosser, A., ... Jacoby, H. D. (2012). Analysis of climate policy targets under uncertainty. *Climatic Change*, *112*(3-4), 569–583. Retrieved 2021-12-28, from `http://link.springer.com/10.1007/s10584-011-0260-0` doi: 10.1007/s10584-011-0260-0

Werner, A. T., & Cannon, A. J. (2016). Hydrologic extremes – An intercomparison of multiple gridded statistical downscaling methods. *Hydrol. Earth Syst. Sci.*, *20*(4), 1483–1508. doi: 10/f8rhnv

Wiel, K., Wanders, N., Selten, F. M., & Bierkens, M. F. P. (2019). Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World. *Geophys. Res. Lett.*, *46*(4), 2093–2102. doi: 10/ggm2q8

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci Data*, *3*(1), 160018. doi: 10/bdd4

Wilson, C., Guivarch, C., Kriegler, E., van Ruijven, B., van Vuuren, D. P., Krey, V., ... Thompson, E. L. (2021). Evaluating process-based integrated assessment models of climate change mitigation. *Climatic Change*, *166*(1-2), 3. doi: 10/gjvt5x

Wong, T. E. (2018). An integration and assessment of multiple covariates of nonstationary storm surge statistical behavior by Bayesian model averaging. *Advances in Statistical Climatology, Meteorology and Oceanography*, *4*(1/2), 53–63. doi: 10/ghjzz4

Wong, T. E., Bakker, A. M. R., & Keller, K. (2017). Impacts of Antarctic fast dynamics on sea-level projections and coastal flood defense. *Climatic Change*, *144*(2), 347–364. doi: 10/gbwjcp

Wong, T. E., Klufas, A., Srikrishnan, V., & Keller, K. (2018). Neglecting model structural uncertainty underestimates upper tails of flood hazard. *Environmental Research Letters*, *13*(7), 074019. doi: 10/ghkp8v

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. (2018). Using stacking to average

Bayesian predictive distributions. *Bayesian Anal.*, *13*(3). doi: 10/gddmx4
Yoon, J., Klassert, C., Selby, P., Lachaut, T., Knox, S., Avisse, N., ... Gorelick, S. M. (2021). A coupled human–natural system analysis of freshwater security under climate and population change. *PNAS*, *118*(14). doi: 10/gjpnp6
Zoraghein, H., & O'Neill, B. (2020). A spatial population downscaling model for integrated human-environment analysis in the United States. *Demographic Research*, *43*, 1483–1526. Retrieved 2021-12-28, from https://www.demographic-research.org/volumes/vol43/54/ doi: 10.4054/DemRes.2020.43.54
Zscheischler, J., Fischer, E. M., & Lange, S. (2019). The effect of univariate bias adjustment on multivariate hazard estimates. *Earth Syst. Dynam.*, *10*(1), 31–43. doi: 10/gm87n7
Zumwald, M., Knüsel, B., Baumberger, C., Hirsch Hadorn, G., Bresch, D. N., & Knutti, R. (2020). Understanding and assessing uncertainty of observational climate datasets for model evaluation using ensembles. *WIREs Climate Change*, *11*(5), e654. doi: 10/gns4jj
Zurek, M. B., & Henrichs, T. (2007). Linking scenarios across geographical scales in international environmental assessments. *Technological Forecasting and Social Change*, *74*(8), 1282–1295. doi: 10/cqvmbq