Posted on 22 Nov 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/essoar.170067181.14466541/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Multi-actor, multi-impact scenario discovery of consequential narrative storylines for human-natural systems planning

Antonia Hadjimichael[1], Patrick M. Reed[2], Julianne D. Quinn[3], Chris R Vernon[4], and Travis Thurber[5]

[1]Penn State University
[2]Cornell University
[3]University of Virginia
[4]Pacific Northwest National Laboratory
[5]Pacific Northwest National Lab

November 22, 2023

Abstract

Scenarios have emerged as valuable tools in managing complex human-natural systems, but the traditional approach of limiting focus on a small number of predetermined scenarios can inadvertently miss consequential dynamics, extremes, and diverse stakeholder impacts. Exploratory modeling approaches have been developed to address these issues by exploring a wide range of possible futures and identifying those that yield consequential vulnerabilities. However, vulnerabilities are typically identified based on aggregate robustness measures that do not take full advantage of the richness of the underlying dynamics in the large ensembles of model simulations and can make it hard to identify key dynamics and/or narrative storylines that can guide planning or further analyses. This study introduces the FRamework for Narrative Scenarios and Impact Classification (FRNSIC; pronounced "forensic'): a scenario discovery framework that addresses these challenges by organizing and investigating consequential scenarios using hierarchical classification of diverse outcomes across actors, sectors, and scales, while also aiding in the selection of narrative storylines, based on system dynamics that drive consequential outcomes. We present an application of this framework to the Upper Colorado River Basin, focusing on decadal droughts and their water scarcity implications for the basin's diverse users and its obligations to downstream states through Lake Powell. We show how FRNSIC can explore alternative sets of impact metrics and drought dynamics and use them to identify narrative drought storylines, that can be used to inform future adaptation planning.

Multi-actor, multi-impact scenario discovery of consequential narrative storylines for human-natural systems planning

Antonia Hadjimichael[1,2], Patrick M. Reed[3], Julianne D. Quinn[4], Chris R. Vernon[5], Travis Thurber[5]

[1]Department of Geosciences, The Pennsylvania State University, State College, PA, USA
[2]Earth and Environmental Systems Institute (EESI), The Pennsylvania State University, State College, PA, USA
[3]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[4]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA
[5]Atmospheric Sciences & Global Change, Pacific Northwest National Laboratory, Richland, WA, USA

Key Points:

- Introduce a hierarchical classification framework for scenario discovery, to identify diverse stakeholder impacts and consequential dynamics.
- Demonstrate the framework in the Upper Colorado River Basin with hundreds of stakeholders and complex human-natural system interactions.
- The framework improves understanding and selection of narrative drought storylines through their effects on user- and basin-scale impacts.

Corresponding author: Antonia Hadjimichael, `hadjimichael@psu.edu`

Abstract

Scenarios have emerged as valuable tools in managing complex human-natural systems, but the traditional approach of limiting focus on a small number of predetermined scenarios can inadvertently miss consequential dynamics, extremes, and diverse stakeholder impacts. Exploratory modeling approaches have been developed to address these issues by exploring a wide range of possible futures and identifying those that yield consequential vulnerabilities. However, vulnerabilities are typically identified based on aggregate robustness measures that do not take full advantage of the richness of the underlying dynamics in the large ensembles of model simulations and can make it hard to identify key dynamics and/or narrative storylines that can guide planning or further analyses. This study introduces the FRamework for Narrative Scenarios and Impact Classification (FRNSIC; pronounced "forensic"): a scenario discovery framework that addresses these challenges by organizing and investigating consequential scenarios using hierarchical classification of diverse outcomes across actors, sectors, and scales, while also aiding in the selection of narrative storylines, based on system dynamics that drive consequential outcomes. We present an application of this framework to the Upper Colorado River Basin, focusing on decadal droughts and their water scarcity implications for the basin's diverse users and its obligations to downstream states through Lake Powell. We show how FRNSIC can explore alternative sets of impact metrics and drought dynamics and use them to identify narrative drought storylines, that can be used to inform future adaptation planning.

Plain Language Summary

Scenario analysis is a useful tool for assessing the impacts of future conditions or alternative strategies. Focusing on a small number of predetermined scenarios can, however, limit our understanding of key uncertainties, and fail to represent diverse stakeholder impacts. Approaches such as exploratory modeling have been developed to address these issues by exploring a wide range of possible futures and system perspectives. These approaches often involve large simulation experiments with their own interpretability challenges. So, on one hand, we recognize the need to utilize large ensembles of hypothesized changes, but on the other hand, each additional dimension considered makes it more difficult to convey actionable insights. We introduce the FRamework for Narrative Scenarios and Impact Classification (FRNSIC; pronounced "forensic"), a scenario discovery framework that helps users identify narrative scenarios that capture key system dynamics and as well as important outcomes. We demonstrate its application to the Upper Colorado River Basin, focusing on decadal droughts and their water scarcity implications for the basin's diverse users and its obligations to downstream states through Lake Powell. We explore alternative impact metrics and dynamics, identifying narrative storylines with significant impacts, which can be used in future planning efforts to adapt to these stressed conditions.

1 Introduction

Understanding and managing human-natural systems confronting change remains an open challenge, as they are highly complex systems with deep uncertainties shaping their candidate futures (Elsawah et al., 2020; Reed, Hadjimichael, Moss, et al., 2022; Schlüter et al., 2012). The interactions and feedbacks between human and natural components, resources, actors, and institutions create nested systems-of-systems that operate at and across multiple scales (Iwanaga et al., 2021). Holistically attending to such complexity and advancing our understanding of such systems requires approaches that transcend disciplinary framings and traditional approaches (Wyborn et al., 2019). Pervasive deep uncertainties are also present in these systems, due to incomplete or contested expert knowledge on system boundaries or key system processes and drivers (Marchau et al., 2019; Moallemi, Zare, et al., 2020). Finally, the multiple and often conflicting objectives of various stakeholders in these systems further complicate the identification of relevant knowledge that engages diverse worldviews to inform their management (Kasprzyk et al., 2013).

Scenario analysis has become increasingly important in understanding and planning for human-natural systems, as scenarios present useful tools in dealing with some of these challenges (Groves

& Lempert, 2007; Moss et al., 2010; O'Neill et al., 2014; Pedersen et al., 2022; Van Ruijven et al., 2023). Scenarios help us assess and communicate the potential severity of hypothesized conditions and deep uncertainties, for example the impacts of a changing climate on local systems (e.g., Vahmani et al. (2022)). They can also act as reference cases for comparison and negotiation of alternative strategies to follow, for example quantifying deviations from historical conditions as a result of different stressors and human actions (e.g., Cohen et al. (2022)). Or they can help capture system complexity in narrative aggregate storylines, for example as they are used by the Intergovernmental Panel on Climate Change to communicate the impacts of alternative emissions pathways (e.g., IPCC (2023)).

An important challenge surrounding the use of scenarios is the number of candidate future states considered, as well as the conditions used to establish their relevance. Using a small number of deterministic future states has well-documented limitations, especially arising from the presence of internal variability (Hawkins & Sutton, 2009; Lehner & Deser, 2023), deep uncertainty about the future (Lempert et al., 2006; Quinn et al., 2020), and the adaptive complexity of human-natural systems (Markolf et al., 2018; Reed, Hadjimichael, Moss, et al., 2022; Simpson et al., 2021). Focusing only on the interests of, or the impacts to, a small number of actors carries its own challenges that undermine successfully engaging with the diverse perspectives of affected stakeholders. Groves and Lempert (2007) point out that *a priori* specification of a small set of "interesting" scenarios to aid narrative clarity, in absence of broader exploratory analysis, might inappropriately narrow the focus to the concerns and values of those involved in crafting them. They might not necessarily be salient with the diverse stakeholders affected, who might view the particular set of selected scenarios as biased or arbitrary. Moreover, the broad array of human as well as natural uncertainties that could shape consequential future outcomes increases the risk that a limited focus on a few specified scenarios would miss key insights (Moallemi, Kwakkel, et al., 2020).

Recognizing the myopic nature of a limited set of pre-specified scenarios or futures, there have been significant advancements in the domain of exploratory modeling (Bankes, 1993) and scenario discovery (Bryant & Lempert, 2010; Groves & Lempert, 2007). As reviewed by Moallemi, Kwakkel, et al. (2020) these approaches focus on the exploration of large ensembles of possible futures and the *a posteriori* identification of consequential scenarios. These approaches have largely been articulated in support of decision making under deep uncertainty methods, such as Robust Decision Making (RDM; Lempert et al. (2003)) and its Many-Objective extension (MORDM; Kasprzyk et al. (2013)), Dynamic Adaptive Policy Pathways (Haasnoot et al., 2013; Schlumberger et al., 2022), Info-Gap (Ben-Haim, 2006), and Decision Scaling (Brown et al., 2012). They structure large exploratory ensemble experiments to investigate diverse hypothesized drivers of change and classify the resulting "states of the world" (SOWs) based on whether they have consequential outcomes for the system's stakeholders. This process of ensemble classification and identification of a subset of consequential SOWs is termed scenario discovery (Bryant & Lempert, 2010; Groves & Lempert, 2007; Steinmann et al., 2020). As such, these exploratory modeling frameworks introduce more quantitative rigor by examining the space of possible future uncertainty and associated consequences more fully (Lempert et al., 2006). Put simply, a broader array of "what if" questions are engaged before selecting scenarios.

Past studies have reviewed and offered taxonomies of these frameworks (Herman et al., 2015; Kwakkel & Haasnoot, 2019; Moallemi, Zare, et al., 2020); at their core they all encompass the following central elements: elucidation or generation of alternative management or planning actions, exploration of alternative SOWs (potential futures or uncertainties), quantification of performance (typically a measure of "robustness"), and vulnerability or tradeoff analysis, where consequential scenarios are identified and strategies are selected, according to the quantified performance. Robustness metrics are used to rank how well systems perform based on their expected value (Wald, 1950), regret (Savage, 1951), or satisficing criteria (Simon, 1956), as extensively reviewed by McPhail et al. (2018). There is an expansive body of literature on scenario discovery that has compared the value and effects of using robustness metrics across a variety of problems and case studies to demonstrate that the choice of metric can have critical implications for which SOWs are deduced as consequential (i.e., which scenarios are selected for further inspection; Herman et al. (2015); Maier et al. (2016); McPhail et al. (2018); Sunkara et al. (2023)). Hadjimichael, Quinn, Wilson, et al. (2020) show that systems with diverse stakeholders introduce additional challenges to defining the appropriate metric to classify consequential SOWs and select a subset of ensemble members that warrant follow-on analysis given their consequential outcomes or challenging dynamics. In systems with many actors, the choice of a singular aggregated metric can ignore asymmetries in stakeholder values and agency (Franssen, 2005), and implicitly suppress the diverse scenario impacts on different users from more explicit consideration in planning (Fletcher et al., 2022). Recognizing this limitation, some studies have looked at multi-actor robustness tradeoffs, by applying the same criterion to the performance of different actors (Gold et al., 2019; Herman et al., 2014; Trindade et al., 2017). Others have applied gradients of a threshold or criterion as a way of capturing different levels of acceptability or relation to past experience to different stakeholders (Bonham et al., 2022; Hadjimichael, Quinn, & Reed, 2020; Hadjimichael, Quinn, Wilson, et al., 2020; Quinn et al., 2020).

A related challenge that arises from aggregation when defining robustness criteria for target levels of system performance is that they can collapse the temporal or spatial dynamics of a scenario into a single outcome by which each scenario is to be classified. For example, there could be a case were two scenarios produce the same average supply of a resource, but one shows substantial temporal variation whereas the other hovers around its mean. One could make the case that we can simply include an additional metric of variance to further disaggregate, but we might be interested in the overall dynamic behavior of the system or other qualitative information, for example common oscillation patterns of different scenarios, the presence of stable equilibria or tipping points. Using metrics that temporally aggregate these dynamics limits the use of this information (Hadjimichael, Reed, & Quinn, 2020). As a result, authors have proposed methods that can temporally classify the simulation dynamics themselves, instead of some aggregated outcome (e.g., Steinmann et al., 2020).

A final important consideration surrounding the development and use of scenarios relates to conveying actionable information. We face challenges in maintaining their narrative capacities (Krauß, 2020; Krauß & Bremer, 2020), encouraging the usability of climate impact findings (Lemos & Morehouse, 2005; Lemos et al., 2012), and producing consequential insights that hold direct beneficial value to the dependent human and environmental systems. Literature on co-production and cognitive research highlights that the way information is presented to and processed by its users is important to how they understand and choose to use it (Calvo et al., 2022; S. Lorenz et al., 2015). Lemos et al. (2012), for example, point out that relating new findings (e.g., potential future impacts on one's crop) to past experiences and memories (e.g., impacts of a past significant drought to one's crop) can help connect that information to their analytical and experiential processing abilities. Highlighting connections to relevant personal experiences also fosters the usability of the new findings. Literature on narrative scenarios highlights that the use of local narratives can give meaning to abstract scientific information and is central to making sense of what it means to live within a changing climate (Krauß & Bremer, 2020).

As such, tools like storylines and narrative scenarios can aid in making connections between new scientific findings and past relevant experiences, as well as form the basis of new analysis iterations (Cork et al., 2006; Krauß, 2020; Lempert et al., 2006; Shepherd et al., 2018). Narrative scenarios can indeed be derived from a RDM analysis (Lempert, 2019). For example, analysts, stakeholders and decision makers can use the discovered scenarios to more closely investigate system processes and dynamics, such as key reasons that lead to failure (e.g., Popper et al. (2009)), or use them as a basis for reiteration and evaluation of new strategies or stressors of interest (e.g., Groves (2005); Lempert and Groves (2010)). Such facilitated reiteration, however, is difficult to achieve with the large and complex ensembles of SOWs that modern state-of-the-art exploratory modeling analyses rely on. For example, in recent past work we generated 10,000 SOWs, within each of which we computed thousands of performance metrics for different stakeholders and different criteria (Hadjimichael, Quinn, Wilson, et al., 2020). Similarly, Gold et al. (2022); Shi et al. (2023); Trindade et al. (2020) and others all use ensemble sizes of thousands to millions of scenarios. As already mentioned, the size of these experiments is an attempt to better capture the space of possible futures and consider relevant uncertainties, recognizing the combinatorial scale of significant factors in highly complex coupled human-natural systems and to better guide a more holistic understanding of highly consequential decision-relevant outcomes.

Large ensemble exploratory modeling therefore creates a tension: on one hand, we understand that there is a large number of interacting processes, candidate futures and alternative framings we should explore, and we thus need to create large ensembles of these hypothesized changes to investigate with our models. On the other hand, each additional dimension considered makes the results of the analysis more intricate and more difficult to convey actionable insights[1]. We argue that making large ensemble experiments more actionable is indeed possible, but requires innovations in how the resulting outcomes and their driving dynamics are organized, investigated, and communicated. This can be complemented with new data visualizations that allow users to navigate hierarchical levels of classification of ensemble outputs, and to zoom in on specific narrative scenarios of interest and investigate their dynamics.

The present study addresses the challenges and needs for large ensemble exploratory modeling discussed above by contributing a new scenario discovery framework: the FRamework for Narrative Scenarios and Impact Classification (FRNSIC)—pronounced "forensic". FRNSIC aims to provide actionable narrative clarity without sacrificing the quantitative rigor of large ensemble experiments. It aids the identification of consequential scenarios through the application of nested criteria that capture hierarchical relationships between sectors, actors, and/or scales, each reflective of different relevant impacts for the stakeholders concerned. We can explore multiple influential system states and hierarchically support the discovery of the diverse conditions that control stakeholder-relevant impacts. The emerging narrative scenarios are clustered not only on their resulting impacts but also on the underlying dynamic scenarios that drive them. As a result, we aid decision makers in discovering smaller sets of narrative scenarios, or dynamic storylines, that represent both complex mappings between a large space of input uncertainty and the large space of resulting outcomes. At the same time, these storylines also maintain a locally-embedded meaning, as well as the potentially critical temporal dynamics that lead to consequential outcomes.

The remaining sections are organized as follows. Section 2 presents the FRNSIC scenario discovery framework and provides an overview of the main component stages of its application. Section 3 details our application of the framework within the Upper Colorado River Basin, with a particular focus on the issue of better understanding plausible drought extremes and their system impacts. Finally, Section 4 presents the outcomes of the application of FRNSIC, and Section 5 provides conclusions as well as opportunities for future extensions.

2 Methodological Framework

Exploratory modeling and its connection to robustness frameworks has been extensively reviewed in several past studies (Herman et al., 2015; Kwakkel & Haasnoot, 2019; Moallemi, Zare, et al., 2020). We refer readers to these publications for a comprehensive introduction to the background literature in this area. Following the terminology established by these authors, this paper introduces a new scenario discovery framework in support of robustness analysis, FRNSIC, begins by following the same broad steps that are common across all exploratory modeling and robustness approaches (framing, system evaluation across many states, quantification of performance, and scenario discovery), and then adds new steps for multi-trait classification and storyline discovery (see Fig. 1).

The *Problem Framing* Stage (I) is critical across all exploratory modeling and robustness frameworks to ensure the decision relevance of their results. During this phase, analysts and stake-

[1] In Aesop's fable about The Fox and the Cat, the fox boasts of hundreds of ways of escaping its enemies, while the cat only has one. When they hear a pack of the hounds approaching, the cat scampers up a tree and hides, while the fox in its confusion gets caught up by the hounds. The moral of the fable is that it is "Better [to have] one safe way than a hundred on which you cannot reckon".

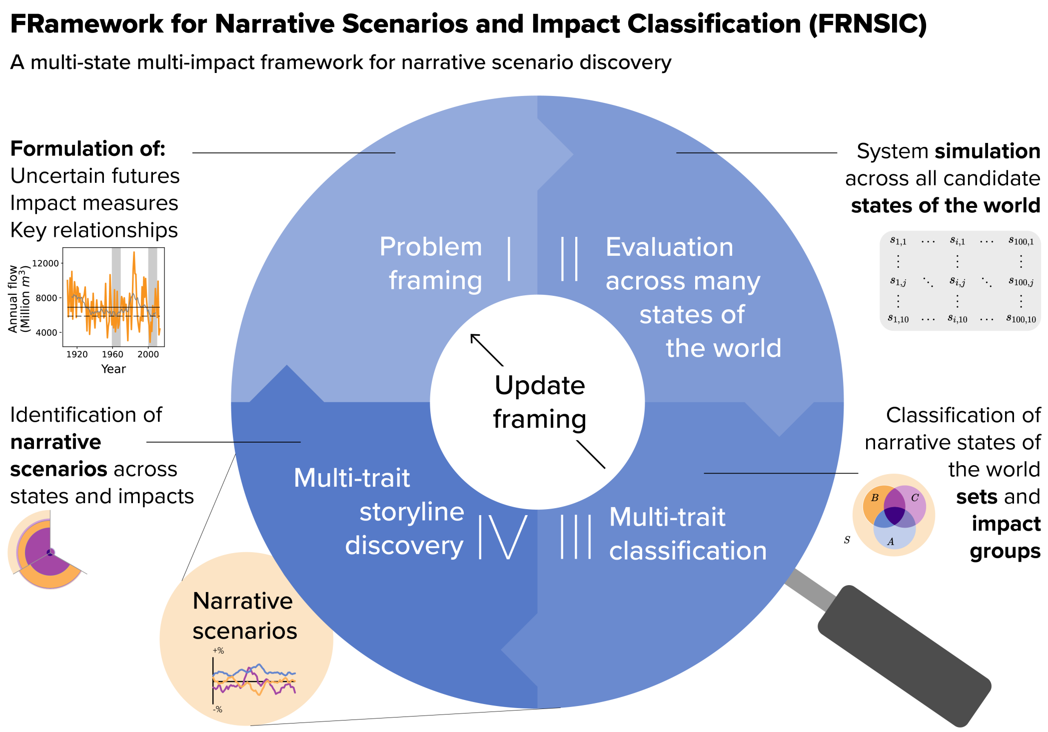

Figure 1. The four stages of the multi-state, multi-impact framework for narrative scenario discovery, FRNSIC.

holders define the key factors in the analysis: system goals (sometimes articulated as objectives) and metrics of performance toward these goals; alternative actions or system configurations that can be taken to affect said metrics; the uncertainties that may affect the connection between actions and metrics; and the relationships (which often take the form of simulation models) between actions, uncertainties, and metrics (Lempert, 2019). Procedures for eliciting these elements have been articulated based on the 'XLRM' matrix (Lempert et al., 2003): exogenous uncertainties ('X'), policy levers ('L'), relationships ('R'), and metrics ('M'). Here, we adopt the same intention behind the problem framing stage. Presenting framing as a distinct stage in these frameworks is intentional; framing choices made during this stage should be transparently articulated, especially as they shape subsequent stages of analysis. The framing could also be updated as performance across states is quantified and consequential conditions are uncovered. In the Upper Colorado River Basin case study, presented in the following section, this stage is used to investigate the water scarcity context of the system and frame how SOWs should be appropriately generated, the dynamic states of consequence (e.g., decadal droughts), and impact metrics.

Exploratory modeling is a central focus of Stage II of FRNSIC (*Evaluation across many states of the world*), evaluating the system, via a simulation model, across alternative actions or policies or system configurations, and across alternative SOWs. Moallemi, Zare, et al. (2020) term these steps "generation of decisions" and "generation of scenarios", respectively. The same authors, as well as others, have also broadly drawn a distinction here between two alternative strategies: exploration and search. Methods that rely on exploration systematically sample points across both the decision space and the SOWs and evaluate their consequences. As such, they rely on the careful designs of experiments which are used to set up simulation frameworks with the minimum computational cost to answer specific questions (Reed, Hadjimichael, Malek, et al., 2022). Exploration techniques produce insights about the global properties of the decision and the uncertainty space (plausible SOWs), such as how much increase in water demand would result in increased supply shortages (e.g., Hadjimichael, Quinn, Wilson, et al. (2020)).

Methodologies that rely on search, in contrast, draw on optimization-based tools to actively identify points with particular properties, such as "how much should we invest in infrastructure to maximize profits?" (searching for high-performing actions) or "how much more warming would cause insufferable heatwaves in our city?" (searching for a subset of consequential SOWs). These approaches typically rely on multi- or many-objective optimization algorithms (Kasprzyk et al., 2013; Kwakkel, 2019). FRNSIC remains agnostic to which of the two strategies is employed at this stage, as both allow us to analyze a system over many of its potential states, and use those states to classify and discover narrative scenarios of interest. If optimization methods were to be used in this case, one would have to ensure that the temporal dynamics of each simulation are carefully maintained, for subsequent analysis in the following stages. In the Upper Colorado River Basin case study, we are using exploration methods.

The core novel contributions of FRNSIC lie in Stages III and IV, where performance is quantified (III *Multi-trait classification*) and consequential scenarios are discovered (IV *Multi-trait storyline discovery*). To clarify these contributions, let us first briefly overview how performance quantification and scenario discovery are traditionally performed. In virtually all applications (see reviews from Marchau et al. (2019); Moallemi, Kwakkel, et al. (2020); Moallemi, Zare, et al. (2020)), the analysts establish one or a set of criteria against which they compare or rank order the performance of different policies or actors across SOWs (i.e., one or more robustness performance metrics). To address some of the challenges brought about by multi-actor systems discussed in Section 1, a variety of robustness metrics or different performance thresholds might also be used (e.g., Hadjimichael, Quinn, Wilson, et al. (2020)). A SOW is then classified as being consequential subject to meeting or failing to meet the specific requirements tied to the robustness metric(s) specified. A tacit effect of using the most commonly employed robustness metrics (e.g., satisficing or regret metrics; see discussions in Herman et al. (2015); McPhail et al. (2018)) is that the temporal dynamics of the underlying sampled SOWs are ignored, and in their place, the analysis is focused on the classification of SOWs as being consequential or not based on a summarizing statistic of those dynamics. A benefit of this approach is that a single quantitative value is much more easily communicated than a vector of them across the duration of the realization. A shortfall of it is that policies or actors achieving similar performance on a particular robustness metric may do so through a diversity of temporal dynamics that lead to tradeoffs on other metrics. Consequently, the temporal dynamics are critical drivers that shape whether or not specified performance metrics are met, and are therefore critical to understanding robustness tradeoffs. The importance of temporal dynamics and their properties is strongly emphasized in the socio-ecological systems and system dynamics bodies of literature (e.g., Gotts et al. (2019); Schlüter et al. (2012)), the data science literature (e.g., Aghabozorgi et al. (2015)), and more recently emphasized in both the exploratory modeling (Steinmann et al., 2020) and the climate risk (de Ruiter & Van Loon, 2022) literature.

In Stage III of FRNSIC (Fig. 1), we use simple set theory to explore the dynamic properties of the sampled SOWs, not restricting focus solely on robustness performance measures (which we also classify, as discussed below). This creates collections of SOWs that exhibit certain dynamic properties (e.g., significant variability, particular equilibria or oscillation patterns) irrespective of the performance outcomes they generate (e.g., impacts to system users). In other words, we create collections of SOWs that specifically focus on the dynamic processes of the system and their defining characteristics, as separate defining properties from the performance in each SOW. The reason this distinction is important is that the same dynamic properties do not always result in the same system impacts, and vice versa. For example, two droughts of the same severity might occur, but have different water scarcity impacts. On the other hand, two SOWs might result in similar outcomes (e.g., 20% of water demands cannot be met), but the underlying dynamics that produce them are different.

These dynamic properties can be identified in several ways. They might be specified *a priori*; for example, if the computational design of experiments is set up to specifically generate them.

Such is the case for some of our prior work evaluating water scarcity, where we used parametric approaches to synthetically generate hydrologic conditions and those conditions were sampled so as to specifically exhibit certain properties (e.g., larger variability; Hadjimichael, Quinn, Wilson, et al. (2020); Quinn et al. (2020)). Dynamic properties can also be discovered *a posteriori*. For example, Steinmann et al. (2020) applied time series clustering to identify collections of SOWs that exhibit similar temporal behaviors. Lastly, dynamic properties can also be analytically or numerically calculated. For example, Hadjimichael, Reed, and Quinn (2020) analytically derived behavioral properties of each SOW that pertained to the system's stability and number of equilibria, and used said properties to create semantically meaningful collections of SOWs that described certain behavior modes. Clarifying the diversity of temporal dynamics that underlie a large ensemble of exploratory modeling simulations using a small number of semantically meaningful sets can facilitate their narrative application later on, when the scenario discovery process identifies consequential SOWs. Utilizing these behavioral properties to discover narrative scenarios in conjunction with using performance criteria to discover impactful scenarios can help analysts illuminate the root causes of vulnerability in a system (Steinmann et al., 2020).

Beyond using set theory to order and better understand the underlying dynamics in sampled SOWs, Stage III of FRNSIC also hierarchically classifies diverse robustness performance measures that can be defined across different actors, scales, and sectors. Hierarchy, as used here, refers to the addition of new criteria (e.g., "$reliability \geq 90\%$" AND "$costs \leq \$100$"), not the preferential weighting of one criterion over another. Even though it is not typically discussed in terms of set theory, classifying sampled SOWs in terms of whether they meet a certain criterion in effect partitions them into specific subsets (or collections) of the broader set of all SOWs, such that for every criterion there exists a conditional set of SOWs for which the condition holds and a complement set for which it does not. For multiple performance criteria, we can therefore create multiple such subsets to denote whether an impact criterion is met, as well as look at the intersections of the conditional sets for the combinations of SOWs where multiple criteria are met simultaneously. This type of algebraic structure is formally referred to as a Boolean algebra or a Boolean lattice and describes relationships between the partitioned subsets of an overall set that result from applying binary classification operations (Drapeau et al., 2016; Priss, 2021). In essence, we can use these binary operations to identify increasingly nested subsets of consequential SOWs that meet or fail to meet additional performance criteria. For complex human-natural systems confronting change that impact a large suite of scales, sectors and stakeholders, FRNSIC's hierarchical classification greatly broadens the diversity of interests and performance concerns that shape our inferences on robustness.

Finally, in Stage IV of FRNSIC (*Multi-trait storyline discovery*), these two sets-of-sets—one created to describe fundamental dynamics and one created to classify the decision-relevant outcomes from hierarchical performance criteria—are combined to guide the discovery of consequential storyline narrative scenarios that can be used to structure further dialogues for the diverse ways a system may confront change. As emphasized in Section 1, achieving narrative meaning in the context of high dimensionality and complexity requires advances in how the information is organized (in our case with hierarchical sets) and presented. For the latter, we contribute a modified version of the stacked hive plot (Krzywinski et al., 2012), which allows us to visualize the resulting sets-of-sets in a single panel figure. Hive plots adapt parallel coordinate plots (Inselberg, 2009; Wegman, 1990) to a radial arrangement, compacting the layout and making the connections easier to follow. Hive plots typically rely on a three-axis model, with the total circle area being uniformly divided between all segments (the areas between two axes). As demonstrated in this study, the three axes we utilize reflect three dynamic properties of the SOWs generated. More than three dimensions can be used, but by having only three axes, hive plots accommodate connections (lines) between each axis pair, without having to cross the axes themselves. With more than three axes this can only be achieved if connections are only drawn between neighboring axes, or if axes are duplicated at multiple positions. This negatively impacts the interpretability of the figure, which defies the aim of creating meaningful and salient narratives, central to our framework. The originators of the figure indeed discourage its use with more than three axes (Krzywinski et al., 2012), and most common applications in network science (e.g., Engle and Whalen (2012)) and gene sequencing (e.g., Yang et al. (2017)) also only use three axes. Furthermore, the compactness of this figure allows us to generate multiple panels reflecting alternative dynamic properties or robustness performance measures, in a "small multiples" visualization (Tufte, 1990). Combining many small visualizations simultaneously allows the reader to compare the separate panels and look for patterns or outliers in the matrix of visuals, and facilitates presentation and storytelling of large amounts of data in a single figure (van den Elzen & van Wijk, 2013).

In the following sections, we present an example application of the key stages of FRNSIC on a multi-actor, institutionally complex human-natural system: the Upper Colorado River Basin within the state of Colorado (henceforth abbreviated to UCRB). Section 3.1 introduces the study area and model utilized. Section 3.2 presents an overview of the problem (FRNSIC Stage I - Problem Framing) and articulates the main challenges surrounding the characterization of drought extremes and investigation of their impacts. Section 3.3 details the generation of hydroclimatic SOWs (FRNSIC Stage II - Evaluation Across Many States of the World) through the use of exploratory modeling, allowing us to account for said challenges. Section 3.4 (FRNSIC Stage III - Multi-trait Classification of States of the World) details how the drought dynamics of the hydroclimatic SOWs are classified into sets of dynamic properties, as illustrated in Fig. 5, as well as how the impacts generated by the SOWs are classified into impact sets, as illustrated in Fig. 7. Finally, Section 3.5 (FRNSIC Stage IV - Multi-trait storyline discovery) describes how the two sets-of-sets come together through the use of hive plots to enable the exploration of narrative drought storylines that summarize both consequential impacts and key drought dynamics.

3 The Upper Colorado River Basin case study implementation

3.1 Study Area and Model

Most of the aforementioned innovations and developments in the domain of exploratory modeling and scenario discovery have been in the area of water resources. Water resources systems are archetypal of the types of challenges we face around understanding and planning in coupled human-natural systems: environmental, social, infrastructural, and institutional complexity; contested views and objectives over how resources should be allocated; increasing stress and deep uncertainty about future stressors. Western river basins in the United States in particular, and the Colorado River more specifically, are under significant hydrologic stress, following decades of aridification (Smith et al., 2022; State of Colorado, 2015; McCoy et al., 2022; Whitney et al., 2023). The Colorado River basin is institutionally complex, with a nested set of compacts, laws, and regulations that dictate water allocation for over 40 million people and 22,000 km^2 of agricultural land (Bureau of Reclamation, 2012). The River has been experiencing prolonged water scarcity and aridification for the past two decades, accumulating to a "crisis" in recent years (Gerlak & Heikkila, 2023). A megadrought that started in 1999 (Overpeck & Udall, 2020), and continues as of the time of writing, has caused major reservoirs on the river to decline to dangerously low levels, prompting the U.S. Department of Interior to call for unprecedented cuts in water usage among the states that depend on it (Flavelle & Rojanasakul, 2023).

Understanding plausible future drought hazards and planning for their impacts in these human-natural systems presents several challenges. First, internal hydroclimatic variability and non-stationarity challenge how we identify extreme events, such as decadal-scale or longer drought hazards (AghaKouchak et al., 2022; Hoylman et al., 2022; Lehner & Deser, 2023; Stevenson et al., 2022). Internal variability, arising from interactions across non-linear processes intrinsic to the hydroclimate, means that any given process has inherent irreducible uncertainty in its manifestation and that our historical observations are only one limited sample of the diverse dynamics that could occur. In the context of hydroclimatic dynamics, internal variability is a fundamentally stochastic process that has been shown to produce magnitudes of variation in flood and drought extremes that exceed historical experiences (Fischer et al., 2021) or that are comparable to anthropogenic climate change at the decadal scale (Deser et al., 2016). Even in regions of the world with long observational records, the full extent of internal variability cannot be estimated from the single realization of the stochastic hydroclimatic process represented by the observed record that exists (Woodhouse & Overpeck, 1998; Woodhouse et al., 2006). Extending the record with reconstructed paleoclimate information can improve on this representation, but has its own methodological limitations, such as underestimating the variance in the data (Quinn et al., 2020), and reducing interpretability (Ault et al., 2014). Lastly, the stochastic nature of internal variability poses important communication challenges, as it necessitates the use of probabilistic descriptions of the occurrence of critical events, instead of simple deterministic predictions of them (Lehner & Deser, 2023).

Non-stationarity in time and space is another a well-recognized challenge. Non-stationarity reflects conditions where the statistical properties of a variable (e.g., its distribution and correlation with other variables) may change over time (Slater et al., 2021). It is especially consequential in how it transforms the occurrence of extreme events like floods, droughts, and heatwaves (AghaKouchak et al., 2022; Berghuijs et al., 2019; R. Lorenz et al., 2019; Sun et al., 2021). Yet, until the recent decade, non-stationarity has not been accounted for in conventional planning for water resources or extreme events. Instead, planners have relied on observed historical time series of streamflow or other hydroclimatic variables for future planning (Yang et al., 2021). In fact, even current drought monitoring products such as the United States Drought Monitor rely on historical distributions of these events to establish their classification (Hoylman et al., 2022), as do the flood maps generated by the Federal Emergency Management Agency (Hobbins et al., 2021). This is largely due to large epistemic uncertainties around the form of future non-stationarity. Even under stationary conditions, when complex systems are concerned, it is often impossible to be in full knowledge of the true model of the system under consideration (Beven, 1993). In the case of non-stationary systems and the development of models for them, the problem is even more challenging because of the larger number of parameters involved (i.e., both the base statistics and also how they are changing) and large number of alternative ways non-stationarity can be included in the analysis (Salas et al., 2018).

Lastly, the complexity of human systems further compounds the challenges in understanding and planning for the potential impacts of droughts. In systems like the Colorado River, institutions, engineered infrastructure, and large numbers of actors come together to shape who gets water, how much, and when, as well as who has to get shorted when conditions are dry. Our understanding of drought-induced water scarcity has evolved to recognize the importance of the feedbacks between anthropogenic and natural system processes, which shape the production and distribution of drought effects and their implications for humans and the environment (AghaKouchak et al., 2023; Lukat et al., 2023; Savelli et al., 2022). Human-natural systems around the world, and especially systems that are heavily managed, have developed strategies to reduce their exposure and vulnerability to drought hazards (Kreibich et al., 2022; Smith et al., 2022). For example, the states that depend on Colorado River water develop and regularly update drought preparedness plans that help them project their water availability and needs, and adjust their operations accordingly (e.g., Arizona Department of Water Resources, 2022; California Natural Resources Agency, 2022; Colorado Water Conservation Board & Department of Natural Resources, 2018). These efforts at higher levels of governance, as well as less-coordinated state or local planning efforts, all must consider the institutional water rights context of the Prior Appropriation Doctrine (Kenney, 2005). Water rights create a complex hierarchy for managing scarcity and strongly shape how a regional drought may differentially affect each water right holder in the river (Hadjimichael, Quinn, Wilson, et al., 2020).

The particular implementation of Prior Appropriation in each state, as well as other local characteristics and needs of each watershed, have prompted states like Colorado to develop water planning and management processes at different scales: at the state-wide scale (i.e., the state of Colorado's Water Plan; State of Colorado (2023)), and the local river basin scale (i.e., the Basin Implementation Plans developed by a local Basin Roundtable for each of the nine basins within the state, e.g., CWCB and CDWR (2022)). To facilitate communication and comparisons, the Colorado Water Plan and the local Basin Implementation Plans all utilize a set of five future scenarios of water scarcity in the state (State of Colorado, 2023), each being a narrative summary of how different drivers of scarcity might evolve in the future (e.g., increased agricultural needs, reduced supply). These five scenarios carry the same challenges discussed in Section 1, but they are not necessarily consequential or relevant at the local level. In other words, each local basin might not necessarily be equally sensitive to the key drivers each scenario assumes, nor have impacts at the same magnitudes. So even though the local impacts of these five scenarios are evaluated in the Basin Implementation Plans, the analysis might inadvertently miss other locally consequential scenarios, that are still plausible but not part of the set of five.

Within this context, we demonstrate how the FRNSIC scenario discovery framework could be utilized by the local Basin Roundtable responsible for water resources planning for the UCRB. The Colorado Basin Roundtable[2] was established in 2005 by Colorado state legislature and is charged with water planning for the UCRB and with implementing the state-wide Water Plan locally. Its members include not only state representatives, like from the Colorado Division of Water Resources and the Colorado Water Conservation Board, but also representatives from the agricultural sector, the industrial sector, domestic water suppliers, environmental and recreation entities, as well as other interested citizens. Besides planning, the Colorado Basin Roundtable also plays a significant role in allocating state funds to enact its water priorities within the UCRB. The diversity of representative members of the Colorado Basin Roundtable is crucial to its ability to address the diverse goals and challenges the UCRB faces.

The UCRB contains the headwaters of the Colorado River with its outflow moving into Utah to deliver water to Lake Powell. As with all western basins in the state, it is bound by the Colorado River Compact, which allocates 9.3 km^3 (7.5 million acre-feet) per year to the Upper Basin states (Colorado, New Mexico, Utah, and Wyoming)—the state of Colorado is allotted 51.75% of that amount. Another 9.3 km^3 is divided among the Lower Basin states (California, Arizona, and Nevada), and Upper Basin states have to deliver water to Lake Powell to meet that requirement. Increasingly frequent and more persistent severe drought conditions inhibit the ability of Upper Basin states and subbasins like the UCRB to make these deliveries. Quantifying the potential effects of future water scarcity and drought on UCRB deliveries to Lake Powell is therefore a key concern for the Colorado Basin Roundtable, as outlined in their Basin Implementation Plan (CWCB & CDWR, 2022). Within the UCRB, several thousand water rights support diversions for agriculture, municipal water supply, industrial production, power generation, as well as recreational uses (Fig. 2). While most of the consumptive use of water within the basin supports agricultural production, large exports of water leave the basin to support urban centers on the east slope, where most of Colorado's population resides. Water to all these users is allocated through the Prior Appropriation Doctrine, which prioritizes users in terms of seniority and limits the received amount of water for each user to their decreed "beneficial use" (Kenney, 2005). Along with the water availability itself, this institutional hierarchical network plays the most fundamental role in shaping the dynamics of water scarcity vulnerabilities across the water rights holders. Given the central importance of the agricultural sector in this basin, quantifying impacts to local agricultural water users is another critical concern highlighted in the Basin Implementation Plan (CWCB & CDWR, 2022).

All these key aspects are captured in Colorado's Decision Support System (CDSS), a collection of databases, data management tools, and models, created to support water resources planning in Colorado's major water basins, including the UCRB (Malers et al., 2001). The principal modeling tool of the CDSS is the State of Colorado's Stream Simulation Model (StateMod), a generic network-based water system model for water accounting and allocation. StateMod was developed to support comprehensive assessments of water demand and supply, as well as reservoir operations, in all the major subbasins within the state of Colorado (Parsons & Bennett, 2006; CWCB, 2012). The model replicates each basin's unique application of the Prior Appropriation doctrine and accounts for all of the consumptive uses of water within each basin. To achieve this, StateMod utilizes detailed historic demand and operation records, which include water right information for all consumptive water diversions, water structures (i.e., wells, ditches, reservoirs, and tunnels), as well as streamflow and other hydroclimatic information. The model also includes estimates of agricultural water consumption based on soil moisture, crop type, irrigated acreage,

[2] https://www.coloradobasinroundtable.org/

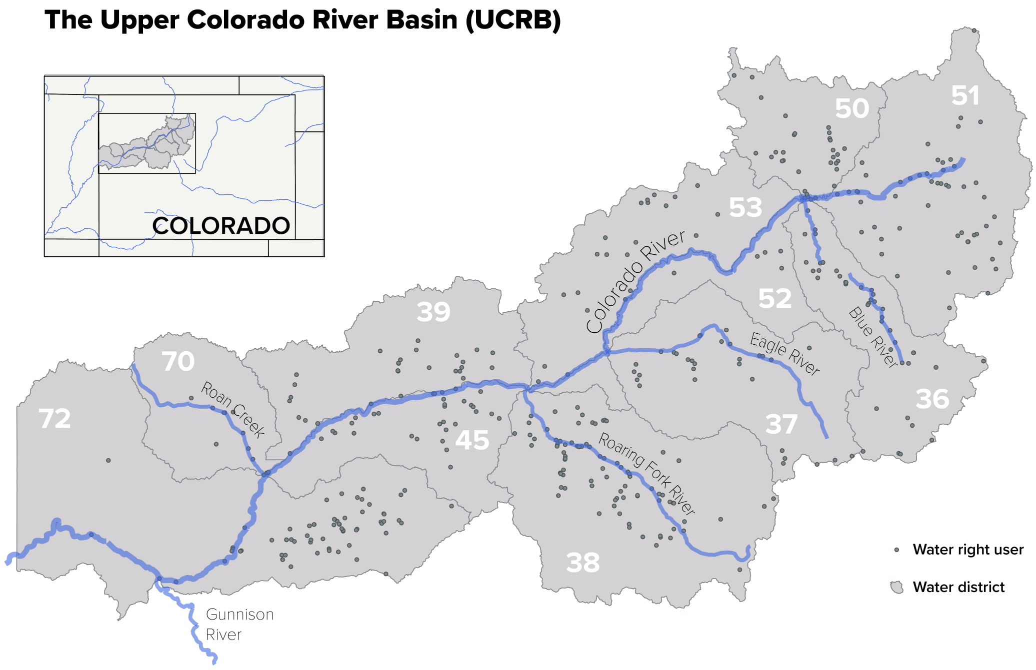

Figure 2. The Upper Colorado River Basin within the state of Colorado (UCRB). The points indicate all modeled diversion points in StateMod (primarily irrigation). The numbered areas indicate water districts.

and conveyance and application efficiencies for each individual irrigation unit in the region. Using these highly-resolved inputs, StateMod accounts for the water consumption of all users in each basin, through their water right allocation. It therefore allows us to simulate and assess the impacts of potential future changes in hydrology, water demands, or operations on all the represented water users in each basin. For the purposes of this study, we focus on the specific StateMod implementation for the UCRB.

The remainder of this section outlines a demonstrative use of FRNSIC that could support the types of coordinated planning studies overseen by groups like the Colorado Basin Roundtable to explore and discover locally consequential and plausible scenarios for their basin. The UCRB system is an ideal testbed to make generalizable advances in exploratory modeling literature, particularly with regard to addressing the dimensionality introduced by multi-actor systems, the importance of capturing behavioral dynamics, and the challenge of providing clarity when selecting consequential drought storyline narratives for further consideration in planning efforts, as discussed in Section 1. The planning application demonstrated here is hypothetical, but stays close to the key water planning concerns articulated in the Basin Implementation Plan, as well as other literature on drought-induced water scarcity in the region, as elaborated below.

3.2 Stage I - Problem Framing

Throughout this study, we classify hydrologic drought conditions as occurring when there is a half a standard deviation departure from the historical average streamflow at the Colorado-Utah state line over the period 1909-2013 (i.e., $\mu - 0.5\sigma$), following the examples of Ault et al. (2014, 2016); Diffenbaugh et al. (2015); Naumann et al. (2018). We apply this classification on naturalized streamflow and identify decadal-scale droughts using an 11-year rolling mean (more details on how the classification is performed are provided in Section 3.4.1). Multidecadal droughts can similarly be identified using longer windows, such as 25 years (Meko et al., 2007) or 35 years (Ault et al., 2014). Applying this classification to the historical streamflow observations for the

UCRB, we see two decadal-scale droughts: one in the 1960s and one starting in the early 2000s (Fig. 3 (a)). This estimate is consistent with other literature sources that classify decadal droughts in the reconstructed paleo record in this region (i.e., one or two instances of decadal drought per century; see Ault et al. (2014); Woodhouse and Overpeck (1998)). The identification of plausible decadal-scale drought hazards is confounded by the presence of: (a) irreducible, internal variability, (b) non-stationarity, and (c) deeply uncertain past and future streamflow dynamics beyond the currently available gauged record (i.e., paleo conditions or future climate change).

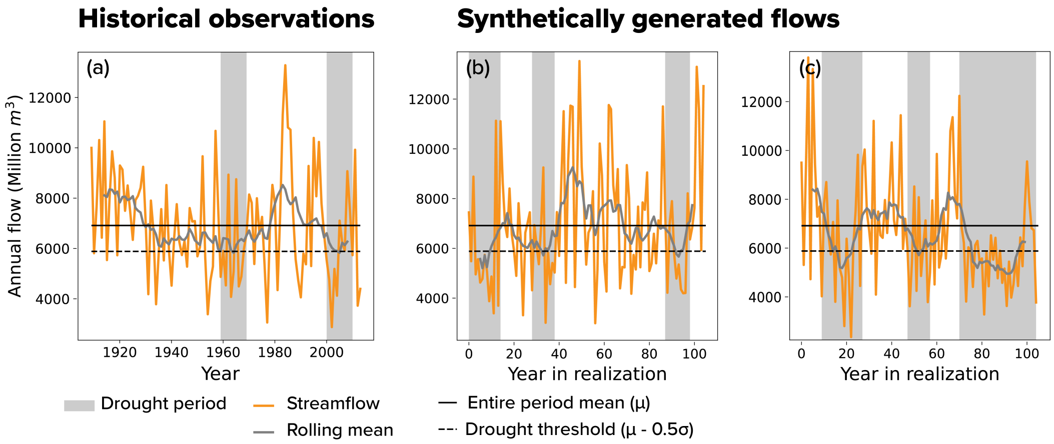

Figure 3. Hydrologic drought identification for the UCRB (a) Decadal-scale droughts identified using historic observations; (b-c) Decadal-scale droughts identified using synthetically generated streamflow. We note that the mean and standard deviation of the distribution remain the same, so does the average annual volumetric drought threshold, at $5,884Mm^3$, computed over the full 105-year record length.

Internal variability complicates the identification of droughts, even in a stationary context (Cook et al., 2022). For example, even if we establish that the moments of the historical streamflow distribution stay the same in the future and use those distributions to inform planning, we might underestimate the true frequency of drought events (i.e., the events that cross the drought threshold in this case). Fig. 3 demonstrates this effect. Here, we compare the drought classification applied to the historic observations of streamflows (Fig. 3 (a)) and the same classification applied to synthetically generated streamflows that have the same base statistical properties as the last century's historical observations (Fig. 3 (b-c)). The synthetic streamflows are created using a synthetic streamflow generator so as to exhibit the same distributional moments for the occurrence of wet years and dry years, as well the probability of transitioning between the two states, through the use of a Hidden Markov Model (see more details in Section 3.3). We see that even though only two decadal droughts are identified in the historical record (using a drought threshold of $5,884Mm^3$), simulating alternative plausible synthetic realizations from the same distributions can give rise to more decades of drought. This undermines the validity of using the historical streamflow observations to deterministically to infer expectations for the frequency of extreme drought conditions (e.g., that only one or two decadal droughts are to be expected in a century), when in fact the same process can give rise to conditions that are much worse.

Non-stationarity makes it challenging to establish appropriate reference conditions (e.g., the drought threshold used above) when seeking to identify decadal drought hazards for a hydroclimatic system with evolving wet and dry regimes (Mondal & Mujumdar, 2015; Slater et al., 2021). The solution often recommended is to use rolling windows of time and establish moving baseline thresholds (Hoylman et al., 2022). Fig. 4 demonstrates this idea and highlights the potential variability of drought thresholds when looking across 60-year rolling windows of streamflows. For reference, the average annual volumetric drought threshold calculated using the entire period of data (105 years) is $5,884Mm^3$ (indicated by the dashed line in Fig. 4 (b)). Starting with the early 1900s, conditions were very wet (top density plot in Fig. 4 (a)) and so the drought threshold established using that early 20th century 60-year window is at a much larger annual average volume (top right point in Fig. 4 (b)). As a result, 30 years in the record since that initial 60-year window would fall below the drought threshold established in this period (Fig. S1). We note that these 30 years are identified in decadal periods, they therefore reflect three decadal droughts, not 30 drought years dispersed throughout the 105-year period. The early 1900s were also the period during which the Colorado River Compact was signed. Moving across time (downward in the figure), we see that the changing streamflow statistics substantially shift the drought thresholds one would establish, down to ≈ 5,540 M m^3 in the most recent window. Using these drier-period thresholds that are substantially lower than that of the entire period (i.e., all points to the left of the dashed line in Fig. 4 (b)) would result in no years classified as droughts (Fig. S1). [3]

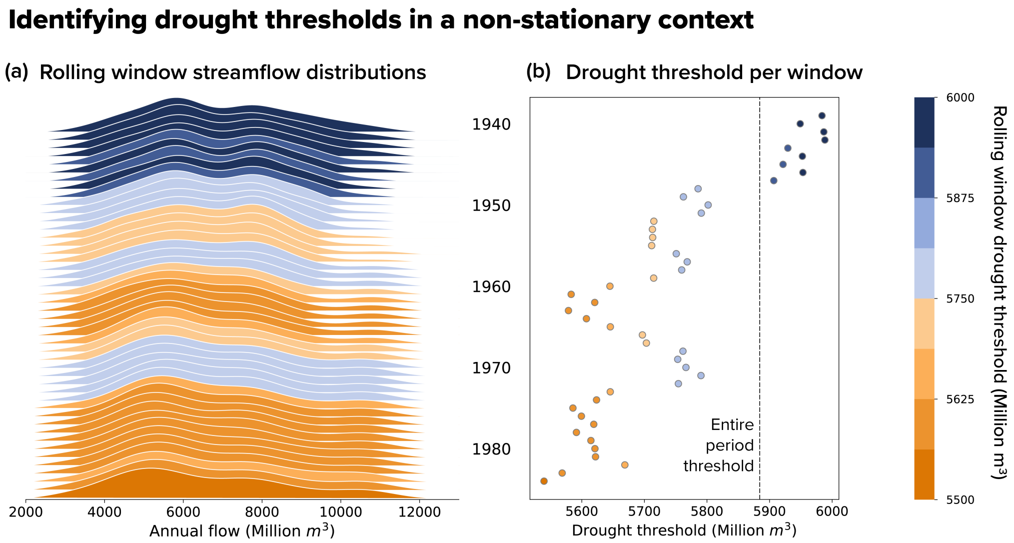

Figure 4. Drought thresholds established using rolling windows (a) Distribution of annual streamflow per 60-year rolling window; (b) Drought threshold established using distribution moments of each 60-year rolling window. The vertical dashed line represents the threshold established using the entire record (same as the threshold in Fig. 3, i.e., $5,884Mm^3$.)

The final type of uncertainty that impacts our understanding of plausible extreme droughts is the inherent deep uncertainty associated with evolving wet and dry dynamic regimes that are beyond the scope of gauged historical streamflow observations. These deeply uncertain regimes can encompass both ungauged historical conditions (e.g., paleo records) and future projections of how the complex human-natural systems may change. Deep uncertainty refers to a lack of consensus over how future events may unfold as well as their associated likelihoods or consequences (Marchau et al., 2019; Walker et al., 2003). Literature focusing on deep uncertainty emphasizes the use of exploratory modeling—the use of intentionally broad hypotheses about future system conditions and the assessment of system outcomes. This allows us to investigate a broader ensemble of states so as to be able to understand system response and inform planning in spite of the presence of these three uncertainty types. Here, we place an explicit focus on exploratory modeling of hydroclimatic factors and their implications for key basin outcomes. As discussed above, increasingly frequent and more persistent severe drought conditions inhibit the ability of basins like the UCRB to meet their obligations to Lower Colorado Basin states through deliveries to Lake

[3] In fact, some have argued the current megadrought should not actually be considered a drought, but a new normal brought about by aridification (Robbins, 2019).

Powell. At the same time, given the central importance of the agricultural sector in the UCRB, quantifying impacts to local agricultural water users is another critical concern. Both these issues are highlighted in the Basin Implementation Plan as key concerns for the Colorado Basin Roundtable (CWCB & CDWR, 2022). Through combinations of hydroclimatic states and these basin impacts, we identify consequential drought storylines that represent complex mappings between the large space of input uncertainty (ensemble of hydroclimatic conditions) and the large space of resulting outcomes for the basin's stakeholders.

3.3 Stage II - Evaluation Across Many States of the World

The system is evaluated under an ensemble of hydrologic SOWs, synthetically generated to reflect different assumptions about future hydroclimatic changes in the region, as well as to explore their internal variability (Fig. 1). Our ensemble of SOWs relies on the Gaussian Hidden Markov Model (HMM) synthetic streamflow generator developed by Quinn et al. (2020). The use of HMMs for the synthetic generation of streamflows has advantages in capturing complex wet-dry hydroclimatic regime dynamics as well as their persistence in Western US drought extremes (Bracken et al., 2014, 2016). We refer the reader to Quinn et al. (2020) for the full details of how the synthetic streamflow ensemble was generated; we summarize key information here. The HMM used comprises two states: one representing wet and the other dry conditions (i.e., higher and lower streamflows). The two states are referred to as 'hidden' because they are not directly observed; rather they are inferred from a time series of continuous flow values, assumed to come from one of two log-normal distributions (one for the distribution of wet years and one for dry years). Fitting an HMM with these characteristics requires the estimation of six parameters: the mean and standard deviation of the dry-state and wet-state Gaussian distributions (μ_d and σ_d, and μ_w and σ_w, respectively), as well as the probabilities of transitioning from a dry state in year t to a dry state in year $t+1$ (p_{dd}), and from a wet state in year t to a wet state in year $t+1$ (p_{ww}). The generator then uses these distributions and the estimated transition probabilities to create synthetic time series of streamflows. Two examples of synthetically generated streamflows using the HMM are shown in Fig. 3 (b-c).

To generate the ensemble, Quinn et al. (2020) fit the HMM to historical observations and then modified its parameters according to several experimental designs, each reflecting different assumptions about how future hydrologic conditions in the basin could change. These different assumptions can all be considered plausible 'rival framings' of future wet-dry regimes. These rival framings were that: (i) streamflow parameters in the future could independently deviate from their stationary historical behavior to a moderate degree, (ii) they could move toward values seen in the past, as inferred from reconstructed paleo data, (iii) they could reflect downscaled climate change projections for the UCRB region, or (iv) they could move toward values generated under any of these assumptions (i.e., the 'all-encompassing' ensemble of candidate futures, which parametrically envelopes all other rival framings of the UCRB's hydroclimate).

In this study, we utilize the all-encompassing experiment. Within the all-encompassing experiment, possible future scenarios consist of multipliers on the dry-state and wet-state means and standard deviations, and delta shifts on the dry-dry and wet-wet transition probabilities. The sets of all scaling factors and the respective ranges for each HMM parameter are given in Eq. 1, which were chosen by Quinn et al. (2020) to span the ranges experienced across all other rival framings. Using these parameter ranges, 100 parameter combinations were generated using Latin hypercube sampling (McKay et al., 1979). The 100-member ensemble size was verified by Quinn et al. (2020) to yield results that are consistent with the results obtained using a larger ensemble of 1,000 parameter combinations.

$$
\begin{aligned}
\mu_d &= \{0.90 \leq \mu_{d_i} \leq 1.03 | i \in I\} \\
\mu_w &= \{0.97 \leq \mu_{w_i} \leq 1.03 | i \in I\} \\
\sigma_d &= \{0.75 \leq \sigma_{d_i} \leq 2.63 | i \in I\} \\
\sigma_w &= \{0.39 \leq \sigma_{w_i} \leq 1.25 | i \in I\} \\
p_{dd} &= \{-0.65 \leq p_{dd_i} \leq 0.30 | i \in I\} \text{ and } p_{dw} = \{1 - p_{dd_i} | i \in I\} \\
p_{ww} &= \{-0.33 \leq p_{ww_i} \leq 0.33 | i \in I\} \text{ and } p_{wd} = \{1 - p_{ww_i} | i \in I\}
\end{aligned}
\tag{1}
$$

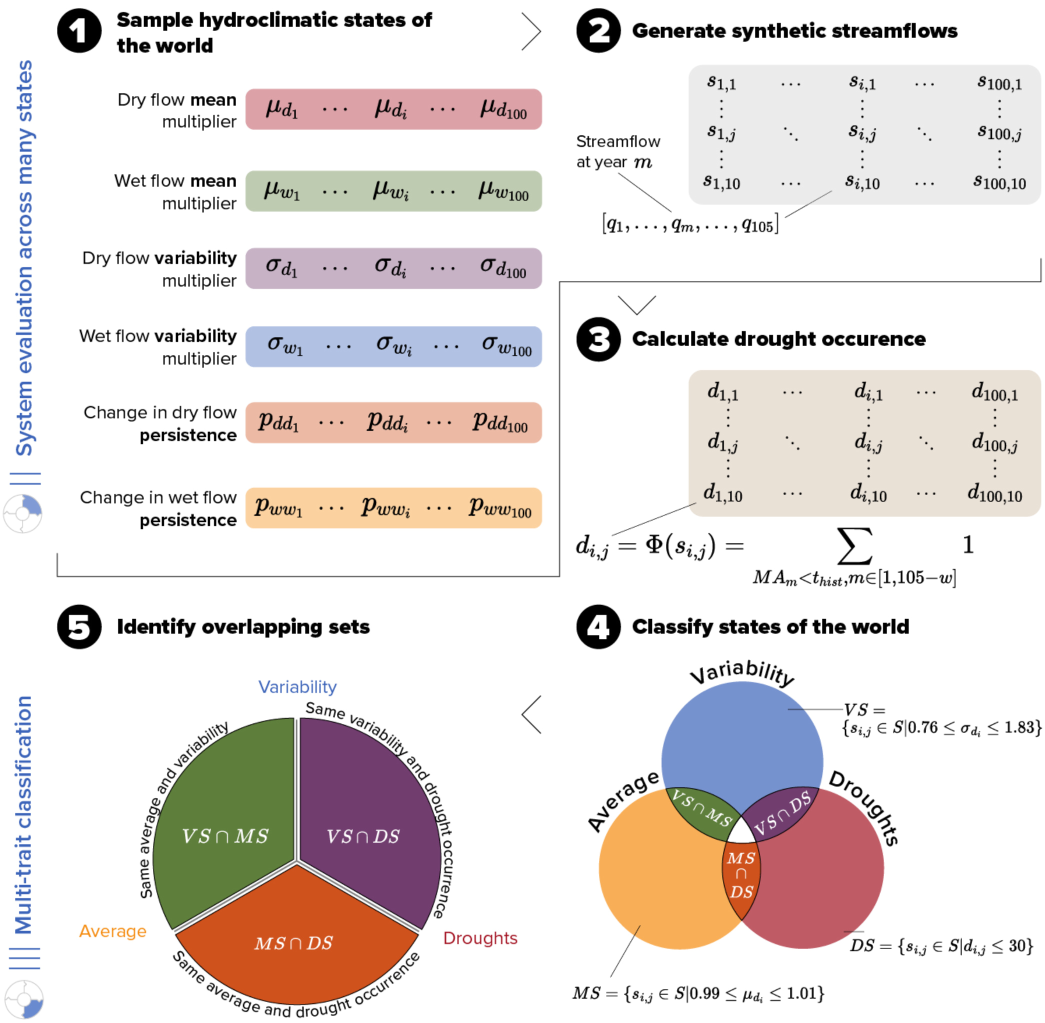

Figure 5. **Applying stages II and III of FRNSIC to the UCRB case study.** Steps 1-2 illustrate the generation and simulation of the hydroclimatic SOWs (Stage II). Steps 3-5 illustrate the classification of behavioral dynamics (Stage III). Sets of dynamic properties are defined as $VS \cap MS$: *Exhibiting the same variability and average annual dry flows*; $MS \cap DS$: *Exhibiting the same average dry flows and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the same variability of annual dry flows and number of decadal drought years.*

For each parameter combination i (i.e., for each combination of $\mu_{d_i}, \mu_{w_i}, \sigma_{d_i}, \sigma_{w_i}, p_{dd_i}, p_{dd_i}$), we generated 10 realizations of 105 years of streamflow, $s_{i,j}$, such that there exists a set of all streamflow SOWs $S = \{s_{i,j} | i \in I \land j \in J\}$ and $J = [1, 2, ..., 10]$. Each SOW $s_{i,j}$ represents a sequence $[q_1, q_2, ..., q_{105}]$, where q_m is the streamflow at year m. In other words, 10 realizations of 105-year-long times series of annual streamflows are created for each of the 100 sampled HMM parameterizations, resulting in a total of 105,000 synthetic years (Fig. 5 Step 2). The annual streamflows are generated in log space for the last node represented in the system model (at the Colorado-Utah state line) and then converted to real space and downscaled to monthly streamflows using a modified version of the proportional scaling method used by Nowak et al. (2010). The same method is also used to identify contributing proportions from all upstream model nodes, as detailed in Hadjimichael, Quinn, Wilson, et al. (2020). We note here that these streamflows are naturalized as required to serve as model input for StateMod water allocation model. The ensemble of streamflows from this all-encompassing experiment span those from all other sets (historical observations, paleo reconstructions, and projections), with values that exceed both sides of the distribution (Fig. S2).

3.4 Stage III - Multi-trait Classification of States of the World

3.4.1 Classification of dynamics

As noted in Section 2, one of the key contributions of our proposed framework is the classification of the dynamic properties of each sampled SOW within an exploratory modeling ensemble, irrespective of its performance on specific impact criteria (Fig. 1). The motivation in capturing these dynamics is largely to help illuminate the behavioral processes that lead to the consequential impacts, something that is often lost when scenario discovery is performed by classifying based on aggregate robustness performance measures. These dynamic properties can be specified *a priori*, if they are part of the design of experiments, or they can be discovered or estimated after each SOW simulation is performed. In our case, we utilize both approaches to capture three dynamic properties of our SOWs: the variability of dry year streamflows, the central tendency (average) of dry year streamflows, and the occurrence of decadal hydrologic drought conditions. With regard to the average and variance of dry years, (μ_d and σ_d, respectively) these properties are part of the sampled HMM parameters used to create each synthetic SOW and are therefore known without additional calculations for each model simulation. We choose to focus on these two properties of the synthetically generated SOWs (as opposed to properties of the wet states of each SOW) to better understand how dry flow dynamics contribute to water scarcity impacts, but any other behavioral property (statistical or otherwise) could also be used, as relevant to the problem under study. We emphasize here that even though these dynamic properties strongly influence impacts (which are classified in Section 3.4.2) the mappings between them are not necessarily known *a priori*, nor are they straightforward to infer. For example, one might intuit that decreasing the average annual streamflow during dry years (i.e., μ_d) will result in more water user impacts, but exactly how much change or how it interacts with other factors to shape impacts are not immediately apparent.

The occurrence of decadal hydrologic drought conditions is identified after the simulations are performed for each of the synthetically generated 105-year streamflow sequences (Fig. 5 Step 3). To do so, we follow Ault et al. (2014) and establish a drought threshold, T, as half a standard deviation from the period average (i.e., $\mu - 0.5\sigma$). For example, in Fig. 3 for the entire period of historical streamflow observations (105 years), we use the threshold $T = 5,884Mm^3$. When a moving average of annual streamflow (q_m) over 11 years falls below this threshold, we identify the period as a decadal-scale drought. Longer windows (e.g., 35 years) can be used to identify multi-decadal droughts, depending on the specific extreme drought application focus. Formally, for each SOW $s_{i,j}$, the total number of decadal drought years $d_{i,j}$ (Fig. 5 Step 3) is given by:

$$\Phi(s_{i,j}) = \sum_{MA_m<T, m\in[1,105-w]} 1, \tag{2}$$

where MA_m is the moving average of annual streamflows at year m given by:

$$MA_m = \frac{1}{w} \sum_{m,m\in[1,105-w]}^{m+w} q_m, \tag{3}$$

and w is the length of the rolling window (11 years in our case). The set of all drought year durations for all SOWs is then defined by:

$$D = \{d_{i,j} | d_{i,j} = \Phi(s_{i,j}) \forall [i \in I \wedge j \in J]\}. \quad (4)$$

We also denote $DY_{i,j}$ as the drought years of SOW $s_{i,j}$, given by:

$$DY_{i,j} = \{m | MA_m < T, m \in [1, 105 - w]\} \quad (5)$$

We therefore use three dynamic properties of each SOW $s_{i,j}$ to classify the dynamics of our SOW ensemble: the variability of dry year streamflows σ_{d_i}, the average of dry year streamflows μ_{d_i}, and the number of decadal drought years $d_{i,j}$. There is a variety of ways one might choose to classify SOW sets using these properties, depending on the specific analysis questions and as informed by the Problem Framing stage. We note in Section 1, that insights from co-production literature highlight that the manner with which information is presented to its users is critical to how they understand and choose to utilize it (Calvo et al., 2022). More specifically, and as it relates to the classification of dynamic properties, Lemos et al. (2012) stress that relating new findings to past experiences can help connect that information to stakeholder analytical and experiential processing abilities, as well as foster the usability of the new findings.

Based on these recommendations, we classify the dynamic properties of the SOWs based on how they relate to the historical experience of basin water users. For example, one might be interested in investigating the impacts of SOWs under the assumption that the future will be similar to the experienced past. In such a case, conditional criteria can be used to separate the SOWs that fall within the bounds of past experiences from the ones that do not. We demonstrate this by focusing on what we will be referring to as "historically-informed" SOWs: synthetic SOWs that exhibit properties within the range of dry year streamflow average and variance values as they appear in 60-year rolling windows of the record of gauged observations, as well as the past drought conditions resulting from said observed streamflow. These history-informed synthetic SOWs of hydrology reflect the assumption that the future will behave like the observed past and can be used to establish plausible stakeholder-relevant impacts that might be unlike those previously experienced. Corollary to this classification, we can identify SOWs that do not meet these criteria (e.g., by exhibiting more dry year streamflow variance relative to what has occurred in the available observed record) as SOWs reflecting a changing system.

To identify historically-informed thresholds for the variability and persistence of dry conditions we utilize the 60-rolling windows of streamflow, shown in Fig. 4 (a). For each window, we estimate its respective μ_d and σ_d and use those estimates to select subsets of our SOW ensemble in which μ_d and σ_d fall within the range of values observed across historical 60-year windows (Fig. S3). The set of SOWs that exhibit dry-flow variability within the bounds of history is therefore defined as:

$$VS = \{s_{i,j} \in S | 0.76 \leq \sigma_{d_i} \leq 1.38\}. \quad (6)$$

Similarly, the set of SOWs that exhibit dry-flow average values within the bounds of history is defined as:

$$MS = \{s_{i,j} \in S | 0.99 \leq \mu_d \leq 1.01\} \quad (7)$$

For a history-informed decadal drought occurrence threshold, we use the same 60-year rolling windows and calculate the number of historical decadal drought years using the drought threshold (T) as defined by the properties of each window (shown in Fig. 4 (b)). Given the varying values of these thresholds ($5,540 \leq T \leq 5,988$), the number of historical hydrologic years out of 105 that are classified as decadal drought years could be as low as zero and as high as 30 (Fig. S1). Assuming that this range of values reflects the range of historical experience of drought, we can use these values as a way to select the SOWs that produce numbers of decadal drought years that fall within the historical experience. The variation in decadal drought years from zero to 30 in this case reflects how drought experience in the basin has historically varied, depending on the different windows of time one may use as reference. To define the set of SOWs exhibiting numbers of decadal drought years within the bounds of historical experience, we therefore use these numbers as the bounds:

$$DS = \{s_{i,j} \in S | d_{i,j} \leq 30\}. \tag{8}$$

In other words, by looking at 60-year rolling windows of historical hydrologic observations (Fig. 4), we are able to deduce a range of values for these dynamic properties as experienced historically. Using these ranges we create three sets of SOWs, each exhibiting these historically-bounded properties. These three sets therefore represent three different dynamic properties of the ensemble of SOWs used in this experiment: VS contains SOWs that fall within the range of the historical variability of dry conditions, MS contains SOWs that fall within the range of the historical average of dry conditions, and DS contains SOWs that fall within the range of drought years experienced in history (Fig. 5 Step 4). We note that these classifications are irrespective of the impacts these SOWs result in (discussed in the following section), and can be used to both uncover the dynamic properties that result in consequential impacts, as well as create narrative storylines of how said impacts come to be. Furthermore, several of our generated SOWs might meet more than one of these conditions. In other words, there exist intersecting sets $VS \cap MS$: *Exhibiting the same variability and average annual dry flows*; $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the same variability in annual dry flows and number of decadal drought years*, as shown in Fig. 5 Step 5. These are simply sets of SOWs where both respective set conditions are met, and might vary in size (discussed in Section 4). All these sets, as well as their intersects, contain SOWs which reflect the hypothesis that the future hydroclimate in the region will be like the past 105 years of observed streamflow conditions. A set where all conditions are met may also exist, and can be further investigated as needed. We do not do so in this current application, largely because the influence of the dynamic conditions is sufficiently demonstrated with the three pairs, and to maintain visual and narrative simplicity.

Corollary to the existence of these sets in our full ensemble of SOWs S, is that for each set of SOWs that meet each dynamic condition there exist complement sets VS', MS', and DS' for which each respective condition does not hold. Specifically: VS' contains SOWs that exhibit dry variability that exceeds the historically observed range, MS' contains SOWs that exhibit average dry values that exceed the historically observed range, and DS' contains SOWs with more drought years than the historically observed range. As such, these sets contain plausible SOWs which reflect the hypothesis that the future hydroclimate in the region will be different from the observed conditions. These SOWs are part of the same ensemble and, even though they exceed historically observed conditions, they remain within plausible future ranges as informed by the extended internal variability based on paleo reconstructed data and changing future conditions simulated under CMIP5 projections (see Section 3.3 and Quinn et al. (2020)). As a result, we create equivalent intersecting sets that capture these plausible, changing dynamic conditions $VS' \cap MS'$: *Changing average and variability in annual dry flows*; $VS' \cap DS'$: *Changing variability in annual dry flows and number of decadal drought years*; and $MS' \cap DS'$: *Changing average of annual dry flows and number of decadal drought years*. It should be noted that the number of decadal drought years only increases relative to historical ranges in these sets (since the lower bound using the historical rolling windows is 0), whereas the average and variability in annual dry flows increases in some and decreases in others.

3.4.2 Classification of impacts

All synthetically generated 105-year timeseries are simulated through StateMod which allocates water to users in the basin according to their rights allocation, the point of their diversion, and the availability of water at each given monthly time step and stream location (CWCB & CDWR, 2016). StateMod allows us to thus assess how these synthetic conditions affect key impacts across all decision-making scales pertinent to the UCRB (Fig. 6). Specifically, the Colorado Basin Roundtable is concerned with meeting the UCRB's obligations for deliveries downstream, as bound by the Colorado River Compact, as well as overall deliveries (or shortages) to the water rights' holders within the basin. Both of these impacts are emphasized as key concerns in Colorado Basin Roundtable's Basin Implementation Plan (CWCB & CDWR, 2022). Within the basin itself, water districts (WDs), are interested in how their own, largely agricultural, users might be affected by future hydroclimatic stress, and individual water rights' holders are primarily concerned with impacts to their own supply.

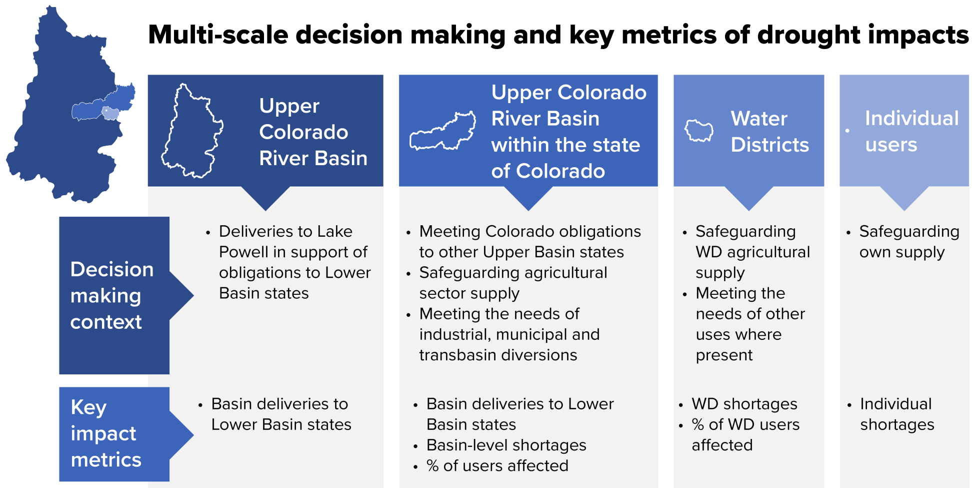

Figure 6. The multi-scale decision making context of the UCRB. Moving from left to right reflects a more localized scale, from the broader multi-state Upper Colorado River Basin region, to the individual water users in the UCRB. Focusing on smaller regions shifts the decision making context and the key metrics of concern with regard to hydrologic drought. These key impacts are reflected in the impact classification scheme (Fig. 7).

We assess these multi-scale impacts by looking at water demands and shortages (undelivered water) to 338 users in the basin during the drought periods of each SOW, as well as basin deliveries downstream (water leaving the UCRB). Water demands per user are a StateMod output, defined here as $W(u, s_{i,j})$, the water demand for user u during the drought periods of SOW $s_{i,j}$. Equivalently, water shortage $G(u, s_{i,j})$ is the undelivered water to user u during the drought periods of SOW $s_{i,j}$ (Fig. 7 Step 6). Using this notation, we can calculate the percentage of shorted users during the drought period of each SOW $s_{i,j}$ as:

$$\Psi(s_{i,j}) = \frac{100}{n_{users}} \sum_{G(u,s_{i,j})>0, u\in[1,\dots,n_{users}]} 1 \tag{9}$$

and the mean shortage across users—during the same drought period—as:

$$X(s_{i,j}) = 100 \sum_{u\in[1,\dots,n_{users}]} \frac{G(u, s_{i,j})}{W(u, s_{i,j})} \tag{10}$$

For both equations we use $n_{users} = 338$ for all consumptive use water users in the basin.

The third key impact metric we are tracking is how delivery obligations to Lake Powell are affected. There is a large number of moments, quantiles, or other distributional measurements we can track here. We are using the rolling 10-year sum of basin deliveries, consistent with how Upper Basin state obligations are typically accounted for (e.g., Bureau of Reclamation (2012); Woodhouse et al. (2021)). For each SOW, we calculate this 10-year rolling sum and estimate the 10^{th} percentile of all values to focus explicitly on the lowest 10-year cumulative deliveries. Formally, we denote qo_m as the basin outflow in year m for each SOW $s_{i,j}$, and $BD_{i,j}$ as the sequence

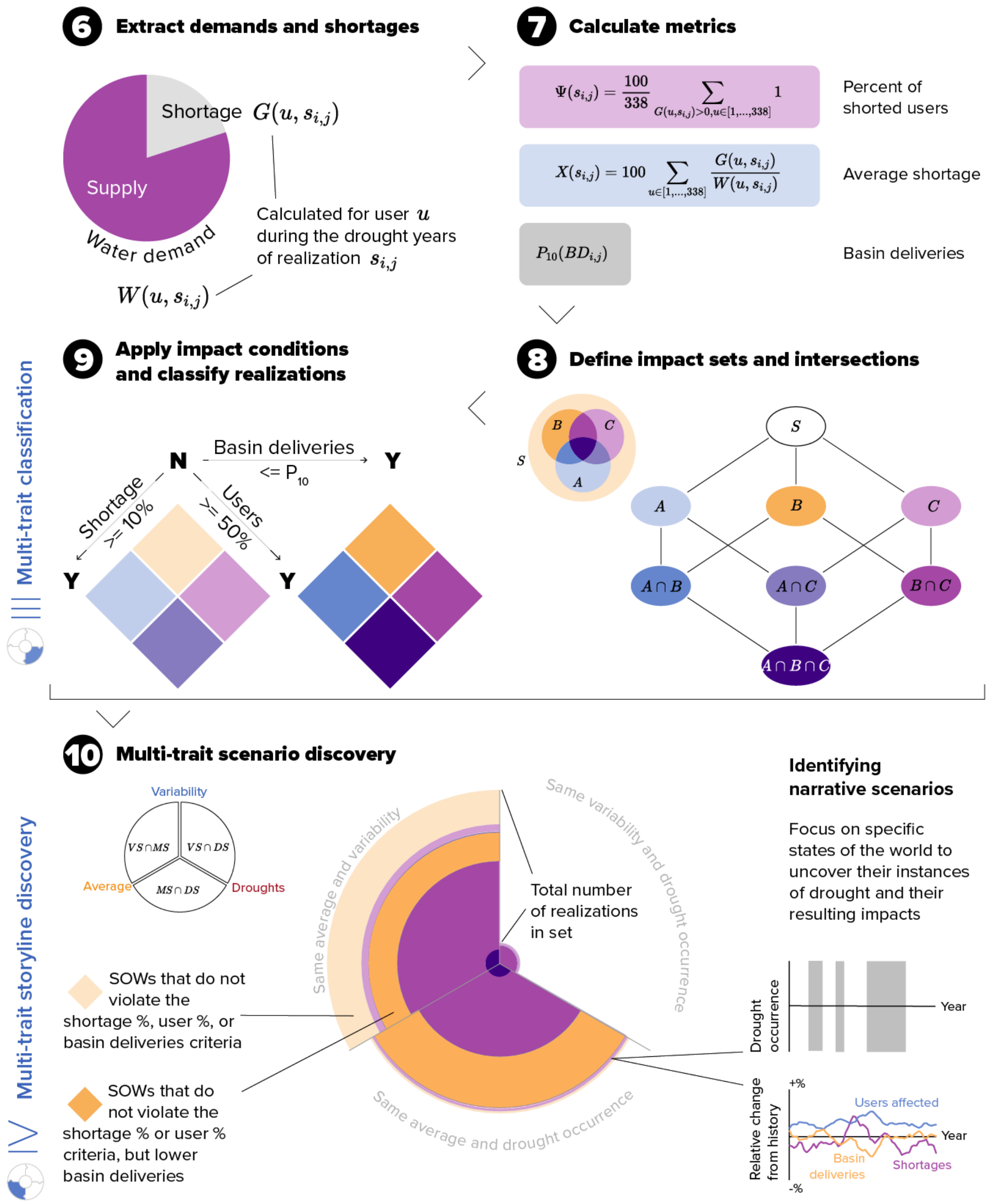

Figure 7. Applying stages III and IV of FRNSIC to the UCRB case study. Steps 6-9 calculation and classification of user- and basin-level impacts (Stage III). Step 10 illustrates the combination of said impacts with behavioral dynamics to identify narrative drought storylines for the UCRB (Stage IV).

798 of all cumulative 10-year sums:

$$BD_{i,j} = (bd_1, ..., bd_m, ..., bd_{95}), \tag{11}$$

799 where:

$$bd_m = \sum_{m,m\in[1,95]}^{m+10} qo_m \tag{12}$$

is the cumulative 10-year sum of deliveries at year m, and $P_{10}(BD_{i,j})$ is the 10$^{\text{th}}$ percentile of all cumulative sums (Fig. 7 Step 7).

Based on these metrics, we identify which of the synthetic SOWs are consequential to the Colorado Basin Roundtable and its stakeholders by quantifying their effects on water deliveries to basin users and downstream. In this manner, the scenarios identified are intrinsically tied to the consequential impacts they generate at the basin itself, overcoming the limitation presented by the limited set of five driver-defined scenarios used by the state (State of Colorado, 2023). Further, through the use of exploratory modeling, we more rigorously investigate the space of plausible future conditions, to then, *a posteriori*, discover the ones that truly matter locally. As overviewed earlier, this process of *a posteriori* scenario classification is formally referred to as scenario discovery (Bryant & Lempert, 2010; Kwakkel, 2019). Traditionally, scenario discovery is a classification process, and categorizes hypothetical scenario conditions as either 'successes' or 'failures' depending on whether they meet a criterion, or a combination of a small number of them. Classification in its simplest form is performed through separating the space using orthogonal subspaces, typically using algorithms such as the Patient Rule Induction Method (PRIM; Friedman and Fisher (1999)) or Classification and Regression Trees (CART; Breiman (1984)). Applying these methods to real complex systems has uncovered several challenges in both the criteria used to identify the scenarios of interest (i.e., what measure to use to select 'failed' SOWs), as well as in the computational methods used to do so, also known as rule induction or factor mapping (i.e., identifying what factors lead to failures). Respective advancements have been made to tackle these challenges. Challenges with regard to rule induction are primarily rooted in the orthogonality (Kwakkel, 2019), linearity (Pruett & Hester, 2016; Quinn et al., 2018), and convexity (Guivarch et al., 2016; Trindade et al., 2019, 2020)—and lack thereof—of the space being separated. We refer the reader to these studies for more information about methodological advancements in this space. The challenges surrounding identification, particularly with regard to complex multi-actor systems with a large number of relevant states, have been broadly articulated in Section 1. Here, we discuss how FRNSIC is addressing them for the UCRB case study.

We utilize three metrics to capture overall impacts to the basin: percentage of shorted users ($\Psi(s_{i,j})$; Eq. 9), mean shortage ($X(s_{i,j})$; Eq. 10) and the 10$^{\text{th}}$ percentile of cumulative basin deliveries ($P_{10}(BD_{i,j})$; Eq. 11), each relevant to the multi-scale decision making context of the UCRB (Fig. 6). As described in Section 2, we utilize a set theory perspective in SOW classification by creating conditional sets based on whether the SOWs meet each impact criterion. For multiple criteria we can also create multiple such subsets and look at the intersections of the conditional sets for combinations of multiple criteria. This mirrors how satisficing metrics are typically used in the robustness analysis stage of RDM or MORDM applications, where more than one performance metric might matter to whether a strategy is considered "robust" (McPhail et al., 2018). In those cases, multiple metrics are used together to assess robustness (e.g., "$reliability \geq 90\%$" AND "$costs \leq \$100$"), but rarely are different subsets and combinations compared. FRNSIC presents an alternative approach, where the hierarchical combination of impact metrics allows for the discovery of robust strategies across all possible combinations of performance metrics. Fig. 7 Step 8 shows an example of this, using three subsets A, B, and C, each corresponding to an impact criterion. This partially ordered set is an algebraic structure formally referred to as a Boolean lattice, often visualized using a Hasse diagram (Priss, 2021), as shown in Step 8. Starting at the top of this graphic, S denotes the entire set of SOWs in our ensemble, of which A, B, and C are subsets. Moving downward, we combine these sets to their intersections indicating two of the conditions being met, with the subset in the very bottom indicating the set where all three conditions are met.

In this application, we establish three criteria based on which conditional SOW sets are created, each using one of the key impact metrics (Fig. 6). Specifically, using the mean shortage experienced during each SOW $X(s_{i,j})$ (Eq. 10), we can define a conditional subset of SOWs that exceed a decision-relevant threshold for water shortage, given by th_{χ}, such that:

$$A = \{s_{i,j} \in S | X(s_{i,j}) >= th_{\chi}\}. \tag{13}$$

For example, using the nominal value of $th_\chi = 10\%$ we select a subset of SOWs A where the mean user shortage exceeds 10% (Fig. 7 Step 9). We can capture higher or lower degrees of risk tolerance in the basin (e.g., a mean shortage of 20% versus 5%) by utilizing shortage thresholds at various levels to establish a different set A conditioned on the threshold used. For reference, the historical average shortage across all years and all basin users is 7%.

Looking at the downstream basin deliveries in each SOW, we compare whether the 10$^{\text{th}}$ percentile of cumulative 10-year streamflows of each SOW ($P_{10}(BD_{i,j})$; Eq. 11) meets or subceeds a critical threshold th_{bd}. This second conditional set B is given by:

$$B = \{s_{i,j} \in S | P_{10}(BD_{i,j}) <= th_{bd}\}. \tag{14}$$

This set identifies SOWs that have their lowest 10% of cumulative deliveries fall below a critical threshold. For instance, using the historical 10$^{\text{th}}$ percentile of cumulative deliveries (46,820 M m^3) as th_{bd}, we select SOWs where the basin is delivering less than its historical 10% worst years.

Lastly, using the percentage of shorted users $\Psi(s_{i,j})$ (Eq. 9), we can identify a conditional subset of SOWs that exceed a consequential threshold of shorted users, given by th_ψ, such that:

$$C = \{s_{i,j} \in S | \Psi(s_{i,j}) >= th_\psi\}. \tag{15}$$

In the FRNSIC illustration in Fig. 7 Step 9, we create subset C by using the nominal value $th_\psi = 50\%$ to select all SOWs where more than 50% of water users are shorted. For reference, historically, an average of 30% of water users is shorted at any given year, with some years reaching up to 66%.

We note that sets A, B, and C are not mutually exclusive and there may exist SOWs in S that meet more than one or all three criteria (Fig. 7 Steps 8-9). By applying each threshold and identifying each conditional subset that meets the condition—including their intersections—we classify every SOW as belonging in either:

- a set where none of the conditions are met (i.e., $(A \cup B \cup C)'$, shown in light yellow ◆),
- three sets where only one of the conditions is met (i.e., set A in light blue ◆ with larger shortages, set B in yellow ◆ with lower deliveries, and set C in lilac ◆ with more shorted users),
- three sets where two conditions are met (i.e., $A \cap B$ in blue ◆ with both larger shortages and lower deliveries, $A \cap C$ in light purple ◆ with both larger shortages and more shorted users, and $B \cap C$ in violet ◆ with both lower deliveries and more shorted users,
- and lastly, one set where all three of the conditions are met (i.e., set $A \cap B \cap C$) in dark purple ◆.

These eight sets are all shown with regard to their partially-ordered relationships in Fig. 7 Step 8 and in how they are applied for impact classification in Step 9. Using these impact sets, we create a hierarchical set-of-sets where impact criteria can be combined to reflect additional stakeholder impacts or conditions. As with the classification of dynamic properties, we only utilize three criteria here, but the proposed method is amenable to larger numbers. We do stress, however, that interpretability and narrative clarity quickly degrade with the addition of more dimensions.

3.5 Stage IV - Multi-trait storyline discovery

The final step in the proposed framework combines the impact classification performed in Step 9 (Fig. 7) with the SOW sets identified in Step 5 (Fig. 5) for the creation of narrative storylines that capture both key behavioral dynamics of SOWs and consequential impact metrics. Fig. 7 Step 10 shows how the SOWs in each overlapping set of dynamic behavior (i.e., $VS \cap MS$: *Exhibiting the same variability and average annual dry flows*; $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the* *same variability of annual dry flows and number of decadal drought years*) can be distributed among the eight impact groups. This graphic is an adapted version of a stacked hive plot (Krzywinski et al., 2012), and allows us to visualize the resulting high-dimensional dataset in a single-panel figure. The three segments of the circle[4] each correspond to the overlapping sets for average and variability of annual dry flows and number of decadal drought years. The radius of each segment (how much it extends from the center point) indicates the total number of SOWs that fall within the overlapping set. For example, in the hive plot shown in Fig. 7 Step 10 the top left set (defined by having the same average and variability of dry years as history) contains the most SOWs, whereas the top right set (defined by having the same dry flow variability and number of decadal drought years as history) contains the least. Within each segment, the width of each band indicates the number of SOWs from that set that result in one of the eight impact groups identified above. Using the same example figure in Step 10, most of the SOWs exhibiting the same variability and average of annual dry flows (in the top left segment) are in the violet impact group ◆ (i.e., they result in both lower basin deliveries and having more in-basin water users shorted).

The reader can use this plot for several insights: to compare the relative size for each overlapping set of dynamic properties (e.g., to make inferences about how the dynamic properties of the SOWs in the ensemble are distributed); and to compare the relative shift in impact groups when moving from one set of dynamics to the other (e.g., starting from the top left segment and moving to the bottom one we can see that fewer SOWs exhibit no impacts at all—the light yellow band goes away). Presenting everything in a condensed single-panel format allows us to combine this with several other panels resulting from other criteria and thresholds combinations, in a "small multiples" visualization (Tufte, 1990). Showing many small visualizations simultaneously allows the reader to compare the separate panels and look for patterns or outliers in the matrix of visuals, and facilitates presentation and storytelling of large amounts of data in a single figure (van den Elzen & van Wijk, 2013). We note here that even though we are only using three types of dynamic sets and three types of impacts, combining them all together means that this single panel figure captures 24 properties in a single panel (3 dynamic sets x 2^3 impact groups). Even though more sets of either kind can be used (i.e., a hive plot can be created with more than three axes and more than eight color bands) the interpretability of the figure greatly diminishes (Krzywinski et al., 2012). We do not consider this a weakness of this specific visual form, as alternative options (e.g., parallel coordinate plots) also struggle from the same limitations, but without the added benefit of being able to be used in a small multiples visualization without further simplification.

In our hypothetical planning context, the Colorado Basin Roundtable can use these plots to examine specific narrative scenarios. The impact sets are organized from most severe in dark purple (all three impact conditions are true) to least severe in light yellow (none of the impact conditions is true) going from the center of the plot outward. In this manner, we illuminate the narrative scenario each SOW can represent, by capturing both the critical impacts it generates and the dynamic properties that lead to it. For example, the Colorado Basin Roundtable users can subselect a segment (e.g., "investigate future SOWs that have the same mean and variance as we've seen in the past") and then subselect a specific SOW from the impact groups of interest (e.g.,"what are the worst impacts we encounter in these futures"). This SOW can then be further investigated for its temporal dynamics and the impacts they result in within the Basin, and be used to frame future planning and adaptation efforts. Even though we do not perform formal scenario discovery in the form of factor mapping in this demonstration (e.g., searching for the specific combinations of σ_d and μ_d values that lead to a mean shortage of more than 10%), one can additionally be performed as needed. We instead highlight the narrative strength of combining sets of dynamic and impact properties in examining candidate futures for the UCRB.

[4] Geometrically, these are in fact sectors of the circle, but we use the term segment here to avoid later confusion with terms like "agricultural sector"

4 Results and Discussion

4.1 Identifying consequential drought storylines at the basin-level

Planning to address drought often starts with an investigation of baseline historical drought hazards. As illustrated in Fig. 3, plausible historical drought extremes can be well beyond those observed in the limited historical streamflow record due to internal variability, even assuming stationarity. We first illustrate a basin-level assessment in which a coordinated planning group such as the Colorado Basin Roundtable is interested in examining futures that remain statistically similar to the last century of observations. In other words, out of our ensemble of hydrologic SOWs (detailed in Section 3.3), they might want to examine ones that exhibit the range of dynamic properties exhibited in the historical streamflow observations. Specifically, they apply the conditional criteria in Eqs. 6-8 to identify intersecting sets of history-informed SOWs ($VS \cap MS$: *Exhibiting the same average and variability in annual dry flows*; $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the same variability in annual dry flows and number of decadal drought years*), shown in Fig. 8 (a).

Several insights can be drawn from this figure. First, in terms of dynamic classification, 100 SOWs exhibit the same average and variability in annual dry flows as in the observed past (top left segment), 82 exhibit the same variability in annual dry flows and number of decadal drought years as in the observed past (top right segment), and 45 SOWs exhibit the same average annual dry flow and number of decadal drought years as in the observed past (bottom segment). The spread of each color in each segment denotes the distribution of each impact group across each set of SOWs, as determined using the classification described in Section 3.4.2, applied at the basin level. Specifically, each SOW is categorized based on whether: (i) it increases the average shortages basin-wide to more than 10% (the yellow to blue dimension), (ii) it increases the number of basin users that experience shortage to above 50% (the yellow to pink dimension), and (iii) it lowers basin deliveries to Lake Powell below the historical 10^{th} percentile (P_{10}) of cumulative 10-year deliveries (the light to dark dimension). If an SOW increases both average shortages and the number of affected users, it is classified in light purple, and if it also decreases deliveries downstream, it is classified in dark purple. Comparing across the segments we see that more SOWs are classified as exhibiting the same average and variability in annual dry flows (top left segment) than other segments, but the impacts in these worlds are minor to moderate (light to dark yellow). The most severe impacts are generated in SOWs that exhibit the same variability in annual dry flows and number of decadal drought years criteria (small violet region in the top right), suggesting these drought characteristics may be more impactful.

In further examining these most severe impacts, a group such as the Colorado Basin Roundtable can zoom in on one of the SOWs that generated them and investigate its temporal dynamics and how they affect the basin as a whole, as well as particular users. For example, Fig. 8 (a) can be further examined by specifically focusing on the small number of SOWs in the top right segment (i.e., those exhibiting the same variability in annual dry flows and number of decadal drought years as observed history) that produce the most extreme impacts. These two SOWs are shown in violet ◆ because they increase the average shortage experienced in the basin to above 50% and also lower cumulative basin deliveries to below the historical 10^{th} percentile. In Fig. 9, we further investigate the dynamics of one of these SOWs: the one that exhibits the fewest drought years. We refer to this drought storyline as "The Unknown Normal". In this narrative storyline, a drought spanning 23 years takes place and affects both the UCRB's downstream deliveries but also the water shortages experienced in the basin. At the basin-wide level, we first compare the basin's 10-year cumulative downstream deliveries to their historical 10^{th} percentile ($46,820 Mm^3$; top left panel in Fig. 9). We see that during the drought period cumulative basin deliveries downstream fall below the historical cumulative 10^{th} percentile for some of the years, down to 80% of that historical threshold ($37,184 Mm^3$) during one of the years. This shows that even non-extreme hydroclimatic changes can have significant impacts in basins like the UCRB and jeopardize their ability to meet their inter-state obligations. Examining impacts within the basin, we look at cumulative basin-wide shortages as they relate to the historical 90^{th} percentile (Fig. 9 top right panel). During this same drought period, we see total shortages in the basin accumulate to almost seven

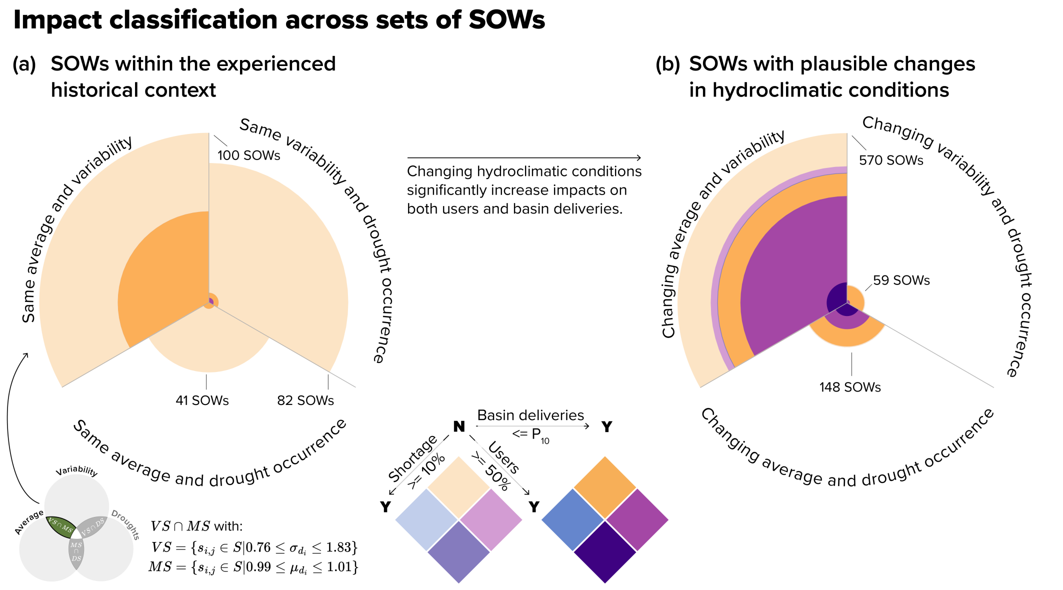

Figure 8. Basin-level impact classification for all states of the world (SOWs) as organized by combinations of dynamic properties. (a) Impacts in SOWs that exhibit dynamic properties within the bounds of the historical context. Starting from the top left: $VS \cap MS$: *Exhibiting the same average and variability in annual dry flows*; $VS \cap DS$: *Exhibiting the same variability in annual dry flows and number of decadal drought years*; and $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; (b) Impacts for SOWs that exhibit dynamic properties outside the bounds of the observed past (changing hydroclimatic context). Starting from the top left: $VS' \cap MS'$: *Changing average and variability in annual dry flows*; $VS' \cap DS'$: *Changing variability in annual dry flows and number of decadal drought years*; and $MS' \cap DS'$: *Changing average of annual dry flows and number of decadal drought years*. All SOWs are categorized based on whether they affect average shortages basin-wide (the blue dimension), they affect the number of basin users that experience shortage (the pink dimension), and they lower basin deliveries below the historical 10^{th} percentile (P_{10}) of cumulative 10-year deliveries (the darkness dimension). Moving from SOWs within the range of historical conditions to the SOWs with changing conditions, experienced impacts become more severe.

times the historical threshold condition and start receding when the drought period is over. We note that there is also a second period during the last 20 years for this simulated future where comparable impacts are seen, but it is not formally classified as a drought period.

As elaborated in Section 3.1 the UCRB supports hundreds of individual water users that use water for many operations: agriculture, municipal water supply, industrial production, power generation, as well as recreational uses (Fig. 2). In prior work in the basin, we have shown that depending on their priority, demands, and location in the basin these users might individually experience very different water scarcity impacts (Hadjimichael, Quinn, Wilson, et al., 2020). We have also shown that aggregate basin impacts (e.g., the mean shortage metric utilized here) can be highly variable across the basin when spatially disaggregated, even at the WD level (Hadjimichael et al., 2023). We therefore further disaggregate these impacts to the UCRB's water districts and users, enabled by StateMod, which traces water allocation and shortage to the individual user level. In Fig. 9 we highlight shortage as a percent of demand for three WDs (39, 37, and 51, moving left to right) in the middle panels with purple lines and four water users in the bottom panels with blue lines. The WD- and user-level shortages show the diverse within-basin experience of this drought storyline, with some WDs and users experiencing very severe shortages

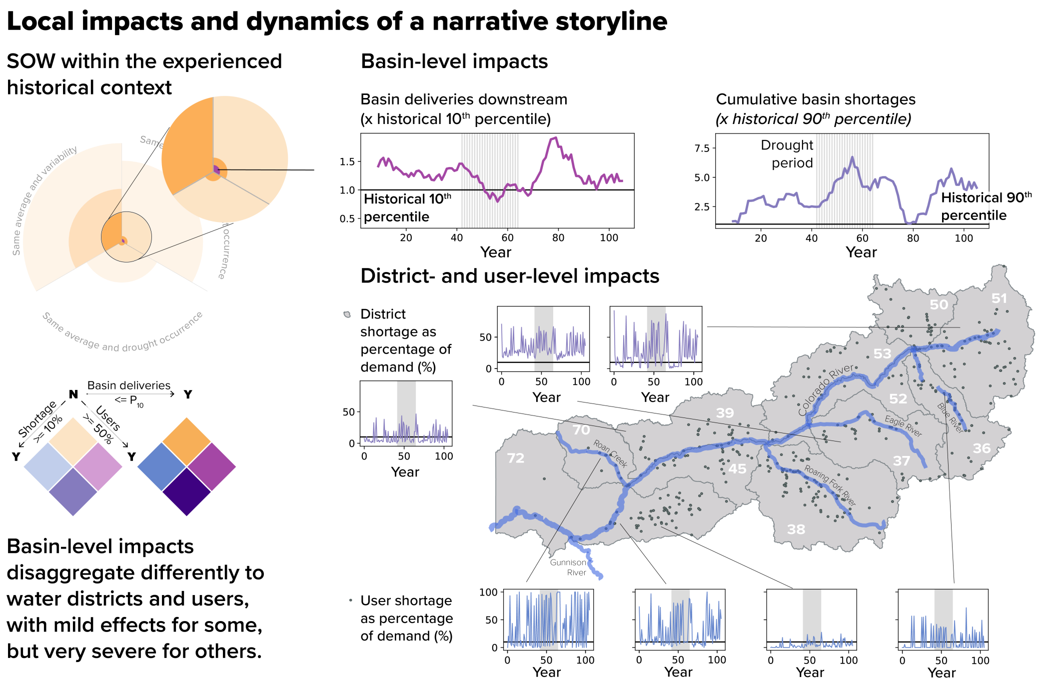

Figure 9. The Unknown Normal: impacts and dynamics of a history-informed drought storyline. The impacts of this state of the world (SOW) are presented for the basin-level at the top, and disaggregated to water districts (middle panels with purple lines) and to individual water users in the basin (bottom panels with blue lines).

and others largely unaffected. These findings align with our prior results while providing a more detailed example of how the same sampled SOW dynamics can yield widely varying shortage impacts subject to the specific characteristics of the various users: their right seniority and decreed allocation, the timing of their demands, and their location in the basin, among others (Hadjimichael, Quinn, Wilson, et al., 2020; Hadjimichael, Quinn, & Reed, 2020; Quinn et al., 2020).

Alternatively, planners might choose to focus on SOWs which reflect assumptions about a changing hydroclimate. In this case the focus would be looking at the complement sets and their intersections (i.e., $VS' \cap MS'$: *Changing average and variability in annual dry flows*; $MS' \cap DS'$: *Changing average of annual dry flows and number of decadal drought years*; and $VS' \cap DS'$: *Changing variability in annual dry flows and number of decadal drought years*). These SOWs and their impacts are shown in Fig. 8 (b). Looking at the changing context sets (Fig. 8 (b)), 570 SOWs exhibit changing average and variability in annual dry flows, 59 SOWs exhibit changing variability in annual dry flows and number of decadal drought years, and 148 SOWs exhibit a changing average of annual dry flows and (increasing) number of decadal drought years. A lot more SOWs meet these dynamic conditions (as compared to Fig. 8 (a)), which is attributed to two main reasons. First, our ensemble of sampled hydroclimatic changes that shape each SOW takes into account projected climate change in the region and how it will change the distributions of streamflow, as well as paleo-reconstructed streamflows (Quinn et al., 2020). This means that several SOWs in our ensemble exhibit statistical properties different from those seen in the gauged record and, in fact, go beyond those distributions (see Fig. S2 and also Fig. S3 (a) for the ranges of mean and variance values). Further, due to these changing properties, the number of drought years in each SOW might also change. In fact, many of the SOWs in our ensemble exhibit more decadal drought years than the maximum of 30 years (or three decades) observed historically based on the highest threshold defined by 60-year rolling windows of streamflow observations (Figs. S1 and S3 (b)), or the deterministic estimate of one or two instances of decadal drought per century, estimated in paleo record studies of Ault et al. (2014); Woodhouse and Overpeck (1998).

This is also related to the second reason we see more SOWs fall outside the historical ranges, especially violating the condition on the number of decadal drought years (Eq. 8). For each sampled change in the average and variability in annual dry flows (i.e., changes in μ_d and σ_d values, as shown in Fig. 5 Step 1), we generate 10 streamflow realizations to capture the internal variability of each hypothesized hydroclimatic change (Fig. 5 Step 2). By better exploring this internal variability we see a wider range of decadal drought years emerge, even between SOWs that exhibit the same statistical properties, as expected (Lehner & Deser, 2023). This is exemplified in Fig. 3 for the internal variability of the recent history. Even though only 22 years of drought were observed (Fig. 3 (a)), this deterministic framing does not represent the true frequency of such events, which may be higher, as seen in Fig. 3 (b). The combined effects of a changing climate and internal variability produce SOWs with many more years of decadal drought than 30 out of 105 (Fig. S2 (b)), classifying them as outside the historical experience of water users in the UCRB under different rolling windows of 60 years (Fig. 4 and S1). These SOWs therefore appear in Fig. 8 (b).

Looking at Fig. 8 (b), SOWs in a changing hydroclimatic context produce much more severe impacts. Whereas most SOWs in the historical context do not produce impacts in any of the impact categories (i.e., no mean shortages more than 10%, no more than 50% of users affected, and no basin deliveries below the historical 10th percentile), most of the SOWs in the changing context produce impacts in at least two. This is seen in how the large bands of light yellow change to bands of yellow , violet , and dark purple . The changing properties of these SOWs to lower average annual dry flows with greater variability and greater number of decadal drought years, leads to more severe impacts to the UCRB's water users. This is especially true for the basin's downstream deliveries: the majority of SOWs are assigned a dark color, indicating basin deliveries falling below the historical 10th percentile of cumulative 10-year deliveries.

Out of the SOWs that belong in the changing context sets (Fig. 8 (b)) 116 of them produce impacts across all impact groups (dark purple band): the average shortage they produce is more than 10%, they affect more than 50% of users, and they reduce basin deliveries below the historical 10th percentile of cumulative deliveries. Relating this to past experiences in the basin, the historical average shortage across all years and all basin users is 7% and has reached up to 26% in exceptionally dry years such as 2002 (the exceptionally dry conditions of 2002 can also be seen in Fig. 3 (a)). Basin-wide shortages of 10% of water demand have historically only been observed during drought periods, and the SOWs represented here capture those conditions. Further, with regard to the 50% of affected users, the historical average number of affected users at any given year in the UCRB is 30%, with the maximum percentage being 65%, again during the exceptionally dry conditions of 2002. Therefore, the SOWs that produce conditions affecting 50% of water users or more reflect plausible impacts of the drought extremes represented in our ensemble.

Fig. 10 examines the impacts and dynamics of one of these SOWs in more detail. In particular, we choose to focus on a SOW that produces impacts across all impact groups under the shortest drought duration. This SOW exhibits changing average and variability in annual dry flows (top left segment of Fig. 8 (b)) and has a total of 20 decadal drought years out of 105. We are referring to this drought storyline as "The Unforeseen Struggles". In the top two panels, we again compare the basin's 10-year cumulative downstream deliveries to their historical 10th percentile (left panel) and the basin-wide 10-year cumulative shortages (right panel). During this drought storyline, a 20-year drought takes place and has dramatic effects on the UCRB: cumulative deliveries drop to below 30% of the historical threshold ($13,862 Mm^3$) and cumulative shortages climb to 11 times more than the historical 90th percentile of shortages. Unfolding these impacts at the finer scale, we compare WDs 70, 37, and 52 in the middle panels, as well as the same four users in the bottom panels, as analyzed in Fig. 9. We again see that the storyline affects the users differently, with some barely affected. Of note is also the fact that even though this storyline is much more severe in aggregate effects compared to "The Unknown Normal" in Fig. 9, impacts to individual users do not necessarily follow the same trend. For example, the leftmost water user

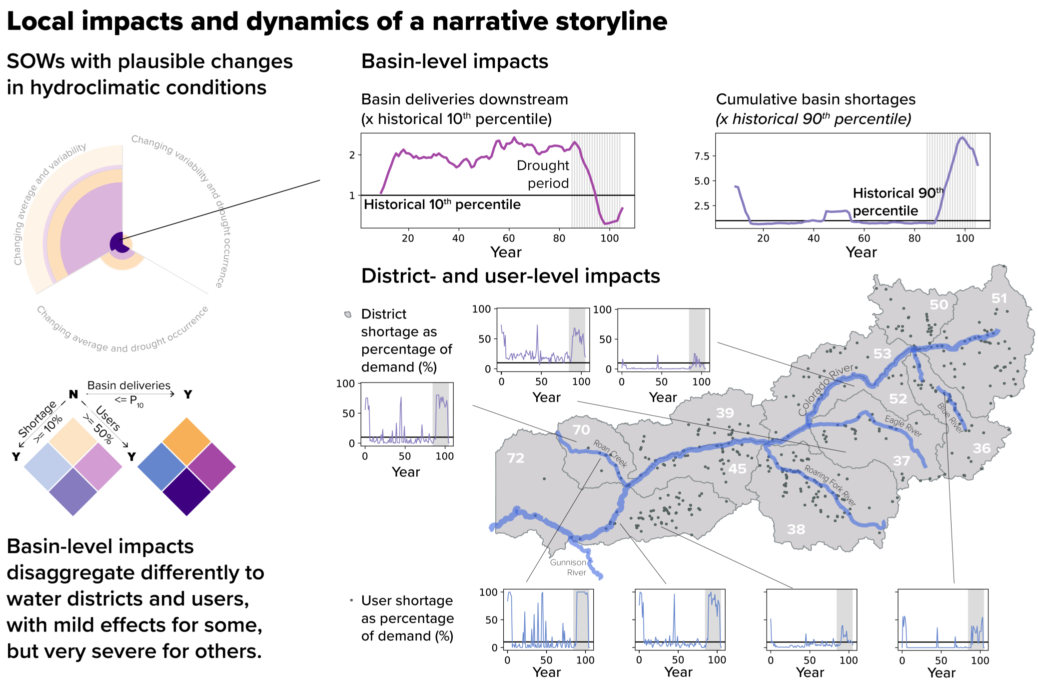

Figure 10. The Unforeseen Struggles: impacts and dynamics of a drought storyline in a changing context. The impacts of this state of the world (SOW) are presented for the basin-level at the top, and disaggregated to water districts (middle panels with purple lines) and to individual water users in the basin (bottom panels with blue lines).

experiences much more severe impacts under "The Unknown Normal" storyline, which falls within the historical bounds. The comparison holds true for other users also, which suggests that the significant aggregate effects we see in Fig. 10 are the result of a larger number of users being affected, not necessarily their larger shortages.

4.2 Examining exploratory ensemble impacts at the sub-basin scale

Beyond the two storylines illustrated in Figs. 9 and 10, we are also interested in how the entire ensemble disaggregates to the subbasin level. For instance, Colorado Basin Roundtable planners might be interested in the distribution of impacts the SOWs generate for a particular WD (Fig. 6). In Fig. 11, we therefore explore what the aggregate basin impacts shown in Fig. 8, look like for each WD in the basin. To do so, we apply Eqs. 9 and 10 to the specific subset of users that divert water in each WD and utilize the same color scheme used in Fig. 8. In this case, each SOW is categorized based on whether: (i) it increases the average shortages at each WD to more than 10% (the yellow to blue dimension), (ii) it increases the number of WD users that experience shortage to above 50% (the yellow to pink dimension), and (iii) it lowers basin deliveries to Lake Powell below the historical 10th percentile (P_{10}) of cumulative 10-year deliveries (the light to dark dimension). If a SOW both increases average shortages and the number of affected users, it is classified in light purple, and if it also decreases deliveries downstream, it is classified in dark purple. In this case, the basin deliveries calculation remains the same, so we do not expect to see any differences in that dimension of impact categories. By calculating mean shortages and the percentage of users shorted for each WD individually, as opposed to the basin as a whole, we therefore expect to see shifts from yellow to lilac or blue (or to purple for both) and vice versa, but we should not observe shifts from light colors to dark colors (or vice versa), as the basin delivery calculation remains the same as that of the aggregate plots (shown in Fig. 8).

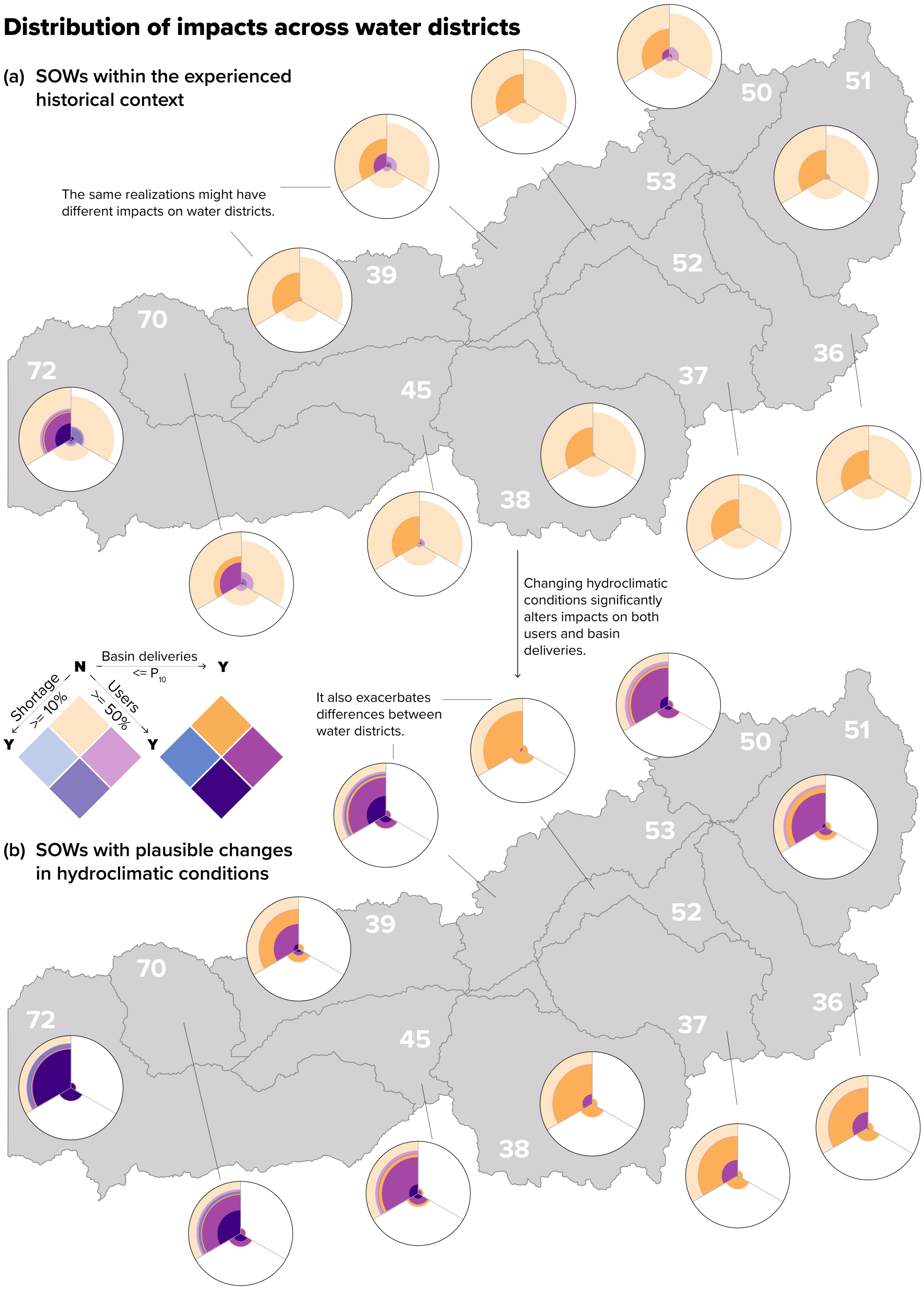

Figure 11. Impact classification for all states of the world (SOWs) as organized by combinations of dynamic properties and calculated for individual water districts. (a) Impacts for SOWs that exhibit dynamic properties within the bounds of the observed past (105 years of gauged streamflow); (b) Impacts for SOWs that exhibit dynamic properties outside the bounds of the observed past (informed by the paleo record and future projections). In both cases, water districts might individually exhibit more severe or less severe impacts than those calculated for the basin in aggregate (shown in Fig. 8.)

It is not entirely unexpected that the same SOWs might have different impacts on the WDs of the UCRB. For example, for the historically-informed SOWs (Fig. 11 (a)), we see that some WDs (36-39, and 52) see no impacts on their users—all bands in the hive plot are shades of yellow. This is better than the basin-wide average conditions shown in Fig. 8 (a). At the same time, some WDs (70 and 72) see their users much more significantly impacted than the basin-level average user of the UCRB, with some historically-informed SOWs producing both larger shortages and for more users (bands in dark purple ◆). SOWs that are outside the historical hydroclimatic context (Fig. 11 (b)) further amplify these differences. For example, users in WD 52 are largely unaffected by all the sets of SOWs, whereas the majority of changing-context SOWs affect both the mean shortages and the number of users affected in WD 72 (dark purple bands). In fact, all other WDs either see their users unaffected by most SOWs with changing hydroclimatic conditions (e.g., WDs 36-39, and 52, which have yellow ◆ as the largest band color) or see only an increase in the number of users affected but not in the mean water shortage (e.g., WDs 45, 50, 51, and 70, which have violet ◆ as the largest band color). This difference in WD experiences is the result of several complex interactions between the number and seniority of rights in each WD, their diversion locations and sources (e.g., the mainstem as opposed to a tributary), and the timing of their demands. These results emphasize that understanding and selecting narrative storylines is critical to capture the natural hydroclimatic drought hazards and their locally consequential impacts as manifested through the UCRB's infrastructure and water governance institutions (i.e., water rights in prior appropriation).

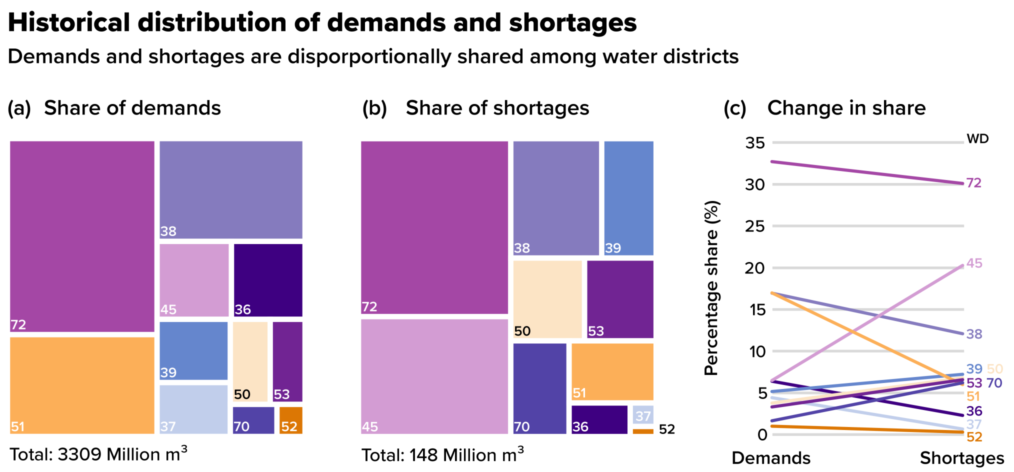

Figure 12. Historical distribution of demands and shortages among water districts. (a-b) Treemaps of (a) the share of water demands as contributed by each water district; and (b) the share of water shortages as contributed by each water district. The treemaps are organized with the largest contributing parts placed at the top left moving first downward and then rightward. (c) Change in relative share between the demands and shortages of each water district.

Specifically, WD 72, which appears to experience the most severe impacts, makes up approximately 33% of all water demands in the UCRB historically, far exceeding the second and third largest demands at 17% by WDs 38 and 51 (Fig. 12 (a)). Compared to the historical data on UCRB shortages (i.e., without any of our sampled hydroclimatic changes imposed on the system), WD 72 indeed represents the largest volumetric share of water shortages in the UCRB (Fig. 12 (b-c)), but their shortages are only 4% of their demands (Fig. 13 (b)), which is below the historical 7% average estimated basin-wide. Indeed, total demand does not explain these impacts on its own (i.e., that the biggest shortages are experienced where the biggest demands are). WD 70, for example, only makes up 1% of the total demands in the basin, yet also sees impacts for its water users that exceed the average (i.e., more violet and purple bands; Fig. 11 (a)), and in the historic observations it exhibits the highest relative ratio of shortages to demands (approximately 16%; Fig. 13 (b)). The historical data also highlights that in general, higher shortages are not necessarily the direct outcome of higher demands (Fig. 12), as some WDs with relatively lower demands experience relatively higher shortages than other WDs (e.g., WD 45), and vice versa (e.g., WD 51). Readers familiar with the region might posit that this difference in impacts can simply be attributed to the number and seniority of rights owned by water users in WD 72; maybe rights in that WD are simply more junior so their demands are not met as much more senior rights in other WDs? Looking at the number of water rights, WD 72 has the same number of actively served consumptive use water rights as WD 38 (296; we note that each water user might own multiple), and its rights are decreed generally larger volumes of water with more senior right ranks on average than WD 38 (Fig. 13 (a)). The differences in impacts can therefore potentially be attributed to the fact that WD 72 (and others) are home to several more junior rights with larger decrees, but it is clear that single factor drivers cannot explain the differences seen.

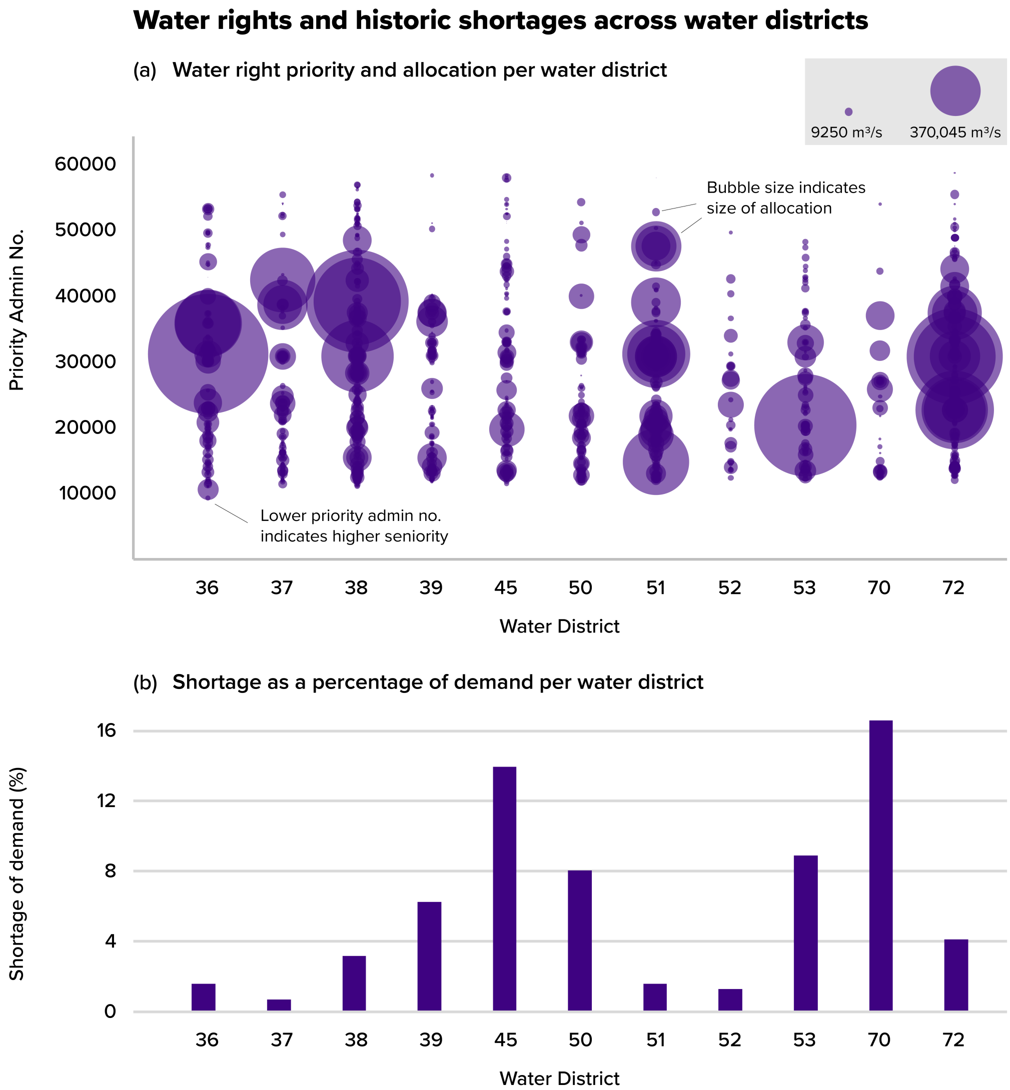

Figure 13. Priority and water allocation per right for each water district. Rights are organized per water district along the horizontal axis and per priority admin number along the vertical axis. Lower priority admin number indicates higher right seniority. Larger bubble size indicates larger water allocation.

4.3 Exploring alternative impact thresholds

Lastly, recognizing the diverse interests represented in the UCRB, we examine more closely how the hierarchical basin-level impact classifications in Fig. 8 are shaped by the assumed problem framing and the impact classification thresholds chosen for basin deliveries downstream, percent of users shorted, and mean shortage (Eqs. 13 - 15). In other words, we would like to know how the classification of these SOWs might change if different shortage risk tolerances were assumed, reflective of the diverse impacts experienced and the different decision-making concerns present in the UCRB (Fig. 6). So in line with the discussion of narrative scenario discovery for multi-actor, multi-sector systems, we repeat the impact classification across different values of each impact threshold (Fig. 14). Specifically, for impact set A containing SOWs that exceed a mean shortage threshold th_{χ}, we use three values of this threshold (5%, 7%, and 10%) and apply them to Eq. 13 to estimate how many SOWs cause the mean shortages in the basin to be above 5%, 7%, and 10% of demand, respectively. Impact set B contains SOWs with their 10^{th} percentile of basin deliveries downstream falling below a critical threshold th_{bd}. In the prior results, we defined th_{bd} using the historical 10^{th} percentile of cumulative deliveries, so B contained SOWs where the basin is delivering less than its historical 10% worst years. Switching th_{bd} to the historical 5^{th} percentile, then B contains SOWs whose low-delivery years are twice as frequent as history. As a result, we are checking if an event that occurred only 5% of the time historically now occurs 10% of the time, in essence doubling its occurrence in the SOWs that meet this criterion. Equivalently, if the threshold used is the historical 1^{st} percentile, then the SOWs in set B have low-delivery years ten times more frequently than history. The 10^{th}, 5^{th}, and 1^{st} percentiles of cumulative 10-year flows are 46,820, 44,896, and 43,776 M m^3, respectively. Lastly, impact set C is the set of all SOWs where more than th_{ψ} of the basin's users are experiencing a shortage. We vary this threshold to 25%, 50%, and 75% to capture SOWs that affect increasing numbers of water users in the basin.

Fig. 14 shows the resulting hive plots for all three thresholds for all three criteria, for the SOWs in the changing hydroclimatic context. This style of small multiples figure allows us to quickly compare the different plots and look for patterns in the matrix of visuals. The following pattern emerges here. Starting at the top left, the hive plot shows the impact classification of all SOWs using the most lenient performance criteria for each impact group (i.e., low basin deliveries occurring as much as history on the vertical axis, mean shortage levels above or equal to 5% of demands on the horizontal axis, and 25% or more users experiencing a shortage along the diagonal axis). Given that these are the most lenient thresholds, they are the easiest criteria to meet, and therefore the majority of SOWs do so (shown in dark purple).

Moving to the right along the horizontal axis, we are increasing the shortage threshold as a percentage of demand so we expect to see fewer blue and purple bands, as fewer SOWs would be classified as causing the larger shortages to water users. Indeed, what we see is a shift from dark purple to a larger lilac band in the top right hive plot. Moving from the top down, we expect to see some of the darker shade classifications turn to lighter colors, as the lower basin deliveries classification is a more extreme condition to meet. Comparing along the three hive plots at the very right, we can indeed see a small number of yellow SOWs turn to light yellow . Finally, moving along the diagonal axis, we are increasing the number of affected users we consider as consequential. In this case, we should expect fewer violet and purple bands as we move diagonally to lower right. This is prominently apparent for the three hive plots at the top right of the figure, where using the 25% threshold, most SOWs are classified as having both more users affected and lower basin deliveries (in violet), but using the 75% threshold, the classifications are largely yellow (only lower basin deliveries).

Even with the more extreme threshold combinations (bottom right hive plot in Fig. 14) most SOWs in the changing context meet at least one of the criteria. Most meet the lower downstream deliveries criterion (yellow band), that their 10^{th} percentile of cumulative 10-year flows fall below the historical 1^{st} percentile (i.e., that low deliveries are occurring ten times as often in these SOWs). Some other SOWs are shown in blue , so they also increase the mean shortage to the basins users to above 10%. We can also compare this hive plot with the one directly to its upper

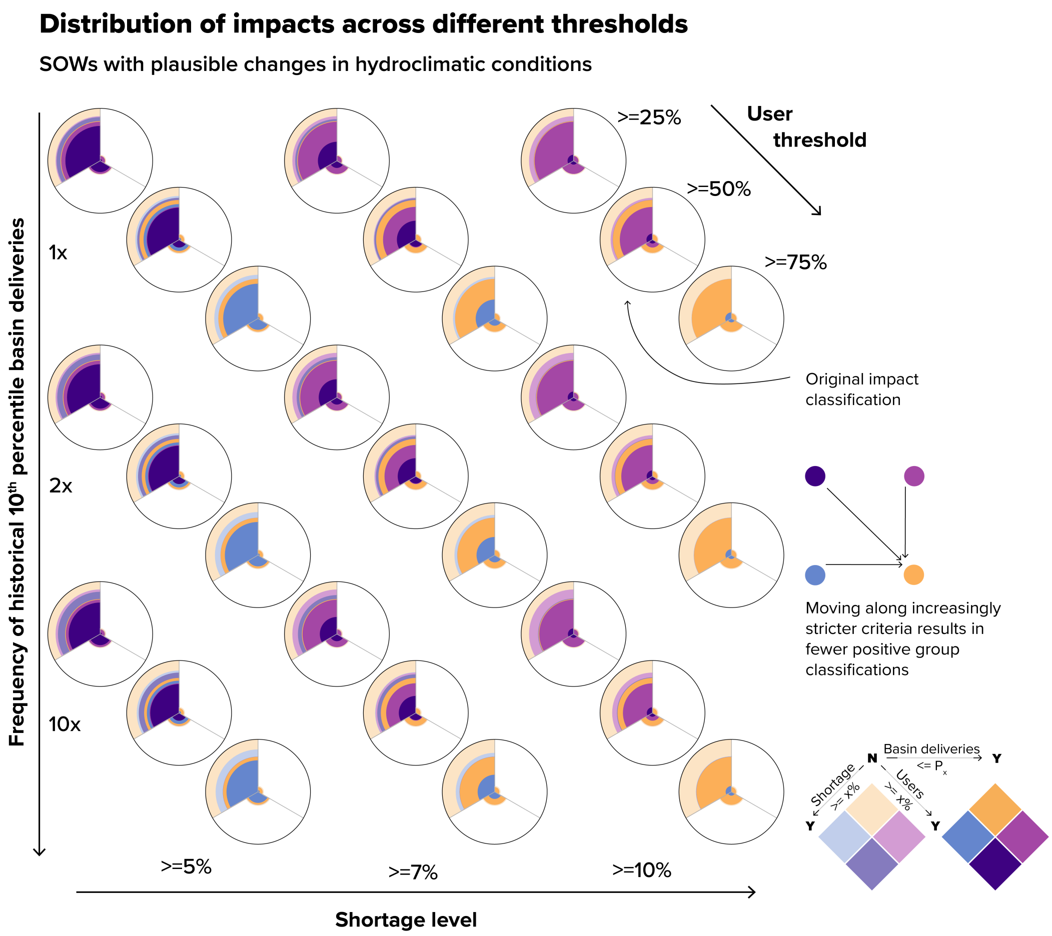

Figure 14. Impact classification for all states of the world as calculated for different thresholds for each impact category. The figure is oriented such the going from the top left to the bottom right, we are moving from more lenient to increasingly stricter criteria.

left, reflecting a change to the user criterion from 75% to 50%, to see that several of the SOWs considered here do affect more than 50% of users in the UCRB (violet and dark purple bands in upper left hive plot) but not more than 75% (same bands disappear when we look back to the lower right hive plot). This shows that even though there might not be a significant increase in the average shortage compared to history (increase from 7% of users to 10%), there is a significant increase in the number of users affected (from 30% historically to above 50%). This further supports the explanation given with regard to the impacts of "The Unforeseen Struggles" storyline (Fig. 10): that they are the result of a larger number of affected users and not necessarily (or only) larger shortages.

Exploring alternative threshold combinations aids with providing an informative feedback to Stage I Framing (Section 3.2) of the FRNSIC assessment of the UCRB, allowing us to address several of the challenges generated by complex human-natural systems more broadly. Namely, as discussed in Section 1, using a small set of scenarios that are considered *a priori* to be "relevant" by the analysts might inadvertently create a very narrow view of what the relevant stakeholder concerns are that is not salient with the diverse views that might exist on the system (Groves & Lempert, 2007). Because each alternative threshold illuminates different SOWs, it allows us to switch to alternative sets of consequential scenarios to focus on, depending on the outcomes they generate. For instance, planners might want to select scenarios from the dark purple SOWs (ones that have impacts across all groups) for further investigation and analysis. The SOWs that fall in these dark purple bands change depending on the thresholds used, so these consequential scenarios can reflect not only varying impact severities, but also different attitudes toward these impacts.

This relates to another complication discussed already, that in systems with many actors making decisions at different scales (Fig. 6), it is difficult to capture their differing priorities, goals and risk aversions with a singular impact metric or threshold imposed on it. We know from prior work (Hadjimichael, Quinn, Wilson, et al., 2020; Quinn et al., 2020), historical estimates (Fig. 12), and also the results here (Fig. 11) that the same conditions imposed on the system can result in diverse impacts for its users. This means that for an SOW with average shortages of 10%, some users or WDs experience shortages lower or higher than that. It follows that some stakeholders in the basin might be more or less conservative about this threshold choice, and the impacts of that change in choice are reflected by moving horizontally in Fig. 14. As a last related point here, in Section 1 we have highlighted recommendations from co-production literature on relating new findings to past experiences as a way to help connect scientific outcomes to stakeholders' analytical and experiential processing (Lemos et al., 2012). Alternative thresholds, especially for the user-level impacts we explore here, can therefore help produce locally-meaningful narratives as they relate the water shortages users and WDs have experienced in the past.

5 Conclusions and Future Work

This paper proposes the FRamework for Narrative Scenarios and Impact Classification (FRNSIC), that enables narrative scenario discovery for multiple states and multiple impacts. The introduced framework is designed to overcome common challenges of scenario discovery with regard to establishing stakeholder-relevant narratives. FRNSIC combines the classification of dynamic behavioral properties of each SOW as well as its impact states in a nested scheme to facilitate hierarchical storyline selection, and produce locally-meaningful narratives from high-dimensional exploratory ensembles. We use a hypothetical planning context—examining the UCRB's potential futures and needing to discover consequential drought storylines to use in planing—and apply FRNSIC to demonstrate its capabilities in a system with multiple actors and institutional complexity. We show that FRNSIC can illuminate the critical dynamic pathways that lead to consequential impacts, by combining a SOW's temporal behavioral properties and the aggregated impacts it results in. The framework therefore addresses several prominent challenges other state-of-the-art scenario discovery frameworks face when applied to complex human-natural systems, and especially institutionally complex systems with many actors like the UCRB.

In applying FRNSIC, several choices must be made on the classification scheme to use (the criteria to use to classify dynamics and impacts, the threshold values to apply, other aggregation choices). This is akin to other scenario discovery applications where consequential or decision-relevant conditions need to be identified, and such choices need to be made transparent from the problem framing stage and throughout the analysis process, as well as reexamined as needed. For example, in the UCRB case study we explore the implications of these choices using gradients of threshold values applied to our criteria. In future work, similar threshold analyses can be applied to the thresholds used to identify the sets of dynamic behaviors exhibited in our ensemble. Changing the criteria through which the dynamics are classified could reflect alternative dynamic behaviors of interest. For example, one could focus on specifically the occurrence of multi-decadal droughts of over 35 years, and this would affect the sizes of the dynamic sets, as well as subsequent results.

The narrative drought storylines produced by FRNSIC can also be utilized in future work in the basin, for example to examine the capacity of adaptive action in modulating the impacts of the drought events seen in each storyline. Specifically, the ensemble of SOWs explored here can be combined with hypothesized policy interventions (e.g., for water conservation) to investigate how said interventions would affect the impacts the basin experiences under each storyline. Just like narrative scenarios and storylines are used in co-production literature, the drought storylines here can also be used in negotiation or stakeholder solicitation contexts to contrast the impacts that WDs or users may potentially experience in the future.

6 Open Research

StateMod is freely available on GitHub `https://github.com/OpenCDSS`. The input files to run StateMod for the UCRB can be found at the CDSS website `https://cdss.colorado.gov/modeling-data/surface-water-statemod`. All the scripts to replicate the analysis performed in this paper and to regenerate all figures can be found at `https://github.com/antonia-had/Hadjimichael-etal_2023_EarthsFuture`. All the output data used in this analysis can be found at `https://doi.org/10.57931/2205512`.

Acknowledgments
This research was supported by the U.S. Department of Energy, Office of Science, as part of research in MultiSector Dynamics, Earth and Environmental System Modeling Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding entities.

References

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015, October). Time-series clustering – A decade review. *Information Systems*, *53*, 16–38. Retrieved 2023-05-15, from `https://www.sciencedirect.com/science/article/pii/S0306437915000733` doi: 10.1016/j.is.2015.04.007

AghaKouchak, A., Huning, L. S., Sadegh, M., Qin, Y., Markonis, Y., Vahedifard, F., ... Kreibich, H. (2023, August). Toward impact-based monitoring of drought and its cascading hazards. *Nature Reviews Earth & Environment*, *4*(8), 582–595. Retrieved 2023-08-08, from `https://www.nature.com/articles/s43017-023-00457-2` (Number: 8 Publisher: Nature Publishing Group) doi: 10.1038/s43017-023-00457-2

AghaKouchak, A., Pan, B., Mazdiyasni, O., Sadegh, M., Jiwa, S., Zhang, W., ... Sorooshian, S. (2022, October). Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *380*(2238), 20210288. Retrieved 2023-01-11, from `https://royalsocietypublishing.org/doi/full/10.1098/rsta.2021.0288` (Publisher: Royal Society) doi: 10.1098/rsta.2021.0288

Arizona Department of Water Resources. (2022). *Arizona Drought Preparedness Plan: 2022 Annual Report* (Tech. Rep.). Retrieved 2023-02-09, from `https://new.azwater.gov/sites/default/files/media/2022ADPAR_0.pdf`

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., & Meko, D. M. (2014, January). Assessing the Risk of Persistent Drought Using Climate Model Simulations and Paleoclimate Data. *Journal of Climate*, *27*(20), 7529–7549. Retrieved 2020-04-28, from `https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-12-00282.1` (Publisher: American Meteorological Society) doi: 10.1175/JCLI-D-12-00282.1

Ault, T. R., Mankin, J. S., Cook, B. I., & Smerdon, J. E. (2016, October). Relative impacts of mitigation, temperature, and precipitation on 21st-century megadrought risk in the American Southwest. *Science Advances*, *2*(10), e1600873. Retrieved 2020-04-28, from `https://advances.sciencemag.org/content/2/10/e1600873` (Publisher: American Association for the Advancement of Science Section: Research Article) doi: 10.1126/sciadv.1600873

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*, *41*(3), 435–449. doi: 10.1287/opre.41.3.435

Ben-Haim, Y. (2006). *Info-gap decision theory: decisions under severe uncertainty*. Elsevier.

Berghuijs, W. R., Allen, S. T., Harrigan, S., & Kirchner, J. W. (2019). Growing spatial scales of synchronous river flooding in Europe. *Geophysical Research Letters*, *46*(3), 1423–1428. (Publisher: Wiley Online Library)

Beven, K. (1993, January). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*(1), 41–51. Retrieved 2020-11-10, from http://www.sciencedirect.com/science/article/pii/030917089390028E doi: 10.1016/0309-1708(93)90028-E

Bonham, N., Kasprzyk, J., & Zagona, E. (2022, November). post-MORDM: Mapping policies to synthesize optimization and robustness results for decision-maker compromise. *Environmental Modelling & Software*, *157*, 105491. Retrieved 2022-09-26, from https://www.sciencedirect.com/science/article/pii/S1364815222001943 doi: 10.1016/j.envsoft.2022.105491

Bracken, C., Rajagopalan, B., & Woodhouse, C. (2016). A Bayesian hierarchical nonhomogeneous hidden Markov model for multisite streamflow reconstructions. *Water Resources Research*, *52*(10), 7837–7850. Retrieved 2023-03-21, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018887 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR018887) doi: 10.1002/2016WR018887

Bracken, C., Rajagopalan, B., & Zagona, E. (2014). A hidden Markov model combined with climate indices for multidecadal streamflow simulation. *Water Resources Research*, *50*(10), 7836–7846. Retrieved 2019-03-26, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015567 doi: 10.1002/2014WR015567

Breiman, L. (1984). *Classification and Regression Trees*. New York: Routledge. Retrieved from https://doi.org/10.1201/9781315139470 (Google-Books-ID: mlZgQgAACAAJ)

Brown, C., Ghile, Y., Laverty, M., & Li, K. (2012). Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. *Water Resources Research*, *48*(9). Retrieved 2019-10-28, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011212 doi: 10.1029/2011WR011212

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. doi: 10.1016/j.techfore.2009.08.002

Bureau of Reclamation. (2012). *Colorado River Basin Water Supply and Demand Study* (Executive Summary). Department of Interior.

California Natural Resources Agency. (2022, August). *California's Water Supply Strategy: Adapting to a Hotter, Drier Future* (Tech. Rep.).

Calvo, L., Christel, I., Terrado, M., Cucchietti, F., & Pérez-Montoro, M. (2022, January). Users' Cognitive Load: A Key Aspect to Successfully Communicate Visual Climate Information. *Bulletin of the American Meteorological Society*, *103*(1), E1–E16. Retrieved 2023-02-17, from https://journals.ametsoc.org/view/journals/bams/103/1/BAMS-D-20-0166.1.xml (Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society) doi: 10.1175/BAMS-D-20-0166.1

Cohen, S. M., Dyreson, A., Turner, S., Tidwell, V., Voisin, N., & Miara, A. (2022, July). A multi-model framework for assessing long- and short-term climate influences on the electric grid. *Applied Energy*, *317*, 119193. Retrieved 2023-01-03, from https://www.sciencedirect.com/science/article/pii/S030626192200561X doi: 10.1016/j.apenergy.2022.119193

Colorado Water Conservation Board, & Department of Natural Resources. (2018). *The Colorado Drought Mitigation and Response Plan* (Tech. Rep.).

Cook, B. I., Smerdon, J. E., Cook, E. R., Williams, A. P., Anchukaitis, K. J., Mankin, J. S., ... Wise, E. K. (2022, October). Megadroughts in the Common Era and the Anthropocene. *Nature Reviews Earth & Environment*, 1–17. Retrieved 2022-10-04, from https://www.nature.com/articles/s43017-022-00329-1 (Publisher: Nature Publishing Group) doi: 10.1038/s43017-022-00329-1

Cork, S. J., Peterson, G. D., Bennett, E. M., Petschel-Held, G., & Zurek, M. (2006). Synthesis of the storylines. *Ecology and Society*, *11*(2), 11. Retrieved 2014-01-24, from http://www.ibcperu.org/doc/isis/10071.pdf

CWCB. (2012). *Colorado River Water Availability Study Phase I Report* (Tech. Rep.). Colorado Water Conservation Board.

CWCB, & CDWR. (2016). *Upper Colorado River Basin Water Resources Planning Model User's Manual* (Tech. Rep.). Colorado Water Conservation Board and Colorado Division of Water Resources. Retrieved 2019-10-02, from https://www.colorado.gov/pacific/cdss/modeling-dataset-documentation

CWCB, & CDWR. (2022). *Colorado Basin Implementation Plan* (Tech. Rep.).

de Ruiter, M. C., & Van Loon, A. F. (2022, July). The challenges of dynamic vulnerability and how to assess it. *iScience*, 104720. Retrieved 2022-07-05, from https://www.sciencedirect.com/science/article/pii/S2589004222009920 doi: 10.1016/j.isci.2022.104720

Deser, C., Terray, L., & Phillips, A. S. (2016, March). Forced and Internal Components of Winter Air Temperature Trends over North America during the past 50 Years: Mechanisms and Implications. *Journal of Climate*, *29*(6), 2237–2258. Retrieved 2023-01-11, from https://journals.ametsoc.org/view/journals/clim/29/6/jcli-d-15-0304.1.xml (Publisher: American Meteorological Society Section: Journal of Climate) doi: 10.1175/JCLI-D-15-0304.1

Diffenbaugh, N. S., Swain, D. L., & Touma, D. (2015). Anthropogenic warming has increased drought risk in California. *Proceedings of the National Academy of Sciences*, *112*(13), 3931–3936. Retrieved from https://www.pnas.org/content/pnas/112/13/3931.full.pdf (Type: Journal Article)

Drapeau, S., Jamneshan, A., Karliczek, M., & Kupper, M. (2016, May). The algebra of conditional sets and the concepts of conditional topology and compactness. *Journal of Mathematical Analysis and Applications*, *437*(1), 561–589. Retrieved 2023-03-28, from https://www.sciencedirect.com/science/article/pii/S0022247X15011300 doi: 10.1016/j.jmaa.2015.11.057

Elsawah, S., Filatova, T., Jakeman, A. J., Kettner, A. J., Zellner, M. L., Athanasiadis, I. N., ... Lade, S. J. (2020, January). Eight grand challenges in socio-environmental systems modeling. *Socio-Environmental Systems Modelling*, *2*, 16226–16226. Retrieved 2020-08-24, from https://sesmo.org/article/view/16226 doi: 10.18174/sesmo.2020a16226

Engle, S., & Whalen, S. (2012, October). Visualizing distributed memory computations with hive plots. In *Proceedings of the Ninth International Symposium on Visualization for Cyber Security* (pp. 56–63). Seattle Washington USA: ACM. Retrieved 2023-06-01, from https://dl.acm.org/doi/10.1145/2379690.2379698 doi: 10.1145/2379690.2379698

Fischer, E. M., Sippel, S., & Knutti, R. (2021, August). Increasing probability of record-shattering climate extremes. *Nature Climate Change*, *11*(8), 689–695. Retrieved 2023-06-01, from https://www.nature.com/articles/s41558-021-01092-9 (Number: 8 Publisher: Nature Publishing Group) doi: 10.1038/s41558-021-01092-9

Flavelle, C., & Rojanasakul, M. (2023, January). As the Colorado River Shrinks, Washington Prepares to Spread the Pain. *The New York Times*. Retrieved 2023-02-09, from https://www.nytimes.com/2023/01/27/climate/colorado-river-biden-cuts.html

Fletcher, S., Hadjimichael, A., Quinn, J., Osman, K., Giuliani, M., Gold, D., ... Gordon, B. (2022, July). Equity in Water Resources Planning: A Path Forward for Decision Support Modelers. *Journal of Water Resources Planning and Management*, *148*(7), 02522005. Retrieved 2022-05-02, from https://ascelibrary.org/doi/full/10.1061/%28ASCE%29WR.1943-5452.0001573 (Publisher: American Society of Civil Engineers) doi: 10.1061/(ASCE)WR.1943-5452.0001573

Franssen, M. (2005, November). Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. *Research in Engineering Design*, *16*(1), 42–56. Retrieved 2019-07-02, from https://doi.org/10.1007/s00163-004-0057-5 doi: 10.1007/s00163-004-0057-5

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, *9*(2), 123–143. doi: doi.org/10.1023/A:1008894516817

Gerlak, A. K., & Heikkila, T. (2023, February). *Navigating the Colorado River crisis: It's time for the federal government to step up* [Text]. Retrieved 2023-02-09, from https://thehill.com/opinion/energy-environment/3847785-navigating-the-colorado-river-crisis-its-time-for-the-federal-government-to-step-up/

Gold, D. F., Reed, P. M., Gorelick, D. E., & Characklis, G. W. (2022). Power and Pathways: Exploring Robustness, Cooperative Stability, and Power Relationships in Regional Infrastructure Investment and Water Supply Management Portfolio Pathways. *Earth's Future*, *10*(2), e2021EF002472. Retrieved 2023-03-21, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002472 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021EF002472) doi: 10.1029/2021EF002472

Gold, D. F., Reed, P. M., Trindade, B. C., & Characklis, G. W. (2019). Identifying Actionable Compromises: Navigating Multi-City Robustness Conflicts to Discover Cooperative Safe Operating Spaces for Regional Water Supply Portfolios. *Water Resources Research*, *n/a*(n/a). Retrieved 2019-12-03, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019WR025462 doi: 10.1029/2019WR025462

Gotts, N. M., Van Voorn, G. A., Polhill, J. G., Jong, E. D., Edmonds, B., Hofstede, G. J., & Meyer, R. (2019, December). Agent-based modelling of socio-ecological systems: Models, projects and ontologies. *Ecological Complexity*, *40*, 100728. Retrieved 2023-05-15, from https://linkinghub.elsevier.com/retrieve/pii/S1476945X18301272 doi: 10.1016/j.ecocom.2018.07.007

Groves, D. G. (2005). *New methods for identifying robust long-term water resources management strategies for California* (Doctoral dissertation, The Pardee RAND Graduate School). Retrieved from https://doi.org/10.7249/RGSD196

Groves, D. G., & Lempert, R. J. (2007). A new analytic method for finding policy-relevant scenarios. *Global Environmental Change*, *17*(1), 73–85. doi: 10.1016/j.gloenvcha.2006.11.006

Guivarch, C., Rozenberg, J., & Schweizer, V. (2016, June). The diversity of socio-economic pathways and CO2 emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling & Software*, *80*, 336–353. Retrieved 2023-01-25, from https://www.sciencedirect.com/science/article/pii/S1364815216300706 doi: 10.1016/j.envsoft.2016.03.006

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, *23*, 485–498. Retrieved 2016-05-10, from http://dx.doi.org/10.1016/j.gloenvcha.2012.12.006 doi: http://dx.doi.org/10.1016/j.gloenvcha.2012.12.006

Hadjimichael, A., Quinn, J., & Reed, P. (2020). Advancing Diagnostic Model Evaluation to Better Understand Water Shortage Mechanisms in Institutionally Complex River Basins. *Water Resources Research*, *56*(10), e2020WR028079. Retrieved 2020-10-16, from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR028079 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020WR028079) doi: 10.1029/2020WR028079

Hadjimichael, A., Quinn, J., Wilson, E., Reed, P., Basdekas, L., Yates, D., & Garrison, M. (2020). Defining Robustness, Vulnerabilities, and Consequential Scenarios for Diverse Stakeholder Interests in Institutionally Complex River Basins.

Earth's Future, *8*(7), e2020EF001503. Retrieved 2020-07-13, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001503 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001503) doi: 10.1029/2020EF001503

Hadjimichael, A., Reed, P. M., & Quinn, J. D. (2020). Navigating Deeply Uncertain Trade-offs in Harvested Predator-Prey Systems. *Complexity*, *2020*, e4170453. Retrieved 2020-03-03, from https://www.hindawi.com/journals/complexity/2020/4170453/ (Publisher: Hindawi) doi: https://doi.org/10.1155/2020/4170453

Hadjimichael, A., Yoon, J., Reed, P., Voisin, N., & Xu, W. (2023, February). Exploring the Consistency of Water Scarcity Inferences between Large-Scale Hydrologic and Node-Based Water System Model Representations of the Upper Colorado River Basin. *Journal of Water Resources Planning and Management*, *149*(2), 04022081. Retrieved 2022-12-08, from https://ascelibrary.org/doi/10.1061/JWRMD5.WRENG-5522 (Publisher: American Society of Civil Engineers) doi: 10.1061/JWRMD5.WRENG-5522

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. Retrieved from https://atoc.colorado.edu/~whan/ATOC4800_5000/Materials/Hawkins_sutton.pdf (Type: Journal Article)

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. doi: 10.1061/(ASCE)WR.1943-5452.0000509

Herman, J. D., Zeff, H. B., Reed, P. M., & Characklis, G. W. (2014, October). Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, *50*(10), 7692–7713. Retrieved 2017-11-29, from http://onlinelibrary.wiley.com/doi/10.1002/2014WR015338/abstract doi: 10.1002/2014WR015338

Hobbins, R., Muñoz-Erickson, T. A., & Miller, C. (2021). Producing and Communicating Flood Risk: A Knowledge System Analysis of FEMA Flood Maps in New York City. In Z. A. Hamstead, D. M. Iwaniec, T. McPhearson, M. Berbés-Blázquez, E. M. Cook, & T. A. Muñoz-Erickson (Eds.), *Resilient Urban Futures* (pp. 67–84). Cham: Springer International Publishing. Retrieved 2023-06-01, from https://doi.org/10.1007/978-3-030-63131-4_5 doi: 10.1007/978-3-030-63131-4_5

Hoylman, Z. H., Bocinsky, R. K., & Jencso, K. G. (2022, May). Drought assessment has been outpaced by climate change: empirical arguments for a paradigm shift. *Nature Communications*, *13*(1), 2715. Retrieved 2022-07-29, from https://www.nature.com/articles/s41467-022-30316-5 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/s41467-022-30316-5

Inselberg, A. (2009). *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. New York, NY: Springer. Retrieved 2023-08-31, from https://link.springer.com/10.1007/978-0-387-68628-8 doi: 10.1007/978-0-387-68628-8

IPCC. (2023). *Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Tech. Rep.). Geneva, Switzerland: IPCC.

Iwanaga, T., Wang, H.-H., Hamilton, S. H., Grimm, V., Koralewski, T. E., Salado, A., ... Little, J. C. (2021). Socio-technical scales in socio-environmental modeling: Managing a system-of-systems modeling approach. *Environmental Modelling & Software*, *135*, 104885. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7537632/pdf/main.pdf (Type: Journal Article) doi: https://doi.org/10.1016/j.envsoft.2020.104885

Kasprzyk, J. R., Nataraj, S., Reed, P. M., & Lempert, R. J. (2013, April). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, *42*, 55–71. Retrieved 2019-09-30, from http://

www.sciencedirect.com/science/article/pii/S1364815212003131 doi: 10.1016/j.envsoft.2012.12.007

Kenney, D. S. (2005). Prior appropriation and water rights reform in the western United States. In B. R. Bruns, C. Ringler, & R. S. Meinzen-Dick (Eds.), *Water Rights Reform: Lessons for Institutional Design* (p. 336). International Food Policy Research Institute.

Krauß, W. (2020, January). Narratives of change and the co-development of climate services for action. *Climate Risk Management*, *28*, 100217. Retrieved 2023-02-16, from https://www.sciencedirect.com/science/article/pii/S2212096320300073 doi: 10.1016/j.crm.2020.100217

Krauß, W., & Bremer, S. (2020, January). The role of place-based narratives of change in climate risk governance. *Climate Risk Management*, *28*, 100221. Retrieved 2023-02-16, from https://www.sciencedirect.com/science/article/pii/S2212096320300115 doi: 10.1016/j.crm.2020.100221

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., . . . Di Baldassarre, G. (2022, August). The challenge of unprecedented floods and droughts in risk management. *Nature*, *608*(7921), 1–7. Retrieved 2022-08-04, from https://www.nature.com/articles/s41586-022-04917-5 (Publisher: Nature Publishing Group) doi: 10.1038/s41586-022-04917-5

Krzywinski, M., Birol, I., Jones, S. J., & Marra, M. A. (2012, September). Hive plots—rational approach to visualizing networks. *Briefings in Bioinformatics*, *13*(5), 627–644. Retrieved 2023-03-29, from https://doi.org/10.1093/bib/bbr069 doi: 10.1093/bib/bbr069

Kwakkel, J. H. (2019). A generalized many-objective optimization approach for scenario discovery. *FUTURES & FORESIGHT SCIENCE*, *1*(2), e8. Retrieved 2019-12-04, from https://onlinelibrary.wiley.com/doi/abs/10.1002/ffo2.8 doi: 10.1002/ffo2.8

Kwakkel, J. H., & Haasnoot, M. (2019). Supporting DMDU: A taxonomy of approaches and tools. In V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, & S. W. Popper (Eds.), *Decision Making under Deep Uncertainty.* Springer.

Lehner, F., & Deser, C. (2023, May). Origin, importance, and predictive limits of internal climate variability. *Environmental Research: Climate*, *2*(2), 023001. Retrieved 2023-06-01, from https://dx.doi.org/10.1088/2752-5295/accf30 (Publisher: IOP Publishing) doi: 10.1088/2752-5295/accf30

Lemos, M. C., Kirchhoff, C. J., & Ramprasad, V. (2012, November). Narrowing the climate information usability gap. *Nature Climate Change*, *2*(11), 789–794. Retrieved 2022-09-29, from https://www.nature.com/articles/nclimate1614 (Number: 11 Publisher: Nature Publishing Group) doi: 10.1038/nclimate1614

Lemos, M. C., & Morehouse, B. J. (2005, April). The co-production of science and policy in integrated climate assessments. *Global Environmental Change*, *15*(1), 57–68. Retrieved 2021-08-12, from https://www.sciencedirect.com/science/article/pii/S0959378004000652 doi: 10.1016/j.gloenvcha.2004.09.004

Lempert, R. J. (2019). Robust Decision Making (RDM). In V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, & S. W. Popper (Eds.), *Decision Making under Deep Uncertainty: From Theory to Practice* (pp. 23–51). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-05252-2_2 doi: 10.1007/978-3-030-05252-2_2

Lempert, R. J., & Groves, D. G. (2010, July). Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the American west. *Technological Forecasting and Social Change*, *77*(6), 960–974. Retrieved 2023-05-13, from https://www.sciencedirect.com/science/article/pii/S0040162510000740 doi: 10.1016/j.techfore.2010.04.007

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006, April). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science*, *52*(4), 514–528. Retrieved 2022-09-29, from https://pubsonline.informs.org/doi/10.1287/mnsc.1050.0472 (Publisher: INFORMS) doi: 10.1287/mnsc.1050.0472

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years*. RAND Corporation. Retrieved 2017-09-14, from https://www.rand.org/pubs/monograph_reports/MR1626.html

Lorenz, R., Stalhandske, Z., & Fischer, E. M. (2019). Detection of a climate change signal in extreme heat, heat stress, and cold in Europe from observations. *Geophysical Research Letters*, *46*(14), 8363–8374. (Publisher: Wiley Online Library)

Lorenz, S., Dessai, S., Forster, P. M., & Paavola, J. (2015, November). Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2055), 20140457. Retrieved 2023-02-17, from https://royalsocietypublishing.org/doi/10.1098/rsta.2014.0457 (Publisher: Royal Society) doi: 10.1098/rsta.2014.0457

Lukat, E., Lenschow, A., Dombrowsky, I., Meergans, F., Schütze, N., Stein, U., & Pahl-Wostl, C. (2023, March). Governance towards coordination for water resources management: The effect of governance modes. *Environmental Science & Policy*, *141*, 50–60. Retrieved 2023-02-09, from https://www.sciencedirect.com/science/article/pii/S1462901122003860 doi: 10.1016/j.envsci.2022.12.016

Maier, H. R., Guillaume, J. H., 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.

Malers, S. A., Ray R. Bennett, & Catherine, N.-L. (2001). Colorado's Decision Support Systems: Data-Centered Water Resources Planning and Administration. *Watershed Management and Operations Management 2000*, 1–9. Retrieved 2019-12-04, from https://ascelibrary.org/doi/abs/10.1061/40499(2000)153 doi: 10.1061/40499(2000)153

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., & Popper, S. W. (Eds.). (2019). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing. Retrieved 2020-08-16, from https://www.springer.com/gp/book/9783030052515 doi: 10.1007/978-3-030-05252-2

Markolf, S. A., Chester, M. V., Eisenberg, D. A., Iwaniec, D. M., Davidson, C. I., Zimmerman, R., . . . Chang, H. (2018). Interdependent Infrastructure as Linked Social, Ecological, and Technological Systems (SETSs) to Address Lock-in and Enhance Resilience. *Earth's Future*, *6*(12), 1638–1659. Retrieved 2023-06-01, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2018EF000926 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018EF000926) doi: 10.1029/2018EF000926

McCoy, A. L., Jacobs, K. L., Vano, J. A., Wilson, J. K., Martin, S., Pendergrass, A. G., & Cifelli, R. (2022). The Press and Pulse of Climate Change: Extreme Events in the Colorado River Basin. *JAWRA Journal of the American Water Resources Association*, *58*(6), 1076–1097. Retrieved 2023-01-25, from https://onlinelibrary.wiley.com/doi/abs/10.1111/1752-1688.13021 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.13021) doi: 10.1111/1752-1688.13021

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2), 239–245. Retrieved from https://www.jstor.org.proxy.library.cornell.edu/stable/1268522 doi: 10.2307/1268522

McPhail, C., Maier, H. R., Kwakkel, J. H., Giuliani, M., Castelletti, A., & Westra, S. (2018). Robustness Metrics: How Are They Calculated, When Should They Be Used and Why Do They Give Different Results? *Earth's Future*, *6*, 169–191. doi: 10.1002/2017EF000649@10.1002/(ISSN)2328-4277.RESDEC1

Meko, D. M., Woodhouse, C. A., Baisan, C. A., Knight, T., Lukas, J. J., Hughes, M. K., & Salzer, M. W. (2007). Medieval drought in the upper Colorado River

Basin. *Geophysical Research Letters*, *34*(10). Retrieved 2023-03-20, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2007GL029988 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007GL029988) doi: 10.1029/2007GL029988

Moallemi, E. A., Kwakkel, J. H., de Haan, F. J., & Bryan, B. A. (2020, November). Exploratory modeling for analyzing coupled human-natural systems under uncertainty. *Global Environmental Change*, *65*, 102186. Retrieved 2020-11-17, from http://www.sciencedirect.com/science/article/pii/S095937802030769X doi: 10.1016/j.gloenvcha.2020.102186

Moallemi, E. A., Zare, F., Reed, P. M., Elsawah, S., Ryan, M. J., & Bryan, B. A. (2020, January). Structuring and evaluating decision support processes to enhance the robustness of complex human–natural systems. *Environmental Modelling & Software*, *123*, 104551. Retrieved 2019-12-03, from http://www.sciencedirect.com/science/article/pii/S1364815219306905 doi: 10.1016/j.envsoft.2019.104551

Mondal, A., & Mujumdar, P. P. (2015, January). Return levels of hydrologic droughts under climate change. *Advances in Water Resources*, *75*, 67–79. Retrieved 2023-06-02, from https://www.sciencedirect.com/science/article/pii/S030917081400219X doi: 10.1016/j.advwatres.2014.11.005

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., ... Wilbanks, T. J. (2010, February). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747–756. Retrieved 2023-06-12, from https://www.nature.com/articles/nature08823 (Number: 7282 Publisher: Nature Publishing Group) doi: 10.1038/nature08823

Naumann, G., Alfieri, L., Wyser, K., Mentaschi, L., Betts, R. A., Carrao, H., ... Feyen, L. (2018). Global Changes in Drought Conditions Under Different Levels of Warming. *Geophysical Research Letters*, *45*(7), 3285–3296. Retrieved 2023-03-20, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076521 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL076521) doi: 10.1002/2017GL076521

Nowak, K., Prairie, J., Rajagopalan, B., & Lall, U. (2010). A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow. *Water Resources Research*, *46*(8). Retrieved 2023-03-21, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008530 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR008530) doi: 10.1029/2009WR008530

Overpeck, J. T., & Udall, B. (2020, June). Climate change and the aridification of North America. *Proceedings of the National Academy of Sciences*, *117*(22), 11856–11858. Retrieved 2023-02-09, from https://www.pnas.org/doi/full/10.1073/pnas.2006323117 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.2006323117

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., ... van Vuuren, D. P. (2014, February). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400. Retrieved 2020-08-14, from https://doi.org/10.1007/s10584-013-0905-2 doi: 10.1007/s10584-013-0905-2

Parsons, R., & Bennett, R. (2006). Reservoir Operations Management Using a Water Resources Model. *Operating Reservoirs in Changing Conditions*, 304–311. Retrieved 2019-07-08, from https://ascelibrary.org/doi/abs/10.1061/40875(212)30 doi: 10.1061/40875(212)30

Pedersen, J. T. S., van Vuuren, D., Gupta, J., Santos, F. D., Edmonds, J., & Swart, R. (2022, July). IPCC emission scenarios: How did critiques affect their quality and relevance 1990–2022? *Global Environmental Change*, *75*, 102538. Retrieved 2023-08-30, from https://www.sciencedirect.com/science/article/pii/S0959378022000760 doi: 10.1016/j.gloenvcha.2022.102538

Popper, S. W., Berrebi, C., Griffin, J., Light, T., Daehner, E. M., & Crane, K. (2009, December). *Natural Gas and Israel's Energy Future: Near-Term Decisions from a Strategic Perspective* (Tech. Rep.). RAND Corporation. Retrieved 2023-05-13, from https://www.rand.org/pubs/monographs/MG927.html

Priss, U. (2021). Set Visualisations with Euler and Hasse Diagrams. In M. Cochez, M. Croitoru, P. Marquis, & S. Rudolph (Eds.), *Graph Structures for Knowledge Representation and Reasoning* (pp. 72–83). Cham: Springer International Publishing. doi: 10.1007/978-3-030-72308-8_5

Pruett, W. A., & Hester, R. L. (2016, June). The Creation of Surrogate Models for Fast Estimation of Complex Model Outcomes. *PLOS ONE*, *11*(6), e0156574. Retrieved 2020-11-16, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0156574 (Publisher: Public Library of Science) doi: 10.1371/journal.pone.0156574

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. Retrieved 2023-01-04, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001650 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001650) doi: 10.1029/2020EF001650

Quinn, J. D., Reed, P. M., Giuliani, M., Castelletti, A., Oyler, J. W., & Nicholas, R. E. (2018, July). 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 2019-10-26, from http://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018WR022743 doi: 10.1029/2018WR022743

Reed, P. M., Hadjimichael, A., Malek, K., Karimi, T., Vernon, C. R., Srikrishnan, V. A., … Rice, J. S. (2022). *Addressing Uncertainty in Multisector Dynamics Research.* Retrieved 2022-03-16, from https://uc-ebook.org/ doi: 10.5281/ZENODO.6110623

Reed, P. M., Hadjimichael, A., Moss, R. H., Monier, E., Alba, S., Brelsford, C., … Yoon, J. (2022, January). *MultiSector Dynamics: Scientific Challenges and a Research Vision for 2030, A Community of Practice Supported by the United States Department of Energy's Office of Science* (Tech. Rep.). MultiSector Dynamics Community of Practice. Retrieved 2022-05-11, from https://zenodo.org/record/6144309 doi: 10.5281/zenodo.6144309

Robbins, J. (2019). On the water-starved colorado river, drought is the new normal. *Yale Environment*, *360*.

Salas, J. D., Obeysekera, J., & Vogel, R. M. (2018, February). Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal*, *63*(3), 325–352. Retrieved 2023-06-01, from https://doi.org/10.1080/02626667.2018.1426858 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2018.1426858) doi: 10.1080/02626667.2018.1426858

Savage, L. J. (1951, March). The Theory of Statistical Decision. *Journal of the American Statistical Association*, *46*(253), 55–67. Retrieved 2019-10-01, from https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1951.10500768 doi: 10.1080/01621459.1951.10500768

Savelli, E., Rusca, M., Cloke, H., & Di Baldassarre, G. (2022). Drought and society: Scientific progress, blind spots, and future prospects. *WIREs Climate Change*, *n/a*(n/a), e761. Retrieved 2022-02-04, from https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.761 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.761) doi: 10.1002/wcc.761

Schlumberger, J., Haasnoot, M., Aerts, J., & de Ruiter, M. (2022, October). Proposing DAPP-MR as a disaster risk management pathways framework for complex, dynamic multi-risk. *iScience*, *25*(10), 105219. Retrieved 2023-06-01, from https://www.sciencedirect.com/science/article/pii/S2589004222014912 doi:

10.1016/j.isci.2022.105219

Schlüter, M., Mcallister, R. R. J., Arlinghaus, R., Bunnefeld, N., Eisenack, K., Hölker, F., . . . Stöven, M. (2012). New Horizons for Managing the Environment: A Review of Coupled Social-Ecological Systems Modeling. *Natural Resource Modeling*, *25*(1), 219–272. Retrieved 2023-05-11, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1939-7445.2011.00108.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1939-7445.2011.00108.x) doi: 10.1111/j.1939-7445.2011.00108.x

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., . . . Zenghelis, D. A. (2018). Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, *151*(3), 555–571. doi: 10.1007/s10584-018-2317-9

Shi, R., Hobbs, B. F., Quinn, J. D., Lempert, R., & Knopman, D. (2023, February). City-Heat Equity Adaptation Tool (City-HEAT): Multi-objective optimization of environmental modifications and human heat exposure reductions for urban heat adaptation under uncertainty. *Environmental Modelling & Software*, *160*, 105607. Retrieved 2023-01-05, from https://www.sciencedirect.com/science/article/pii/S1364815222003073 doi: 10.1016/j.envsoft.2022.105607

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129. doi: https://doi.org/10.1037/h0042769

Simpson, N. P., Mach, K. J., Constable, A., Hess, J., Hogarth, R., Howden, M., . . . Trisos, C. H. (2021, April). A framework for complex climate change risk assessment. *One Earth*, *4*(4), 489–501. Retrieved 2021-04-28, from https://www.cell.com/one-earth/abstract/S2590-3322(21)00179-2 (Publisher: Elsevier) doi: 10.1016/j.oneear.2021.03.005

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., . . . Wilby, R. L. (2021, July). Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, *25*(7), 3897–3935. Retrieved 2023-06-01, from https://hess.copernicus.org/articles/25/3897/2021/ (Publisher: Copernicus GmbH) doi: 10.5194/hess-25-3897-2021

Smith, R., Zagona, E., Kasprzyk, J., Bonham, N., Alexander, E., Butler, A., . . . Jerla, C. (2022, January). Decision Science Can Help Address the Challenges of Long-Term Planning in the Colorado River Basin. *JAWRA Journal of the American Water Resources Association*, *n/a*(n/a). Retrieved 2022-01-25, from https://onlinelibrary.wiley.com/doi/abs/10.1111/1752-1688.12985 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.12985) doi: 10.1111/1752-1688.12985

State of Colorado. (2015). *Colorado's water plan* (Tech. Rep.). Denver, Colorado.

State of Colorado. (2023). *Colorado's Water Plan* (Tech. Rep.).

Steinmann, P., Auping, W. L., & Kwakkel, J. H. (2020, July). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change*, *156*, 120052. Retrieved 2020-05-04, from http://www.sciencedirect.com/science/article/pii/S0040162519302380 doi: 10.1016/j.techfore.2020.120052

Stevenson, S., Coats, S., Touma, D., Cole, J., Lehner, F., Fasullo, J., & Otto-Bliesner, B. (2022, March). Twenty-first century hydroclimate: A continually changing baseline, with more frequent extremes. *Proceedings of the National Academy of Sciences*, *119*(12), e2108124119. Retrieved 2022-04-04, from https://www.pnas.org/doi/10.1073/pnas.2108124119 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.2108124119

Sun, Q., Zhang, X., Zwiers, F., Westra, S., & Alexander, L. V. (2021). A global, continental, and regional analysis of changes in extreme precipitation. *Journal of Climate*, *34*(1), 243–258.

Sunkara, S. V., Singh, R., Gold, D., Reed, P., & Bhave, A. (2023). How Should Diverse Stakeholder Interests Shape Evaluations of Complex Water Resources Systems Robustness When Confronting Deeply Uncertain Changes? *Earth's Future*, *11*(8), e2022EF003469. Retrieved 2023-08-30, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2022EF003469 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022EF003469) doi: 10.1029/2022EF003469

Trindade, B. C., Gold, D. F., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2020, October). Water pathways: An open source stochastic simulation system for integrated water supply portfolio management and infrastructure investment planning. *Environmental Modelling & Software*, *132*, 104772. Retrieved 2020-08-21, from http://www.sciencedirect.com/science/article/pii/S1364815220301511 doi: 10.1016/j.envsoft.2020.104772

Trindade, B. C., Reed, P. M., & Characklis, G. W. (2019, December). Deeply uncertain pathways: Integrated multi-city regional water supply infrastructure investment and portfolio management. *Advances in Water Resources*, *134*, 103442. Retrieved 2020-03-31, from http://www.sciencedirect.com/science/article/pii/S0309170819306475 doi: 10.1016/j.advwatres.2019.103442

Trindade, B. C., Reed, P. M., Herman, J. D., Zeff, H. B., & Characklis, G. W. (2017, June). Reducing regional drought vulnerabilities and multi-city robustness conflicts using many-objective optimization under deep uncertainty. *Advances in Water Resources*, *104*(Supplement C), 195–209. Retrieved from http://www.sciencedirect.com/science/article/pii/S0309170816307333 doi: 10.1016/j.advwatres.2017.03.023

Tufte, E. R. (1990). *Envisioning Information* (Vol. 6410). Cheshire, Connecticut: Graphics Press.

Vahmani, P., Jones, A. D., & Li, D. (2022). Will Anthropogenic Warming Increase Evapotranspiration? Examining Irrigation Water Demand Implications of Climate Change in California. *Earth's Future*, *10*(1), e2021EF002221. Retrieved 2023-05-11, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002221 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021EF002221) doi: 10.1029/2021EF002221

van den Elzen, S., & van Wijk, J. J. (2013). Small Multiples, Large Singles: A New Approach for Visual Data Exploration. *Computer Graphics Forum*, *32*(3pt2), 191–200. Retrieved 2023-03-29, from https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12106 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12106) doi: 10.1111/cgf.12106

Van Ruijven, B., Carlsen, H., Chaturvedi, V., Ebi, K., Fuglestvedt, J., Gasalla, M., . . . Leininger, J. (2023). The SSP-RCP scenario framework: progress, needs, and next steps–Insights from the Scenarios Forum 2022. Bordeaux, France.

Wald, A. (1950). *Statistical decision functions*. Wiley.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated assessment*, *4*(1), 5–17. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.7495Źrep=rep1Źtype=pdf (Type: Journal Article)

Wegman, E. J. (1990, September). Hyperdimensional Data Analysis Using Parallel Coordinates. *Journal of the American Statistical Association*, *85*(411), 664–675. Retrieved 2023-06-01, from https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10474926 (Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1990.10474926) doi: 10.1080/01621459.1990.10474926

Whitney, K. M., Vivoni, E. R., Bohn, T. J., Mascaro, G., Wang, Z., Xiao, M., . . . White, D. D. (2023, February). Spatial attribution of declining Colorado River streamflow under future warming. *Journal of Hydrology*, *617*, 129125. Retrieved 2023-01-25, from https://www.sciencedirect.com/science/article/pii/

S0022169423000677 doi: 10.1016/j.jhydrol.2023.129125

Woodhouse, C. A., Gray, S. T., & Meko, D. M. (2006). Updated streamflow reconstructions for the Upper Colorado River Basin. *Water Resources Research*, *42*(5). Retrieved 2022-04-06, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2005WR004455 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005WR004455) doi: 10.1029/2005WR004455

Woodhouse, C. A., & Overpeck, J. T. (1998, December). 2000 Years of Drought Variability in the Central United States. *Bulletin of the American Meteorological Society*, *79*(12), 2693–2714. Retrieved 2022-04-14, from https://journals.ametsoc.org/view/journals/bams/79/12/1520-0477_1998_079_2693_yodvit_2_0_co_2.xml (Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society) doi: 10.1175/1520-0477(1998)079<2693:YODVIT>2.0.CO;2

Woodhouse, C. A., Smith, R. M., McAfee, S. A., Pederson, G. T., McCabe, G. J., Miller, W. P., & Csank, A. (2021, January). Upper Colorado River Basin 20th century droughts under 21st century warming: Plausible scenarios for the future. *Climate Services*, *21*, 100206. Retrieved 2022-04-01, from https://www.sciencedirect.com/science/article/pii/S2405880720300583 doi: 10.1016/j.cliser.2020.100206

Wyborn, C., Datta, A., Montana, J., Ryan, M., Leith, P., Chaffin, B., . . . van Kerkhoff, L. (2019). Co-Producing Sustainability: Reordering the Governance of Science, Policy, and Practice. *Annual Review of Environment and Resources*, *44*(1), 319–346. Retrieved 2023-05-11, from https://doi.org/10.1146/annurev-environ-101718-033103 (_eprint: https://doi.org/10.1146/annurev-environ-101718-033103) doi: 10.1146/annurev-environ-101718-033103

Yang, Y., Botton, M. R., Scott, E. R., & Scott, S. A. (2017, May). Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics*, *18*(7), 673–685. Retrieved 2023-06-01, from https://www.futuremedicine.com/doi/abs/10.2217/pgs-2017-0033 (Publisher: Future Medicine) doi: 10.2217/pgs-2017-0033

Yang, Y., Roderick, M. L., Yang, D., Wang, Z., Ruan, F., McVicar, T. R., . . . Beck, H. E. (2021, June). Streamflow stationarity in a changing world. *Environmental Research Letters*, *16*(6), 064096. Retrieved 2023-06-01, from https://dx.doi.org/10.1088/1748-9326/ac08c1 (Publisher: IOP Publishing) doi: 10.1088/1748-9326/ac08c1

Multi-actor, multi-impact scenario discovery of consequential narrative storylines for human-natural systems planning

Antonia Hadjimichael[1,2], Patrick M. Reed[3], Julianne D. Quinn[4], Chris R. Vernon[5], Travis Thurber[5]

[1]Department of Geosciences, The Pennsylvania State University, State College, PA, USA
[2]Earth and Environmental Systems Institute (EESI), The Pennsylvania State University, State College, PA, USA
[3]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[4]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA
[5]Atmospheric Sciences & Global Change, Pacific Northwest National Laboratory, Richland, WA, USA

Key Points:

- Introduce a hierarchical classification framework for scenario discovery, to identify diverse stakeholder impacts and consequential dynamics.
- Demonstrate the framework in the Upper Colorado River Basin with hundreds of stakeholders and complex human-natural system interactions.
- The framework improves understanding and selection of narrative drought storylines through their effects on user- and basin-scale impacts.

Corresponding author: Antonia Hadjimichael, hadjimichael@psu.edu

Abstract

Scenarios have emerged as valuable tools in managing complex human-natural systems, but the traditional approach of limiting focus on a small number of predetermined scenarios can inadvertently miss consequential dynamics, extremes, and diverse stakeholder impacts. Exploratory modeling approaches have been developed to address these issues by exploring a wide range of possible futures and identifying those that yield consequential vulnerabilities. However, vulnerabilities are typically identified based on aggregate robustness measures that do not take full advantage of the richness of the underlying dynamics in the large ensembles of model simulations and can make it hard to identify key dynamics and/or narrative storylines that can guide planning or further analyses. This study introduces the FRamework for Narrative Scenarios and Impact Classification (FRNSIC; pronounced "forensic"): a scenario discovery framework that addresses these challenges by organizing and investigating consequential scenarios using hierarchical classification of diverse outcomes across actors, sectors, and scales, while also aiding in the selection of narrative storylines, based on system dynamics that drive consequential outcomes. We present an application of this framework to the Upper Colorado River Basin, focusing on decadal droughts and their water scarcity implications for the basin's diverse users and its obligations to downstream states through Lake Powell. We show how FRNSIC can explore alternative sets of impact metrics and drought dynamics and use them to identify narrative drought storylines, that can be used to inform future adaptation planning.

Plain Language Summary

Scenario analysis is a useful tool for assessing the impacts of future conditions or alternative strategies. Focusing on a small number of predetermined scenarios can, however, limit our understanding of key uncertainties, and fail to represent diverse stakeholder impacts. Approaches such as exploratory modeling have been developed to address these issues by exploring a wide range of possible futures and system perspectives. These approaches often involve large simulation experiments with their own interpretability challenges. So, on one hand, we recognize the need to utilize large ensembles of hypothesized changes, but on the other hand, each additional dimension considered makes it more difficult to convey actionable insights. We introduce the FRamework for Narrative Scenarios and Impact Classification (FRNSIC; pronounced "forensic"), a scenario discovery framework that helps users identify narrative scenarios that capture key system dynamics and as well as important outcomes. We demonstrate its application to the Upper Colorado River Basin, focusing on decadal droughts and their water scarcity implications for the basin's diverse users and its obligations to downstream states through Lake Powell. We explore alternative impact metrics and dynamics, identifying narrative storylines with significant impacts, which can be used in future planning efforts to adapt to these stressed conditions.

1 Introduction

Understanding and managing human-natural systems confronting change remains an open challenge, as they are highly complex systems with deep uncertainties shaping their candidate futures (Elsawah et al., 2020; Reed, Hadjimichael, Moss, et al., 2022; Schlüter et al., 2012). The interactions and feedbacks between human and natural components, resources, actors, and institutions create nested systems-of-systems that operate at and across multiple scales (Iwanaga et al., 2021). Holistically attending to such complexity and advancing our understanding of such systems requires approaches that transcend disciplinary framings and traditional approaches (Wyborn et al., 2019). Pervasive deep uncertainties are also present in these systems, due to incomplete or contested expert knowledge on system boundaries or key system processes and drivers (Marchau et al., 2019; Moallemi, Zare, et al., 2020). Finally, the multiple and often conflicting objectives of various stakeholders in these systems further complicate the identification of relevant knowledge that engages diverse worldviews to inform their management (Kasprzyk et al., 2013).

Scenario analysis has become increasingly important in understanding and planning for human-natural systems, as scenarios present useful tools in dealing with some of these challenges (Groves

& Lempert, 2007; Moss et al., 2010; O'Neill et al., 2014; Pedersen et al., 2022; Van Ruijven et al., 2023). Scenarios help us assess and communicate the potential severity of hypothesized conditions and deep uncertainties, for example the impacts of a changing climate on local systems (e.g., Vahmani et al. (2022)). They can also act as reference cases for comparison and negotiation of alternative strategies to follow, for example quantifying deviations from historical conditions as a result of different stressors and human actions (e.g., Cohen et al. (2022)). Or they can help capture system complexity in narrative aggregate storylines, for example as they are used by the Intergovernmental Panel on Climate Change to communicate the impacts of alternative emissions pathways (e.g., IPCC (2023)).

An important challenge surrounding the use of scenarios is the number of candidate future states considered, as well as the conditions used to establish their relevance. Using a small number of deterministic future states has well-documented limitations, especially arising from the presence of internal variability (Hawkins & Sutton, 2009; Lehner & Deser, 2023), deep uncertainty about the future (Lempert et al., 2006; Quinn et al., 2020), and the adaptive complexity of human-natural systems (Markolf et al., 2018; Reed, Hadjimichael, Moss, et al., 2022; Simpson et al., 2021). Focusing only on the interests of, or the impacts to, a small number of actors carries its own challenges that undermine successfully engaging with the diverse perspectives of affected stakeholders. Groves and Lempert (2007) point out that *a priori* specification of a small set of "interesting" scenarios to aid narrative clarity, in absence of broader exploratory analysis, might inappropriately narrow the focus to the concerns and values of those involved in crafting them. They might not necessarily be salient with the diverse stakeholders affected, who might view the particular set of selected scenarios as biased or arbitrary. Moreover, the broad array of human as well as natural uncertainties that could shape consequential future outcomes increases the risk that a limited focus on a few specified scenarios would miss key insights (Moallemi, Kwakkel, et al., 2020).

Recognizing the myopic nature of a limited set of pre-specified scenarios or futures, there have been significant advancements in the domain of exploratory modeling (Bankes, 1993) and scenario discovery (Bryant & Lempert, 2010; Groves & Lempert, 2007). As reviewed by Moallemi, Kwakkel, et al. (2020) these approaches focus on the exploration of large ensembles of possible futures and the *a posteriori* identification of consequential scenarios. These approaches have largely been articulated in support of decision making under deep uncertainty methods, such as Robust Decision Making (RDM; Lempert et al. (2003)) and its Many-Objective extension (MORDM; Kasprzyk et al. (2013)), Dynamic Adaptive Policy Pathways (Haasnoot et al., 2013; Schlumberger et al., 2022), Info-Gap (Ben-Haim, 2006), and Decision Scaling (Brown et al., 2012). They structure large exploratory ensemble experiments to investigate diverse hypothesized drivers of change and classify the resulting "states of the world" (SOWs) based on whether they have consequential outcomes for the system's stakeholders. This process of ensemble classification and identification of a subset of consequential SOWs is termed scenario discovery (Bryant & Lempert, 2010; Groves & Lempert, 2007; Steinmann et al., 2020). As such, these exploratory modeling frameworks introduce more quantitative rigor by examining the space of possible future uncertainty and associated consequences more fully (Lempert et al., 2006). Put simply, a broader array of "what if" questions are engaged before selecting scenarios.

Past studies have reviewed and offered taxonomies of these frameworks (Herman et al., 2015; Kwakkel & Haasnoot, 2019; Moallemi, Zare, et al., 2020); at their core they all encompass the following central elements: elucidation or generation of alternative management or planning actions, exploration of alternative SOWs (potential futures or uncertainties), quantification of performance (typically a measure of "robustness"), and vulnerability or tradeoff analysis, where consequential scenarios are identified and strategies are selected, according to the quantified performance. Robustness metrics are used to rank how well systems perform based on their expected value (Wald, 1950), regret (Savage, 1951), or satisficing criteria (Simon, 1956), as extensively reviewed by McPhail et al. (2018). There is an expansive body of literature on scenario discovery that has compared the value and effects of using robustness metrics across a variety of problems and case studies to demonstrate that the choice of metric can have critical implications for which SOWs are deduced as consequential (i.e., which scenarios are selected for further inspection; Herman et al. (2015); Maier et al. (2016); McPhail et al. (2018); Sunkara et al. (2023)). Hadjimichael, Quinn, Wilson, et al. (2020) show that systems with diverse stakeholders introduce additional challenges to defining the appropriate metric to classify consequential SOWs and select a subset of ensemble members that warrant follow-on analysis given their consequential outcomes or challenging dynamics. In systems with many actors, the choice of a singular aggregated metric can ignore asymmetries in stakeholder values and agency (Franssen, 2005), and implicitly suppress the diverse scenario impacts on different users from more explicit consideration in planning (Fletcher et al., 2022). Recognizing this limitation, some studies have looked at multi-actor robustness tradeoffs, by applying the same criterion to the performance of different actors (Gold et al., 2019; Herman et al., 2014; Trindade et al., 2017). Others have applied gradients of a threshold or criterion as a way of capturing different levels of acceptability or relation to past experience to different stakeholders (Bonham et al., 2022; Hadjimichael, Quinn, & Reed, 2020; Hadjimichael, Quinn, Wilson, et al., 2020; Quinn et al., 2020).

A related challenge that arises from aggregation when defining robustness criteria for target levels of system performance is that they can collapse the temporal or spatial dynamics of a scenario into a single outcome by which each scenario is to be classified. For example, there could be a case were two scenarios produce the same average supply of a resource, but one shows substantial temporal variation whereas the other hovers around its mean. One could make the case that we can simply include an additional metric of variance to further disaggregate, but we might be interested in the overall dynamic behavior of the system or other qualitative information, for example common oscillation patterns of different scenarios, the presence of stable equilibria or tipping points. Using metrics that temporally aggregate these dynamics limits the use of this information (Hadjimichael, Reed, & Quinn, 2020). As a result, authors have proposed methods that can temporally classify the simulation dynamics themselves, instead of some aggregated outcome (e.g., Steinmann et al., 2020).

A final important consideration surrounding the development and use of scenarios relates to conveying actionable information. We face challenges in maintaining their narrative capacities (Krauß, 2020; Krauß & Bremer, 2020), encouraging the usability of climate impact findings (Lemos & Morehouse, 2005; Lemos et al., 2012), and producing consequential insights that hold direct beneficial value to the dependent human and environmental systems. Literature on co-production and cognitive research highlights that the way information is presented to and processed by its users is important to how they understand and choose to use it (Calvo et al., 2022; S. Lorenz et al., 2015). Lemos et al. (2012), for example, point out that relating new findings (e.g., potential future impacts on one's crop) to past experiences and memories (e.g., impacts of a past significant drought to one's crop) can help connect that information to their analytical and experiential processing abilities. Highlighting connections to relevant personal experiences also fosters the usability of the new findings. Literature on narrative scenarios highlights that the use of local narratives can give meaning to abstract scientific information and is central to making sense of what it means to live within a changing climate (Krauß & Bremer, 2020).

As such, tools like storylines and narrative scenarios can aid in making connections between new scientific findings and past relevant experiences, as well as form the basis of new analysis iterations (Cork et al., 2006; Krauß, 2020; Lempert et al., 2006; Shepherd et al., 2018). Narrative scenarios can indeed be derived from a RDM analysis (Lempert, 2019). For example, analysts, stakeholders and decision makers can use the discovered scenarios to more closely investigate system processes and dynamics, such as key reasons that lead to failure (e.g., Popper et al. (2009)), or use them as a basis for reiteration and evaluation of new strategies or stressors of interest (e.g., Groves (2005); Lempert and Groves (2010)). Such facilitated reiteration, however, is difficult to achieve with the large and complex ensembles of SOWs that modern state-of-the-art exploratory modeling analyses rely on. For example, in recent past work we generated 10,000 SOWs, within each of which we computed thousands of performance metrics for different stakeholders and different criteria (Hadjimichael, Quinn, Wilson, et al., 2020). Similarly, Gold et al. (2022); Shi et al. (2023); Trindade et al. (2020) and others all use ensemble sizes of thousands to millions of scenarios. As already mentioned, the size of these experiments is an attempt to better capture the space of possible futures and consider relevant uncertainties, recognizing the combinatorial scale of significant factors in highly complex coupled human-natural systems and to better guide a more holistic understanding of highly consequential decision-relevant outcomes.

Large ensemble exploratory modeling therefore creates a tension: on one hand, we understand that there is a large number of interacting processes, candidate futures and alternative framings we should explore, and we thus need to create large ensembles of these hypothesized changes to investigate with our models. On the other hand, each additional dimension considered makes the results of the analysis more intricate and more difficult to convey actionable insights[1]. We argue that making large ensemble experiments more actionable is indeed possible, but requires innovations in how the resulting outcomes and their driving dynamics are organized, investigated, and communicated. This can be complemented with new data visualizations that allow users to navigate hierarchical levels of classification of ensemble outputs, and to zoom in on specific narrative scenarios of interest and investigate their dynamics.

The present study addresses the challenges and needs for large ensemble exploratory modeling discussed above by contributing a new scenario discovery framework: the FRamework for Narrative Scenarios and Impact Classification (FRNSIC)—pronounced "forensic". FRNSIC aims to provide actionable narrative clarity without sacrificing the quantitative rigor of large ensemble experiments. It aids the identification of consequential scenarios through the application of nested criteria that capture hierarchical relationships between sectors, actors, and/or scales, each reflective of different relevant impacts for the stakeholders concerned. We can explore multiple influential system states and hierarchically support the discovery of the diverse conditions that control stakeholder-relevant impacts. The emerging narrative scenarios are clustered not only on their resulting impacts but also on the underlying dynamic scenarios that drive them. As a result, we aid decision makers in discovering smaller sets of narrative scenarios, or dynamic storylines, that represent both complex mappings between a large space of input uncertainty and the large space of resulting outcomes. At the same time, these storylines also maintain a locally-embedded meaning, as well as the potentially critical temporal dynamics that lead to consequential outcomes.

The remaining sections are organized as follows. Section 2 presents the FRNSIC scenario discovery framework and provides an overview of the main component stages of its application. Section 3 details our application of the framework within the Upper Colorado River Basin, with a particular focus on the issue of better understanding plausible drought extremes and their system impacts. Finally, Section 4 presents the outcomes of the application of FRNSIC, and Section 5 provides conclusions as well as opportunities for future extensions.

2 Methodological Framework

Exploratory modeling and its connection to robustness frameworks has been extensively reviewed in several past studies (Herman et al., 2015; Kwakkel & Haasnoot, 2019; Moallemi, Zare, et al., 2020). We refer readers to these publications for a comprehensive introduction to the background literature in this area. Following the terminology established by these authors, this paper introduces a new scenario discovery framework in support of robustness analysis, FRNSIC, begins by following the same broad steps that are common across all exploratory modeling and robustness approaches (framing, system evaluation across many states, quantification of performance, and scenario discovery), and then adds new steps for multi-trait classification and storyline discovery (see Fig. 1).

The *Problem Framing* Stage (I) is critical across all exploratory modeling and robustness frameworks to ensure the decision relevance of their results. During this phase, analysts and stake-

[1] In Aesop's fable about The Fox and the Cat, the fox boasts of hundreds of ways of escaping its enemies, while the cat only has one. When they hear a pack of the hounds approaching, the cat scampers up a tree and hides, while the fox in its confusion gets caught up by the hounds. The moral of the fable is that it is "Better [to have] one safe way than a hundred on which you cannot reckon".

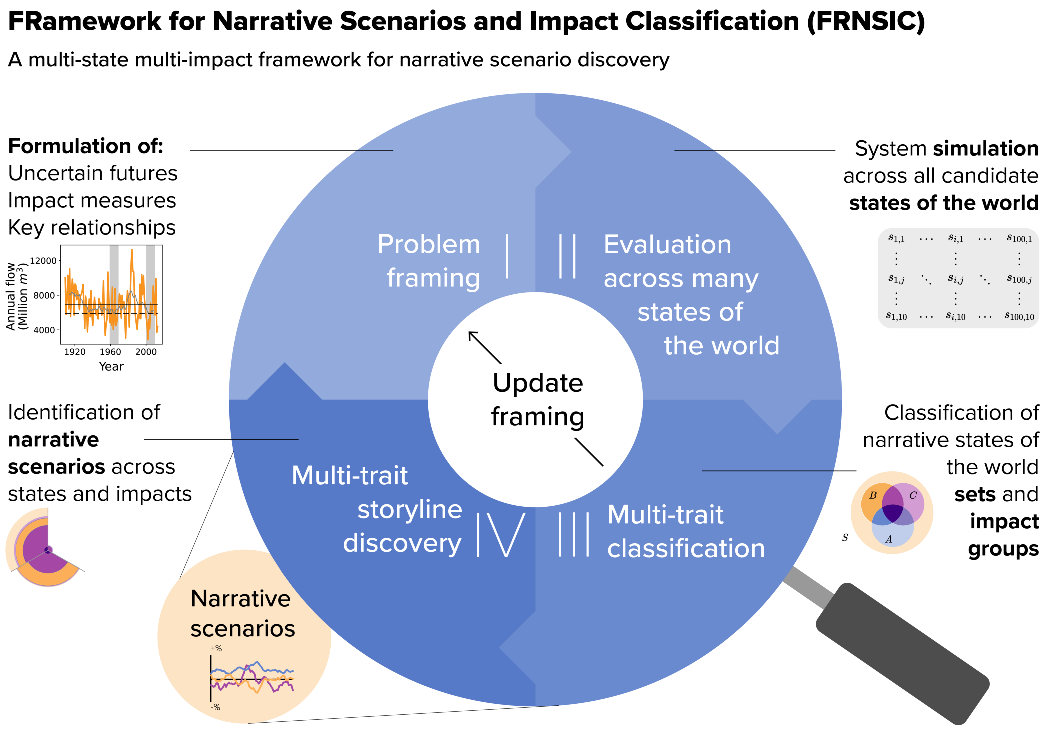

Figure 1. **The four stages of the multi-state, multi-impact framework for narrative scenario discovery, FRNSIC.**

holders define the key factors in the analysis: system goals (sometimes articulated as objectives) and metrics of performance toward these goals; alternative actions or system configurations that can be taken to affect said metrics; the uncertainties that may affect the connection between actions and metrics; and the relationships (which often take the form of simulation models) between actions, uncertainties, and metrics (Lempert, 2019). Procedures for eliciting these elements have been articulated based on the 'XLRM' matrix (Lempert et al., 2003): exogenous uncertainties ('X'), policy levers ('L'), relationships ('R'), and metrics ('M'). Here, we adopt the same intention behind the problem framing stage. Presenting framing as a distinct stage in these frameworks is intentional; framing choices made during this stage should be transparently articulated, especially as they shape subsequent stages of analysis. The framing could also be updated as performance across states is quantified and consequential conditions are uncovered. In the Upper Colorado River Basin case study, presented in the following section, this stage is used to investigate the water scarcity context of the system and frame how SOWs should be appropriately generated, the dynamic states of consequence (e.g., decadal droughts), and impact metrics.

Exploratory modeling is a central focus of Stage II of FRNSIC (*Evaluation across many states of the world*), evaluating the system, via a simulation model, across alternative actions or policies or system configurations, and across alternative SOWs. Moallemi, Zare, et al. (2020) term these steps "generation of decisions" and "generation of scenarios", respectively. The same authors, as well as others, have also broadly drawn a distinction here between two alternative strategies: exploration and search. Methods that rely on exploration systematically sample points across both the decision space and the SOWs and evaluate their consequences. As such, they rely on the careful designs of experiments which are used to set up simulation frameworks with the minimum computational cost to answer specific questions (Reed, Hadjimichael, Malek, et al., 2022). Exploration techniques produce insights about the global properties of the decision and the uncertainty space (plausible SOWs), such as how much increase in water demand would result in increased supply shortages (e.g., Hadjimichael, Quinn, Wilson, et al. (2020)).

Methodologies that rely on search, in contrast, draw on optimization-based tools to actively identify points with particular properties, such as "how much should we invest in infrastructure to maximize profits?" (searching for high-performing actions) or "how much more warming would cause insufferable heatwaves in our city?" (searching for a subset of consequential SOWs). These approaches typically rely on multi- or many-objective optimization algorithms (Kasprzyk et al., 2013; Kwakkel, 2019). FRNSIC remains agnostic to which of the two strategies is employed at this stage, as both allow us to analyze a system over many of its potential states, and use those states to classify and discover narrative scenarios of interest. If optimization methods were to be used in this case, one would have to ensure that the temporal dynamics of each simulation are carefully maintained, for subsequent analysis in the following stages. In the Upper Colorado River Basin case study, we are using exploration methods.

The core novel contributions of FRNSIC lie in Stages III and IV, where performance is quantified (III *Multi-trait classification*) and consequential scenarios are discovered (IV *Multi-trait storyline discovery*). To clarify these contributions, let us first briefly overview how performance quantification and scenario discovery are traditionally performed. In virtually all applications (see reviews from Marchau et al. (2019); Moallemi, Kwakkel, et al. (2020); Moallemi, Zare, et al. (2020)), the analysts establish one or a set of criteria against which they compare or rank order the performance of different policies or actors across SOWs (i.e., one or more robustness performance metrics). To address some of the challenges brought about by multi-actor systems discussed in Section 1, a variety of robustness metrics or different performance thresholds might also be used (e.g., Hadjimichael, Quinn, Wilson, et al. (2020)). A SOW is then classified as being consequential subject to meeting or failing to meet the specific requirements tied to the robustness metric(s) specified. A tacit effect of using the most commonly employed robustness metrics (e.g., satisficing or regret metrics; see discussions in Herman et al. (2015); McPhail et al. (2018)) is that the temporal dynamics of the underlying sampled SOWs are ignored, and in their place, the analysis is focused on the classification of SOWs as being consequential or not based on a summarizing statistic of those dynamics. A benefit of this approach is that a single quantitative value is much more easily communicated than a vector of them across the duration of the realization. A shortfall of it is that policies or actors achieving similar performance on a particular robustness metric may do so through a diversity of temporal dynamics that lead to tradeoffs on other metrics. Consequently, the temporal dynamics are critical drivers that shape whether or not specified performance metrics are met, and are therefore critical to understanding robustness tradeoffs. The importance of temporal dynamics and their properties is strongly emphasized in the socio-ecological systems and system dynamics bodies of literature (e.g., Gotts et al. (2019); Schlüter et al. (2012)), the data science literature (e.g., Aghabozorgi et al. (2015)), and more recently emphasized in both the exploratory modeling (Steinmann et al., 2020) and the climate risk (de Ruiter & Van Loon, 2022) literature.

In Stage III of FRNSIC (Fig. 1), we use simple set theory to explore the dynamic properties of the sampled SOWs, not restricting focus solely on robustness performance measures (which we also classify, as discussed below). This creates collections of SOWs that exhibit certain dynamic properties (e.g., significant variability, particular equilibria or oscillation patterns) irrespective of the performance outcomes they generate (e.g., impacts to system users). In other words, we create collections of SOWs that specifically focus on the dynamic processes of the system and their defining characteristics, as separate defining properties from the performance in each SOW. The reason this distinction is important is that the same dynamic properties do not always result in the same system impacts, and vice versa. For example, two droughts of the same severity might occur, but have different water scarcity impacts. On the other hand, two SOWs might result in similar outcomes (e.g., 20% of water demands cannot be met), but the underlying dynamics that produce them are different.

These dynamic properties can be identified in several ways. They might be specified *a priori*; for example, if the computational design of experiments is set up to specifically generate them.

Such is the case for some of our prior work evaluating water scarcity, where we used parametric approaches to synthetically generate hydrologic conditions and those conditions were sampled so as to specifically exhibit certain properties (e.g., larger variability; Hadjimichael, Quinn, Wilson, et al. (2020); Quinn et al. (2020)). Dynamic properties can also be discovered *a posteriori*. For example, Steinmann et al. (2020) applied time series clustering to identify collections of SOWs that exhibit similar temporal behaviors. Lastly, dynamic properties can also be analytically or numerically calculated. For example, Hadjimichael, Reed, and Quinn (2020) analytically derived behavioral properties of each SOW that pertained to the system's stability and number of equilibria, and used said properties to create semantically meaningful collections of SOWs that described certain behavior modes. Clarifying the diversity of temporal dynamics that underlie a large ensemble of exploratory modeling simulations using a small number of semantically meaningful sets can facilitate their narrative application later on, when the scenario discovery process identifies consequential SOWs. Utilizing these behavioral properties to discover narrative scenarios in conjunction with using performance criteria to discover impactful scenarios can help analysts illuminate the root causes of vulnerability in a system (Steinmann et al., 2020).

Beyond using set theory to order and better understand the underlying dynamics in sampled SOWs, Stage III of FRNSIC also hierarchically classifies diverse robustness performance measures that can be defined across different actors, scales, and sectors. Hierarchy, as used here, refers to the addition of new criteria (e.g., "$reliability \geq 90\%$" AND "$costs \leq \$100$"), not the preferential weighting of one criterion over another. Even though it is not typically discussed in terms of set theory, classifying sampled SOWs in terms of whether they meet a certain criterion in effect partitions them into specific subsets (or collections) of the broader set of all SOWs, such that for every criterion there exists a conditional set of SOWs for which the condition holds and a complement set for which it does not. For multiple performance criteria, we can therefore create multiple such subsets to denote whether an impact criterion is met, as well as look at the intersections of the conditional sets for the combinations of SOWs where multiple criteria are met simultaneously. This type of algebraic structure is formally referred to as a Boolean algebra or a Boolean lattice and describes relationships between the partitioned subsets of an overall set that result from applying binary classification operations (Drapeau et al., 2016; Priss, 2021). In essence, we can use these binary operations to identify increasingly nested subsets of consequential SOWs that meet or fail to meet additional performance criteria. For complex human-natural systems confronting change that impact a large suite of scales, sectors and stakeholders, FRNSIC's hierarchical classification greatly broadens the diversity of interests and performance concerns that shape our inferences on robustness.

Finally, in Stage IV of FRNSIC (*Multi-trait storyline discovery*), these two sets-of-sets—one created to describe fundamental dynamics and one created to classify the decision-relevant outcomes from hierarchical performance criteria—are combined to guide the discovery of consequential storyline narrative scenarios that can be used to structure further dialogues for the diverse ways a system may confront change. As emphasized in Section 1, achieving narrative meaning in the context of high dimensionality and complexity requires advances in how the information is organized (in our case with hierarchical sets) and presented. For the latter, we contribute a modified version of the stacked hive plot (Krzywinski et al., 2012), which allows us to visualize the resulting sets-of-sets in a single panel figure. Hive plots adapt parallel coordinate plots (Inselberg, 2009; Wegman, 1990) to a radial arrangement, compacting the layout and making the connections easier to follow. Hive plots typically rely on a three-axis model, with the total circle area being uniformly divided between all segments (the areas between two axes). As demonstrated in this study, the three axes we utilize reflect three dynamic properties of the SOWs generated. More than three dimensions can be used, but by having only three axes, hive plots accommodate connections (lines) between each axis pair, without having to cross the axes themselves. With more than three axes this can only be achieved if connections are only drawn between neighboring axes, or if axes are duplicated at multiple positions. This negatively impacts the interpretability of the figure, which defies the aim of creating meaningful and salient narratives, central to our framework. The originators of the figure indeed discourage its use with more than three axes (Krzywinski et al., 2012), and most common applications in network science (e.g., Engle and Whalen (2012)) and gene sequencing (e.g., Yang et al. (2017)) also only use three axes. Furthermore, the compactness of this figure allows us to generate multiple panels reflecting alternative dynamic properties or robustness performance measures, in a "small multiples" visualization (Tufte, 1990). Combining many small visualizations simultaneously allows the reader to compare the separate panels and look for patterns or outliers in the matrix of visuals, and facilitates presentation and storytelling of large amounts of data in a single figure (van den Elzen & van Wijk, 2013).

In the following sections, we present an example application of the key stages of FRNSIC on a multi-actor, institutionally complex human-natural system: the Upper Colorado River Basin within the state of Colorado (henceforth abbreviated to UCRB). Section 3.1 introduces the study area and model utilized. Section 3.2 presents an overview of the problem (FRNSIC Stage I - Problem Framing) and articulates the main challenges surrounding the characterization of drought extremes and investigation of their impacts. Section 3.3 details the generation of hydroclimatic SOWs (FRNSIC Stage II - Evaluation Across Many States of the World) through the use of exploratory modeling, allowing us to account for said challenges. Section 3.4 (FRNSIC Stage III - Multi-trait Classification of States of the World) details how the drought dynamics of the hydroclimatic SOWs are classified into sets of dynamic properties, as illustrated in Fig. 5, as well as how the impacts generated by the SOWs are classified into impact sets, as illustrated in Fig. 7. Finally, Section 3.5 (FRNSIC Stage IV - Multi-trait storyline discovery) describes how the two sets-of-sets come together through the use of hive plots to enable the exploration of narrative drought storylines that summarize both consequential impacts and key drought dynamics.

3 The Upper Colorado River Basin case study implementation

3.1 Study Area and Model

Most of the aforementioned innovations and developments in the domain of exploratory modeling and scenario discovery have been in the area of water resources. Water resources systems are archetypal of the types of challenges we face around understanding and planning in coupled human-natural systems: environmental, social, infrastructural, and institutional complexity; contested views and objectives over how resources should be allocated; increasing stress and deep uncertainty about future stressors. Western river basins in the United States in particular, and the Colorado River more specifically, are under significant hydrologic stress, following decades of aridification (Smith et al., 2022; State of Colorado, 2015; McCoy et al., 2022; Whitney et al., 2023). The Colorado River basin is institutionally complex, with a nested set of compacts, laws, and regulations that dictate water allocation for over 40 million people and 22,000 km^2 of agricultural land (Bureau of Reclamation, 2012). The River has been experiencing prolonged water scarcity and aridification for the past two decades, accumulating to a "crisis" in recent years (Gerlak & Heikkila, 2023). A megadrought that started in 1999 (Overpeck & Udall, 2020), and continues as of the time of writing, has caused major reservoirs on the river to decline to dangerously low levels, prompting the U.S. Department of Interior to call for unprecedented cuts in water usage among the states that depend on it (Flavelle & Rojanasakul, 2023).

Understanding plausible future drought hazards and planning for their impacts in these human-natural systems presents several challenges. First, internal hydroclimatic variability and non-stationarity challenge how we identify extreme events, such as decadal-scale or longer drought hazards (AghaKouchak et al., 2022; Hoylman et al., 2022; Lehner & Deser, 2023; Stevenson et al., 2022). Internal variability, arising from interactions across non-linear processes intrinsic to the hydroclimate, means that any given process has inherent irreducible uncertainty in its manifestation and that our historical observations are only one limited sample of the diverse dynamics that could occur. In the context of hydroclimatic dynamics, internal variability is a fundamentally stochastic process that has been shown to produce magnitudes of variation in flood and drought extremes that exceed historical experiences (Fischer et al., 2021) or that are comparable to anthropogenic climate change at the decadal scale (Deser et al., 2016). Even in regions of the world with long observational records, the full extent of internal variability cannot be estimated from the single realization of the stochastic hydroclimatic process represented by the observed record that exists (Woodhouse & Overpeck, 1998; Woodhouse et al., 2006). Extending the record with reconstructed paleoclimate information can improve on this representation, but has its own methodological limitations, such as underestimating the variance in the data (Quinn et al., 2020), and reducing interpretability (Ault et al., 2014). Lastly, the stochastic nature of internal variability poses important communication challenges, as it necessitates the use of probabilistic descriptions of the occurrence of critical events, instead of simple deterministic predictions of them (Lehner & Deser, 2023).

Non-stationarity in time and space is another a well-recognized challenge. Non-stationarity reflects conditions where the statistical properties of a variable (e.g., its distribution and correlation with other variables) may change over time (Slater et al., 2021). It is especially consequential in how it transforms the occurrence of extreme events like floods, droughts, and heatwaves (AghaKouchak et al., 2022; Berghuijs et al., 2019; R. Lorenz et al., 2019; Sun et al., 2021). Yet, until the recent decade, non-stationarity has not been accounted for in conventional planning for water resources or extreme events. Instead, planners have relied on observed historical time series of streamflow or other hydroclimatic variables for future planning (Yang et al., 2021). In fact, even current drought monitoring products such as the United States Drought Monitor rely on historical distributions of these events to establish their classification (Hoylman et al., 2022), as do the flood maps generated by the Federal Emergency Management Agency (Hobbins et al., 2021). This is largely due to large epistemic uncertainties around the form of future non-stationarity. Even under stationary conditions, when complex systems are concerned, it is often impossible to be in full knowledge of the true model of the system under consideration (Beven, 1993). In the case of non-stationary systems and the development of models for them, the problem is even more challenging because of the larger number of parameters involved (i.e., both the base statistics and also how they are changing) and large number of alternative ways non-stationarity can be included in the analysis (Salas et al., 2018).

Lastly, the complexity of human systems further compounds the challenges in understanding and planning for the potential impacts of droughts. In systems like the Colorado River, institutions, engineered infrastructure, and large numbers of actors come together to shape who gets water, how much, and when, as well as who has to get shorted when conditions are dry. Our understanding of drought-induced water scarcity has evolved to recognize the importance of the feedbacks between anthropogenic and natural system processes, which shape the production and distribution of drought effects and their implications for humans and the environment (AghaKouchak et al., 2023; Lukat et al., 2023; Savelli et al., 2022). Human-natural systems around the world, and especially systems that are heavily managed, have developed strategies to reduce their exposure and vulnerability to drought hazards (Kreibich et al., 2022; Smith et al., 2022). For example, the states that depend on Colorado River water develop and regularly update drought preparedness plans that help them project their water availability and needs, and adjust their operations accordingly (e.g., Arizona Department of Water Resources, 2022; California Natural Resources Agency, 2022; Colorado Water Conservation Board & Department of Natural Resources, 2018). These efforts at higher levels of governance, as well as less-coordinated state or local planning efforts, all must consider the institutional water rights context of the Prior Appropriation Doctrine (Kenney, 2005). Water rights create a complex hierarchy for managing scarcity and strongly shape how a regional drought may differentially affect each water right holder in the river (Hadjimichael, Quinn, Wilson, et al., 2020).

The particular implementation of Prior Appropriation in each state, as well as other local characteristics and needs of each watershed, have prompted states like Colorado to develop water planning and management processes at different scales: at the state-wide scale (i.e., the state of Colorado's Water Plan; State of Colorado (2023)), and the local river basin scale (i.e., the Basin Implementation Plans developed by a local Basin Roundtable for each of the nine basins within the state, e.g., CWCB and CDWR (2022)). To facilitate communication and comparisons, the Colorado Water Plan and the local Basin Implementation Plans all utilize a set of five future scenarios of water scarcity in the state (State of Colorado, 2023), each being a narrative summary of how different drivers of scarcity might evolve in the future (e.g., increased agricultural needs, reduced supply). These five scenarios carry the same challenges discussed in Section 1, but they are not necessarily consequential or relevant at the local level. In other words, each local basin might not necessarily be equally sensitive to the key drivers each scenario assumes, nor have impacts at the same magnitudes. So even though the local impacts of these five scenarios are evaluated in the Basin Implementation Plans, the analysis might inadvertently miss other locally consequential scenarios, that are still plausible but not part of the set of five.

Within this context, we demonstrate how the FRNSIC scenario discovery framework could be utilized by the local Basin Roundtable responsible for water resources planning for the UCRB. The Colorado Basin Roundtable[2] was established in 2005 by Colorado state legislature and is charged with water planning for the UCRB and with implementing the state-wide Water Plan locally. Its members include not only state representatives, like from the Colorado Division of Water Resources and the Colorado Water Conservation Board, but also representatives from the agricultural sector, the industrial sector, domestic water suppliers, environmental and recreation entities, as well as other interested citizens. Besides planning, the Colorado Basin Roundtable also plays a significant role in allocating state funds to enact its water priorities within the UCRB. The diversity of representative members of the Colorado Basin Roundtable is crucial to its ability to address the diverse goals and challenges the UCRB faces.

The UCRB contains the headwaters of the Colorado River with its outflow moving into Utah to deliver water to Lake Powell. As with all western basins in the state, it is bound by the Colorado River Compact, which allocates 9.3 km^3 (7.5 million acre-feet) per year to the Upper Basin states (Colorado, New Mexico, Utah, and Wyoming)—the state of Colorado is allotted 51.75% of that amount. Another 9.3 km^3 is divided among the Lower Basin states (California, Arizona, and Nevada), and Upper Basin states have to deliver water to Lake Powell to meet that requirement. Increasingly frequent and more persistent severe drought conditions inhibit the ability of Upper Basin states and subbasins like the UCRB to make these deliveries. Quantifying the potential effects of future water scarcity and drought on UCRB deliveries to Lake Powell is therefore a key concern for the Colorado Basin Roundtable, as outlined in their Basin Implementation Plan (CWCB & CDWR, 2022). Within the UCRB, several thousand water rights support diversions for agriculture, municipal water supply, industrial production, power generation, as well as recreational uses (Fig. 2). While most of the consumptive use of water within the basin supports agricultural production, large exports of water leave the basin to support urban centers on the east slope, where most of Colorado's population resides. Water to all these users is allocated through the Prior Appropriation Doctrine, which prioritizes users in terms of seniority and limits the received amount of water for each user to their decreed "beneficial use" (Kenney, 2005). Along with the water availability itself, this institutional hierarchical network plays the most fundamental role in shaping the dynamics of water scarcity vulnerabilities across the water rights holders. Given the central importance of the agricultural sector in this basin, quantifying impacts to local agricultural water users is another critical concern highlighted in the Basin Implementation Plan (CWCB & CDWR, 2022).

All these key aspects are captured in Colorado's Decision Support System (CDSS), a collection of databases, data management tools, and models, created to support water resources planning in Colorado's major water basins, including the UCRB (Malers et al., 2001). The principal modeling tool of the CDSS is the State of Colorado's Stream Simulation Model (StateMod), a generic network-based water system model for water accounting and allocation. StateMod was developed to support comprehensive assessments of water demand and supply, as well as reservoir operations, in all the major subbasins within the state of Colorado (Parsons & Bennett, 2006; CWCB, 2012). The model replicates each basin's unique application of the Prior Appropriation doctrine and accounts for all of the consumptive uses of water within each basin. To achieve this, StateMod utilizes detailed historic demand and operation records, which include water right information for all consumptive water diversions, water structures (i.e., wells, ditches, reservoirs, and tunnels), as well as streamflow and other hydroclimatic information. The model also includes estimates of agricultural water consumption based on soil moisture, crop type, irrigated acreage,

[2] https://www.coloradobasinroundtable.org/

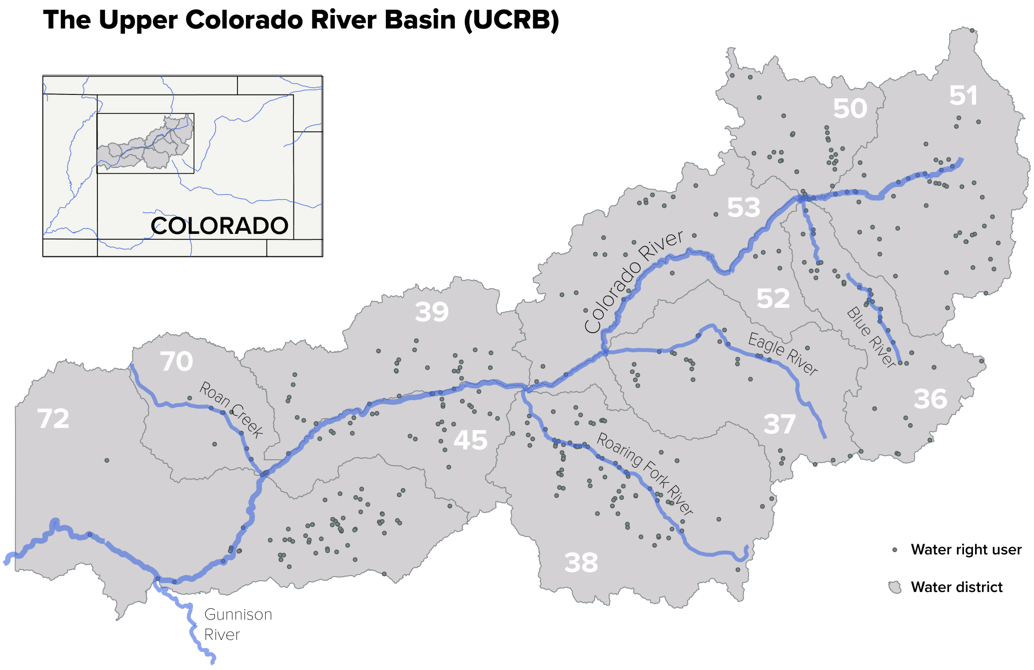

Figure 2. The Upper Colorado River Basin within the state of Colorado (UCRB). The points indicate all modeled diversion points in StateMod (primarily irrigation). The numbered areas indicate water districts.

and conveyance and application efficiencies for each individual irrigation unit in the region. Using these highly-resolved inputs, StateMod accounts for the water consumption of all users in each basin, through their water right allocation. It therefore allows us to simulate and assess the impacts of potential future changes in hydrology, water demands, or operations on all the represented water users in each basin. For the purposes of this study, we focus on the specific StateMod implementation for the UCRB.

The remainder of this section outlines a demonstrative use of FRNSIC that could support the types of coordinated planning studies overseen by groups like the Colorado Basin Roundtable to explore and discover locally consequential and plausible scenarios for their basin. The UCRB system is an ideal testbed to make generalizable advances in exploratory modeling literature, particularly with regard to addressing the dimensionality introduced by multi-actor systems, the importance of capturing behavioral dynamics, and the challenge of providing clarity when selecting consequential drought storyline narratives for further consideration in planning efforts, as discussed in Section 1. The planning application demonstrated here is hypothetical, but stays close to the key water planning concerns articulated in the Basin Implementation Plan, as well as other literature on drought-induced water scarcity in the region, as elaborated below.

3.2 Stage I - Problem Framing

Throughout this study, we classify hydrologic drought conditions as occurring when there is a half a standard deviation departure from the historical average streamflow at the Colorado-Utah state line over the period 1909-2013 (i.e., $\mu - 0.5\sigma$), following the examples of Ault et al. (2014, 2016); Diffenbaugh et al. (2015); Naumann et al. (2018). We apply this classification on naturalized streamflow and identify decadal-scale droughts using an 11-year rolling mean (more details on how the classification is performed are provided in Section 3.4.1). Multidecadal droughts can similarly be identified using longer windows, such as 25 years (Meko et al., 2007) or 35 years (Ault et al., 2014). Applying this classification to the historical streamflow observations for the

UCRB, we see two decadal-scale droughts: one in the 1960s and one starting in the early 2000s (Fig. 3 (a)). This estimate is consistent with other literature sources that classify decadal droughts in the reconstructed paleo record in this region (i.e., one or two instances of decadal drought per century; see Ault et al. (2014); Woodhouse and Overpeck (1998)). The identification of plausible decadal-scale drought hazards is confounded by the presence of: (a) irreducible, internal variability, (b) non-stationarity, and (c) deeply uncertain past and future streamflow dynamics beyond the currently available gauged record (i.e., paleo conditions or future climate change).

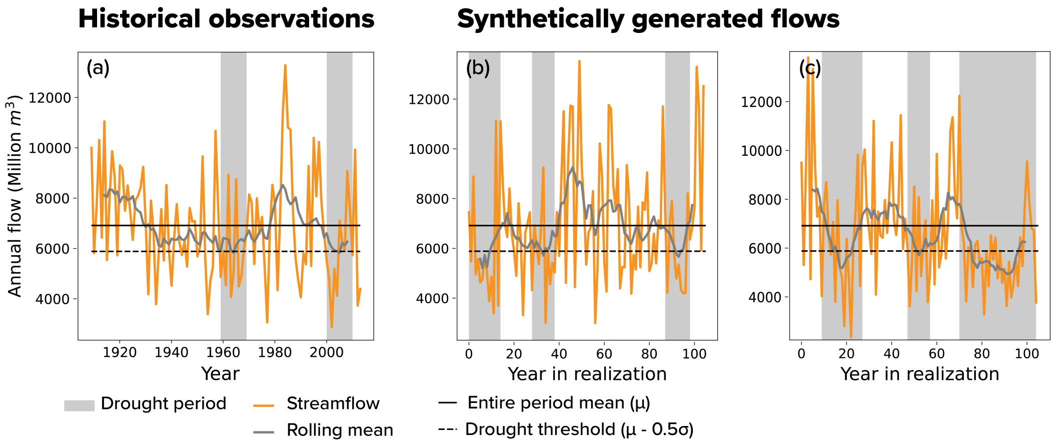

Figure 3. Hydrologic drought identification for the UCRB (a) Decadal-scale droughts identified using historic observations; (b-c) Decadal-scale droughts identified using synthetically generated streamflow. We note that the mean and standard deviation of the distribution remain the same, so does the average annual volumetric drought threshold, at $5,884 Mm^3$, computed over the full 105-year record length.

Internal variability complicates the identification of droughts, even in a stationary context (Cook et al., 2022). For example, even if we establish that the moments of the historical streamflow distribution stay the same in the future and use those distributions to inform planning, we might underestimate the true frequency of drought events (i.e., the events that cross the drought threshold in this case). Fig. 3 demonstrates this effect. Here, we compare the drought classification applied to the historic observations of streamflows (Fig. 3 (a)) and the same classification applied to synthetically generated streamflows that have the same base statistical properties as the last century's historical observations (Fig. 3 (b-c)). The synthetic streamflows are created using a synthetic streamflow generator so as to exhibit the same distributional moments for the occurrence of wet years and dry years, as well the probability of transitioning between the two states, through the use of a Hidden Markov Model (see more details in Section 3.3). We see that even though only two decadal droughts are identified in the historical record (using a drought threshold of $5,884 Mm^3$), simulating alternative plausible synthetic realizations from the same distributions can give rise to more decades of drought. This undermines the validity of using the historical streamflow observations to deterministically to infer expectations for the frequency of extreme drought conditions (e.g., that only one or two decadal droughts are to be expected in a century), when in fact the same process can give rise to conditions that are much worse.

Non-stationarity makes it challenging to establish appropriate reference conditions (e.g., the drought threshold used above) when seeking to identify decadal drought hazards for a hydroclimatic system with evolving wet and dry regimes (Mondal & Mujumdar, 2015; Slater et al., 2021). The solution often recommended is to use rolling windows of time and establish moving baseline thresholds (Hoylman et al., 2022). Fig. 4 demonstrates this idea and highlights the potential variability of drought thresholds when looking across 60-year rolling windows of streamflows. For reference, the average annual volumetric drought threshold calculated using the entire period of data (105 years) is $5,884 Mm^3$ (indicated by the dashed line in Fig. 4 (b)). Starting with the early 1900s, conditions were very wet (top density plot in Fig. 4 (a)) and so the drought threshold established using that early 20th century 60-year window is at a much larger annual average volume (top right point in Fig. 4 (b)). As a result, 30 years in the record since that initial 60-year window would fall below the drought threshold established in this period (Fig. S1). We note that these 30 years are identified in decadal periods, they therefore reflect three decadal droughts, not 30 drought years dispersed throughout the 105-year period. The early 1900s were also the period during which the Colorado River Compact was signed. Moving across time (downward in the figure), we see that the changing streamflow statistics substantially shift the drought thresholds one would establish, down to ≈ 5,540 M m^3 in the most recent window. Using these drier-period thresholds that are substantially lower than that of the entire period (i.e., all points to the left of the dashed line in Fig. 4 (b)) would result in no years classified as droughts (Fig. S1). [3]

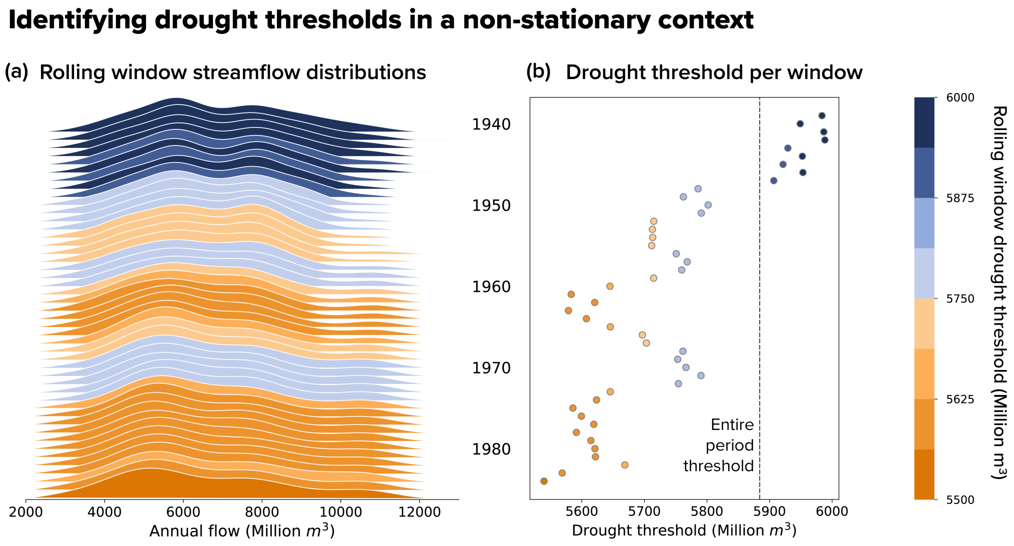

Figure 4. Drought thresholds established using rolling windows (a) Distribution of annual streamflow per 60-year rolling window; (b) Drought threshold established using distribution moments of each 60-year rolling window. The vertical dashed line represents the threshold established using the entire record (same as the threshold in Fig. 3, i.e., $5,884Mm^3$.)

The final type of uncertainty that impacts our understanding of plausible extreme droughts is the inherent deep uncertainty associated with evolving wet and dry dynamic regimes that are beyond the scope of gauged historical streamflow observations. These deeply uncertain regimes can encompass both ungauged historical conditions (e.g., paleo records) and future projections of how the complex human-natural systems may change. Deep uncertainty refers to a lack of consensus over how future events may unfold as well as their associated likelihoods or consequences (Marchau et al., 2019; Walker et al., 2003). Literature focusing on deep uncertainty emphasizes the use of exploratory modeling—the use of intentionally broad hypotheses about future system conditions and the assessment of system outcomes. This allows us to investigate a broader ensemble of states so as to be able to understand system response and inform planning in spite of the presence of these three uncertainty types. Here, we place an explicit focus on exploratory modeling of hydroclimatic factors and their implications for key basin outcomes. As discussed above, increasingly frequent and more persistent severe drought conditions inhibit the ability of basins like the UCRB to meet their obligations to Lower Colorado Basin states through deliveries to Lake

[3] In fact, some have argued the current megadrought should not actually be considered a drought, but a new normal brought about by aridification (Robbins, 2019).

Powell. At the same time, given the central importance of the agricultural sector in the UCRB, quantifying impacts to local agricultural water users is another critical concern. Both these issues are highlighted in the Basin Implementation Plan as key concerns for the Colorado Basin Roundtable (CWCB & CDWR, 2022). Through combinations of hydroclimatic states and these basin impacts, we identify consequential drought storylines that represent complex mappings between the large space of input uncertainty (ensemble of hydroclimatic conditions) and the large space of resulting outcomes for the basin's stakeholders.

3.3 Stage II - Evaluation Across Many States of the World

The system is evaluated under an ensemble of hydrologic SOWs, synthetically generated to reflect different assumptions about future hydroclimatic changes in the region, as well as to explore their internal variability (Fig. 1). Our ensemble of SOWs relies on the Gaussian Hidden Markov Model (HMM) synthetic streamflow generator developed by Quinn et al. (2020). The use of HMMs for the synthetic generation of streamflows has advantages in capturing complex wet-dry hydroclimatic regime dynamics as well as their persistence in Western US drought extremes (Bracken et al., 2014, 2016). We refer the reader to Quinn et al. (2020) for the full details of how the synthetic streamflow ensemble was generated; we summarize key information here. The HMM used comprises two states: one representing wet and the other dry conditions (i.e., higher and lower streamflows). The two states are referred to as 'hidden' because they are not directly observed; rather they are inferred from a time series of continuous flow values, assumed to come from one of two log-normal distributions (one for the distribution of wet years and one for dry years). Fitting an HMM with these characteristics requires the estimation of six parameters: the mean and standard deviation of the dry-state and wet-state Gaussian distributions (μ_d and σ_d, and μ_w and σ_w, respectively), as well as the probabilities of transitioning from a dry state in year t to a dry state in year t+1 (p_{dd}), and from a wet state in year t to a wet state in year $t+1$ (p_{ww}). The generator then uses these distributions and the estimated transition probabilities to create synthetic time series of streamflows. Two examples of synthetically generated streamflows using the HMM are shown in Fig. 3 (b-c).

To generate the ensemble, Quinn et al. (2020) fit the HMM to historical observations and then modified its parameters according to several experimental designs, each reflecting different assumptions about how future hydrologic conditions in the basin could change. These different assumptions can all be considered plausible 'rival framings' of future wet-dry regimes. These rival framings were that: (i) streamflow parameters in the future could independently deviate from their stationary historical behavior to a moderate degree, (ii) they could move toward values seen in the past, as inferred from reconstructed paleo data, (iii) they could reflect downscaled climate change projections for the UCRB region, or (iv) they could move toward values generated under any of these assumptions (i.e., the 'all-encompassing' ensemble of candidate futures, which parametrically envelopes all other rival framings of the UCRB's hydroclimate).

In this study, we utilize the all-encompassing experiment. Within the all-encompassing experiment, possible future scenarios consist of multipliers on the dry-state and wet-state means and standard deviations, and delta shifts on the dry-dry and wet-wet transition probabilities. The sets of all scaling factors and the respective ranges for each HMM parameter are given in Eq. 1, which were chosen by Quinn et al. (2020) to span the ranges experienced across all other rival framings. Using these parameter ranges, 100 parameter combinations were generated using Latin hypercube sampling (McKay et al., 1979). The 100-member ensemble size was verified by Quinn et al. (2020) to yield results that are consistent with the results obtained using a larger ensemble of 1,000 parameter combinations.

$$
\begin{aligned}
\mu_d &= \{0.90 \leq \mu_{d_i} \leq 1.03 | i \in I\} \\
\mu_w &= \{0.97 \leq \mu_{w_i} \leq 1.03 | i \in I\} \\
\sigma_d &= \{0.75 \leq \sigma_{d_i} \leq 2.63 | i \in I\} \\
\sigma_w &= \{0.39 \leq \sigma_{w_i} \leq 1.25 | i \in I\} \\
p_{dd} &= \{-0.65 \leq p_{dd_i} \leq 0.30 | i \in I\} \text{ and } p_{dw} = \{1 - p_{dd_i} | i \in I\} \\
p_{ww} &= \{-0.33 \leq p_{ww_i} \leq 0.33 | i \in I\} \text{ and } p_{wd} = \{1 - p_{ww_i} | i \in I\}
\end{aligned}
\tag{1}
$$

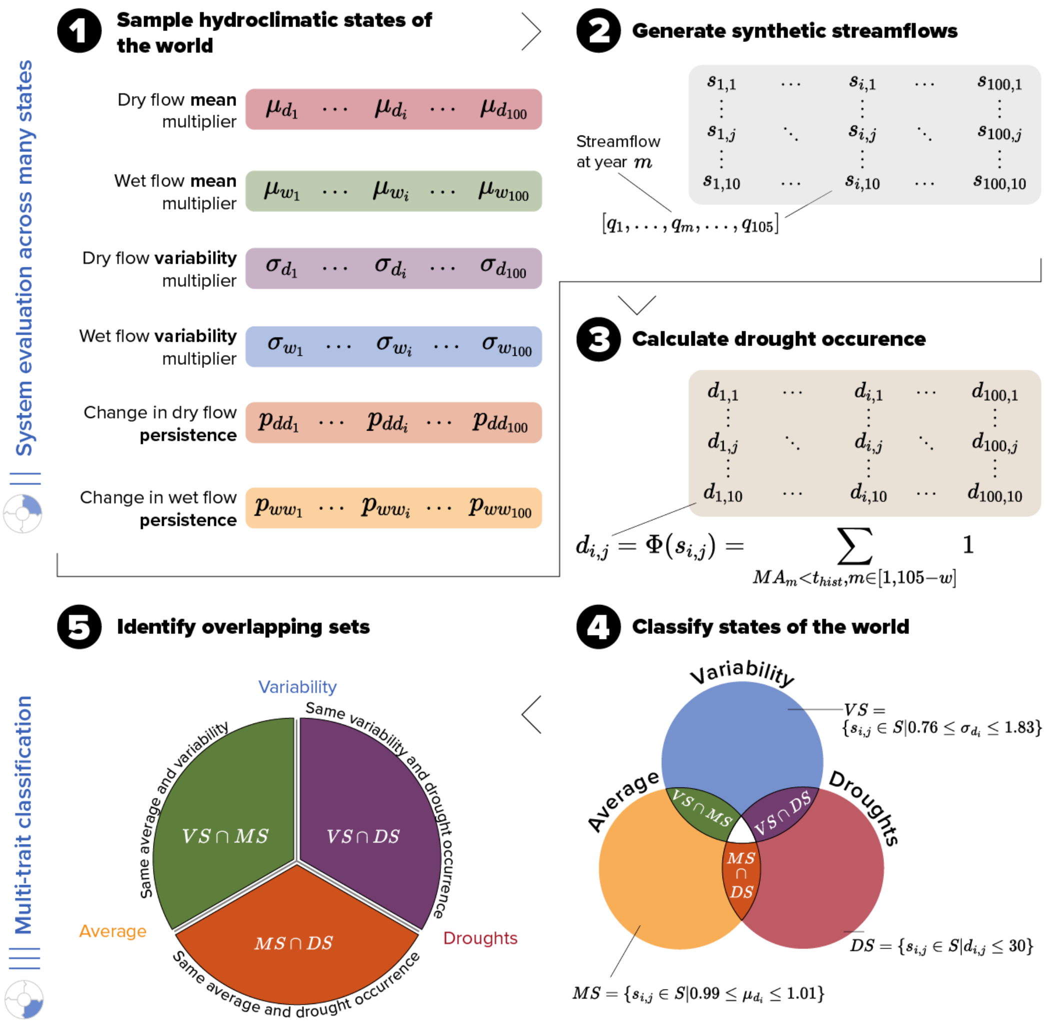

Figure 5. Applying stages II and III of FRNSIC to the UCRB case study. Steps 1-2 illustrate the generation and simulation of the hydroclimatic SOWs (Stage II). Steps 3-5 illustrate the classification of behavioral dynamics (Stage III). Sets of dynamic properties are defined as $VS \cap MS$: *Exhibiting the same variability and average annual dry flows*; $MS \cap DS$: *Exhibiting the same average dry flows and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the same variability of annual dry flows and number of decadal drought years.*

For each parameter combination i (i.e., for each combination of $\mu_{d_i}, \mu_{w_i}, \sigma_{d_i}, \sigma_{w_i}, p_{dd_i}, p_{dd_i}$), we generated 10 realizations of 105 years of streamflow, $s_{i,j}$, such that there exists a set of all streamflow SOWs $S = \{s_{i,j} | i \in I \wedge j \in J\}$ and $J = [1, 2, ..., 10]$. Each SOW $s_{i,j}$ represents a sequence $[q_1, q_2, ..., q_{105}]$, where q_m is the streamflow at year m. In other words, 10 realizations of 105-year-long times series of annual streamflows are created for each of the 100 sampled HMM parameterizations, resulting in a total of 105,000 synthetic years (Fig. 5 Step 2). The annual streamflows are generated in log space for the last node represented in the system model (at the Colorado-Utah state line) and then converted to real space and downscaled to monthly streamflows using a modified version of the proportional scaling method used by Nowak et al. (2010). The same method is also used to identify contributing proportions from all upstream model nodes, as detailed in Hadjimichael, Quinn, Wilson, et al. (2020). We note here that these streamflows are naturalized as required to serve as model input for StateMod water allocation model. The ensemble of streamflows from this all-encompassing experiment span those from all other sets (historical observations, paleo reconstructions, and projections), with values that exceed both sides of the distribution (Fig. S2).

3.4 Stage III - Multi-trait Classification of States of the World

3.4.1 Classification of dynamics

As noted in Section 2, one of the key contributions of our proposed framework is the classification of the dynamic properties of each sampled SOW within an exploratory modeling ensemble, irrespective of its performance on specific impact criteria (Fig. 1). The motivation in capturing these dynamics is largely to help illuminate the behavioral processes that lead to the consequential impacts, something that is often lost when scenario discovery is performed by classifying based on aggregate robustness performance measures. These dynamic properties can be specified *a priori*, if they are part of the design of experiments, or they can be discovered or estimated after each SOW simulation is performed. In our case, we utilize both approaches to capture three dynamic properties of our SOWs: the variability of dry year streamflows, the central tendency (average) of dry year streamflows, and the occurrence of decadal hydrologic drought conditions. With regard to the average and variance of dry years, (μ_d and σ_d, respectively) these properties are part of the sampled HMM parameters used to create each synthetic SOW and are therefore known without additional calculations for each model simulation. We choose to focus on these two properties of the synthetically generated SOWs (as opposed to properties of the wet states of each SOW) to better understand how dry flow dynamics contribute to water scarcity impacts, but any other behavioral property (statistical or otherwise) could also be used, as relevant to the problem under study. We emphasize here that even though these dynamic properties strongly influence impacts (which are classified in Section 3.4.2) the mappings between them are not necessarily known *a priori*, nor are they straightforward to infer. For example, one might intuit that decreasing the average annual streamflow during dry years (i.e., μ_d) will result in more water user impacts, but exactly how much change or how it interacts with other factors to shape impacts are not immediately apparent.

The occurrence of decadal hydrologic drought conditions is identified after the simulations are performed for each of the synthetically generated 105-year streamflow sequences (Fig. 5 Step 3). To do so, we follow Ault et al. (2014) and establish a drought threshold, T, as half a standard deviation from the period average (i.e., $\mu - 0.5\sigma$). For example, in Fig. 3 for the entire period of historical streamflow observations (105 years), we use the threshold $T = 5,884Mm^3$. When a moving average of annual streamflow (q_m) over 11 years falls below this threshold, we identify the period as a decadal-scale drought. Longer windows (e.g., 35 years) can be used to identify multi-decadal droughts, depending on the specific extreme drought application focus. Formally, for each SOW $s_{i,j}$, the total number of decadal drought years $d_{i,j}$ (Fig. 5 Step 3) is given by:

$$\Phi(s_{i,j}) = \sum_{MA_m<T, m\in[1,105-w]} 1, \tag{2}$$

where MA_m is the moving average of annual streamflows at year m given by:

$$MA_m = \frac{1}{w} \sum_{m,m\in[1,105-w]}^{m+w} q_m, \tag{3}$$

and w is the length of the rolling window (11 years in our case). The set of all drought year durations for all SOWs is then defined by:

$$D = \{d_{i,j} | d_{i,j} = \Phi(s_{i,j}) \forall [i \in I \wedge j \in J]\}. \tag{4}$$

We also denote $DY_{i,j}$ as the drought years of SOW $s_{i,j}$, given by:

$$DY_{i,j} = \{m | MA_m < T, m \in [1, 105 - w]\} \tag{5}$$

We therefore use three dynamic properties of each SOW $s_{i,j}$ to classify the dynamics of our SOW ensemble: the variability of dry year streamflows σ_{d_i}, the average of dry year streamflows μ_{d_i}, and the number of decadal drought years $d_{i,j}$. There is a variety of ways one might choose to classify SOW sets using these properties, depending on the specific analysis questions and as informed by the Problem Framing stage. We note in Section 1, that insights from co-production literature highlight that the manner with which information is presented to its users is critical to how they understand and choose to utilize it (Calvo et al., 2022). More specifically, and as it relates to the classification of dynamic properties, Lemos et al. (2012) stress that relating new findings to past experiences can help connect that information to stakeholder analytical and experiential processing abilities, as well as foster the usability of the new findings.

Based on these recommendations, we classify the dynamic properties of the SOWs based on how they relate to the historical experience of basin water users. For example, one might be interested in investigating the impacts of SOWs under the assumption that the future will be similar to the experienced past. In such a case, conditional criteria can be used to separate the SOWs that fall within the bounds of past experiences from the ones that do not. We demonstrate this by focusing on what we will be referring to as "historically-informed" SOWs: synthetic SOWs that exhibit properties within the range of dry year streamflow average and variance values as they appear in 60-year rolling windows of the record of gauged observations, as well as the past drought conditions resulting from said observed streamflow. These history-informed synthetic SOWs of hydrology reflect the assumption that the future will behave like the observed past and can be used to establish plausible stakeholder-relevant impacts that might be unlike those previously experienced. Corollary to this classification, we can identify SOWs that do not meet these criteria (e.g., by exhibiting more dry year streamflow variance relative to what has occurred in the available observed record) as SOWs reflecting a changing system.

To identify historically-informed thresholds for the variability and persistence of dry conditions we utilize the 60-rolling windows of streamflow, shown in Fig. 4 (a). For each window, we estimate its respective μ_d and σ_d and use those estimates to select subsets of our SOW ensemble in which μ_d and σ_d fall within the range of values observed across historical 60-year windows (Fig. S3). The set of SOWs that exhibit dry-flow variability within the bounds of history is therefore defined as:

$$VS = \{s_{i,j} \in S | 0.76 \leq \sigma_{d_i} \leq 1.38\}. \tag{6}$$

Similarly, the set of SOWs that exhibit dry-flow average values within the bounds of history is defined as:

$$MS = \{s_{i,j} \in S | 0.99 \leq \mu_d \leq 1.01\} \tag{7}$$

For a history-informed decadal drought occurrence threshold, we use the same 60-year rolling windows and calculate the number of historical decadal drought years using the drought threshold (T) as defined by the properties of each window (shown in Fig. 4 (b)). Given the varying values of these thresholds ($5,540 \leq T \leq 5,988$), the number of historical hydrologic years out of 105 that are classified as decadal drought years could be as low as zero and as high as 30 (Fig. S1). Assuming that this range of values reflects the range of historical experience of drought, we can use these values as a way to select the SOWs that produce numbers of decadal drought years that fall within the historical experience. The variation in decadal drought years from zero to 30 in this case reflects how drought experience in the basin has historically varied, depending on the different windows of time one may use as reference. To define the set of SOWs exhibiting numbers of decadal drought years within the bounds of historical experience, we therefore use these numbers as the bounds:

$$DS = \{s_{i,j} \in S | d_{i,j} \leq 30\}. \tag{8}$$

In other words, by looking at 60-year rolling windows of historical hydrologic observations (Fig. 4), we are able to deduce a range of values for these dynamic properties as experienced historically. Using these ranges we create three sets of SOWs, each exhibiting these historically-bounded properties. These three sets therefore represent three different dynamic properties of the ensemble of SOWs used in this experiment: VS contains SOWs that fall within the range of the historical variability of dry conditions, MS contains SOWs that fall within the range of the historical average of dry conditions, and DS contains SOWs that fall within the range of drought years experienced in history (Fig. 5 Step 4). We note that these classifications are irrespective of the impacts these SOWs result in (discussed in the following section), and can be used to both uncover the dynamic properties that result in consequential impacts, as well as create narrative storylines of how said impacts come to be. Furthermore, several of our generated SOWs might meet more than one of these conditions. In other words, there exist intersecting sets $VS \cap MS$: *Exhibiting the same variability and average annual dry flows*; $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the same variability in annual dry flows and number of decadal drought years*, as shown in Fig. 5 Step 5. These are simply sets of SOWs where both respective set conditions are met, and might vary in size (discussed in Section 4). All these sets, as well as their intersects, contain SOWs which reflect the hypothesis that the future hydroclimate in the region will be like the past 105 years of observed streamflow conditions. A set where all conditions are met may also exist, and can be further investigated as needed. We do not do so in this current application, largely because the influence of the dynamic conditions is sufficiently demonstrated with the three pairs, and to maintain visual and narrative simplicity.

Corollary to the existence of these sets in our full ensemble of SOWs S, is that for each set of SOWs that meet each dynamic condition there exist complement sets VS', MS', and DS' for which each respective condition does not hold. Specifically: VS' contains SOWs that exhibit dry variability that exceeds the historically observed range, MS' contains SOWs that exhibit average dry values that exceed the historically observed range, and DS' contains SOWs with more drought years than the historically observed range. As such, these sets contain plausible SOWs which reflect the hypothesis that the future hydroclimate in the region will be different from the observed conditions. These SOWs are part of the same ensemble and, even though they exceed historically observed conditions, they remain within plausible future ranges as informed by the extended internal variability based on paleo reconstructed data and changing future conditions simulated under CMIP5 projections (see Section 3.3 and Quinn et al. (2020)). As a result, we create equivalent intersecting sets that capture these plausible, changing dynamic conditions $VS' \cap MS'$: *Changing average and variability in annual dry flows*; $VS' \cap DS'$: *Changing variability in annual dry flows and number of decadal drought years*; and $MS' \cap DS'$: *Changing average of annual dry flows and number of decadal drought years*. It should be noted that the number of decadal drought years only increases relative to historical ranges in these sets (since the lower bound using the historical rolling windows is 0), whereas the average and variability in annual dry flows increases in some and decreases in others.

3.4.2 Classification of impacts

All synthetically generated 105-year timeseries are simulated through StateMod which allocates water to users in the basin according to their rights allocation, the point of their diversion, and the availability of water at each given monthly time step and stream location (CWCB & CDWR, 2016). StateMod allows us to thus assess how these synthetic conditions affect key impacts across all decision-making scales pertinent to the UCRB (Fig. 6). Specifically, the Colorado Basin Roundtable is concerned with meeting the UCRB's obligations for deliveries downstream, as bound by the Colorado River Compact, as well as overall deliveries (or shortages) to the water rights' holders within the basin. Both of these impacts are emphasized as key concerns in Colorado Basin Roundtable's Basin Implementation Plan (CWCB & CDWR, 2022). Within the basin itself, water districts (WDs), are interested in how their own, largely agricultural, users might be affected by future hydroclimatic stress, and individual water rights' holders are primarily concerned with impacts to their own supply.

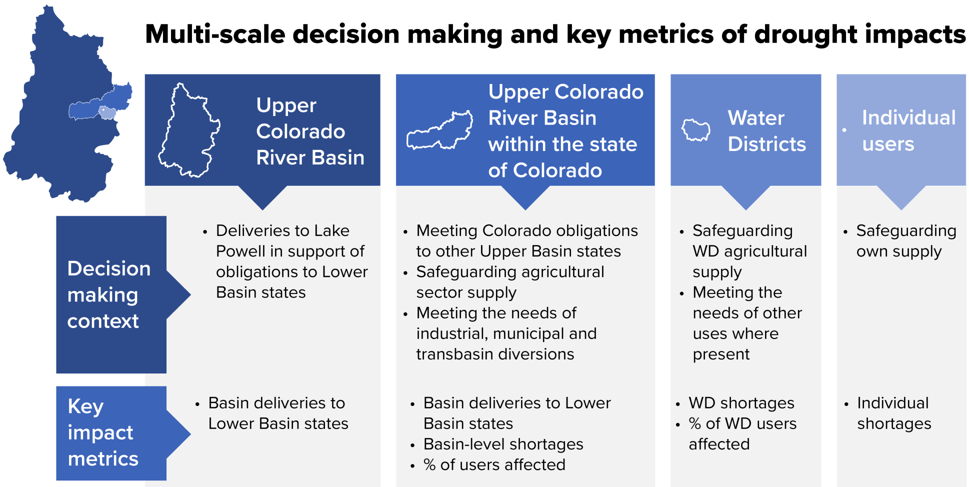

Figure 6. The multi-scale decision making context of the UCRB. Moving from left to right reflects a more localized scale, from the broader multi-state Upper Colorado River Basin region, to the individual water users in the UCRB. Focusing on smaller regions shifts the decision making context and the key metrics of concern with regard to hydrologic drought. These key impacts are reflected in the impact classification scheme (Fig. 7).

We assess these multi-scale impacts by looking at water demands and shortages (undelivered water) to 338 users in the basin during the drought periods of each SOW, as well as basin deliveries downstream (water leaving the UCRB). Water demands per user are a StateMod output, defined here as $W(u, s_{i,j})$, the water demand for user u during the drought periods of SOW $s_{i,j}$. Equivalently, water shortage $G(u, s_{i,j})$ is the undelivered water to user u during the drought periods of SOW $s_{i,j}$ (Fig. 7 Step 6). Using this notation, we can calculate the percentage of shorted users during the drought period of each SOW $s_{i,j}$ as:

$$\Psi(s_{i,j}) = \frac{100}{n_{users}} \sum_{G(u,s_{i,j})>0,u\in[1,\ldots,n_{users}]} 1 \tag{9}$$

and the mean shortage across users—during the same drought period—as:

$$X(s_{i,j}) = 100 \sum_{u\in[1,\ldots,n_{users}]} \frac{G(u, s_{i,j})}{W(u, s_{i,j})} \tag{10}$$

For both equations we use $n_{users} = 338$ for all consumptive use water users in the basin.

The third key impact metric we are tracking is how delivery obligations to Lake Powell are affected. There is a large number of moments, quantiles, or other distributional measurements we can track here. We are using the rolling 10-year sum of basin deliveries, consistent with how Upper Basin state obligations are typically accounted for (e.g., Bureau of Reclamation (2012); Woodhouse et al. (2021)). For each SOW, we calculate this 10-year rolling sum and estimate the 10^{th} percentile of all values to focus explicitly on the lowest 10-year cumulative deliveries. Formally, we denote qo_m as the basin outflow in year m for each SOW $s_{i,j}$, and $BD_{i,j}$ as the sequence

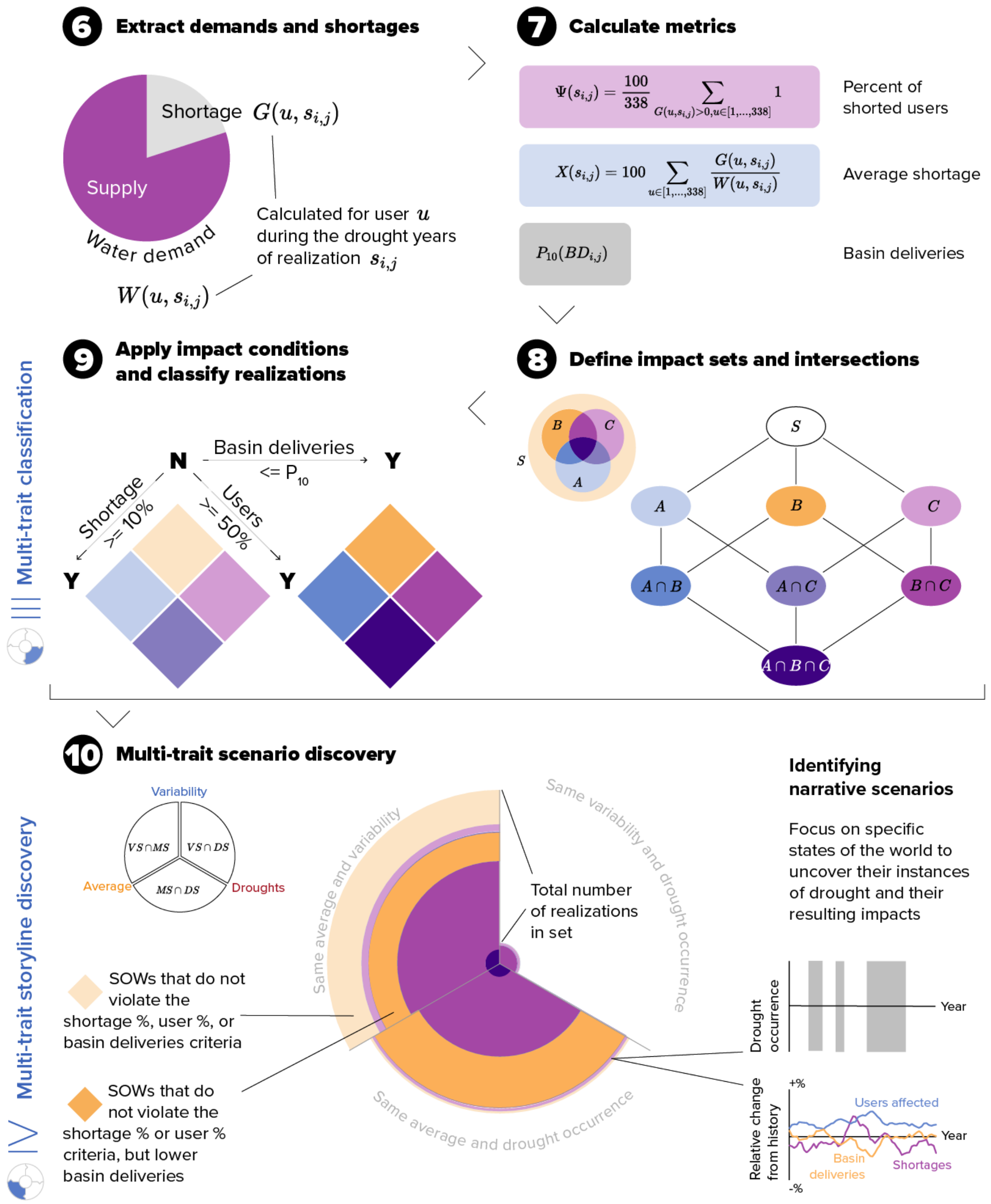

Figure 7. Applying stages III and IV of FRNSIC to the UCRB case study. Steps 6-9 calculation and classification of user- and basin-level impacts (Stage III). Step 10 illustrates the combination of said impacts with behavioral dynamics to identify narrative drought storylines for the UCRB (Stage IV).

of all cumulative 10-year sums:

$$BD_{i,j} = (bd_1, ..., bd_m, ..., bd_{95}), \tag{11}$$

where:

$$bd_m = \sum_{m,m\in[1,95]}^{m+10} qo_m \tag{12}$$

is the cumulative 10-year sum of deliveries at year m, and $P_{10}(BD_{i,j})$ is the 10$^{\text{th}}$ percentile of all cumulative sums (Fig. 7 Step 7).

Based on these metrics, we identify which of the synthetic SOWs are consequential to the Colorado Basin Roundtable and its stakeholders by quantifying their effects on water deliveries to basin users and downstream. In this manner, the scenarios identified are intrinsically tied to the consequential impacts they generate at the basin itself, overcoming the limitation presented by the limited set of five driver-defined scenarios used by the state (State of Colorado, 2023). Further, through the use of exploratory modeling, we more rigorously investigate the space of plausible future conditions, to then, *a posteriori*, discover the ones that truly matter locally. As overviewed earlier, this process of *a posteriori* scenario classification is formally referred to as scenario discovery (Bryant & Lempert, 2010; Kwakkel, 2019). Traditionally, scenario discovery is a classification process, and categorizes hypothetical scenario conditions as either 'successes' or 'failures' depending on whether they meet a criterion, or a combination of a small number of them. Classification in its simplest form is performed through separating the space using orthogonal subspaces, typically using algorithms such as the Patient Rule Induction Method (PRIM; Friedman and Fisher (1999)) or Classification and Regression Trees (CART; Breiman (1984)). Applying these methods to real complex systems has uncovered several challenges in both the criteria used to identify the scenarios of interest (i.e., what measure to use to select 'failed' SOWs), as well as in the computational methods used to do so, also known as rule induction or factor mapping (i.e., identifying what factors lead to failures). Respective advancements have been made to tackle these challenges. Challenges with regard to rule induction are primarily rooted in the orthogonality (Kwakkel, 2019), linearity (Pruett & Hester, 2016; Quinn et al., 2018), and convexity (Guivarch et al., 2016; Trindade et al., 2019, 2020)—and lack thereof—of the space being separated. We refer the reader to these studies for more information about methodological advancements in this space. The challenges surrounding identification, particularly with regard to complex multi-actor systems with a large number of relevant states, have been broadly articulated in Section 1. Here, we discuss how FRNSIC is addressing them for the UCRB case study.

We utilize three metrics to capture overall impacts to the basin: percentage of shorted users ($\Psi(s_{i,j})$; Eq. 9), mean shortage ($X(s_{i,j})$; Eq. 10) and the 10$^{\text{th}}$ percentile of cumulative basin deliveries ($P_{10}(BD_{i,j})$; Eq. 11), each relevant to the multi-scale decision making context of the UCRB (Fig. 6). As described in Section 2, we utilize a set theory perspective in SOW classification by creating conditional sets based on whether the SOWs meet each impact criterion. For multiple criteria we can also create multiple such subsets and look at the intersections of the conditional sets for combinations of multiple criteria. This mirrors how satisficing metrics are typically used in the robustness analysis stage of RDM or MORDM applications, where more than one performance metric might matter to whether a strategy is considered "robust" (McPhail et al., 2018). In those cases, multiple metrics are used together to assess robustness (e.g., "$reliability \geq 90\%$" AND "$costs \leq \$100$"), but rarely are different subsets and combinations compared. FRNSIC presents an alternative approach, where the hierarchical combination of impact metrics allows for the discovery of robust strategies across all possible combinations of performance metrics. Fig. 7 Step 8 shows an example of this, using three subsets A, B, and C, each corresponding to an impact criterion. This partially ordered set is an algebraic structure formally referred to as a Boolean lattice, often visualized using a Hasse diagram (Priss, 2021), as shown in Step 8. Starting at the top of this graphic, S denotes the entire set of SOWs in our ensemble, of which A, B, and C are subsets. Moving downward, we combine these sets to their intersections indicating two of the conditions being met, with the subset in the very bottom indicating the set where all three conditions are met.

In this application, we establish three criteria based on which conditional SOW sets are created, each using one of the key impact metrics (Fig. 6). Specifically, using the mean shortage experienced during each SOW $X(s_{i,j})$ (Eq. 10), we can define a conditional subset of SOWs that exceed a decision-relevant threshold for water shortage, given by th_χ, such that:

$$A = \{s_{i,j} \in S | X(s_{i,j}) >= th_\chi\}. \tag{13}$$

For example, using the nominal value of $th_\chi = 10\%$ we select a subset of SOWs A where the mean user shortage exceeds 10% (Fig. 7 Step 9). We can capture higher or lower degrees of risk tolerance in the basin (e.g., a mean shortage of 20% versus 5%) by utilizing shortage thresholds at various levels to establish a different set A conditioned on the threshold used. For reference, the historical average shortage across all years and all basin users is 7%.

Looking at the downstream basin deliveries in each SOW, we compare whether the 10^{th} percentile of cumulative 10-year streamflows of each SOW ($P_{10}(BD_{i,j})$; Eq. 11) meets or subceeds a critical threshold th_{bd}. This second conditional set B is given by:

$$B = \{s_{i,j} \in S | P_{10}(BD_{i,j}) <= th_{bd}\}. \tag{14}$$

This set identifies SOWs that have their lowest 10% of cumulative deliveries fall below a critical threshold. For instance, using the historical 10^{th} percentile of cumulative deliveries (46,820 M m^3) as th_{bd}, we select SOWs where the basin is delivering less than its historical 10% worst years.

Lastly, using the percentage of shorted users $\Psi(s_{i,j})$ (Eq. 9), we can identify a conditional subset of SOWs that exceed a consequential threshold of shorted users, given by th_ψ, such that:

$$C = \{s_{i,j} \in S | \Psi(s_{i,j}) >= th_\psi\}. \tag{15}$$

In the FRNSIC illustration in Fig. 7 Step 9, we create subset C by using the nominal value $th_\psi = 50\%$ to select all SOWs where more than 50% of water users are shorted. For reference, historically, an average of 30% of water users is shorted at any given year, with some years reaching up to 66%.

We note that sets A, B, and C are not mutually exclusive and there may exist SOWs in S that meet more than one or all three criteria (Fig. 7 Steps 8-9). By applying each threshold and identifying each conditional subset that meets the condition—including their intersections—we classify every SOW as belonging in either:

- a set where none of the conditions are met (i.e., $(A \cup B \cup C)'$, shown in light yellow ◆),
- three sets where only one of the conditions is met (i.e., set A in light blue ◆ with larger shortages, set B in yellow ◆ with lower deliveries, and set C in lilac ◆ with more shorted users),
- three sets where two conditions are met (i.e., $A \cap B$ in blue ◆ with both larger shortages and lower deliveries, $A \cap C$ in light purple ◆ with both larger shortages and more shorted users, and $B \cap C$ in violet ◆ with both lower deliveries and more shorted users,
- and lastly, one set where all three of the conditions are met (i.e., set $A \cap B \cap C$) in dark purple ◆.

These eight sets are all shown with regard to their partially-ordered relationships in Fig. 7 Step 8 and in how they are applied for impact classification in Step 9. Using these impact sets, we create a hierarchical set-of-sets where impact criteria can be combined to reflect additional stakeholder impacts or conditions. As with the classification of dynamic properties, we only utilize three criteria here, but the proposed method is amenable to larger numbers. We do stress, however, that interpretability and narrative clarity quickly degrade with the addition of more dimensions.

3.5 Stage IV - Multi-trait storyline discovery

The final step in the proposed framework combines the impact classification performed in Step 9 (Fig. 7) with the SOW sets identified in Step 5 (Fig. 5) for the creation of narrative storylines that capture both key behavioral dynamics of SOWs and consequential impact metrics. Fig. 7 Step 10 shows how the SOWs in each overlapping set of dynamic behavior (i.e., $VS \cap MS$: *Exhibiting the same variability and average annual dry flows*; $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the* *same variability of annual dry flows and number of decadal drought years*) can be distributed among the eight impact groups. This graphic is an adapted version of a stacked hive plot (Krzywinski et al., 2012), and allows us to visualize the resulting high-dimensional dataset in a single-panel figure. The three segments of the circle[4] each correspond to the overlapping sets for average and variability of annual dry flows and number of decadal drought years. The radius of each segment (how much it extends from the center point) indicates the total number of SOWs that fall within the overlapping set. For example, in the hive plot shown in Fig. 7 Step 10 the top left set (defined by having the same average and variability of dry years as history) contains the most SOWs, whereas the top right set (defined by having the same dry flow variability and number of decadal drought years as history) contains the least. Within each segment, the width of each band indicates the number of SOWs from that set that result in one of the eight impact groups identified above. Using the same example figure in Step 10, most of the SOWs exhibiting the same variability and average of annual dry flows (in the top left segment) are in the violet impact group ◆ (i.e., they result in both lower basin deliveries and having more in-basin water users shorted).

The reader can use this plot for several insights: to compare the relative size for each overlapping set of dynamic properties (e.g., to make inferences about how the dynamic properties of the SOWs in the ensemble are distributed); and to compare the relative shift in impact groups when moving from one set of dynamics to the other (e.g., starting from the top left segment and moving to the bottom one we can see that fewer SOWs exhibit no impacts at all—the light yellow band goes away). Presenting everything in a condensed single-panel format allows us to combine this with several other panels resulting from other criteria and thresholds combinations, in a "small multiples" visualization (Tufte, 1990). Showing many small visualizations simultaneously allows the reader to compare the separate panels and look for patterns or outliers in the matrix of visuals, and facilitates presentation and storytelling of large amounts of data in a single figure (van den Elzen & van Wijk, 2013). We note here that even though we are only using three types of dynamic sets and three types of impacts, combining them all together means that this single panel figure captures 24 properties in a single panel (3 dynamic sets x 2^3 impact groups). Even though more sets of either kind can be used (i.e., a hive plot can be created with more than three axes and more than eight color bands) the interpretability of the figure greatly diminishes (Krzywinski et al., 2012). We do not consider this a weakness of this specific visual form, as alternative options (e.g., parallel coordinate plots) also struggle from the same limitations, but without the added benefit of being able to be used in a small multiples visualization without further simplification.

In our hypothetical planning context, the Colorado Basin Roundtable can use these plots to examine specific narrative scenarios. The impact sets are organized from most severe in dark purple (all three impact conditions are true) to least severe in light yellow (none of the impact conditions is true) going from the center of the plot outward. In this manner, we illuminate the narrative scenario each SOW can represent, by capturing both the critical impacts it generates and the dynamic properties that lead to it. For example, the Colorado Basin Roundtable users can subselect a segment (e.g., "investigate future SOWs that have the same mean and variance as we've seen in the past") and then subselect a specific SOW from the impact groups of interest (e.g.,"what are the worst impacts we encounter in these futures"). This SOW can then be further investigated for its temporal dynamics and the impacts they result in within the Basin, and be used to frame future planning and adaptation efforts. Even though we do not perform formal scenario discovery in the form of factor mapping in this demonstration (e.g., searching for the specific combinations of σ_d and μ_d values that lead to a mean shortage of more than 10%), one can additionally be performed as needed. We instead highlight the narrative strength of combining sets of dynamic and impact properties in examining candidate futures for the UCRB.

[4] Geometrically, these are in fact sectors of the circle, but we use the term segment here to avoid later confusion with terms like "agricultural sector"

4 Results and Discussion

4.1 Identifying consequential drought storylines at the basin-level

Planning to address drought often starts with an investigation of baseline historical drought hazards. As illustrated in Fig. 3, plausible historical drought extremes can be well beyond those observed in the limited historical streamflow record due to internal variability, even assuming stationarity. We first illustrate a basin-level assessment in which a coordinated planning group such as the Colorado Basin Roundtable is interested in examining futures that remain statistically similar to the last century of observations. In other words, out of our ensemble of hydrologic SOWs (detailed in Section 3.3), they might want to examine ones that exhibit the range of dynamic properties exhibited in the historical streamflow observations. Specifically, they apply the conditional criteria in Eqs. 6-8 to identify intersecting sets of history-informed SOWs ($VS \cap MS$: *Exhibiting the same average and variability in annual dry flows*; $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; and $VS \cap DS$: *Exhibiting the same variability in annual dry flows and number of decadal drought years*), shown in Fig. 8 (a).

Several insights can be drawn from this figure. First, in terms of dynamic classification, 100 SOWs exhibit the same average and variability in annual dry flows as in the observed past (top left segment), 82 exhibit the same variability in annual dry flows and number of decadal drought years as in the observed past (top right segment), and 45 SOWs exhibit the same average annual dry flow and number of decadal drought years as in the observed past (bottom segment). The spread of each color in each segment denotes the distribution of each impact group across each set of SOWs, as determined using the classification described in Section 3.4.2, applied at the basin level. Specifically, each SOW is categorized based on whether: (i) it increases the average shortages basin-wide to more than 10% (the yellow to blue dimension), (ii) it increases the number of basin users that experience shortage to above 50% (the yellow to pink dimension), and (iii) it lowers basin deliveries to Lake Powell below the historical 10$^{\text{th}}$ percentile (P_{10}) of cumulative 10-year deliveries (the light to dark dimension). If an SOW increases both average shortages and the number of affected users, it is classified in light purple, and if it also decreases deliveries downstream, it is classified in dark purple. Comparing across the segments we see that more SOWs are classified as exhibiting the same average and variability in annual dry flows (top left segment) than other segments, but the impacts in these worlds are minor to moderate (light to dark yellow). The most severe impacts are generated in SOWs that exhibit the same variability in annual dry flows and number of decadal drought years criteria (small violet region in the top right), suggesting these drought characteristics may be more impactful.

In further examining these most severe impacts, a group such as the Colorado Basin Roundtable can zoom in on one of the SOWs that generated them and investigate its temporal dynamics and how they affect the basin as a whole, as well as particular users. For example, Fig. 8 (a) can be further examined by specifically focusing on the small number of SOWs in the top right segment (i.e., those exhibiting the same variability in annual dry flows and number of decadal drought years as observed history) that produce the most extreme impacts. These two SOWs are shown in violet because they increase the average shortage experienced in the basin to above 50% and also lower cumulative basin deliveries to below the historical 10$^{\text{th}}$ percentile. In Fig. 9, we further investigate the dynamics of one of these SOWs: the one that exhibits the fewest drought years. We refer to this drought storyline as "The Unknown Normal". In this narrative storyline, a drought spanning 23 years takes place and affects both the UCRB's downstream deliveries but also the water shortages experienced in the basin. At the basin-wide level, we first compare the basin's 10-year cumulative downstream deliveries to their historical 10$^{\text{th}}$ percentile ($46,820 Mm^3$; top left panel in Fig. 9). We see that during the drought period cumulative basin deliveries downstream fall below the historical cumulative 10$^{\text{th}}$ percentile for some of the years, down to 80% of that historical threshold ($37,184 Mm^3$) during one of the years. This shows that even non-extreme hydroclimatic changes can have significant impacts in basins like the UCRB and jeopardize their ability to meet their inter-state obligations. Examining impacts within the basin, we look at cumulative basin-wide shortages as they relate to the historical 90$^{\text{th}}$ percentile (Fig. 9 top right panel). During this same drought period, we see total shortages in the basin accumulate to almost seven

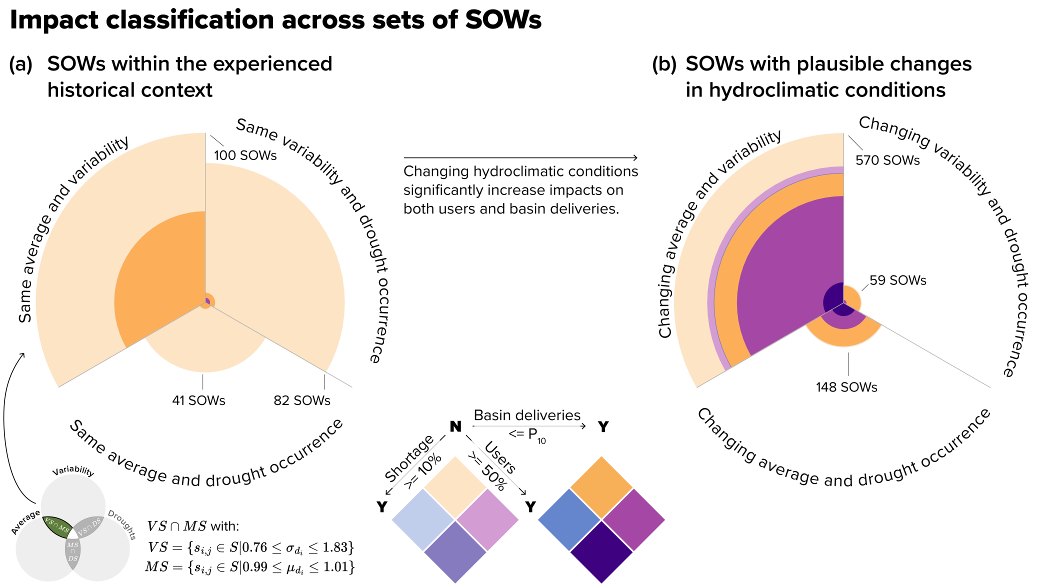

Figure 8. Basin-level impact classification for all states of the world (SOWs) as organized by combinations of dynamic properties. (a) Impacts in SOWs that exhibit dynamic properties within the bounds of the historical context. Starting from the top left: $VS \cap MS$: *Exhibiting the same average and variability in annual dry flows*; $VS \cap DS$: *Exhibiting the same variability in annual dry flows and number of decadal drought years*; and $MS \cap DS$: *Exhibiting the same average annual dry flow and number of decadal drought years*; (b) Impacts for SOWs that exhibit dynamic properties outside the bounds of the observed past (changing hydroclimatic context). Starting from the top left: $VS' \cap MS'$: *Changing average and variability in annual dry flows*; $VS' \cap DS'$: *Changing variability in annual dry flows and number of decadal drought years*; and $MS' \cap DS'$: *Changing average of annual dry flows and number of decadal drought years*. All SOWs are categorized based on whether they affect average shortages basin-wide (the blue dimension), they affect the number of basin users that experience shortage (the pink dimension), and they lower basin deliveries below the historical 10$^{\text{th}}$ percentile (P_{10}) of cumulative 10-year deliveries (the darkness dimension). Moving from SOWs within the range of historical conditions to the SOWs with changing conditions, experienced impacts become more severe.

times the historical threshold condition and start receding when the drought period is over. We note that there is also a second period during the last 20 years for this simulated future where comparable impacts are seen, but it is not formally classified as a drought period.

As elaborated in Section 3.1 the UCRB supports hundreds of individual water users that use water for many operations: agriculture, municipal water supply, industrial production, power generation, as well as recreational uses (Fig. 2). In prior work in the basin, we have shown that depending on their priority, demands, and location in the basin these users might individually experience very different water scarcity impacts (Hadjimichael, Quinn, Wilson, et al., 2020). We have also shown that aggregate basin impacts (e.g., the mean shortage metric utilized here) can be highly variable across the basin when spatially disaggregated, even at the WD level (Hadjimichael et al., 2023). We therefore further disaggregate these impacts to the UCRB's water districts and users, enabled by StateMod, which traces water allocation and shortage to the individual user level. In Fig. 9 we highlight shortage as a percent of demand for three WDs (39, 37, and 51, moving left to right) in the middle panels with purple lines and four water users in the bottom panels with blue lines. The WD- and user-level shortages show the diverse within-basin experience of this drought storyline, with some WDs and users experiencing very severe shortages

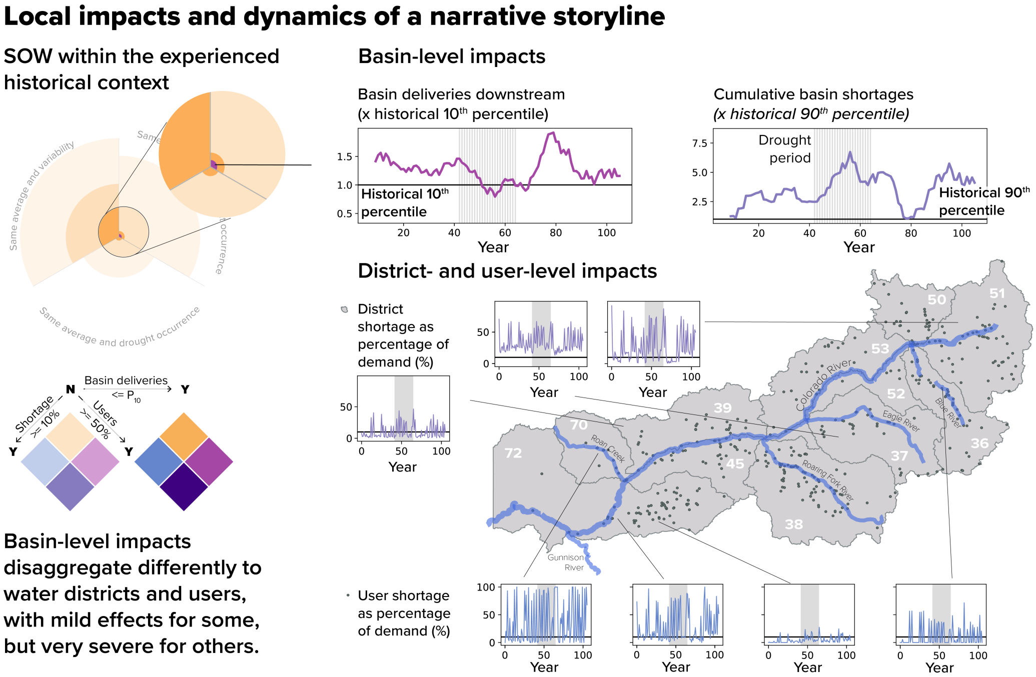

Figure 9. The Unknown Normal: impacts and dynamics of a history-informed drought storyline. The impacts of this state of the world (SOW) are presented for the basin-level at the top, and disaggregated to water districts (middle panels with purple lines) and to individual water users in the basin (bottom panels with blue lines).

and others largely unaffected. These findings align with our prior results while providing a more detailed example of how the same sampled SOW dynamics can yield widely varying shortage impacts subject to the specific characteristics of the various users: their right seniority and decreed allocation, the timing of their demands, and their location in the basin, among others (Hadjimichael, Quinn, Wilson, et al., 2020; Hadjimichael, Quinn, & Reed, 2020; Quinn et al., 2020).

Alternatively, planners might choose to focus on SOWs which reflect assumptions about a changing hydroclimate. In this case the focus would be looking at the complement sets and their intersections (i.e., $VS' \cap MS'$: *Changing average and variability in annual dry flows*; $MS' \cap DS'$: *Changing average of annual dry flows and number of decadal drought years*; and $VS' \cap DS'$: *Changing variability in annual dry flows and number of decadal drought years*). These SOWs and their impacts are shown in Fig. 8 (b). Looking at the changing context sets (Fig. 8 (b)), 570 SOWs exhibit changing average and variability in annual dry flows, 59 SOWs exhibit changing variability in annual dry flows and number of decadal drought years, and 148 SOWs exhibit a changing average of annual dry flows and (increasing) number of decadal drought years. A lot more SOWs meet these dynamic conditions (as compared to Fig. 8 (a)), which is attributed to two main reasons. First, our ensemble of sampled hydroclimatic changes that shape each SOW takes into account projected climate change in the region and how it will change the distributions of streamflow, as well as paleo-reconstructed streamflows (Quinn et al., 2020). This means that several SOWs in our ensemble exhibit statistical properties different from those seen in the gauged record and, in fact, go beyond those distributions (see Fig. S2 and also Fig. S3 (a) for the ranges of mean and variance values). Further, due to these changing properties, the number of drought years in each SOW might also change. In fact, many of the SOWs in our ensemble exhibit more decadal drought years than the maximum of 30 years (or three decades) observed historically based on the highest threshold defined by 60-year rolling windows of streamflow observations (Figs. S1 and S3 (b)), or the deterministic estimate of one or two instances of decadal drought per century, estimated in paleo record studies of Ault et al. (2014); Woodhouse and Overpeck (1998).

This is also related to the second reason we see more SOWs fall outside the historical ranges, especially violating the condition on the number of decadal drought years (Eq. 8). For each sampled change in the average and variability in annual dry flows (i.e., changes in μ_d and σ_d values, as shown in Fig. 5 Step 1), we generate 10 streamflow realizations to capture the internal variability of each hypothesized hydroclimatic change (Fig. 5 Step 2). By better exploring this internal variability we see a wider range of decadal drought years emerge, even between SOWs that exhibit the same statistical properties, as expected (Lehner & Deser, 2023). This is exemplified in Fig. 3 for the internal variability of the recent history. Even though only 22 years of drought were observed (Fig. 3 (a)), this deterministic framing does not represent the true frequency of such events, which may be higher, as seen in Fig. 3 (b). The combined effects of a changing climate and internal variability produce SOWs with many more years of decadal drought than 30 out of 105 (Fig. S2 (b)), classifying them as outside the historical experience of water users in the UCRB under different rolling windows of 60 years (Fig. 4 and S1). These SOWs therefore appear in Fig. 8 (b).

Looking at Fig. 8 (b), SOWs in a changing hydroclimatic context produce much more severe impacts. Whereas most SOWs in the historical context do not produce impacts in any of the impact categories (i.e., no mean shortages more than 10%, no more than 50% of users affected, and no basin deliveries below the historical 10$^{\text{th}}$ percentile), most of the SOWs in the changing context produce impacts in at least two. This is seen in how the large bands of light yellow ◆ change to bands of yellow ◆, violet ◆, and dark purple ◆. The changing properties of these SOWs to lower average annual dry flows with greater variability and greater number of decadal drought years, leads to more severe impacts to the UCRB's water users. This is especially true for the basin's downstream deliveries: the majority of SOWs are assigned a dark color, indicating basin deliveries falling below the historical 10$^{\text{th}}$ percentile of cumulative 10-year deliveries.

Out of the SOWs that belong in the changing context sets (Fig. 8 (b)) 116 of them produce impacts across all impact groups (dark purple ◆ band): the average shortage they produce is more than 10%, they affect more than 50% of users, and they reduce basin deliveries below the historical 10$^{\text{th}}$ percentile of cumulative deliveries. Relating this to past experiences in the basin, the historical average shortage across all years and all basin users is 7% and has reached up to 26% in exceptionally dry years such as 2002 (the exceptionally dry conditions of 2002 can also be seen in Fig. 3 (a)). Basin-wide shortages of 10% of water demand have historically only been observed during drought periods, and the SOWs represented here capture those conditions. Further, with regard to the 50% of affected users, the historical average number of affected users at any given year in the UCRB is 30%, with the maximum percentage being 65%, again during the exceptionally dry conditions of 2002. Therefore, the SOWs that produce conditions affecting 50% of water users or more reflect plausible impacts of the drought extremes represented in our ensemble.

Fig. 10 examines the impacts and dynamics of one of these SOWs in more detail. In particular, we choose to focus on a SOW that produces impacts across all impact groups under the shortest drought duration. This SOW exhibits changing average and variability in annual dry flows (top left segment of Fig. 8 (b)) and has a total of 20 decadal drought years out of 105. We are referring to this drought storyline as "The Unforeseen Struggles". In the top two panels, we again compare the basin's 10-year cumulative downstream deliveries to their historical 10$^{\text{th}}$ percentile (left panel) and the basin-wide 10-year cumulative shortages (right panel). During this drought storyline, a 20-year drought takes place and has dramatic effects on the UCRB: cumulative deliveries drop to below 30% of the historical threshold ($13,862Mm^3$) and cumulative shortages climb to 11 times more than the historical 90$^{\text{th}}$ percentile of shortages. Unfolding these impacts at the finer scale, we compare WDs 70, 37, and 52 in the middle panels, as well as the same four users in the bottom panels, as analyzed in Fig. 9. We again see that the storyline affects the users differently, with some barely affected. Of note is also the fact that even though this storyline is much more severe in aggregate effects compared to "The Unknown Normal" in Fig. 9, impacts to individual users do not necessarily follow the same trend. For example, the leftmost water user

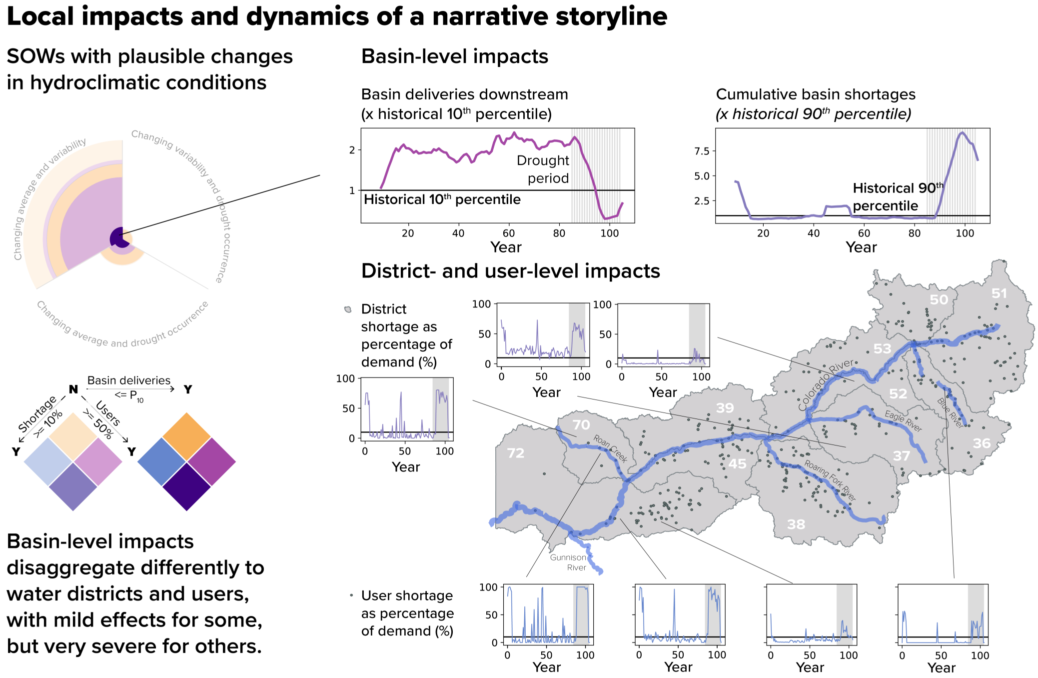

Figure 10. The Unforeseen Struggles: impacts and dynamics of a drought storyline in a changing context. The impacts of this state of the world (SOW) are presented for the basin-level at the top, and disaggregated to water districts (middle panels with purple lines) and to individual water users in the basin (bottom panels with blue lines).

experiences much more severe impacts under "The Unknown Normal" storyline, which falls within the historical bounds. The comparison holds true for other users also, which suggests that the significant aggregate effects we see in Fig. 10 are the result of a larger number of users being affected, not necessarily their larger shortages.

4.2 Examining exploratory ensemble impacts at the sub-basin scale

Beyond the two storylines illustrated in Figs. 9 and 10, we are also interested in how the entire ensemble disaggregates to the subbasin level. For instance, Colorado Basin Roundtable planners might be interested in the distribution of impacts the SOWs generate for a particular WD (Fig. 6). In Fig. 11, we therefore explore what the aggregate basin impacts shown in Fig. 8, look like for each WD in the basin. To do so, we apply Eqs. 9 and 10 to the specific subset of users that divert water in each WD and utilize the same color scheme used in Fig. 8. In this case, each SOW is categorized based on whether: (i) it increases the average shortages at each WD to more than 10% (the yellow to blue dimension), (ii) it increases the number of WD users that experience shortage to above 50% (the yellow to pink dimension), and (iii) it lowers basin deliveries to Lake Powell below the historical 10^{th} percentile (P_{10}) of cumulative 10-year deliveries (the light to dark dimension). If a SOW both increases average shortages and the number of affected users, it is classified in light purple, and if it also decreases deliveries downstream, it is classified in dark purple. In this case, the basin deliveries calculation remains the same, so we do not expect to see any differences in that dimension of impact categories. By calculating mean shortages and the percentage of users shorted for each WD individually, as opposed to the basin as a whole, we therefore expect to see shifts from yellow to lilac or blue (or to purple for both) and vice versa, but we should not observe shifts from light colors to dark colors (or vice versa), as the basin delivery calculation remains the same as that of the aggregate plots (shown in Fig. 8).

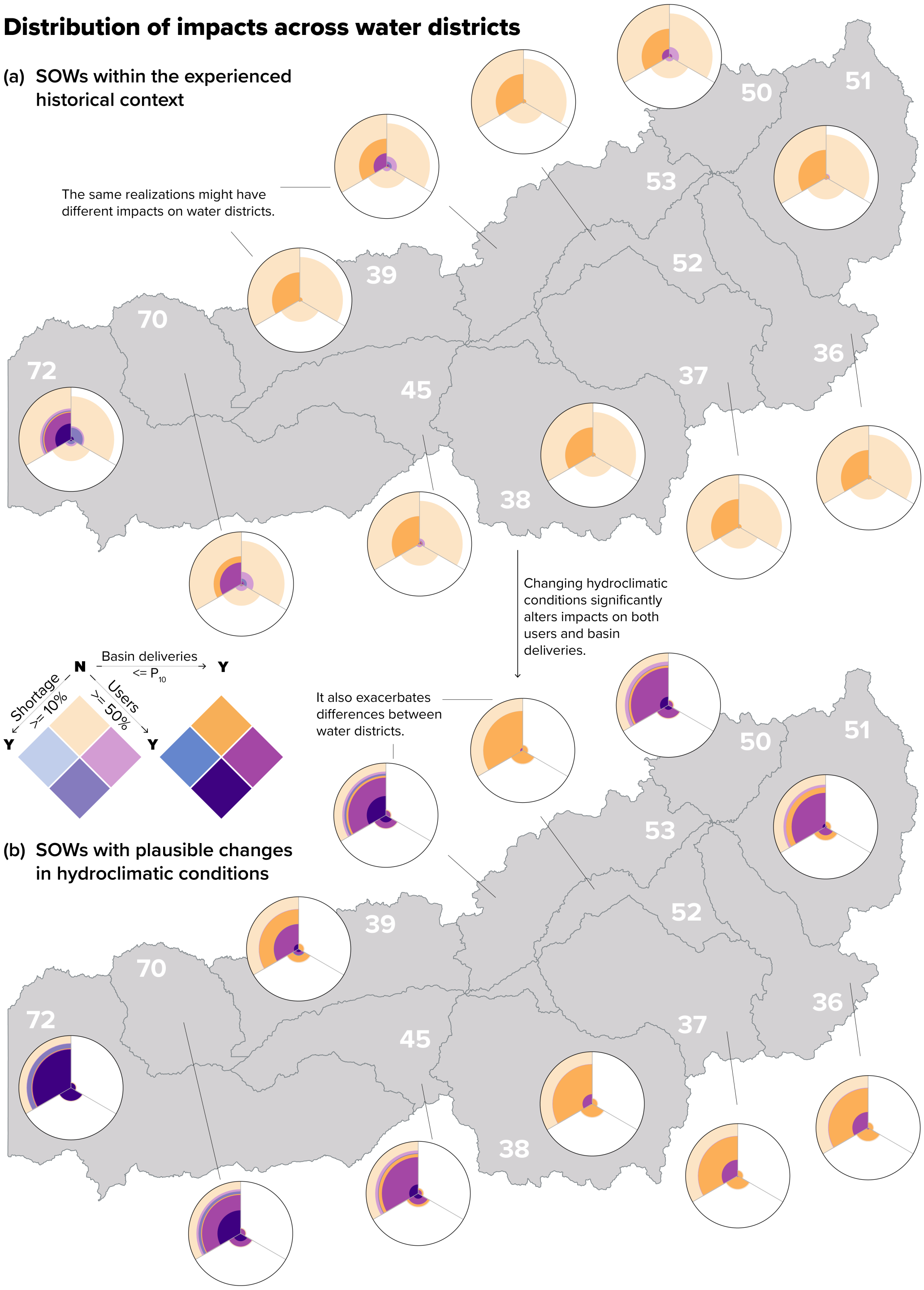

Figure 11. Impact classification for all states of the world (SOWs) as organized by combinations of dynamic properties and calculated for individual water districts. (a) Impacts for SOWs that exhibit dynamic properties within the bounds of the observed past (105 years of gauged streamflow); (b) Impacts for SOWs that exhibit dynamic properties outside the bounds of the observed past (informed by the paleo record and future projections). In both cases, water districts might individually exhibit more severe or less severe impacts than those calculated for the basin in aggregate (shown in Fig. 8.)

It is not entirely unexpected that the same SOWs might have different impacts on the WDs of the UCRB. For example, for the historically-informed SOWs (Fig. 11 (a)), we see that some WDs (36-39, and 52) see no impacts on their users—all bands in the hive plot are shades of yellow. This is better than the basin-wide average conditions shown in Fig. 8 (a). At the same time, some WDs (70 and 72) see their users much more significantly impacted than the basin-level average user of the UCRB, with some historically-informed SOWs producing both larger shortages and for more users (bands in dark purple ◆). SOWs that are outside the historical hydroclimatic context (Fig. 11 (b)) further amplify these differences. For example, users in WD 52 are largely unaffected by all the sets of SOWs, whereas the majority of changing-context SOWs affect both the mean shortages and the number of users affected in WD 72 (dark purple bands). In fact, all other WDs either see their users unaffected by most SOWs with changing hydroclimatic conditions (e.g., WDs 36-39, and 52, which have yellow ◆ as the largest band color) or see only an increase in the number of users affected but not in the mean water shortage (e.g., WDs 45, 50, 51, and 70, which have violet ◆ as the largest band color). This difference in WD experiences is the result of several complex interactions between the number and seniority of rights in each WD, their diversion locations and sources (e.g., the mainstem as opposed to a tributary), and the timing of their demands. These results emphasize that understanding and selecting narrative storylines is critical to capture the natural hydroclimatic drought hazards and their locally consequential impacts as manifested through the UCRB's infrastructure and water governance institutions (i.e., water rights in prior appropriation).

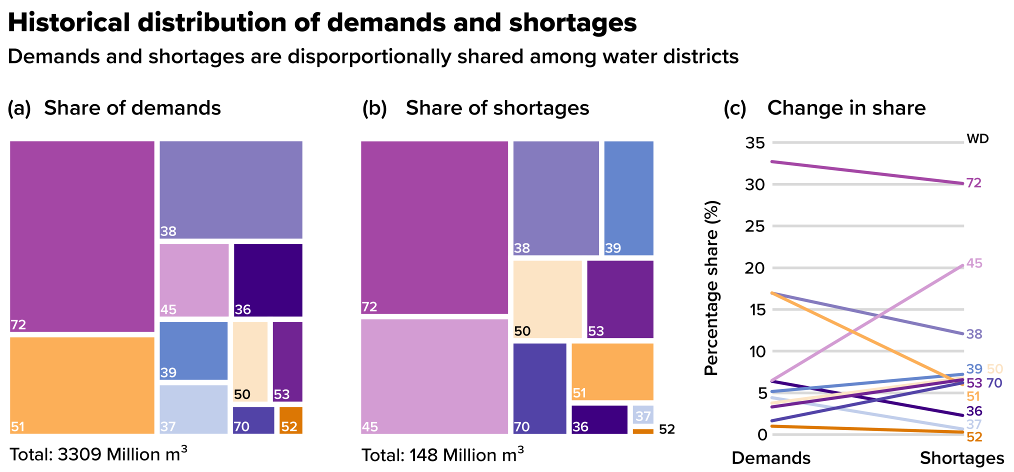

Figure 12. Historical distribution of demands and shortages among water districts. (a-b) Treemaps of (a) the share of water demands as contributed by each water district; and (b) the share of water shortages as contributed by each water district. The treemaps are organized with the largest contributing parts placed at the top left moving first downward and then rightward. (c) Change in relative share between the demands and shortages of each water district.

Specifically, WD 72, which appears to experience the most severe impacts, makes up approximately 33% of all water demands in the UCRB historically, far exceeding the second and third largest demands at 17% by WDs 38 and 51 (Fig. 12 (a)). Compared to the historical data on UCRB shortages (i.e., without any of our sampled hydroclimatic changes imposed on the system), WD 72 indeed represents the largest volumetric share of water shortages in the UCRB (Fig. 12 (b-c)), but their shortages are only 4% of their demands (Fig. 13 (b)), which is below the historical 7% average estimated basin-wide. Indeed, total demand does not explain these impacts on its own (i.e., that the biggest shortages are experienced where the biggest demands are). WD 70, for example, only makes up 1% of the total demands in the basin, yet also sees impacts for its water users that exceed the average (i.e., more violet and purple bands; Fig. 11 (a)), and in the historic observations it exhibits the highest relative ratio of shortages to demands (approximately 16%; Fig. 13 (b)). The historical data also highlights that in general, higher shortages are not necessarily the direct outcome of higher demands (Fig. 12), as some WDs with relatively lower demands experience relatively higher shortages than other WDs (e.g., WD 45), and vice versa (e.g., WD 51). Readers familiar with the region might posit that this difference in impacts can simply be attributed to the number and seniority of rights owned by water users in WD 72; maybe rights in that WD are simply more junior so their demands are not met as much more senior rights in other WDs? Looking at the number of water rights, WD 72 has the same number of actively served consumptive use water rights as WD 38 (296; we note that each water user might own multiple), and its rights are decreed generally larger volumes of water with more senior right ranks on average than WD 38 (Fig. 13 (a)). The differences in impacts can therefore potentially be attributed to the fact that WD 72 (and others) are home to several more junior rights with larger decrees, but it is clear that single factor drivers cannot explain the differences seen.

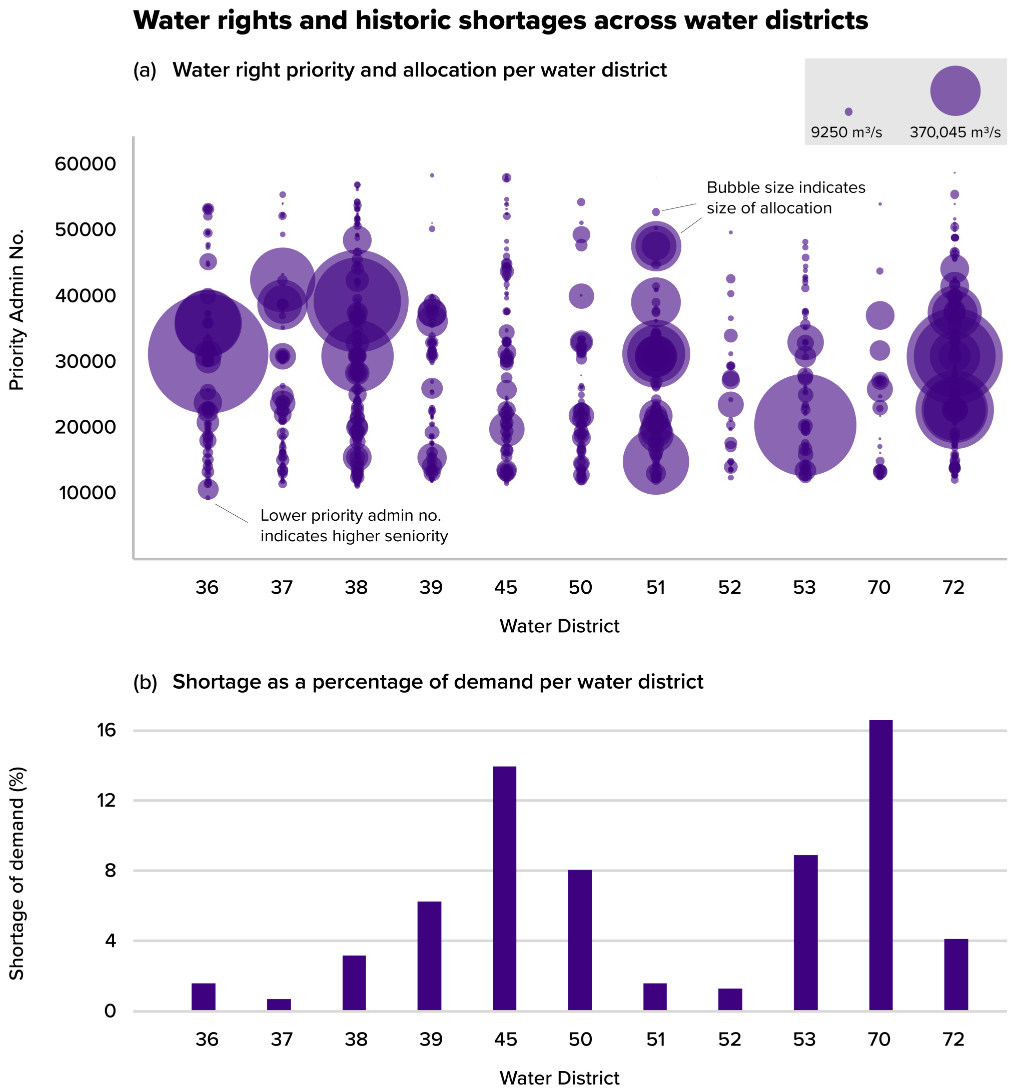

Figure 13. Priority and water allocation per right for each water district. Rights are organized per water district along the horizontal axis and per priority admin number along the vertical axis. Lower priority admin number indicates higher right seniority. Larger bubble size indicates larger water allocation.

4.3 Exploring alternative impact thresholds

Lastly, recognizing the diverse interests represented in the UCRB, we examine more closely how the hierarchical basin-level impact classifications in Fig. 8 are shaped by the assumed problem framing and the impact classification thresholds chosen for basin deliveries downstream, percent of users shorted, and mean shortage (Eqs. 13 - 15). In other words, we would like to know how the classification of these SOWs might change if different shortage risk tolerances were assumed, reflective of the diverse impacts experienced and the different decision-making concerns present in the UCRB (Fig. 6). So in line with the discussion of narrative scenario discovery for multi-actor, multi-sector systems, we repeat the impact classification across different values of each impact threshold (Fig. 14). Specifically, for impact set A containing SOWs that exceed a mean shortage threshold th_{χ}, we use three values of this threshold (5%, 7%, and 10%) and apply them to Eq. 13 to estimate how many SOWs cause the mean shortages in the basin to be above 5%, 7%, and 10% of demand, respectively. Impact set B contains SOWs with their 10^{th} percentile of basin deliveries downstream falling below a critical threshold th_{bd}. In the prior results, we defined th_{bd} using the historical 10^{th} percentile of cumulative deliveries, so B contained SOWs where the basin is delivering less than its historical 10% worst years. Switching th_{bd} to the historical 5^{th} percentile, then B contains SOWs whose low-delivery years are twice as frequent as history. As a result, we are checking if an event that occurred only 5% of the time historically now occurs 10% of the time, in essence doubling its occurrence in the SOWs that meet this criterion. Equivalently, if the threshold used is the historical 1^{st} percentile, then the SOWs in set B have low-delivery years ten times more frequently than history. The 10^{th}, 5^{th}, and 1^{st} percentiles of cumulative 10-year flows are 46,820, 44,896, and 43,776 M m^3, respectively. Lastly, impact set C is the set of all SOWs where more than th_{ψ} of the basin's users are experiencing a shortage. We vary this threshold to 25%, 50%, and 75% to capture SOWs that affect increasing numbers of water users in the basin.

Fig. 14 shows the resulting hive plots for all three thresholds for all three criteria, for the SOWs in the changing hydroclimatic context. This style of small multiples figure allows us to quickly compare the different plots and look for patterns in the matrix of visuals. The following pattern emerges here. Starting at the top left, the hive plot shows the impact classification of all SOWs using the most lenient performance criteria for each impact group (i.e., low basin deliveries occurring as much as history on the vertical axis, mean shortage levels above or equal to 5% of demands on the horizontal axis, and 25% or more users experiencing a shortage along the diagonal axis). Given that these are the most lenient thresholds, they are the easiest criteria to meet, and therefore the majority of SOWs do so (shown in dark purple ◆).

Moving to the right along the horizontal axis, we are increasing the shortage threshold as a percentage of demand so we expect to see fewer blue and purple bands, as fewer SOWs would be classified as causing the larger shortages to water users. Indeed, what we see is a shift from dark purple to a larger lilac ◆ band in the top right hive plot. Moving from the top down, we expect to see some of the darker shade classifications turn to lighter colors, as the lower basin deliveries classification is a more extreme condition to meet. Comparing along the three hive plots at the very right, we can indeed see a small number of yellow ◆ SOWs turn to light yellow ◆. Finally, moving along the diagonal axis, we are increasing the number of affected users we consider as consequential. In this case, we should expect fewer violet ◆ and purple bands ◆ as we move diagonally to lower right. This is prominently apparent for the three hive plots at the top right of the figure, where using the 25% threshold, most SOWs are classified as having both more users affected and lower basin deliveries (in violet), but using the 75% threshold, the classifications are largely yellow (only lower basin deliveries).

Even with the more extreme threshold combinations (bottom right hive plot in Fig. 14) most SOWs in the changing context meet at least one of the criteria. Most meet the lower downstream deliveries criterion (yellow band ◆), that their 10^{th} percentile of cumulative 10-year flows fall below the historical 1^{st} percentile (i.e., that low deliveries are occurring ten times as often in these SOWs). Some other SOWs are shown in blue ◆, so they also increase the mean shortage to the basins users to above 10%. We can also compare this hive plot with the one directly to its upper

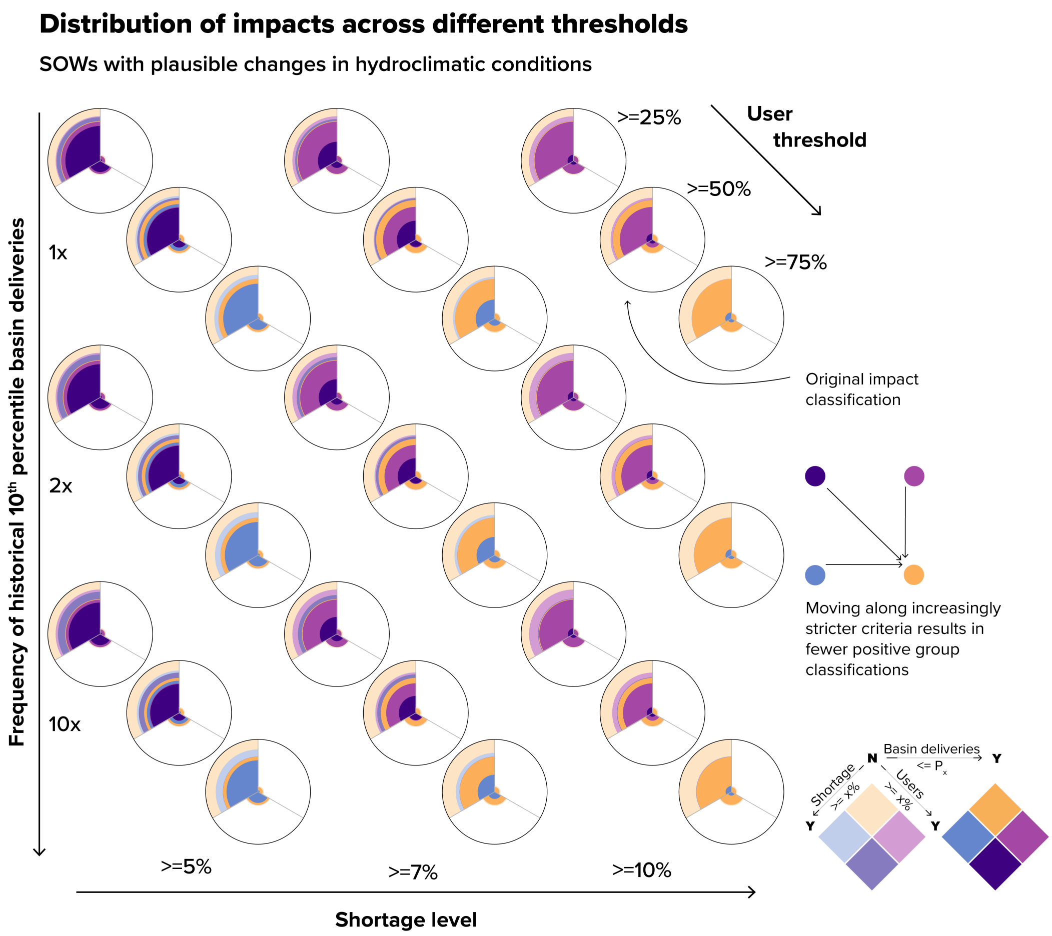

Figure 14. Impact classification for all states of the world as calculated for different thresholds for each impact category. The figure is oriented such the going from the top left to the bottom right, we are moving from more lenient to increasingly stricter criteria.

left, reflecting a change to the user criterion from 75% to 50%, to see that several of the SOWs considered here do affect more than 50% of users in the UCRB (violet and dark purple bands in upper left hive plot) but not more than 75% (same bands disappear when we look back to the lower right hive plot). This shows that even though there might not be a significant increase in the average shortage compared to history (increase from 7% of users to 10%), there is a significant increase in the number of users affected (from 30% historically to above 50%). This further supports the explanation given with regard to the impacts of "The Unforeseen Struggles" storyline (Fig. 10): that they are the result of a larger number of affected users and not necessarily (or only) larger shortages.

Exploring alternative threshold combinations aids with providing an informative feedback to Stage I Framing (Section 3.2) of the FRNSIC assessment of the UCRB, allowing us to address several of the challenges generated by complex human-natural systems more broadly. Namely, as discussed in Section 1, using a small set of scenarios that are considered *a priori* to be "relevant" by the analysts might inadvertently create a very narrow view of what the relevant stakeholder concerns are that is not salient with the diverse views that might exist on the system (Groves & Lempert, 2007). Because each alternative threshold illuminates different SOWs, it allows us to switch to alternative sets of consequential scenarios to focus on, depending on the outcomes they generate. For instance, planners might want to select scenarios from the dark purple SOWs (ones that have impacts across all groups) for further investigation and analysis. The SOWs that fall in these dark purple bands change depending on the thresholds used, so these consequential scenarios can reflect not only varying impact severities, but also different attitudes toward these impacts.

This relates to another complication discussed already, that in systems with many actors making decisions at different scales (Fig. 6), it is difficult to capture their differing priorities, goals and risk aversions with a singular impact metric or threshold imposed on it. We know from prior work (Hadjimichael, Quinn, Wilson, et al., 2020; Quinn et al., 2020), historical estimates (Fig. 12), and also the results here (Fig. 11) that the same conditions imposed on the system can result in diverse impacts for its users. This means that for an SOW with average shortages of 10%, some users or WDs experience shortages lower or higher than that. It follows that some stakeholders in the basin might be more or less conservative about this threshold choice, and the impacts of that change in choice are reflected by moving horizontally in Fig. 14. As a last related point here, in Section 1 we have highlighted recommendations from co-production literature on relating new findings to past experiences as a way to help connect scientific outcomes to stakeholders' analytical and experiential processing (Lemos et al., 2012). Alternative thresholds, especially for the user-level impacts we explore here, can therefore help produce locally-meaningful narratives as they relate the water shortages users and WDs have experienced in the past.

5 Conclusions and Future Work

This paper proposes the FRamework for Narrative Scenarios and Impact Classification (FRNSIC), that enables narrative scenario discovery for multiple states and multiple impacts. The introduced framework is designed to overcome common challenges of scenario discovery with regard to establishing stakeholder-relevant narratives. FRNSIC combines the classification of dynamic behavioral properties of each SOW as well as its impact states in a nested scheme to facilitate hierarchical storyline selection, and produce locally-meaningful narratives from high-dimensional exploratory ensembles. We use a hypothetical planning context—examining the UCRB's potential futures and needing to discover consequential drought storylines to use in planing—and apply FRNSIC to demonstrate its capabilities in a system with multiple actors and institutional complexity. We show that FRNSIC can illuminate the critical dynamic pathways that lead to consequential impacts, by combining a SOW's temporal behavioral properties and the aggregated impacts it results in. The framework therefore addresses several prominent challenges other state-of-the-art scenario discovery frameworks face when applied to complex human-natural systems, and especially institutionally complex systems with many actors like the UCRB.

In applying FRNSIC, several choices must be made on the classification scheme to use (the criteria to use to classify dynamics and impacts, the threshold values to apply, other aggregation choices). This is akin to other scenario discovery applications where consequential or decision-relevant conditions need to be identified, and such choices need to be made transparent from the problem framing stage and throughout the analysis process, as well as reexamined as needed. For example, in the UCRB case study we explore the implications of these choices using gradients of threshold values applied to our criteria. In future work, similar threshold analyses can be applied to the thresholds used to identify the sets of dynamic behaviors exhibited in our ensemble. Changing the criteria through which the dynamics are classified could reflect alternative dynamic behaviors of interest. For example, one could focus on specifically the occurrence of multi-decadal droughts of over 35 years, and this would affect the sizes of the dynamic sets, as well as subsequent results.

The narrative drought storylines produced by FRNSIC can also be utilized in future work in the basin, for example to examine the capacity of adaptive action in modulating the impacts of the drought events seen in each storyline. Specifically, the ensemble of SOWs explored here can be combined with hypothesized policy interventions (e.g., for water conservation) to investigate how said interventions would affect the impacts the basin experiences under each storyline. Just like narrative scenarios and storylines are used in co-production literature, the drought storylines here can also be used in negotiation or stakeholder solicitation contexts to contrast the impacts that WDs or users may potentially experience in the future.

6 Open Research

StateMod is freely available on GitHub `https://github.com/OpenCDSS`. The input files to run StateMod for the UCRB can be found at the CDSS website `https://cdss.colorado.gov/modeling-data/surface-water-statemod`. All the scripts to replicate the analysis performed in this paper and to regenerate all figures can be found at `https://github.com/antonia-had/Hadjimichael-etal_2023_EarthsFuture`. All the output data used in this analysis can be found at `https://doi.org/10.57931/2205512`.

Acknowledgments
This research was supported by the U.S. Department of Energy, Office of Science, as part of research in MultiSector Dynamics, Earth and Environmental System Modeling Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding entities.

References

Aghabozorgi, S., Seyed Shirkhorshidi, A., & Ying Wah, T. (2015, October). Time-series clustering – A decade review. *Information Systems*, *53*, 16–38. Retrieved 2023-05-15, from `https://www.sciencedirect.com/science/article/pii/S0306437915000733` doi: 10.1016/j.is.2015.04.007

AghaKouchak, A., Huning, L. S., Sadegh, M., Qin, Y., Markonis, Y., Vahedifard, F., ... Kreibich, H. (2023, August). Toward impact-based monitoring of drought and its cascading hazards. *Nature Reviews Earth & Environment*, *4*(8), 582–595. Retrieved 2023-08-08, from `https://www.nature.com/articles/s43017-023-00457-2` (Number: 8 Publisher: Nature Publishing Group) doi: 10.1038/s43017-023-00457-2

AghaKouchak, A., Pan, B., Mazdiyasni, O., Sadegh, M., Jiwa, S., Zhang, W., ... Sorooshian, S. (2022, October). Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *380*(2238), 20210288. Retrieved 2023-01-11, from `https://royalsocietypublishing.org/doi/full/10.1098/rsta.2021.0288` (Publisher: Royal Society) doi: 10.1098/rsta.2021.0288

Arizona Department of Water Resources. (2022). *Arizona Drought Preparedness Plan: 2022 Annual Report* (Tech. Rep.). Retrieved 2023-02-09, from `https://new.azwater.gov/sites/default/files/media/2022ADPAR_0.pdf`

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., & Meko, D. M. (2014, January). Assessing the Risk of Persistent Drought Using Climate Model Simulations and Paleoclimate Data. *Journal of Climate*, *27*(20), 7529–7549. Retrieved 2020-04-28, from `https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-12-00282.1` (Publisher: American Meteorological Society) doi: 10.1175/JCLI-D-12-00282.1

Ault, T. R., Mankin, J. S., Cook, B. I., & Smerdon, J. E. (2016, October). Relative impacts of mitigation, temperature, and precipitation on 21st-century megadrought risk in the American Southwest. *Science Advances*, *2*(10), e1600873. Retrieved 2020-04-28, from `https://advances.sciencemag.org/content/2/10/e1600873` (Publisher: American Association for the Advancement of Science Section: Research Article) doi: 10.1126/sciadv.1600873

Bankes, S. C. (1993). Exploratory Modeling for Policy Analysis. *Operations Research*, *41*(3), 435–449. doi: 10.1287/opre.41.3.435

Ben-Haim, Y. (2006). *Info-gap decision theory: decisions under severe uncertainty*. Else-

vier.
Berghuijs, W. R., Allen, S. T., Harrigan, S., & Kirchner, J. W. (2019). Growing spatial scales of synchronous river flooding in Europe. *Geophysical Research Letters*, *46*(3), 1423–1428. (Publisher: Wiley Online Library)
Beven, K. (1993, January). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, *16*(1), 41–51. Retrieved 2020-11-10, from http://www.sciencedirect.com/science/article/pii/030917089390028E doi: 10.1016/0309-1708(93)90028-E
Bonham, N., Kasprzyk, J., & Zagona, E. (2022, November). post-MORDM: Mapping policies to synthesize optimization and robustness results for decision-maker compromise. *Environmental Modelling & Software*, *157*, 105491. Retrieved 2022-09-26, from https://www.sciencedirect.com/science/article/pii/S1364815222001943 doi: 10.1016/j.envsoft.2022.105491
Bracken, C., Rajagopalan, B., & Woodhouse, C. (2016). A Bayesian hierarchical nonhomogeneous hidden Markov model for multisite streamflow reconstructions. *Water Resources Research*, *52*(10), 7837–7850. Retrieved 2023-03-21, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018887 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR018887) doi: 10.1002/2016WR018887
Bracken, C., Rajagopalan, B., & Zagona, E. (2014). A hidden Markov model combined with climate indices for multidecadal streamflow simulation. *Water Resources Research*, *50*(10), 7836–7846. Retrieved 2019-03-26, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015567 doi: 10.1002/2014WR015567
Breiman, L. (1984). *Classification and Regression Trees*. New York: Routledge. Retrieved from https://doi.org/10.1201/9781315139470 (Google-Books-ID: mlZgQgAACAAJ)
Brown, C., Ghile, Y., Laverty, M., & Li, K. (2012). Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. *Water Resources Research*, *48*(9). Retrieved 2019-10-28, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011212 doi: 10.1029/2011WR011212
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. doi: 10.1016/j.techfore.2009.08.002
Bureau of Reclamation. (2012). *Colorado River Basin Water Supply and Demand Study* (Executive Summary). Department of Interior.
California Natural Resources Agency. (2022, August). *California's Water Supply Strategy: Adapting to a Hotter, Drier Future* (Tech. Rep.).
Calvo, L., Christel, I., Terrado, M., Cucchietti, F., & Pérez-Montoro, M. (2022, January). Users' Cognitive Load: A Key Aspect to Successfully Communicate Visual Climate Information. *Bulletin of the American Meteorological Society*, *103*(1), E1–E16. Retrieved 2023-02-17, from https://journals.ametsoc.org/view/journals/bams/103/1/BAMS-D-20-0166.1.xml (Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society) doi: 10.1175/BAMS-D-20-0166.1
Cohen, S. M., Dyreson, A., Turner, S., Tidwell, V., Voisin, N., & Miara, A. (2022, July). A multi-model framework for assessing long- and short-term climate influences on the electric grid. *Applied Energy*, *317*, 119193. Retrieved 2023-01-03, from https://www.sciencedirect.com/science/article/pii/S030626192200561X doi: 10.1016/j.apenergy.2022.119193
Colorado Water Conservation Board, & Department of Natural Resources. (2018). *The Colorado Drought Mitigation and Response Plan* (Tech. Rep.).
Cook, B. I., Smerdon, J. E., Cook, E. R., Williams, A. P., Anchukaitis, K. J., Mankin, J. S., . . . Wise, E. K. (2022, October). Megadroughts in the Common Era and the Anthropocene. *Nature Reviews Earth & Environment*, 1–17. Retrieved 2022-10-04, from

https://www.nature.com/articles/s43017-022-00329-1 (Publisher: Nature Publishing Group) doi: 10.1038/s43017-022-00329-1

Cork, S. J., Peterson, G. D., Bennett, E. M., Petschel-Held, G., & Zurek, M. (2006). Synthesis of the storylines. *Ecology and Society*, *11*(2), 11. Retrieved 2014-01-24, from http://www.ibcperu.org/doc/isis/10071.pdf

CWCB. (2012). *Colorado River Water Availability Study Phase I Report* (Tech. Rep.). Colorado Water Conservation Board.

CWCB, & CDWR. (2016). *Upper Colorado River Basin Water Resources Planning Model User's Manual* (Tech. Rep.). Colorado Water Conservation Board and Colorado Division of Water Resources. Retrieved 2019-10-02, from https://www.colorado.gov/pacific/cdss/modeling-dataset-documentation

CWCB, & CDWR. (2022). *Colorado Basin Implementation Plan* (Tech. Rep.).

de Ruiter, M. C., & Van Loon, A. F. (2022, July). The challenges of dynamic vulnerability and how to assess it. *iScience*, 104720. Retrieved 2022-07-05, from https://www.sciencedirect.com/science/article/pii/S2589004222009920 doi: 10.1016/j.isci.2022.104720

Deser, C., Terray, L., & Phillips, A. S. (2016, March). Forced and Internal Components of Winter Air Temperature Trends over North America during the past 50 Years: Mechanisms and Implications. *Journal of Climate*, *29*(6), 2237–2258. Retrieved 2023-01-11, from https://journals.ametsoc.org/view/journals/clim/29/6/jcli-d-15-0304.1.xml (Publisher: American Meteorological Society Section: Journal of Climate) doi: 10.1175/JCLI-D-15-0304.1

Diffenbaugh, N. S., Swain, D. L., & Touma, D. (2015). Anthropogenic warming has increased drought risk in California. *Proceedings of the National Academy of Sciences*, *112*(13), 3931–3936. Retrieved from https://www.pnas.org/content/pnas/112/13/3931.full.pdf (Type: Journal Article)

Drapeau, S., Jamneshan, A., Karliczek, M., & Kupper, M. (2016, May). The algebra of conditional sets and the concepts of conditional topology and compactness. *Journal of Mathematical Analysis and Applications*, *437*(1), 561–589. Retrieved 2023-03-28, from https://www.sciencedirect.com/science/article/pii/S0022247X15011300 doi: 10.1016/j.jmaa.2015.11.057

Elsawah, S., Filatova, T., Jakeman, A. J., Kettner, A. J., Zellner, M. L., Athanasiadis, I. N., ... Lade, S. J. (2020, January). Eight grand challenges in socio-environmental systems modeling. *Socio-Environmental Systems Modelling*, *2*, 16226–16226. Retrieved 2020-08-24, from https://sesmo.org/article/view/16226 doi: 10.18174/sesmo.2020a16226

Engle, S., & Whalen, S. (2012, October). Visualizing distributed memory computations with hive plots. In *Proceedings of the Ninth International Symposium on Visualization for Cyber Security* (pp. 56–63). Seattle Washington USA: ACM. Retrieved 2023-06-01, from https://dl.acm.org/doi/10.1145/2379690.2379698 doi: 10.1145/2379690.2379698

Fischer, E. M., Sippel, S., & Knutti, R. (2021, August). Increasing probability of record-shattering climate extremes. *Nature Climate Change*, *11*(8), 689–695. Retrieved 2023-06-01, from https://www.nature.com/articles/s41558-021-01092-9 (Number: 8 Publisher: Nature Publishing Group) doi: 10.1038/s41558-021-01092-9

Flavelle, C., & Rojanasakul, M. (2023, January). As the Colorado River Shrinks, Washington Prepares to Spread the Pain. *The New York Times*. Retrieved 2023-02-09, from https://www.nytimes.com/2023/01/27/climate/colorado-river-biden-cuts.html

Fletcher, S., Hadjimichael, A., Quinn, J., Osman, K., Giuliani, M., Gold, D., ... Gordon, B. (2022, July). Equity in Water Resources Planning: A Path Forward for Decision Support Modelers. *Journal of Water Resources Planning and Management*, *148*(7), 02522005. Retrieved 2022-05-02, from https://ascelibrary.org/doi/full/10.1061/%28ASCE%29WR.1943-5452.0001573 (Publisher: American Society of Civil Engineers) doi: 10.1061/(ASCE)WR.1943-5452.0001573

Franssen, M. (2005, November). Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. *Research in Engineering Design*, *16*(1), 42–56. Retrieved 2019-07-02, from https://doi.org/10.1007/s00163-004-0057-5 doi: 10.1007/s00163-004-0057-5

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, *9*(2), 123–143. doi: doi.org/10.1023/A:1008894516817

Gerlak, A. K., & Heikkila, T. (2023, February). *Navigating the Colorado River crisis: It's time for the federal government to step up* [Text]. Retrieved 2023-02-09, from https://thehill.com/opinion/energy-environment/3847785-navigating-the-colorado-river-crisis-its-time-for-the-federal-government-to-step-up/

Gold, D. F., Reed, P. M., Gorelick, D. E., & Characklis, G. W. (2022). Power and Pathways: Exploring Robustness, Cooperative Stability, and Power Relationships in Regional Infrastructure Investment and Water Supply Management Portfolio Pathways. *Earth's Future*, *10*(2), e2021EF002472. Retrieved 2023-03-21, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002472 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021EF002472) doi: 10.1029/2021EF002472

Gold, D. F., Reed, P. M., Trindade, B. C., & Characklis, G. W. (2019). Identifying Actionable Compromises: Navigating Multi-City Robustness Conflicts to Discover Cooperative Safe Operating Spaces for Regional Water Supply Portfolios. *Water Resources Research*, *n/a*(n/a). Retrieved 2019-12-03, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019WR025462 doi: 10.1029/2019WR025462

Gotts, N. M., Van Voorn, G. A., Polhill, J. G., Jong, E. D., Edmonds, B., Hofstede, G. J., & Meyer, R. (2019, December). Agent-based modelling of socio-ecological systems: Models, projects and ontologies. *Ecological Complexity*, *40*, 100728. Retrieved 2023-05-15, from https://linkinghub.elsevier.com/retrieve/pii/S1476945X18301272 doi: 10.1016/j.ecocom.2018.07.007

Groves, D. G. (2005). *New methods for identifying robust long-term water resources management strategies for California* (Doctoral dissertation, The Pardee RAND Graduate School). Retrieved from https://doi.org/10.7249/RGSD196

Groves, D. G., & Lempert, R. J. (2007). A new analytic method for finding policy-relevant scenarios. *Global Environmental Change*, *17*(1), 73–85. doi: 10.1016/j.gloenvcha.2006.11.006

Guivarch, C., Rozenberg, J., & Schweizer, V. (2016, June). The diversity of socio-economic pathways and CO2 emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling & Software*, *80*, 336–353. Retrieved 2023-01-25, from https://www.sciencedirect.com/science/article/pii/S1364815216300706 doi: 10.1016/j.envsoft.2016.03.006

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, *23*, 485–498. Retrieved 2016-05-10, from http://dx.doi.org/10.1016/j.gloenvcha.2012.12.006 doi: http://dx.doi.org/10.1016/j.gloenvcha.2012.12.006

Hadjimichael, A., Quinn, J., & Reed, P. (2020). Advancing Diagnostic Model Evaluation to Better Understand Water Shortage Mechanisms in Institutionally Complex River Basins. *Water Resources Research*, *56*(10), e2020WR028079. Retrieved 2020-10-16, from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR028079 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020WR028079) doi: 10.1029/2020WR028079

Hadjimichael, A., Quinn, J., Wilson, E., Reed, P., Basdekas, L., Yates, D., & Garrison, M. (2020). Defining Robustness, Vulnerabilities, and Consequential Scenarios for Diverse Stakeholder Interests in Institutionally Complex River Basins.

Earth's Future, *8*(7), e2020EF001503. Retrieved 2020-07-13, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001503 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001503) doi: 10.1029/2020EF001503

Hadjimichael, A., Reed, P. M., & Quinn, J. D. (2020). Navigating Deeply Uncertain Tradeoffs in Harvested Predator-Prey Systems. *Complexity*, *2020*, e4170453. Retrieved 2020-03-03, from https://www.hindawi.com/journals/complexity/2020/4170453/ (Publisher: Hindawi) doi: https://doi.org/10.1155/2020/4170453

Hadjimichael, A., Yoon, J., Reed, P., Voisin, N., & Xu, W. (2023, February). Exploring the Consistency of Water Scarcity Inferences between Large-Scale Hydrologic and Node-Based Water System Model Representations of the Upper Colorado River Basin. *Journal of Water Resources Planning and Management*, *149*(2), 04022081. Retrieved 2022-12-08, from https://ascelibrary.org/doi/10.1061/JWRMD5.WRENG-5522 (Publisher: American Society of Civil Engineers) doi: 10.1061/JWRMD5.WRENG-5522

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. Retrieved from https://atoc.colorado.edu/~whan/ATOC4800_5000/Materials/Hawkins_sutton.pdf (Type: Journal Article)

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. doi: 10.1061/(ASCE)WR.1943-5452.0000509

Herman, J. D., Zeff, H. B., Reed, P. M., & Characklis, G. W. (2014, October). Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, *50*(10), 7692–7713. Retrieved 2017-11-29, from http://onlinelibrary.wiley.com/doi/10.1002/2014WR015338/abstract doi: 10.1002/2014WR015338

Hobbins, R., Muñoz-Erickson, T. A., & Miller, C. (2021). Producing and Communicating Flood Risk: A Knowledge System Analysis of FEMA Flood Maps in New York City. In Z. A. Hamstead, D. M. Iwaniec, T. McPhearson, M. Berbés-Blázquez, E. M. Cook, & T. A. Muñoz-Erickson (Eds.), *Resilient Urban Futures* (pp. 67–84). Cham: Springer International Publishing. Retrieved 2023-06-01, from https://doi.org/10.1007/978-3-030-63131-4_5 doi: 10.1007/978-3-030-63131-4_5

Hoylman, Z. H., Bocinsky, R. K., & Jencso, K. G. (2022, May). Drought assessment has been outpaced by climate change: empirical arguments for a paradigm shift. *Nature Communications*, *13*(1), 2715. Retrieved 2022-07-29, from https://www.nature.com/articles/s41467-022-30316-5 (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/s41467-022-30316-5

Inselberg, A. (2009). *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. New York, NY: Springer. Retrieved 2023-08-31, from https://link.springer.com/10.1007/978-0-387-68628-8 doi: 10.1007/978-0-387-68628-8

IPCC. (2023). *Climate Change 2023: Synthesis Report. A Report of the Intergovernmental Panel on Climate Change. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Tech. Rep.). Geneva, Switzerland: IPCC.

Iwanaga, T., Wang, H.-H., Hamilton, S. H., Grimm, V., Koralewski, T. E., Salado, A., . . . Little, J. C. (2021). Socio-technical scales in socio-environmental modeling: Managing a system-of-systems modeling approach. *Environmental Modelling & Software*, *135*, 104885. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7537632/pdf/main.pdf (Type: Journal Article) doi: https://doi.org/10.1016/j.envsoft.2020.104885

Kasprzyk, J. R., Nataraj, S., Reed, P. M., & Lempert, R. J. (2013, April). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling & Software*, *42*, 55–71. Retrieved 2019-09-30, from http://

www.sciencedirect.com/science/article/pii/S1364815212003131 doi: 10.1016/j.envsoft.2012.12.007

Kenney, D. S. (2005). Prior appropriation and water rights reform in the western United States. In B. R. Bruns, C. Ringler, & R. S. Meinzen-Dick (Eds.), *Water Rights Reform: Lessons for Institutional Design* (p. 336). International Food Policy Research Institute.

Krauß, W. (2020, January). Narratives of change and the co-development of climate services for action. *Climate Risk Management*, *28*, 100217. Retrieved 2023-02-16, from https://www.sciencedirect.com/science/article/pii/S2212096320300073 doi: 10.1016/j.crm.2020.100217

Krauß, W., & Bremer, S. (2020, January). The role of place-based narratives of change in climate risk governance. *Climate Risk Management*, *28*, 100221. Retrieved 2023-02-16, from https://www.sciencedirect.com/science/article/pii/S2212096320300115 doi: 10.1016/j.crm.2020.100221

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., ... Di Baldassarre, G. (2022, August). The challenge of unprecedented floods and droughts in risk management. *Nature*, *608*(7921), 1–7. Retrieved 2022-08-04, from https://www.nature.com/articles/s41586-022-04917-5 (Publisher: Nature Publishing Group) doi: 10.1038/s41586-022-04917-5

Krzywinski, M., Birol, I., Jones, S. J., & Marra, M. A. (2012, September). Hive plots—rational approach to visualizing networks. *Briefings in Bioinformatics*, *13*(5), 627–644. Retrieved 2023-03-29, from https://doi.org/10.1093/bib/bbr069 doi: 10.1093/bib/bbr069

Kwakkel, J. H. (2019). A generalized many-objective optimization approach for scenario discovery. *FUTURES & FORESIGHT SCIENCE*, *1*(2), e8. Retrieved 2019-12-04, from https://onlinelibrary.wiley.com/doi/abs/10.1002/ffo2.8 doi: 10.1002/ffo2.8

Kwakkel, J. H., & Haasnoot, M. (2019). Supporting DMDU: A taxonomy of approaches and tools. In V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, & S. W. Popper (Eds.), *Decision Making under Deep Uncertainty.* Springer.

Lehner, F., & Deser, C. (2023, May). Origin, importance, and predictive limits of internal climate variability. *Environmental Research: Climate*, *2*(2), 023001. Retrieved 2023-06-01, from https://dx.doi.org/10.1088/2752-5295/accf30 (Publisher: IOP Publishing) doi: 10.1088/2752-5295/accf30

Lemos, M. C., Kirchhoff, C. J., & Ramprasad, V. (2012, November). Narrowing the climate information usability gap. *Nature Climate Change*, *2*(11), 789–794. Retrieved 2022-09-29, from https://www.nature.com/articles/nclimate1614 (Number: 11 Publisher: Nature Publishing Group) doi: 10.1038/nclimate1614

Lemos, M. C., & Morehouse, B. J. (2005, April). The co-production of science and policy in integrated climate assessments. *Global Environmental Change*, *15*(1), 57–68. Retrieved 2021-08-12, from https://www.sciencedirect.com/science/article/pii/S0959378004000652 doi: 10.1016/j.gloenvcha.2004.09.004

Lempert, R. J. (2019). Robust Decision Making (RDM). In V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, & S. W. Popper (Eds.), *Decision Making under Deep Uncertainty: From Theory to Practice* (pp. 23–51). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-05252-2_2 doi: 10.1007/978-3-030-05252-2_2

Lempert, R. J., & Groves, D. G. (2010, July). Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the American west. *Technological Forecasting and Social Change*, *77*(6), 960–974. Retrieved 2023-05-13, from https://www.sciencedirect.com/science/article/pii/S0040162510000740 doi: 10.1016/j.techfore.2010.04.007

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006, April). A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios. *Management Science*, *52*(4), 514–528. Retrieved 2022-09-29, from https://pubsonline.informs.org/doi/10.1287/mnsc.1050.0472 (Publisher: INFORMS) doi: 10.1287/mnsc.1050.0472

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the Next One Hundred Years*. RAND Corporation. Retrieved 2017-09-14, from https://www.rand.org/pubs/monograph_reports/MR1626.html

Lorenz, R., Stalhandske, Z., & Fischer, E. M. (2019). Detection of a climate change signal in extreme heat, heat stress, and cold in Europe from observations. *Geophysical Research Letters*, *46*(14), 8363–8374. (Publisher: Wiley Online Library)

Lorenz, S., Dessai, S., Forster, P. M., & Paavola, J. (2015, November). Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2055), 20140457. Retrieved 2023-02-17, from https://royalsocietypublishing.org/doi/10.1098/rsta.2014.0457 (Publisher: Royal Society) doi: 10.1098/rsta.2014.0457

Lukat, E., Lenschow, A., Dombrowsky, I., Meergans, F., Schütze, N., Stein, U., & Pahl-Wostl, C. (2023, March). Governance towards coordination for water resources management: The effect of governance modes. *Environmental Science & Policy*, *141*, 50–60. Retrieved 2023-02-09, from https://www.sciencedirect.com/science/article/pii/S1462901122003860 doi: 10.1016/j.envsci.2022.12.016

Maier, H. R., Guillaume, J. H., 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.

Malers, S. A., Ray R. Bennett, & Catherine, N.-L. (2001). Colorado's Decision Support Systems: Data-Centered Water Resources Planning and Administration. *Watershed Management and Operations Management 2000*, 1–9. Retrieved 2019-12-04, from https://ascelibrary.org/doi/abs/10.1061/40499(2000)153 doi: 10.1061/40499(2000)153

Marchau, V. A. W. J., Walker, W. E., Bloemen, P. J. T. M., & Popper, S. W. (Eds.). (2019). *Decision Making under Deep Uncertainty: From Theory to Practice*. Springer International Publishing. Retrieved 2020-08-16, from https://www.springer.com/gp/book/9783030052515 doi: 10.1007/978-3-030-05252-2

Markolf, S. A., Chester, M. V., Eisenberg, D. A., Iwaniec, D. M., Davidson, C. I., Zimmerman, R., . . . Chang, H. (2018). Interdependent Infrastructure as Linked Social, Ecological, and Technological Systems (SETSs) to Address Lock-in and Enhance Resilience. *Earth's Future*, *6*(12), 1638–1659. Retrieved 2023-06-01, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2018EF000926 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018EF000926) doi: 10.1029/2018EF000926

McCoy, A. L., Jacobs, K. L., Vano, J. A., Wilson, J. K., Martin, S., Pendergrass, A. G., & Cifelli, R. (2022). The Press and Pulse of Climate Change: Extreme Events in the Colorado River Basin. *JAWRA Journal of the American Water Resources Association*, *58*(6), 1076–1097. Retrieved 2023-01-25, from https://onlinelibrary.wiley.com/doi/abs/10.1111/1752-1688.13021 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.13021) doi: 10.1111/1752-1688.13021

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2), 239–245. Retrieved from https://www.jstor.org.proxy.library.cornell.edu/stable/1268522 doi: 10.2307/1268522

McPhail, C., Maier, H. R., Kwakkel, J. H., Giuliani, M., Castelletti, A., & Westra, S. (2018). Robustness Metrics: How Are They Calculated, When Should They Be Used and Why Do They Give Different Results? *Earth's Future*, *6*, 169–191. doi: 10.1002/2017EF000649@10.1002/(ISSN)2328-4277.RESDEC1

Meko, D. M., Woodhouse, C. A., Baisan, C. A., Knight, T., Lukas, J. J., Hughes, M. K., & Salzer, M. W. (2007). Medieval drought in the upper Colorado River

Basin. *Geophysical Research Letters*, *34*(10). Retrieved 2023-03-20, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2007GL029988 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2007GL029988) doi: 10.1029/2007GL029988

Moallemi, E. A., Kwakkel, J. H., de Haan, F. J., & Bryan, B. A. (2020, November). Exploratory modeling for analyzing coupled human-natural systems under uncertainty. *Global Environmental Change*, *65*, 102186. Retrieved 2020-11-17, from http://www.sciencedirect.com/science/article/pii/S095937802030769X doi: 10.1016/j.gloenvcha.2020.102186

Moallemi, E. A., Zare, F., Reed, P. M., Elsawah, S., Ryan, M. J., & Bryan, B. A. (2020, January). Structuring and evaluating decision support processes to enhance the robustness of complex human–natural systems. *Environmental Modelling & Software*, *123*, 104551. Retrieved 2019-12-03, from http://www.sciencedirect.com/science/article/pii/S1364815219306905 doi: 10.1016/j.envsoft.2019.104551

Mondal, A., & Mujumdar, P. P. (2015, January). Return levels of hydrologic droughts under climate change. *Advances in Water Resources*, *75*, 67–79. Retrieved 2023-06-02, from https://www.sciencedirect.com/science/article/pii/S030917081400219X doi: 10.1016/j.advwatres.2014.11.005

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., ... Wilbanks, T. J. (2010, February). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747–756. Retrieved 2023-06-12, from https://www.nature.com/articles/nature08823 (Number: 7282 Publisher: Nature Publishing Group) doi: 10.1038/nature08823

Naumann, G., Alfieri, L., Wyser, K., Mentaschi, L., Betts, R. A., Carrao, H., ... Feyen, L. (2018). Global Changes in Drought Conditions Under Different Levels of Warming. *Geophysical Research Letters*, *45*(7), 3285–3296. Retrieved 2023-03-20, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2017GL076521 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017GL076521) doi: 10.1002/2017GL076521

Nowak, K., Prairie, J., Rajagopalan, B., & Lall, U. (2010). A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow. *Water Resources Research*, *46*(8). Retrieved 2023-03-21, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008530 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR008530) doi: 10.1029/2009WR008530

Overpeck, J. T., & Udall, B. (2020, June). Climate change and the aridification of North America. *Proceedings of the National Academy of Sciences*, *117*(22), 11856–11858. Retrieved 2023-02-09, from https://www.pnas.org/doi/full/10.1073/pnas.2006323117 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.2006323117

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., ... van Vuuren, D. P. (2014, February). A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400. Retrieved 2020-08-14, from https://doi.org/10.1007/s10584-013-0905-2 doi: 10.1007/s10584-013-0905-2

Parsons, R., & Bennett, R. (2006). Reservoir Operations Management Using a Water Resources Model. *Operating Reservoirs in Changing Conditions*, 304–311. Retrieved 2019-07-08, from https://ascelibrary.org/doi/abs/10.1061/40875(212)30 doi: 10.1061/40875(212)30

Pedersen, J. T. S., van Vuuren, D., Gupta, J., Santos, F. D., Edmonds, J., & Swart, R. (2022, July). IPCC emission scenarios: How did critiques affect their quality and relevance 1990–2022? *Global Environmental Change*, *75*, 102538. Retrieved 2023-08-30, from https://www.sciencedirect.com/science/article/pii/S0959378022000760 doi: 10.1016/j.gloenvcha.2022.102538

Popper, S. W., Berrebi, C., Griffin, J., Light, T., Daehner, E. M., & Crane, K. (2009, December). *Natural Gas and Israel's Energy Future: Near-Term Decisions from a Strategic Perspective* (Tech. Rep.). RAND Corporation. Retrieved 2023-05-13, from https://www.rand.org/pubs/monographs/MG927.html

Priss, U. (2021). Set Visualisations with Euler and Hasse Diagrams. In M. Cochez, M. Croitoru, P. Marquis, & S. Rudolph (Eds.), *Graph Structures for Knowledge Representation and Reasoning* (pp. 72–83). Cham: Springer International Publishing. doi: 10.1007/978-3-030-72308-8_5

Pruett, W. A., & Hester, R. L. (2016, June). The Creation of Surrogate Models for Fast Estimation of Complex Model Outcomes. *PLOS ONE*, *11*(6), e0156574. Retrieved 2020-11-16, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0156574 (Publisher: Public Library of Science) doi: 10.1371/journal.pone.0156574

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. Retrieved 2023-01-04, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001650 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001650) doi: 10.1029/2020EF001650

Quinn, J. D., Reed, P. M., Giuliani, M., Castelletti, A., Oyler, J. W., & Nicholas, R. E. (2018, July). 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 2019-10-26, from http://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018WR022743 doi: 10.1029/2018WR022743

Reed, P. M., Hadjimichael, A., Malek, K., Karimi, T., Vernon, C. R., Srikrishnan, V. A., ... Rice, J. S. (2022). *Addressing Uncertainty in Multisector Dynamics Research.* Retrieved 2022-03-16, from https://uc-ebook.org/ doi: 10.5281/ZENODO.6110623

Reed, P. M., Hadjimichael, A., Moss, R. H., Monier, E., Alba, S., Brelsford, C., ... Yoon, J. (2022, January). *MultiSector Dynamics: Scientific Challenges and a Research Vision for 2030, A Community of Practice Supported by the United States Department of Energy's Office of Science* (Tech. Rep.). MultiSector Dynamics Community of Practice. Retrieved 2022-05-11, from https://zenodo.org/record/6144309 doi: 10.5281/zenodo.6144309

Robbins, J. (2019). On the water-starved colorado river, drought is the new normal. *Yale Environment*, *360*.

Salas, J. D., Obeysekera, J., & Vogel, R. M. (2018, February). Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal*, *63*(3), 325–352. Retrieved 2023-06-01, from https://doi.org/10.1080/02626667.2018.1426858 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02626667.2018.1426858) doi: 10.1080/02626667.2018.1426858

Savage, L. J. (1951, March). The Theory of Statistical Decision. *Journal of the American Statistical Association*, *46*(253), 55–67. Retrieved 2019-10-01, from https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1951.10500768 doi: 10.1080/01621459.1951.10500768

Savelli, E., Rusca, M., Cloke, H., & Di Baldassarre, G. (2022). Drought and society: Scientific progress, blind spots, and future prospects. *WIREs Climate Change*, *n/a*(n/a), e761. Retrieved 2022-02-04, from https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.761 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.761) doi: 10.1002/wcc.761

Schlumberger, J., Haasnoot, M., Aerts, J., & de Ruiter, M. (2022, October). Proposing DAPP-MR as a disaster risk management pathways framework for complex, dynamic multi-risk. *iScience*, *25*(10), 105219. Retrieved 2023-06-01, from https://www.sciencedirect.com/science/article/pii/S2589004222014912 doi:

10.1016/j.isci.2022.105219

Schlüter, M., Mcallister, R. R. J., Arlinghaus, R., Bunnefeld, N., Eisenack, K., Hölker, F., ... Stöven, M. (2012). New Horizons for Managing the Environment: A Review of Coupled Social-Ecological Systems Modeling. *Natural Resource Modeling*, *25*(1), 219–272. Retrieved 2023-05-11, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1939-7445.2011.00108.x (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1939-7445.2011.00108.x) doi: 10.1111/j.1939-7445.2011.00108.x

Shepherd, T. G., Boyd, E., Calel, R. A., Chapman, S. C., Dessai, S., Dima-West, I. M., ... Zenghelis, D. A. (2018). Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, *151*(3), 555–571. doi: 10.1007/s10584-018-2317-9

Shi, R., Hobbs, B. F., Quinn, J. D., Lempert, R., & Knopman, D. (2023, February). City-Heat Equity Adaptation Tool (City-HEAT): Multi-objective optimization of environmental modifications and human heat exposure reductions for urban heat adaptation under uncertainty. *Environmental Modelling & Software*, *160*, 105607. Retrieved 2023-01-05, from https://www.sciencedirect.com/science/article/pii/S1364815222003073 doi: 10.1016/j.envsoft.2022.105607

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, *63*(2), 129. doi: https://doi.org/10.1037/h0042769

Simpson, N. P., Mach, K. J., Constable, A., Hess, J., Hogarth, R., Howden, M., ... Trisos, C. H. (2021, April). A framework for complex climate change risk assessment. *One Earth*, *4*(4), 489–501. Retrieved 2021-04-28, from https://www.cell.com/one-earth/abstract/S2590-3322(21)00179-2 (Publisher: Elsevier) doi: 10.1016/j.oneear.2021.03.005

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., ... Wilby, R. L. (2021, July). Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, *25*(7), 3897–3935. Retrieved 2023-06-01, from https://hess.copernicus.org/articles/25/3897/2021/ (Publisher: Copernicus GmbH) doi: 10.5194/hess-25-3897-2021

Smith, R., Zagona, E., Kasprzyk, J., Bonham, N., Alexander, E., Butler, A., ... Jerla, C. (2022, January). Decision Science Can Help Address the Challenges of Long-Term Planning in the Colorado River Basin. *JAWRA Journal of the American Water Resources Association*, *n/a*(n/a). Retrieved 2022-01-25, from https://onlinelibrary.wiley.com/doi/abs/10.1111/1752-1688.12985 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.12985) doi: 10.1111/1752-1688.12985

State of Colorado. (2015). *Colorado's water plan* (Tech. Rep.). Denver, Colorado.

State of Colorado. (2023). *Colorado's Water Plan* (Tech. Rep.).

Steinmann, P., Auping, W. L., & Kwakkel, J. H. (2020, July). Behavior-based scenario discovery using time series clustering. *Technological Forecasting and Social Change*, *156*, 120052. Retrieved 2020-05-04, from http://www.sciencedirect.com/science/article/pii/S0040162519302380 doi: 10.1016/j.techfore.2020.120052

Stevenson, S., Coats, S., Touma, D., Cole, J., Lehner, F., Fasullo, J., & Otto-Bliesner, B. (2022, March). Twenty-first century hydroclimate: A continually changing baseline, with more frequent extremes. *Proceedings of the National Academy of Sciences*, *119*(12), e2108124119. Retrieved 2022-04-04, from https://www.pnas.org/doi/10.1073/pnas.2108124119 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.2108124119

Sun, Q., Zhang, X., Zwiers, F., Westra, S., & Alexander, L. V. (2021). A global, continental, and regional analysis of changes in extreme precipitation. *Journal of Climate*, *34*(1), 243–258.

Sunkara, S. V., Singh, R., Gold, D., Reed, P., & Bhave, A. (2023). How Should Diverse Stakeholder Interests Shape Evaluations of Complex Water Resources Systems Robustness When Confronting Deeply Uncertain Changes? *Earth's Future*, *11*(8), e2022EF003469. Retrieved 2023-08-30, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2022EF003469 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022EF003469) doi: 10.1029/2022EF003469

Trindade, B. C., Gold, D. F., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2020, October). Water pathways: An open source stochastic simulation system for integrated water supply portfolio management and infrastructure investment planning. *Environmental Modelling & Software*, *132*, 104772. Retrieved 2020-08-21, from http://www.sciencedirect.com/science/article/pii/S1364815220301511 doi: 10.1016/j.envsoft.2020.104772

Trindade, B. C., Reed, P. M., & Characklis, G. W. (2019, December). Deeply uncertain pathways: Integrated multi-city regional water supply infrastructure investment and portfolio management. *Advances in Water Resources*, *134*, 103442. Retrieved 2020-03-31, from http://www.sciencedirect.com/science/article/pii/S0309170819306475 doi: 10.1016/j.advwatres.2019.103442

Trindade, B. C., Reed, P. M., Herman, J. D., Zeff, H. B., & Characklis, G. W. (2017, June). Reducing regional drought vulnerabilities and multi-city robustness conflicts using many-objective optimization under deep uncertainty. *Advances in Water Resources*, *104*(Supplement C), 195–209. Retrieved from http://www.sciencedirect.com/science/article/pii/S0309170816307333 doi: 10.1016/j.advwatres.2017.03.023

Tufte, E. R. (1990). *Envisioning Information* (Vol. 6410). Cheshire, Connecticut: Graphics Press.

Vahmani, P., Jones, A. D., & Li, D. (2022). Will Anthropogenic Warming Increase Evapotranspiration? Examining Irrigation Water Demand Implications of Climate Change in California. *Earth's Future*, *10*(1), e2021EF002221. Retrieved 2023-05-11, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2021EF002221 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021EF002221) doi: 10.1029/2021EF002221

van den Elzen, S., & van Wijk, J. J. (2013). Small Multiples, Large Singles: A New Approach for Visual Data Exploration. *Computer Graphics Forum*, *32*(3pt2), 191–200. Retrieved 2023-03-29, from https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12106 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12106) doi: 10.1111/cgf.12106

Van Ruijven, B., Carlsen, H., Chaturvedi, V., Ebi, K., Fuglestvedt, J., Gasalla, M., ... Leininger, J. (2023). The SSP-RCP scenario framework: progress, needs, and next steps–Insights from the Scenarios Forum 2022. Bordeaux, France.

Wald, A. (1950). *Statistical decision functions*. Wiley.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated assessment*, *4*(1), 5–17. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.469.7495Źrep=rep1Źtype=pdf (Type: Journal Article)

Wegman, E. J. (1990, September). Hyperdimensional Data Analysis Using Parallel Coordinates. *Journal of the American Statistical Association*, *85*(411), 664–675. Retrieved 2023-06-01, from https://www.tandfonline.com/doi/abs/10.1080/01621459.1990.10474926 (Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1990.10474926) doi: 10.1080/01621459.1990.10474926

Whitney, K. M., Vivoni, E. R., Bohn, T. J., Mascaro, G., Wang, Z., Xiao, M., ... White, D. D. (2023, February). Spatial attribution of declining Colorado River streamflow under future warming. *Journal of Hydrology*, *617*, 129125. Retrieved 2023-01-25, from https://www.sciencedirect.com/science/article/pii/

S0022169423000677 doi: 10.1016/j.jhydrol.2023.129125

Woodhouse, C. A., Gray, S. T., & Meko, D. M. (2006). Updated streamflow reconstructions for the Upper Colorado River Basin. *Water Resources Research*, *42*(5). Retrieved 2022-04-06, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2005WR004455 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005WR004455) doi: 10.1029/2005WR004455

Woodhouse, C. A., & Overpeck, J. T. (1998, December). 2000 Years of Drought Variability in the Central United States. *Bulletin of the American Meteorological Society*, *79*(12), 2693–2714. Retrieved 2022-04-14, from https://journals.ametsoc.org/view/journals/bams/79/12/1520-0477_1998_079_2693_yodvit_2_0_co_2.xml (Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society) doi: 10.1175/1520-0477(1998)079<2693:YODVIT>2.0.CO;2

Woodhouse, C. A., Smith, R. M., McAfee, S. A., Pederson, G. T., McCabe, G. J., Miller, W. P., & Csank, A. (2021, January). Upper Colorado River Basin 20th century droughts under 21st century warming: Plausible scenarios for the future. *Climate Services*, *21*, 100206. Retrieved 2022-04-01, from https://www.sciencedirect.com/science/article/pii/S2405880720300583 doi: 10.1016/j.cliser.2020.100206

Wyborn, C., Datta, A., Montana, J., Ryan, M., Leith, P., Chaffin, B., . . . van Kerkhoff, L. (2019). Co-Producing Sustainability: Reordering the Governance of Science, Policy, and Practice. *Annual Review of Environment and Resources*, *44*(1), 319–346. Retrieved 2023-05-11, from https://doi.org/10.1146/annurev-environ-101718-033103 (_eprint: https://doi.org/10.1146/annurev-environ-101718-033103) doi: 10.1146/annurev-environ-101718-033103

Yang, Y., Botton, M. R., Scott, E. R., & Scott, S. A. (2017, May). Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics*, *18*(7), 673–685. Retrieved 2023-06-01, from https://www.futuremedicine.com/doi/abs/10.2217/pgs-2017-0033 (Publisher: Future Medicine) doi: 10.2217/pgs-2017-0033

Yang, Y., Roderick, M. L., Yang, D., Wang, Z., Ruan, F., McVicar, T. R., . . . Beck, H. E. (2021, June). Streamflow stationarity in a changing world. *Environmental Research Letters*, *16*(6), 064096. Retrieved 2023-06-01, from https://dx.doi.org/10.1088/1748-9326/ac08c1 (Publisher: IOP Publishing) doi: 10.1088/1748-9326/ac08c1

Supporting Information for "Multi-actor, multi-impact scenario discovery of consequential narrative storylines in human-natural systems"

Antonia Hadjimichael[1,2], Patrick M. Reed[3], Julianne D. Quinn[4], Chris R. Vernon[5], Travis Thurber[5]

[1]Department of Geosciences, The Pennsylvania State University, State College, PA, USA

[2]Earth and Environmental Systems Institute (EESI), The Pennsylvania State University, State College, PA, USA

[3]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA

[4]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA

[5]Atmospheric Sciences & Global Change, Pacific Northwest National Laboratory, Richland, WA, USA

Contents of this file

1. Figures S1 to S3: Drought year classification by historical 60-year rolling windows; distribution of streamflows in exploratory ensemble, as created by Quinn, Hadjimichael, Reed, and Steinschneider (2020); thresholds used to classify the states of the world as within the historical context

References

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., & Meko, D. M. (2014, January). Assessing the Risk of Persistent Drought Using Climate Model Simulations and Paleoclimate Data. *Journal of Climate*, *27*(20), 7529–7549. Retrieved 2020-04-28, from https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-12-00282.1 (Publisher: American Meteorological Society) doi: 10.1175/JCLI-D-12-00282.1

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. Retrieved 2023-01-04, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2020EF001650 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001650) doi: 10.1029/2020EF001650

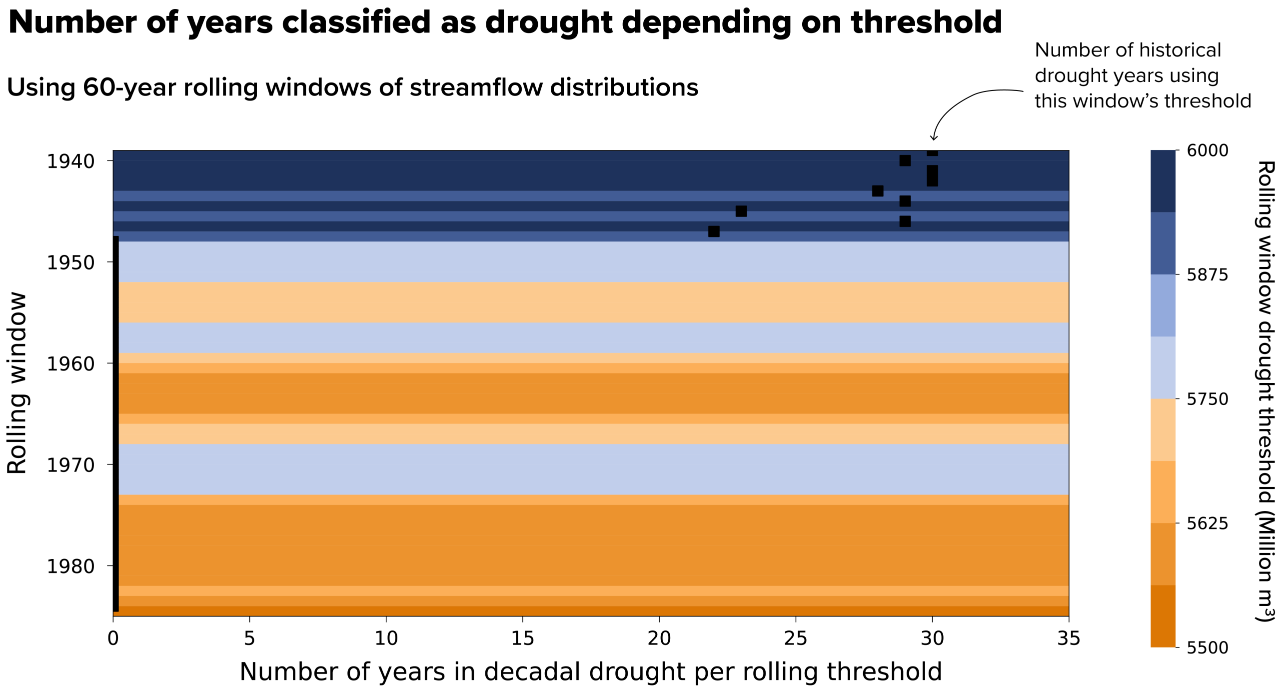

Figure S1. Number of years classified as drought depending on each rolling-window threshold.

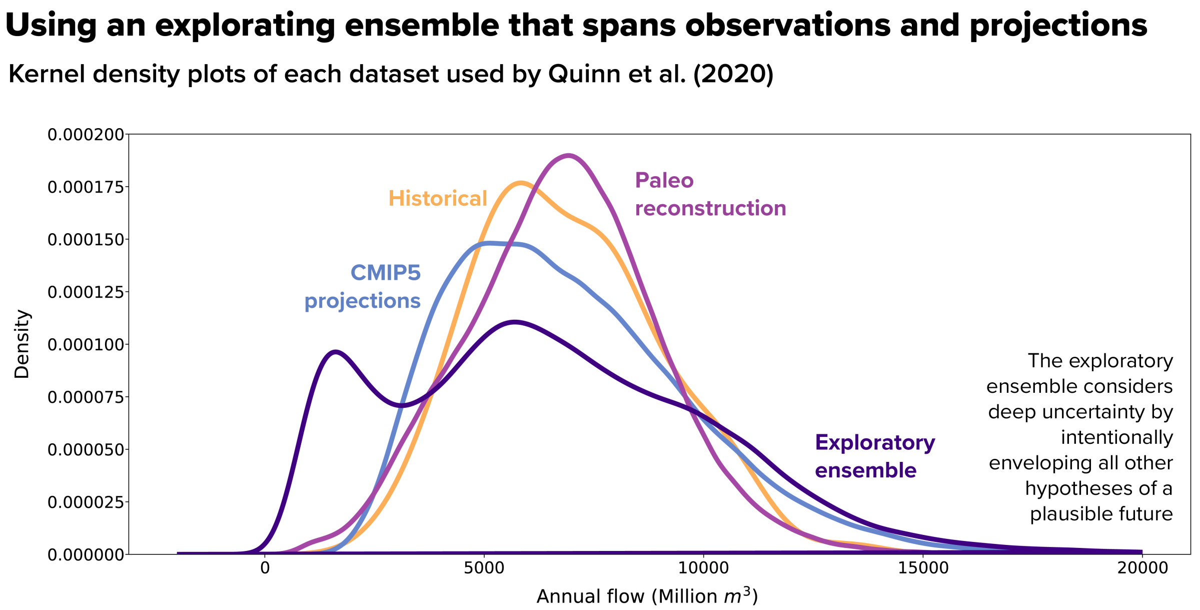

Figure S2. Distribution of streamflows in the exploratory ensemble used by this experiment, as it relates to other 'rival framings' of plausible future streamflow. The ensemble used is created by Quinn et al. (2020) and all the data are provided by that paper and accompanying online repository (`https://github.com/julianneq/UCRB_analysis`).

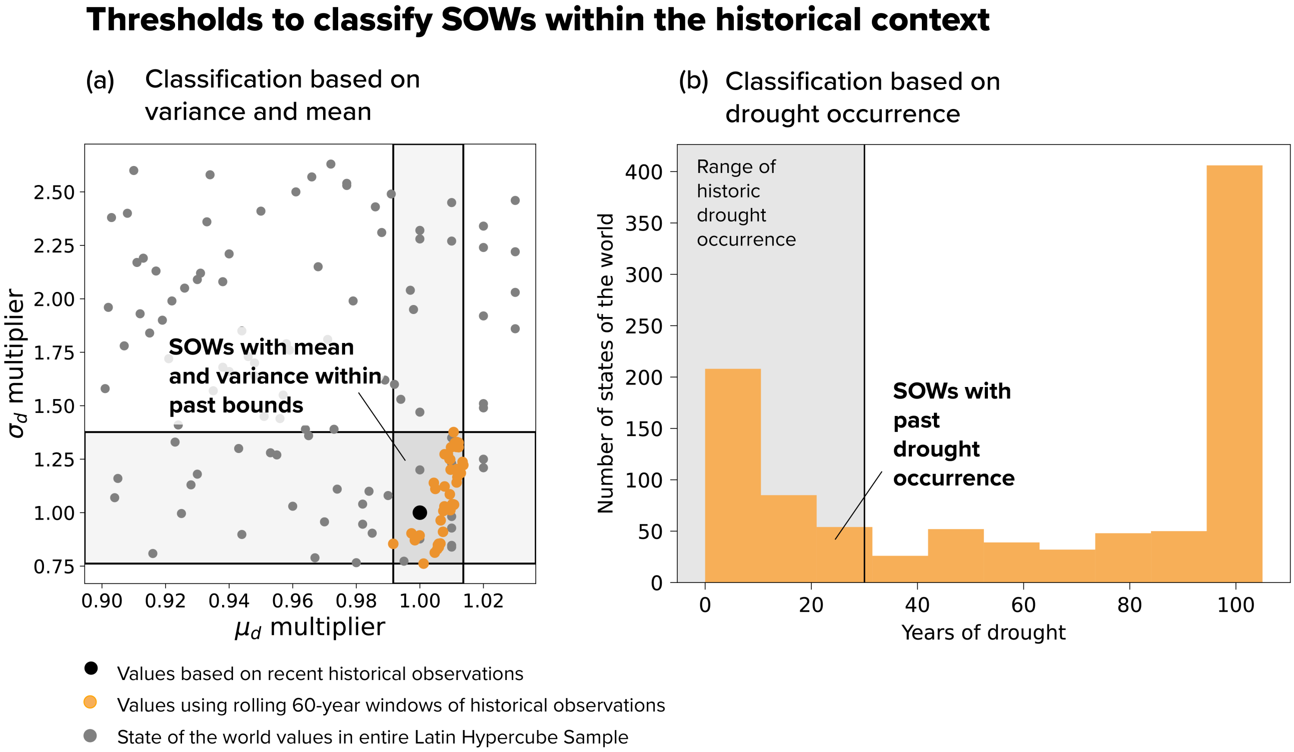

Figure S3. Identification of states of the world (SOWs) within the bounds of the past. (a) Variability (σ_d) and persistence (p_{dd}) properties of each SOW in the ensemble. These properties are determined by fitting the Gaussian Hidden Markov Model to the historical observations (resulting in the black point) and then sampling changes to these properties to represent alternative SOWs for the basin, as elaborated in Quinn et al. (2020). Each orange point represents 10 realizations of streamflow that exhibit the same sampled statistical properties, for a total of 1000 SOWs. Each grey point represents the variability and persistence properties of one of the 100 reconstructions of paleo streamflow with added noise, following the same procedure as Quinn et al. (2020). The mean values of both the variability and persistence properties are used to select SOWs that fall within the bounds of the past (recent history and paleo reconstructions). (b) Histogram of drought years occurring in each sampled SOW. The black vertical line represents the number of droughts that have occurred per century in both the historic and paleo record, using the threshold-based classification of Ault et al. (2014) and others.