



RESEARCH ARTICLE

10.1029/2020EF001503

Key Points:

- Expand bottom-up vulnerability assessments to highly resolved multi-user systems governed by prior appropriation
- Advance robustness analysis methods to account for both multiple stakeholders, and multiple definitions of vulnerability for each
- Illustrate how user-specific scenario discovery reveals the diversity of stakeholder vulnerabilities in institutionally complex river basins

Supporting Information:

- Supporting Information

Correspondence to:

A. Hadjimichael,
ah986@cornell.edu

Citation:

Hadjimichael, A., Quinn, J. D., 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*, 7, e2020EF001503. <https://doi.org/10.1029/2020EF001503>

Received 30 JAN 2020

Accepted 4 MAY 2020

Accepted article online 12 MAY 2020

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Defining Robustness, Vulnerabilities, and Consequential Scenarios for Diverse Stakeholder Interests in Institutionally Complex River Basins

Antonia Hadjimichael¹ , Julianne Quinn² , Erin Wilson³, Patrick Reed¹ , Leon Basdekas⁴, David Yates⁵, and Michelle Garrison⁶

¹School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA, ²Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA, ³Wilson Water Group LLC, Lakewood, CO, USA, ⁴Black and Veatch, Colorado Springs, CO, USA, ⁵National Center for Atmospheric Research, Boulder, CO, USA, ⁶Colorado Water Conservation Board, Denver, CO, USA

Abstract The Upper Basin of the Colorado River in the southwestern United States supports municipal, industrial, agricultural, and recreational activities worth an estimated \$300 billion/year within the state of Colorado alone. The allocation of water to these activities is fundamentally shaped by water rights that in turn distribute risks among a diverse suite of sectors and stakeholders. In this study, we assess the vulnerabilities faced by the hierarchy of hundreds of water users in the portion of the Upper Basin within the state of Colorado, as a result of changes in hydrologic extremes, demand growth, and institutional and physical infrastructure in the basin. We also determine how robust the different users are to these potential changes, examining how their sensitivity to these factors depends on the magnitude and frequency of shortage they are unwilling to exceed. This advances previous robustness evaluation methods by formalizing an a posteriori exploration of alternative definitions of robustness tailored to each user's unique context. Our analysis reveals that robustness varies significantly not only across users but also across different definitions of acceptable performance for each user. We further show the importance of using scenario discovery to evaluate how the factors that drive consequential scenarios differ depending on the definition of robustness. Our results highlight the need for robustness and vulnerability frameworks to avoid broad categorical aggregations of stakeholder groups, to carefully capture complex institutional dependencies (e.g., water rights), and to facilitate a more transparent illustration of the implications of alternative definitions of robustness for specific stakeholders.

1. Introduction

Water resources systems are inherently complex human-natural systems, characterized by multisector interactions between the environment, physical infrastructure, and societal needs (Clarke et al., 2018; Moss et al., 2016). The availability and provision of water to its multiple uses is shaped by the human systems (water institutions, societal values, and infrastructure requirements) as well as by their natural contexts (hydrology, climate, and ecosystems). In institutionally complex river basins, the institutions governing water (e.g., systems of allocation and legal rights) manage its dynamics and, by extension, distribute the risks and vulnerabilities innate to increasingly variable extremes (Hall et al., 2014). This is especially true for the American West, where the doctrine of “prior appropriation” fundamentally shapes water resources allocations, vulnerabilities, and dynamic responses to hydrologic extremes. Prior appropriation distributes resources across water users in order of priority (“first-in-time, first-in-right”) and limits the decreed volume of water of each right to the amount that could be put to a “beneficial use” (Kenney, 2005). The seniority of water rights hierarchically defines the relationships between stakeholders in these systems, where the senior water rights holders fundamentally shape the dynamics of water entitlements for more junior water rights holders.

In Colorado, rapid population growth and increasing demands are stressing the state's water supplies (State of Colorado, 2015). Climate change is expected to increase the strain on supply with rising temperatures, earlier peak runoff, and reduced water availability (Chavarria & Gutzler, 2018; Fyfe et al., 2017; Solander et al., 2017). Yet, while the direction of these temperature-induced impacts is understood, the complex feedbacks of uncertain global carbon emissions on precipitation make the magnitude of hydrologic impacts from

climate change far more difficult to project (Elias et al., 2015; Lehner et al., 2019; Vano et al., 2014). These drivers are examples of what some have referred to as “deep uncertainties” (Knight, 1921; Polasky et al., 2011; Walker et al., 2003). More formally, deep uncertainties can be defined as component factors within a system for which expert opinion cannot know or cannot agree on the full set of outcomes and their associated likelihoods (Kwakkel et al., 2010; Lempert, 2002; Lempert et al., 2003). The challenges and potential high consequences of deep uncertainties have motivated a rapidly growing body of literature focused on bottom-up approaches, which employ exploratory modeling (Bankes, 1993) to simulate large ensembles of plausible scenarios and thereupon identify those with consequential effects on the system (see Dittrich et al., 2016; Herman et al., 2015; Kwakkel & Haasnoot, 2019; Moallemi et al., 2020, for recent reviews of these approaches). In this manner, exploratory modeling shifts the focus from predicting future conditions to discovering which future conditions lead to (un)desirable consequences. The relative likelihoods or decision relevance of the scenarios that lead to these outcomes can then be evaluated by the relevant decision makers a posteriori in the richer context of knowing their potential consequences (Dessai et al., 2009).

In bottom-up methods, the identification of consequential scenarios (also known as scenario discovery) presupposes a metric-based classification of the acceptability of system performance, most commonly by exploiting one or more candidate measures of robustness. Common families of robustness measures include expected value metrics (Wald, 1950), regret-based metrics (Savage, 1951), or satisficing metrics (Simon, 1956). Recent literature has highlighted that there can be substantial implications and insights when choosing among different candidate robustness metrics (Herman et al., 2015; Maier et al., 2016; McPhail et al., 2018), with the shared conclusion across these works that each different metric choice reflects potentially very different aspects of what makes the system robust. These studies, however, do not address the underlying complexities associated with better resolving highly diverse multistakeholder perspectives. In multistakeholder systems, the choice of robustness measures bears the additional burden of promoting agency in individuals' choices while also clearly mapping what makes a system robust for whom. Highly aggregated system-wide robustness measures potentially ignore asymmetries in values and adaptive capacities across individual actors in the system (Franssen, 2005). These concerns are especially relevant for water resources systems supporting demands across multiple sectors, such as agriculture, manufacturing, power generation, and recreation (Kasprzyk et al., 2016). Relatively few recent studies have focused on clarifying the implications of multistakeholder trade-offs in robustness (Gold et al., 2019; Herman et al., 2014; Trindade et al., 2017) to actively reduce inadvertent consequences for particular actors. Moreover, these studies are limited in their focus on a single sector (i.e., water supply) and only four significant actors.

In more institutionally complex contexts, such as the Upper Basin of the Colorado River, bottom-up vulnerability assessments should resolve water rights at a high resolution, especially when assessing how particular users are affected. The Colorado Water Conservation Board (CWCB) and the Division of Water Resources (DWR) have jointly developed Colorado's Decision Support System (CDSS), a collection of databases, data viewing and management tools, and models to support water resources planning in Colorado's major water basins (Malers et al., 2001). The CDSS is made up of a central database with water resources data (HydroBase), a public website where the data can be accessed, a Geographic Information System (GIS) for viewing and analyzing the data, and a consumptive use model (StateCU) that estimates consumptive use by each irrigation unit in a basin. The outputs from StateCU are then input to the State of Colorado's Stream Simulation Model (StateMod), a generic network-based water system model for water accounting and allocation, and the final component of CDSS. StateMod was developed to support comprehensive assessments of water demand, allocation, and use, as well as reservoir operations. It represents all of the major sub-basins within the state of Colorado (i.e., Parsons & Bennett, 2006; White, Yampa, Upper Colorado, Gunnison, Dolores, San Juan, and San Miguel CWCB, 2012). StateMod replicates each basin's unique application and enforcement of the prior appropriation doctrine and accounts for all of the consumptive use within the basins. To do so, it relies on the detailed historic demand and operation records contained in HydroBase that include individual water right information for all consumptive use, data on water structures (wells, ditches, reservoirs, and tunnels), and streamflow data. Further, StateMod uses irrigation consumptive use data output from StateCU, which calculates water consumption based on soil moisture, crop type, irrigated acreage, and conveyance and application efficiencies for each individual irrigation unit in the region. These detailed, fine-scale inputs allow StateMod to resolve the effect of all users and water rights on water availability, which can in turn be used to simulate hypothetical scenarios to assess the impacts of changes in hydrology, water rights, or infrastructure on any and all represented water rights.

Both the DWR and the CWCB, as well as several other water resources entities, have been actively using StateMod to model and assess impacts to water users by changing water demand, allocation, and operations since the late 1980s (Parsons & Bennett, 2006). StateMod has been used to support informed planning for the basin including studies on annual consumptive water use, as well as the possible system impacts of having to meet interstate river compact obligations under potential changes in stream flows, reservoir storage capacity, and reservoir operations. StateMod has also been instrumental in the development of the *Colorado River Water Availability Study* (CWCB, 2012), investigating how climate change may affect water availability at the water user and water rights level, as well as reservoir use and operations in the four major tributary river basins of the Colorado River in the state of Colorado. The study was considered at the time the “most detailed look at how specific water users may be impacted by climate change performed anywhere in the world” (CWCB, 2012).

The Upper Colorado basin within the state of Colorado (henceforth abbreviated to UCRB) is the largest of the five basins, containing the headwaters of the Colorado River, and is the basin of interest in this study. The outflow of this subbasin of the Upper Basin of the Colorado River flows into Utah and continues to deliver water downstream to Lake Powell. The Colorado River Compact allocates 9.3 km^3 (7.5 million acre feet) per year to the Upper Basin states (Colorado, Utah, Wyoming, New Mexico, and Arizona), and the state of Colorado is allotted 51.75% of that amount. The UCRB contains several thousand water rights for agriculture, municipal water supply, and recreational uses, power generation, and fisheries. The basin lies west of the Continental Divide so major diversions of water are necessary to support urban centers on the east slope, where most of Colorado's population lives. Amid a historic 20-year drought (Rhee et al., 2018; Schwartz, 2019; Udall & Overpeck, 2017), state-level planning efforts must carefully balance the increasing potential for resource competition across users while accounting for potential vulnerabilities to climate change, demand growth, and infrastructure changes. In this paper, we perform a bottom-up water-supply vulnerability assessment for hundreds of stakeholders in the UCRB, seeking to identify the factors that most significantly shape their ability to meet their water demands. To do so, we pair StateMod with an exploratory modeling approach, through which an ensemble of scenarios reflecting potential changes in hydrologic conditions, evaporation, water demand, infrastructure, and water rights are simulated. We investigate how these changes manifest through the system to impact its users and how these impacts are shaped by the presence of strong institutional governance and intricate allocation-priority relationships.

The objectives of this paper are therefore the following: (1) perform a bottom-up vulnerability assessment of the diverse water users in UCRB, resolving their water shortage vulnerabilities resulting from an ensemble of deeply uncertain hydrologic, social, and institutional changes; and (2) evaluate their robustness for a spectrum of potential definitions of vulnerability to inform scenario discovery that identifies the most consequential scenarios shaping users' water shortages. This study advances previous applications of the Robust Decision Making (RDM) framework (Lempert et al., 2006, 2010) by formalizing an a posteriori exploration of alternative history-informed performance preferences for large multiactor systems that clarifies the implications for defining robustness and the identification of key factors that control vulnerabilities.

2. Study Area and Model

The UCRB spans $25,682 \text{ km}^2$ ($9,915 \text{ mi}^2$) in western Colorado, from its headwaters at the Continental Divide to its outflows located at the Colorado-Utah state line (Figure 1). The following description of the study area will be centered around the main uses of water in the basin, as well as several specific elements that were deemed to be of particular interest by the CWCB and other stakeholders. The area supports a moderate but growing population of approximately 300,000 inhabitants, with the major water use in the basin being irrigation. Several thousand irrigation ditches divert water from the main stream and its tributaries to irrigate approximately $1,012 \text{ km}^2$ (391 mi^2) in the basin. Eighty percent of Colorado's population resides on the east slope, which necessitates major transbasin (and transmountain) diversions from the UCRB. These diversions amount to approximately $567,400,000 \text{ m}^3$ (460,000 acre feet) a year and export water for irrigation, industrial, and municipal use in northern and eastern Colorado, including the cities of Denver and Colorado Springs (State of Colorado, 2015). The major tunnels serving these diversions are indicated in Figure 1.

The other major water use in the basin is for power generation, with the Shoshone Power Plant, the Grand Valley Power Plant, and the Molina Power Plant being the primary power producers. Water for power generation is generally nonconsumptive and does not significantly deplete the water resources in the basin.

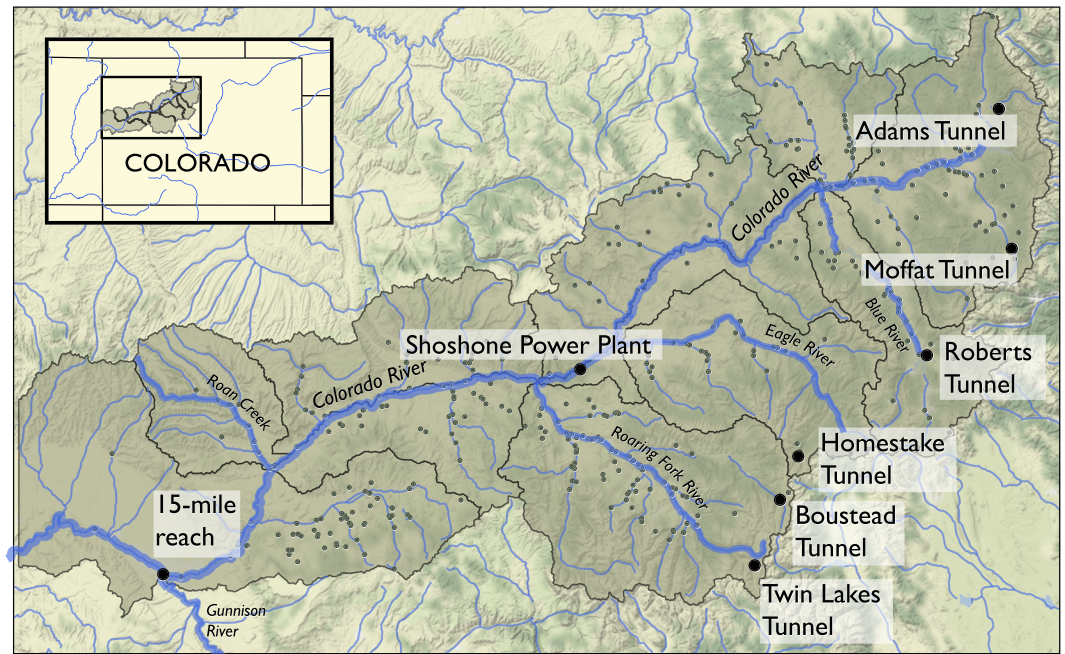


Figure 1. The Upper Colorado River Basin within the state of Colorado (UCRB) with structures of interest indicated. The Shoshone Power Plant owns one of the largest and oldest water rights in the basin. The 15-mile reach is critical for the recovery of endangered species. The highlighted tunnels are the most major transbasin diversions, exporting water to the eastern plains. The smaller points indicate all other modeled diversion points in StateMod (primarily irrigation).

Of note, the Shoshone Power Plant (indicated in Figure 1) owns one of the most important water rights in the basin, in terms of size and seniority. The oldest right owned by the plant dates back to 1902 with a large decree of $39.40 \text{ m}^3/\text{s}$ (1,391 cf/s), representing approximately 68% of the median flow at that location (USGS, 2019; Yates et al., 2015). Generally, when the Shoshone Power Plant calls their decreed water right, upstream junior reservoirs, transbasin diversions, and irrigation must either cease or offset their diversion and storage to satisfy the call. Given that the Shoshone Power Plant use is largely nonconsumptive, a majority of the water is immediately returned to the stream, sustaining an important part of the downstream recreation economy (rafting, kayaking, and fishing). A key aspect of the Shoshone Power Plant call is that the decreed water needs to be put into beneficial use. The 100-year-old plant had to shut down for repairs and maintenance in 2007 and in 2019 (Gardner-Smith, 2019; Proctor, 2008), and when not operating, its right does not have to be honored. This can significantly affect downstream users that benefit from the water called by the plant, as well as the upstream projects that could potentially use the water if their water right priority allows. For these reasons, the potential effects of its nonoperation were included in our study, since it could impact local stakeholders.

Another notable feature is a 24-kilometer river segment, referred to locally as “the 15-mile reach,” extending from Palisade to the confluence of the Colorado River with the Gunnison River in Grand Junction (Figure 1). This reach is considered critical for the recovery of four endangered fish species: the razorback sucker (*Xyrauchen texanus*), the Colorado pikeminnow (*Ptychocheilus lucius*), humpback chub (*Gila cypha*), and bonytail (*Gila elegans*) (IUCN, 2012a, 2012b, 2012a, 2012b; USFWS, 2012b). As a result, the U.S. Fish and Wildlife Service has made recommendations with regard to necessary flow in this reach, depending on annual hydrologic conditions (USFWS, 1999; a more detailed discussion of these recommended flow rates is presented in section 4). Following these recommendations, state agencies and other stakeholders have been particularly interested in the effects of the seniority of the flow right at this location.

All of the critical river segments as well as the consumptive and nonconsumptive water users are represented within the UCRB StateMod platform. The model simulates stream flow, stream diversions, instream flow demands, and reservoir operations as determined by the basin’s hydrology, the decreed water rights in the geographical area of the basin, and their associated structures and operating rights. Each water right is

defined by an administration number representing its seniority, a decreed water flow, and a stream location. The model then resolves river flow and operations for each (monthly) time step in order of water right priority. All consumptive use of water that takes place in the basin is accounted for and represented. For reasons of practicality, the model does not represent the several thousand diversions that take place in the basin explicitly but does represent 75% of all use in the basin in strictly correct locations. The remaining structures are grouped into aggregates based on use location, tributary boundaries, and instream flow reaches. As a result, the UCRB model includes over 400 explicit structures and 65 aggregate structures, illustrated as points in Figure 1. Likewise, the UCRB model does not explicitly represent every single reservoir and stock pond existent in the basin; only those with decreed capacities of above 4,934,000 m³ (4,000 acre feet) are represented. The reservoirs explicitly represented at strict locations (18 in total) make up 94% of the total storage capacity in the basin. The remaining storage capacity is represented using ten aggregate reservoirs and one aggregate stock pond. Additional information on how the aggregation of structures was performed during model development can be found in the manual for the UCRB model (CWCB & CDWR, 2016).

Demands at these structures are defined by best-estimate levels of irrigation, population, and reservoir capacity up to 2010. Irrigation structure water demands are not limited by water right but are set to the irrigation water requirement for the specific time step divided by the historical efficiency for that month of the year. Historical efficiencies reflect irrigation practices and application methods, conveyance losses, and timing of the diversion (e.g., early versus late season). Municipal structure demands and transbasin diversion demands are based on average diversions per month of the year, during the period of 1998–2005. The 15-mile reach is represented by a minimum streamflow demand that varies monthly and annually based on hydrologic conditions. Reservoir demands are modeled using minimum and maximum reservoir target storage limits. Maximum targets for reservoirs that serve primarily agricultural and municipal diversions are set to capacity. Maximum targets for reservoirs that include operations for hydropower or flood control purposes are set to operational targets according to rule curves provided by the reservoir operating agencies. Additional information on how historical water demands were estimated and represented for all consumptive use in the basin can be found in the model's manual (CWCB & CDWR, 2016).

When simulating water allocations and reservoir operations, the model needs to first represent natural flows—the flow that would have been in the stream had none of the diversions and reservoir operations taken place. Natural flows are estimated by superimposing historical diversion data, end-of-month contents of modeled reservoirs, and estimated consumption and return flows on historical streamflow data from U.S. Geological Survey (USGS) gauges. For the ungauged locations represented in the model, flow is distributed using proration factors that account for the fraction of the drainage area contributing to each gauged natural flow. The record represented in the model spans water years 1909–2013, which includes extended dry and wet periods, as well as years with high runoff and extreme drought. The term water year refers to the 12-month period starting on October 1st and ending on 30 September the following year.

Water year 2002 (spanning October 2001 to September 2002) is a recent severe drought period with exceptional water shortages across the state of Colorado (Pielke et al., 2005). Figure 2 presents the historical shortages experienced by six users in the basin, highlighting those experienced during water year 2002: (a) a senior-right irrigation user, (b) a median-right irrigation user, (c) a junior-right irrigation user, (d) a senior irrigation user with a large decree of water allocation, (e) a transbasin diversion, and (f) the 15-mile reach. Each black line in each panel represents a water year in the historical record of 1909–2013, with 12 points each, representing each month. Monthly shortages for each water year are sorted by magnitude, with the values at the far left of each panel showing the minimum monthly shortage for each water year (0 m³ for all users), and the values at the far right showing the maximum monthly shortage for each water year. Water Year 2002 is shown in red. The six users presented are selected to represent a range of activities, geographic locations, right seniority, and demand levels. The respective decreed flows for each user are (a) 0.47 m³/s (1.22 × 10⁶ m³/month), (b) 0.35 m³/s (0.91 × 10⁶ m³/month), (c) 0.96 m³/s (2.49 × 10⁶ m³/month), (d) 26.63 m³/s (69.02 × 10⁶ m³/month), (e) 8.5 m³/s (22.03 × 10⁶ m³/month), and (f) 24.95 m³/s (64.67 × 10⁶ m³/month). User (a) is located upstream in the basin, diverting water from Blue River. User (b) is located downstream in the basin, diverting water from Roan Creek in Garfield County, and users (c, d), also located downstream, are diverting from Lower Colorado River. The transbasin diversion (e) is located midstream, diverting from Eagle River. Finally, the 15-mile reach (f) is located in the Lower Colorado River (also indicated in Figure 1). The experienced shortages throughout the record and particularly during 2002 indicate large differences in how shortages are experienced by users in the basin with regard to their frequency and

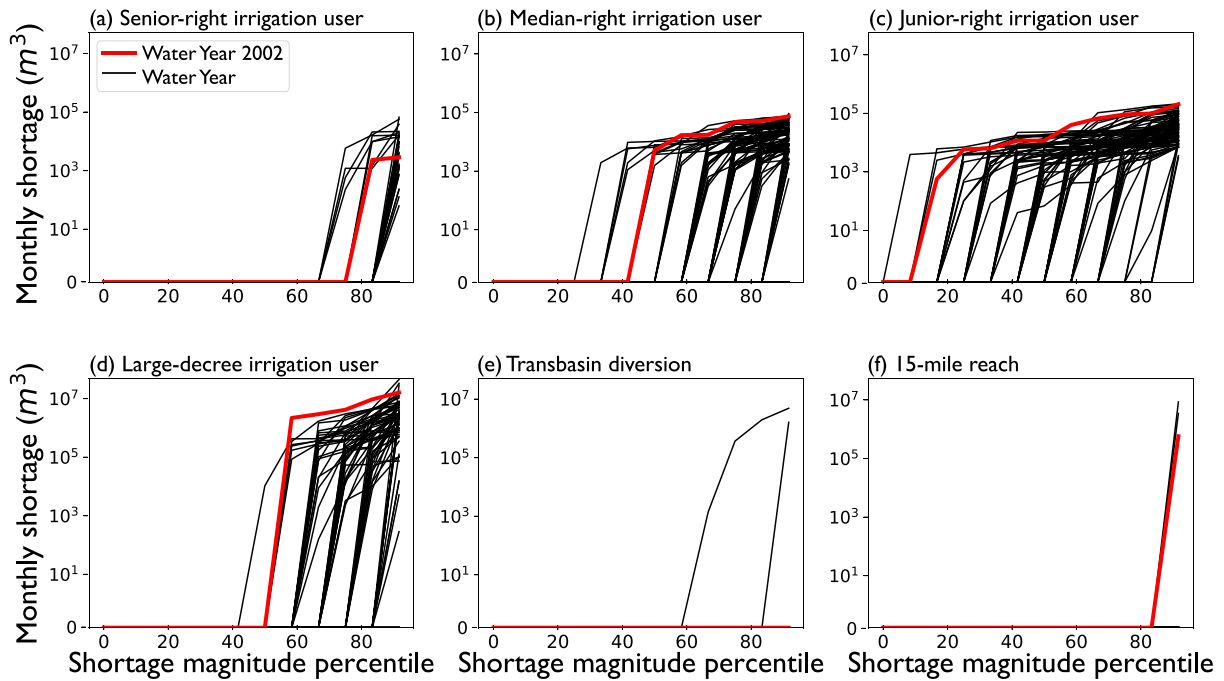


Figure 2. Historical monthly shortages for six users in the basin. Every line in each panel represents monthly shortages for every water year of the historical record of 1909–2013, with shortages during water year 2002 indicated in red. Each panel represents shortages experienced by a specific user in the basin: (a) a senior-right irrigation user, (b) a median-right irrigation user, (c) a junior-right irrigation user, (d) an irrigation user with a large decreed flow of water, (e) a transbasin diversion, and (f) the 15-mile reach.

magnitude. Specifically, panels (a)–(c) show three irrigation users at three different levels of right seniority, with the most senior user (a) experiencing shortages at most three months out of a year and the most junior user (c) experiencing shortages for up to 11 months (indicated by the percentile of magnitude per year). The shortages experienced during the 2002 water year appear to be some of the most volumetrically severe in the record for the median- and junior-right irrigation users (b and c), the large-decree irrigation user (d), and the 15-mile reach (f). However, the senior-right irrigation user (a) experienced severe shortages for only two months, and they were an order of magnitude lower than the user’s worst recorded shortages. The transbasin diversion (e) did not experience shortages during that year. These data suggest that these users are vulnerable to different drought characteristics. Resolving the relative severity of a particular drought for each user requires a careful accounting for how their demands and place in the basin’s rights hierarchy amplify or modulate shortages of varying timing, duration, frequency, and magnitude.

3. Methodological Framework

3.1. Experimental Design

This study contributes a bottom up exploratory assessment for how vulnerable UCRB users are to changes in droughts, demands, institutional rights, and physical storage. Individually and collectively the changes in these factors are considered deeply uncertain. This experiment identifies which of these factors are most critical for different users through a three-step process: (1) generating alternative “states of the world” (SOWs) represented by different combinations of these uncertain factors and 10 streamflow realizations for each SOW, (2) simulating StateMod over these alternative realizations to compute different users’ shortage distributions in each world, and (3) using alternative definitions of robustness in combination with scenario discovery to determine which combinations of changes in the uncertain factors lead to poor performance for different users. Our experimental design is presented in Figure 3, using notation consistent with McPhail et al. (2018). Panel I illustrates Step (1): We generate 1,000 samples of uncertain factors (Ψ), representing the set of SOWs, each producing ten streamflow realizations (s). Step (2) is illustrated in Panels II and III. The performance of the system is evaluated for each realization using StateMod (Panel II) producing shortages for all users (U) in the basin ($f(U, s)$). Panel III shows how $f(U, s)$ is a set of performances for each user (u) across all realizations ($f(u, S)$). Step (3) is illustrated in Panels IV and V. The performance for each user

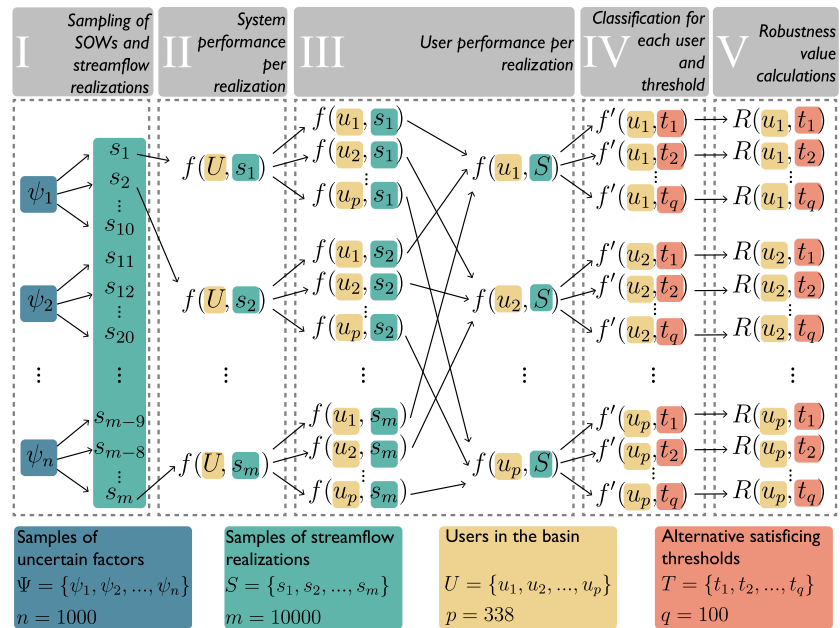


Figure 3. Experimental design of this study. Panel I shows how the samples of uncertain factors (Ψ) relate to the generation of streamflow realizations (s), that is, 10 streamflows for each sample Ψ . The performance of the system is evaluated for each realization using StateMod (Panel II) producing shortages for all users (U) in the basin ($f(U, s)$). The performance for each user (u) across all realizations (S) is then represented by $f(u, S)$ (Panel III). Panel IV shows how that performance is classified using alternative satisfying thresholds t , resulting in an individual robustness value R for each user and satisfying threshold (Panel V). This figure follows the same mathematical notation for robustness analyses as defined by (McPhail et al., 2018).

is classified using alternative satisfying thresholds t (Panel IV), each resulting in an individual robustness value R for each user and satisfying threshold (Panel V).

Table 1 lists the 14 deeply uncertain factors that are sampled in this experiment: eight hydrologic parameters, three demand parameters, and three parameters related to environmental, structural, or institutional changes. To capture changes in the hydrologic factors, a two-state Gaussian Hidden Markov Model (HMM) is used, as detailed in section 3.1.1. A Latin Hypercube sample (McKay et al., 1979) of 1,000 parameter combinations (SOWs) is generated across the ranges given in Table 1 assuming uniformity and independence. This is represented by set Ψ in Figure 3. While these parameters are likely correlated and not uniformly distributed in reality, this bottom-up analysis is simply intended as a screening analysis that identifies consequential parameter combinations. As such, the ranges assumed for these factors are intentionally wide for exploratory purposes. Given the complex and high-dimensional nature of the vulnerability assessment, our experiment shifts focus from predicting the likelihood of a given SOW to instead characterizing which of those SOWs are consequential to users. Users can then use the dimensionally reduced consequential scenarios in an a posteriori assessment of which changes are of most concern. We have grounded our experimental bottom-up sampling design in detailed recent assessments and in collaboration with basin management experts. We present the details of our sampling and analysis steps below.

3.1.1. Alternative Hydrologic Conditions

Managing climate variability and its socioenvironmental impacts is a familiar challenge to water users in the UCRB. The region is currently in the midst a nearly 20-year drought in which streamflows have decreased despite increased precipitation, as increasing temperatures have resulted in greater evapotranspiration and earlier snowmelt, reducing summertime flows (Xiao et al., 2018). Stakeholders in the basin are beginning to wonder if this is the “new normal,” or worse, only the beginning of a downhill trajectory. These concerns are unfortunately consistent with many regional climate projections that on average predict decreased deliveries to Lake Powell at the outlet of the entire Upper Colorado River Basin (Christensen & Lettenmaier, 2006; Christensen et al., 2004; Rasmussen et al., 2014). However, there is great variability across these projections, making the mean an inappropriate projection that masks the true deep uncertainty (Harding et al., 2012). This variability is as great from decade to decade within individual general circulation model (GCM) runs

Table 1
Uncertain Factors and Sampling Ranges

Parameter	Current value	Lower bound	Upper bound
<i>Hydrologic factors</i>			
Change in evaporation (cm/month)	0.0	−15.24	30.48
Shift in timing of snowmelt (days earlier)	0	0	60
Log-space dry flow mean (m ³) multiplier	1.0	0.98	1.02
Log-space dry flow standard deviation (m ³) multiplier	1.0	0.75	1.25
Log-space wet flow mean (m ³) multiplier	1.0	0.98	1.02
Log-space wet flow standard deviation (m ³) multiplier	1.0	0.75	1.25
Change in dry-to-dry transition probability	0.0	−0.3	0.3
Change in wet-to-wet transition probability	0.0	−0.3	0.3
<i>Demand factors</i>			
Irrigation demand multiplier	1.0	0.5	1.5
Transbasin demand multiplier	1.0	0.5	1.5
Municipal and industrial demand multiplier	1.0	0.5	1.5
<i>Environmental and institutional factors</i>			
Reservoir storage	1.0	0.8	1.0
Operation of Shoshone Power Plant	1	0	1
Seniority of environmental flows	0	0	1

Note. Using a Latin Hypercube Sample, 1,000 parameter combinations (SOWs) are generated, under which the system performance is evaluated for every user.

as across them (Harding et al., 2012). Such variability may be the result of significant climatic persistence in the basin leading to long wet periods and long dry periods, such as the current bidecadal drought. Paleoreconstructions in the basin suggest these decadal, or even multidecadal droughts, are not uncommon in the UCRB (Ault et al., 2013, 2014; Woodhouse et al., 2006).

Given that extreme hydrologic variability and persistence has been observed historically and is projected in the future, our exploratory analysis exploits a synthetic streamflow generator that can explicitly model the effects of potential changes in drought frequency, severity, and persistence. Following Bracken et al. (2014), we capture these characteristics with a two-state Gaussian HMM fit to the log-space natural annual flows at the last node in StateMod, the CO-UT state line. This model consists of two “hidden” climate states representing wet and dry conditions, respectively. HMMs can have more states, but Bracken et al. (2014) found two to be optimal in this system. Within each of these states, log-space flows are generated from Gaussian distributions with different parameters. We observe the flows generated from these distributions, but not the states themselves, hence the term “hidden.”

Fitting the Gaussian HMM requires estimation of the mean and standard deviation of each of these Gaussian distributions, as well as the probabilities of transitioning between wet states and dry states. Changes in these transition probabilities can capture changes in drought persistence, changes in the mean and standard deviation of the distributions can capture changes in severity, and changes in both sets of parameters can collectively capture changes in drought frequency. Bracken et al. (2014) show that these parameters, in particular the transition probabilities, are nonstationary and can be estimated as a function of large-scale climate phenomena such as the Pacific Decadal Oscillations (PDO) and El Niño–Southern Oscillations (ENSO). Thus, by changing these parameters in our scenario discovery experiment, we can determine how sensitive users are to changes in these large scale climate phenomena.

Given known nonstationarity in hydrologic conditions in the UCRB, we fit the Gaussian HMM to only the last two thirds of the 105-year record (70 years) to capture recent conditions. This is also consistent with the 64-year records that have been used for future climate simulations in the basin (CWCB, 2012). The parameter estimation for the Gaussian HMM was performed using Expectation-Maximization with Python's `hmmlearn` package (Lebedev, 2015). The parameter estimates are reported in supporting information (SI) Table S1 and the Gaussian fits are shown in SI Figure S1. Note that the parameters reported in Table S1

are based on fitting the Gaussian HMM to the log of annual flows in m^3 but that the multipliers in Table 1 are applied to parameters estimated from fitting the Gaussian HMM to the log of annual flows in acre feet. Across the Latin Hypercube sample, this results in real-space multipliers of about 0.73–1.37 for the dry-state mean, 0.64–1.54 for the dry-state standard deviation, 0.72–1.39 for the wet-state mean, and 0.63–1.58 for the wet-state standard deviation. A visual depiction of how these streamflow parameter ranges map to changes in the real-space annual flow distribution is provided in SI Figure S3.

State identification of wet and dry years, shown in Figure S2, was performed using the Viterbi algorithm with the same Python library. As can be seen from this figure and the transition probabilities in Table S1, there is considerable persistence in the system under recent historical conditions, with the probability of wet-to-wet and dry-to-dry transitions ranging from 65–68%. More details on the fitting method and validation of the generator are provided in the SI.

For our exploratory modeling experiment, we synthetically generate log-space annual flows at the last node from a two-state Gaussian HMM for parameterized changes from the historical record by modifying the historical transition probabilities with delta shifts and the historical means and standard deviations with multipliers over the ranges in Table 1. The log-space annual flows are converted to real space and temporally downscaled to monthly flows using a modification of the proportional scaling method used by Nowak et al. (2010). First, a historical year is probabilistically selected based on its “nearness” to the synthetic flow at the last node in terms of annual total. The daily hydrograph observed at that site in the selected year is then altered to model earlier, dissipated peaks, controlled by a peak shift parameter listed in Table 1. This change is intended to capture the effects of increasing temperatures, exacerbated by dust on snow, on reducing snow cover duration by several weeks and bringing peak runoff earlier in the year (Livneh et al., 2015; Painter et al., 2012; Skiles & Painter, 2019). The shifted monthly flows at the last node are then downscaled to all other nodes using the same ratios of monthly flows at the upstream nodes to the last node as in the historically selected year. A detailed explanation of the methods used for the seasonal shift, as well as validation of the generator's ability to capture spatial correlation with this approach, is provided in the SI, with links to the code for the synthetic streamflow generator.

Combining changes sampled in our 1,000 SOWs across all of these seven streamflow parameters, we observe a wide range of streamflows beyond what has been observed historically. For every 1 of the 1,000 parameter samples (ψ), 10 realizations of synthetic 105-year streamflow records are generated (s) using that SOW's streamflow parameters, resulting in ten thousand 105-year sequences making up the entire ensemble S (Figure 3). Figure 4 shows the ranges of (a) monthly and (b) annual flows at the last simulated node in the basin across these 10,000 simulations compared with other observed and synthetic series. The range of historical flows is shown in light red, with the extreme drought year of 2002 indicated in dark red. Shown in dark cyan are synthetically generated flows for 10 sequences of 105 years using the stationary parameterization of the HMM, fit to the last 70 years of the historical record. Shown in green and yellow are 112 and 97 projected streamflows associated with downscaled and bias-corrected precipitation and temperature projections from the Coupled Model Intercomparison Project 3 (CMIP3) and 5 (CMIP5), respectively. These scenarios were developed through the *Colorado River Water Availability Study* (CWCB, 2012). In that study, the downscaled projections were used as forcing to the Variable Infiltration Capacity (VIC) model to quantify the effect of climate changes on naturalized streamflows in the basin. Differences between VIC streamflow projections under historical and projected climates were applied to natural historical flows to create the scenarios run through StateMod, which are what is plotted in Figure 4. Finally, the flows shown in gray in Figure 4 are the synthetically generated streamflows across all SOWs used in this study. The ranges of the parameters sampled were chosen so that the resulting flows would span beyond both the (a) monthly and (b) total annual flows of historical, stationary synthetic, and CMIP-projected flows, both for wetter and drier conditions.

The final climate-related parameter sampled in our scenario discovery experiment is the reservoir evaporation rate, which is expected to increase with rising temperatures. StateMod represents reservoir evaporation using net average monthly evaporation rates (feet/month) assigned to each modeled reservoir. Historical net annual evaporation rates were estimated by subtracting weighted average effective precipitation from estimated gross water surface evaporation. The annual estimates of evaporation were distributed to monthly values based on the elevation of each reservoir. More details on how the values were estimated and their distributions can be found in the manual for the UCRB model (CWCB & CDWR, 2016). To represent changes

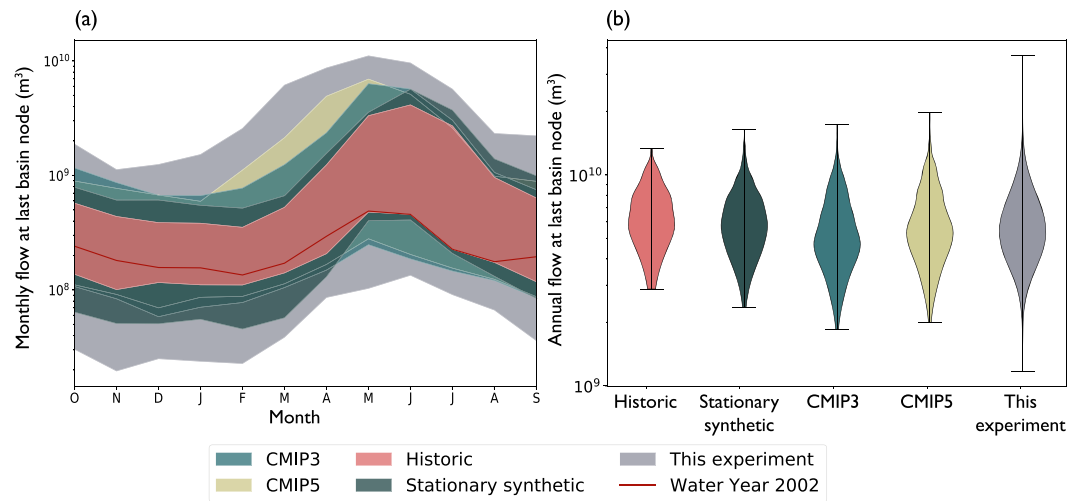


Figure 4. Range of (a) monthly and (b) annual flows at the end of the basin. Observed historic flows (soft red), stationary synthetic (soft cyan), CMIP3 downscaled (green), CMIP5 downscaled (yellow), and nonstationary synthetically generated flows used in this study (gray). The 2002 drought year is also indicated in dark red.

in evaporation within the experiment, we sample additive changes to the monthly evaporation rates and apply them to all modeled reservoirs.

3.1.2. Scenarios for Changing Demands

Colorado's population has seen a fivefold increase in the past century, a growth trend that is expected to continue (DOLA, 2015). Maximum projection levels anticipate population to double by the mid-21st century, further increasing municipal and industrial water demands. Even though population growth is inevitable, state and local governments can influence how and where the population grows (State of Colorado, 2015), as well as how much water is actually needed to support this growth. Since 2000, conservation and efficiency programs implemented throughout the state have reduced per capita water needs by as much as 30% in some communities (CWCB, 2010). Additional conservation measures planned throughout the state are expected to further reduce industrial and municipal per capita water demand (State of Colorado, 2015).

Throughout Colorado and in the UCRB particularly, agriculture represents the largest sectoral water demand. Increasing temperatures are expected to increase the number of days in the growing season for perennial crops (most predominant in the UCRB), as well as the crop demand for irrigation water. As a result, crop irrigation requirements may increase by up to 30% by 2040 (CWCB, 2012). In contrast, the CWCB expects total irrigated acreage to decline, as many municipalities meet their increasing demands by acquiring agricultural water rights. Real estate developers have also been purchasing farmland to expand urban areas. Reduced irrigated land might result in crop irrigation requirements reducing by up to 20% by 2050 (State of Colorado, 2015).

Given these conflicting estimates, it is unclear exactly how water demands across these sectors might change in the future. In this study, we explore the effects of both growth and decline in each type of demand in the UCRB (municipal and industrial, transbasin, and irrigation) of up to 50%. For transbasin diversions, we apply these rates of growth or decline to their maximum historical monthly demands observed in the historical record, considering this the ideal amount of supply. It should be noted that any increase in demand is limited by the flow and storage rights associated with each diversion. As we are not considering the purchase of additional rights by the users in this study, the shortages reported in section 4 are in part due to the fact that the demands might be exceeding the right associated with each user, not necessarily due to water being unavailable. In reality, users, and especially the large municipal users diverting water to the eastern plains, could consider the purchase of additional water rights, among other measures, to meet their growing demand.

However, within this exploratory experiment, streamflows and irrigation demands should be anticorrelated, as evapotranspirative demands increase in drought years when flows are low. To ensure consistent irrigation and streamflow time series within each SOW, we compute the synthetic annual flow anomaly at the

last node in every simulated year. We then use a regression of historical annual total irrigation anomalies versus historical annual flow anomalies at the last node to determine an appropriate annual total irrigation anomaly for each year of that simulation (with added noise to preserve variance). We then add this time series of irrigation demand anomalies to the mean irrigation demand for that SOW. These total irrigation demands are then distributed to all of the irrigation structures using their average historical proportions of the total demand. Additional details on the regression used to anticorrelate streamflow anomalies and irrigation anomalies are provided in the SI.

3.1.3. Environmental and Institutional Changes

The last three factors sampled in our exploratory modeling experiment represent environmental and institutional changes in the basin with potentially consequential impacts to some users. The first is an environmental factor related to erosion and sedimentation of reservoirs in the UCRB, resulting in reduced storage. As shown in Table 1, for each SOW ψ we sample potential losses of up to 20% of capacity, a number informed by Graf et al. (2010).

We sample the institutional effect of whether the Shoshone Power Plant continues its operation using a binary variable indicating whether or not it maintains its “beneficial use” of water. If the Shoshone Power Plant shuts down (Gardner-Smith, 2019; Proctor, 2008), its decreed volume of water can no longer be put into beneficial use, allowing otherwise out-of-priority users to take advantage of the available water instead. However, junior users downstream of the Shoshone Power Plant might be negatively affected, as water that would have otherwise been delivered by the Shoshone call, can now be diverted to more senior upstream users.

The final institutional change we sample is the effect of changing the flow right at the 15-mile reach to be a senior right. This is again a binary factor where a sampled value of 1 enforces the potential legal change of assigning the 15-mile reach a first priority senior right with a decree of 22.94 m³/s (810 cf/s, the minimum flow rate recommended during dry conditions USFWS, 1999). Including this binary variable in our experiment allows us to investigate the effects of legally prioritizing the system's ability to meet the minimum flow requirement.

3.2. Robustness Analysis and Scenario Discovery

To assess robustness, one needs to define the outcome of interest, measured by some performance metric (e.g., volume of delivered water and cost), and aggregate the value of that performance metric across the scenarios being considered (McPhail et al., 2018). This paper explores how different definitions of acceptable water shortages affect relative robustness values across users (Panels IV and V in Figure 3). In this case, we focus on satisficing as the primary measure.

The domain criterion satisficing metric (Starr, 1962) has been widely used in the robustness literature and quantifies the fraction of scenarios in which a desired performance is met. The desired performance is typically defined by some performance threshold (e.g., a maximum budget if performance is concerned with cost); the greater the number of scenarios in which this threshold is met, the more robust the system. Satisficing metrics are particularly applicable to water supply problems, as water supply management tends to be concerned with assuring that the necessary quantities of water are provided (e.g., to sustain agricultural production), rather than maximizing water consumption or minimizing deficit (Beh et al., 2017; Hall & Borgomeo, 2013). Using this robustness metric, we investigate how different performance thresholds affect both the interpretation of robustness and the scenarios under which the system performance is considered acceptable. Since the UCRB is a system of diverse stakeholders, we perform this analysis for all users represented in StateMod.

The different performance thresholds used in this study have been defined using combinations of magnitude and frequency of shortage. Magnitude of shortage is defined as the percentage of user demand that is not met, while frequency is defined as the percentage of time (in months) that the user experiences such shortage. Both measures of shortage (magnitude and frequency) are binned by percentiles with a width of ten (e.g., one bin might include 0th to 10th percentile of shortage magnitude and 10th to 20th percentile of shortage frequency), resulting in 100 combinations of performance thresholds for each user (t in Figure 3). Applying these performance thresholds for each user (indicated as $f'(u, t)$ in Figure 3) results in a satisficing

surface for each user, indicating their robustness value as determined by all threshold combinations (indicated as $R(u, t)$ in Figure 3). Scenario discovery is then performed for select performance thresholds for each user, informed by the frequency with which they historically experienced shortage of that magnitude.

Scenario discovery is most typically performed using the Patient Rule Induction Method (PRIM) (Friedman & Fisher, 1999) or Classification and Regression Trees (CART) (Breiman, 1984). One of the main criticisms of these methods is that they rely on orthogonal subspaces to classify the parameter space into success or failure, making their application to nonlinear systems inadequate (Kwakkel, 2019). Logistic regression is a common technique in binary classification problems that overcomes this limitation by using a nonlinear function to describe the probability that a SOW belongs to the scenarios that lead to failure (Lamontagne et al., 2019; Trindade et al., 2019). This also allows users to define regions of success based on the estimated probabilities of success in those worlds, a more intuitive way of capturing different levels of risk aversion than coverage and density used in PRIM. Following several recent examples (Gold et al., 2019; Lamontagne et al., 2019; Quinn et al., 2018; Trindade et al., 2019), we use logistic regression to identify combinations of uncertain factors under which each performance criterion is likely to be met. Logistic regression provides us with an estimate of this probability of success as a function of combinations of uncertain factors. As the factors considered in this study are assumed to be deeply uncertain, the probability that a parameter combination (or SOW) will occur cannot be known. However, with logistic regression we can predict the probability that, should a SOW occur, the performance threshold will be met. The resulting probability estimates can then be used to discover parametric combinations (scenarios) that lead to failure. The scenario discovery procedure is performed independently for each user and is detailed below.

For each performance threshold (combination of shortage magnitude and frequency), several logistic regression models are built using a different covariate (uncertain factor) each time. Every model estimates the probability that a SOW will be classified as a success (1) or a failure (0) in meeting the performance threshold, given the value of the uncertain factor value in that SOW. Each logistic regression model is defined by

$$\ln \left(\frac{p_i}{1 - p_i} \right) = \mathbf{X}_i^T \boldsymbol{\beta} \quad (1)$$

where p_i is the probability the performance in the i th SOW will be classified as a success, \mathbf{X}_i is the vector of covariates describing the i th SOW, and $\boldsymbol{\beta}$ is the vector of coefficients describing the relationship between the covariates and the response, estimated using maximum likelihood estimation. Iterating through all possible covariates, one at a time, we identify those with the highest McFadden's pseudo- R^2 . McFadden's pseudo- R^2 is given by

$$R_{McFadden}^2 = 1 - \frac{\ln \hat{L}(M_{full})}{\ln \hat{L}(M_{intercept})} \quad (2)$$

where $\ln \hat{L}(M_{full})$ is the log likelihood of the full model (model with predictor) and $\ln \hat{L}(M_{intercept})$ is the log likelihood of the intercept model (model with no predictor besides the intercept). The intercept model predicts the mean probability of success across all SOWs, and McFadden's pseudo- R^2 is a measure of improvement of the model with the predictor variable over the intercept model. Higher McFadden's pseudo- R^2 values can therefore be used to rank all candidate covariates in terms of significance in predicting success or failure (Quinn et al., 2018). After identifying the top two individual predictor variables for the performance threshold being modeled, we then build a new logistic regression model with both as covariates. While this approach could be problematic with historic observations, multicollinearity is not a concern in this experiment since all uncertain factors being used as covariates were sampled independently in the Latin hypercube design. From our new bivariate logistic regression model, a probability plane can be plotted to create a "factor map" illustrating the probability of success given the values of the two variables. An interaction term was also added to the logistic regression model, to capture potential interactions between the top two factors. The factor maps show how users' sensitivities to the deeply uncertain factors differ depending on the satisficing threshold (i.e., shortage magnitude and frequency) used in the robustness definition.

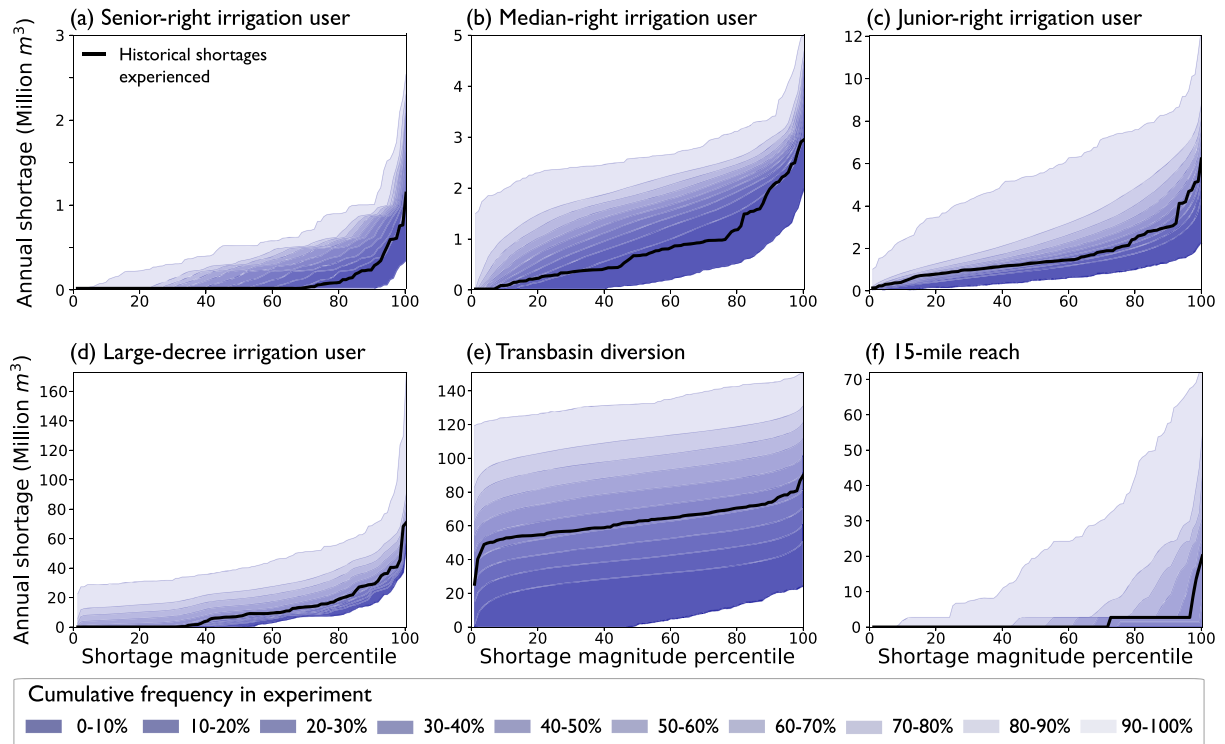


Figure 5. Annual shortages for six users in the basin. Each panel represents shortages experienced by a specific user in the basin: (a) a senior-right irrigation user, (b) a median-right irrigation user, (c) a junior-right irrigation user, (d) an irrigation user with a large decreed flow of water, (e) a transbasin diversion, and (f) the 15-mile reach. The black line in every panel indicates the percent of time different magnitudes of annual shortage were experienced historically for each user. The shaded areas represent the frequency with which different magnitudes of shortage were experienced at each percentile across the simulated ensemble. Lighter shades indicate increased cumulative frequency within the ensemble.

4. Results and Discussion

4.1. Impacts of Changing Conditions on UCRB Users

Across the synthetic ensemble of 10,000 realizations (each one encompassing 105 years of streamflow), water users in the UCRB are affected in very different ways by deeply uncertain changes in streamflows, demands, and institutions, experiencing broadly different shortages (unmet demands) over time and across simulations. Figure 5 shows annual shortages experienced historically and across the ensemble for the same users presented in Figure 2 (i.e., $f(u, S)$ for each user in Figure 3). The black line in every panel indicates magnitude of annual shortage (y axis) and the percent of time it was experienced historically (x axis) for each user. The shaded areas represent the frequency with which shortages of different magnitudes were experienced at each percentile across the ensemble. For example, looking at panel (Figure 5e), the annual shortage at the 20th percentile historically was approximately 50M m^3 (40,536 acre feet). Within each realization of the ensemble experiment, annual shortages at the 20th percentile ranged between near 0 and 120M m^3 (9,729 acre feet). Lighter shades within these areas indicate increased cumulative frequency within the ensemble.

Looking across the panels in this figure, it is apparent that users in the UCRB experience diverse impacts as a result of these changing conditions. For the senior- and median-right irrigation users (Figures 5a and 5b), historical experiences appear to lie at the lower bands of cumulative frequency in the ensemble (0–10% and 10–20%), meaning that in the majority of realizations simulated (80–90% of simulated SOWs) they experience higher and more frequent shortage magnitudes across all percentiles of shortage. The historic shortages for the large-decree user (irrigation right with the largest decreed flow of water) and the transbasin diversion (Figures 5d and 5e) lie in the 50–60% and 40–50% bands of cumulative frequency for all shortage magnitude percentiles, respectively. This means that in approximately half the SOWs, shortage magnitudes and frequencies increase, whereas in the rest they are reduced, in some cases down to 0. For the 15-mile reach (Figure 5f), the historic shortages lie near the 60% band of cumulative frequency, which means that in most

SOWs the shortages for this diversion are reduced in both magnitude and frequency (to no shortages at all for the entire 105-year sequence). One also notes that deducing relative sensitivities of the six users to changing conditions based on historical observations, reported in Figure 2 in monthly magnitudes (also in black lines in Figure 5), would have been misleading. For example, in Figure 2a, the senior-right irrigation user appears to have historically had less frequent shortages (maximum three months in a year) of smaller magnitudes—less than 100,000 m³ (or 81 acre feet) per month—with the 2002 drought year only having a moderate impact on their water yield. The large-decree irrigation user, on the other hand (Figure 2d), had experienced shortages for up to six months in a year, with magnitudes over 10,000,000 m³ (8,107 acre feet) per month. Yet the historically more vulnerable user (Figure 5d) appears less sensitive to changing conditions in Figure 5 than the historically less vulnerable user (Figure 5a). This highlights that if the state is concerned with addressing the vulnerabilities of various users in the basin, their past performance is not necessarily a direct indicator of their vulnerability under changing conditions.

With the exception of the large-decree irrigation user and the transbasin diversion (Figures 5d and 5e), it is also apparent that the magnitude impacts are not uniformly distributed across shortage percentiles. Looking at panel (Figure 5c), the variance in absolute magnitudes of shortage across the ensemble appears to be much larger for higher percentiles of shortage than for lower percentiles. The inverse appears to be true for the user in panel (Figure 5b). There is also significant variability in how the frequencies of shortage are affected. For example, for the senior-right irrigation user (Figure 5a), no shortages were experienced 70% of the time historically but are experienced up to 90% of the time in the worst-case ensemble member, making shortages up to three times more likely in that scenario. On the other hand, their worst historical shortage at the 99th percentile has moved to around the 90th percentile in the worst case scenario, making it up to 10 times more likely to be exceeded. The large-decree irrigation user (Figure 5d) historically experienced a shortage less than 40% of the time, but under the ensemble they experience some level of shortage at all times in the more challenging realizations. It is apparent from these comparisons that general inferences by users individually or aggregated by type would not sufficiently capture the nonlinear differences in how shortage magnitude and frequency may change for UCRB stakeholders. Water right seniority and level of extraction are also not indicative a priori of how the various users might be affected, as can be deduced by comparing across panels (Figures 5a–5c), and (Figure 5d) and (Figure 5e), respectively.

4.2. Translating User Impacts Into Robustness

To understand how robust different UCRB users are to the changing hydrologic conditions, human systems demands, and environmental institutional factors sampled in this study, the impacts in Figure 5 need to be transformed into a robustness metric, in this case, the domain criterion satisficing metric (Starr, 1962). As discussed in McPhail et al. (2018), the calculation of a robustness metric involves three steps: (i) a transformation of the shortages experienced by each user to an aggregate measure of performance over time in each simulation (e.g., number of periods with shortage, maximum shortage); (ii) a scenario selection over which robustness is to be assessed; and (iii) an aggregation of the metric in (i) across the scenarios in (ii). In this study, we use a binary transformation in Step (i) to define realizations as “successes” or “failures” if a performance threshold is met (i.e., $f^j(u, t)$ for each user and threshold). For Step (ii), we use the full set of realizations to assess robustness (10,000), and for Step (iii), we calculate robustness by counting the number of realizations that meet the performance threshold (i.e., the number of successes, $R(u, t)$ for each user and threshold). For each user a matrix of satisficing performance thresholds is used that encompasses 100 combinations of shortage magnitude and frequency. Each combination of magnitude and frequency of occurrence is a potential performance threshold over which robustness is evaluated.

Figure 6 presents these matrices of performance thresholds and the resulting robustness values for the same six users presented in Figures 2 and 5. Robustness for a shortage magnitude and frequency is calculated as the percentage of SOWs where shortages of that magnitude do not occur more often than the specified frequency. The magnitude and frequency of shortage are classified on their annual scales. For example, looking at Figure 6a (the upstream senior-right irrigation user) and taking the third square from the left in the top row: This performance threshold is defined by shortages of magnitude at least 30% of the annual demand occurring at a frequency of 10% of the years or fewer in a given realization. Realizations that meet this criterion are considered successes, while those that do not are considered failures. The color of the square indicates the percentage of realizations that are considered successes (in this case, approximately 30%). In each matrix, moving from the left to the right indicates performance concern with shortages of

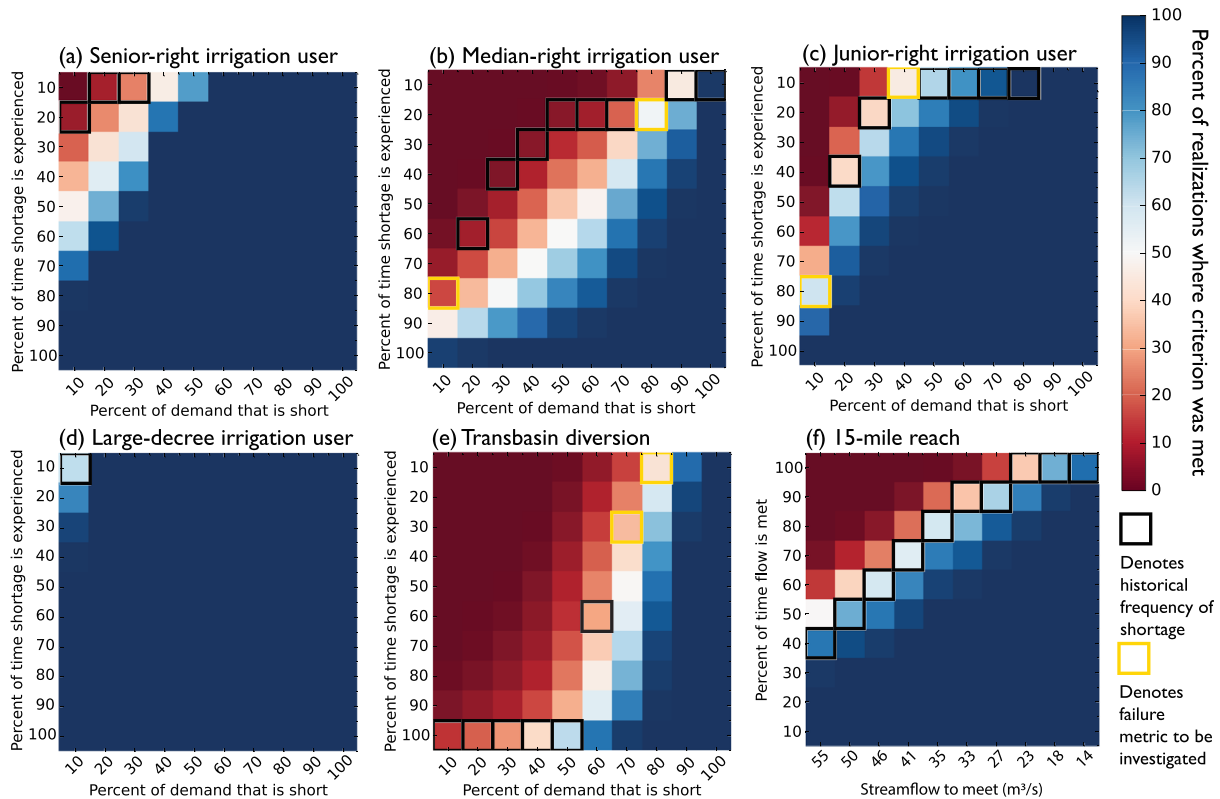


Figure 6. Threshold-varying robustness for six users in the basin. Each panel represents robustness for a different user in the basin: (a) a senior-right irrigation user, (b) a median-right irrigation user, (c) a junior-right irrigation user, (d) an irrigation user with large decree, (e) a transbasin diversion, and (f) the 15-mile reach. Users in panels (a)–(c) have small decreed flows, (d) and (f) have large decreed flows, and (e) has a moderate decreed flow. Robustness in each panel is calculated using a matrix of performance thresholds defined by combinations of shortage magnitudes and frequencies. Black borders indicate the historical frequency with which each magnitude of shortage has been experienced by each user. Yellow borders indicate performance thresholds to be used for scenario discovery.

higher magnitude, and moving from the top to the bottom indicates performance concern with shortages of higher frequency. Consequently, moving from the top left to the bottom right represents a shift in concern from low-magnitude, low-frequency shortages to high-magnitude, high-frequency shortages. This translates to moving from a focus on less severe droughts to far more extreme events (in terms of shortage magnitude and frequency).

Each user's potential willingness to accept a magnitude and frequency of shortage can be contextualized in each Figure 6 panel by the squares with black borders. Black borders indicate the historical frequency with which each magnitude of shortage was experienced by each user. To facilitate presentation, frequencies have been binned in brackets of tens: moving from the top to the bottom, the top row indicates frequencies of 1–10% of the time, 11–20% of the time, and so on. For example, looking at Figure 6a again, this user has historically experienced shortages at a magnitude of at least 10% of demand at a frequency of 11–20% of the time. They also experienced a shortage magnitude of at least 20% of demand at a frequency of 1–10% of the time, and a shortage magnitude of at least 30% of demand (one square to the right) at the same frequency interval. Shortages of magnitudes higher than 40% of demand were never experienced.

The median-right irrigation user (Figure 6b) experienced shortage magnitudes of up to 100% historically, albeit infrequently. A shortage magnitude of at least 10% of their demand took place 70–80% of the time. The same comparison for the large-decree irrigation user (Figure 6d), suggests that even though some shortage took place 60% of the time (Figure 5d), all of those shortages were of magnitudes below 10% of demand. The transbasin diversion (Figure 6e) appears to have experienced shortages throughout the historical record (matching the corresponding Figure 5e), which were at magnitudes of up to 50% of demand in magnitude. Shortages of higher magnitudes were also experienced at reduced frequencies. Finally, results for the 15-mile

reach are presented slightly differently (Figure 6f). Moving from left to right, each column indicates decreasing flow rates to be met in the river, and moving from top to bottom, each row indicates the percent of time that flow rate is met. These discrete flow rate levels were informed by the U.S. Fish and Wildlife Service recommendations for 46, 35, and 23 m³ per second (or 1,630, 1,240, and 810 cubic feet per second) under wet, average, and dry hydrologic conditions, respectively (USFWS, 1999). The minimum official requirement of 23 m³ per second was met 90–100% of the time historically. This flow rate was only met at the same frequency in approximately 40% of the sampled realizations (as indicated by the square's color). If alternatively, one chooses to have the same flow rate but only 60–70% of the time (i.e., move three squares down), the robustness value changes to 100% of the realizations being able to meet this performance criterion (square in dark blue).

Looking at the color gradients across all of the panels in Figure 6, robustness values vary significantly subject to the magnitude-frequency combination of threshold selected, as expected. Even though generally minimizing shortages is clearly a concern for all stakeholders in this system, this figure highlights the complexity of choosing a single magnitude-frequency combination threshold (e.g., to not have shortages of magnitude of 50% at a frequency of more than 20% of the time) that would be appropriate and relevant across the diverse stakeholders in the UCRB. Additionally, using a single threshold for individual stakeholders or stakeholder groups is inappropriate as the relationship between the satisficing thresholds and robustness values is neither simple nor consistent across users. Across the panels in Figure 6, the slope of the transition plane from thresholds resulting in 0% to 100% robustness values (from deep red to deep blue) appears to be equally dominated by both the magnitude and frequency dimensions in Figures 6b and 6f but more strongly affected by its magnitude dimension in Figures 6c and 6e. This slope also appears to vary in steepness, for example, comparing the transition in Figure 6b with that in Figure 6f.

This complex relationship between performance thresholds and robustness would not be known a priori: It is highly dependent on the interactions between all the considered uncertainties (hydrologic, demands, and infrastructure), the institutional water rights allocation hierarchy, and physical constraints on water availability in different regions of the basin. For example, senior users located on small upstream tributaries may be more vulnerable than junior users located downstream on the mainstem, as water right seniority cannot protect against physical supply limitations. Even though history can provide some information on how these interactions impact different users (the black lines in Figure 5 and the black borders in Figure 6), our exploratory modeling illustrates the complex relationship between alternative magnitude-frequency performance thresholds and robustness values for each user, as shaped by the combination of their water right seniority, extraction levels, and location in the basin. For example, robustness values for the large-decree irrigation user (Figure 6d) appear to be insensitive to the choice of performance threshold, in contrast to all the other users in this figure. Differences are also observed in the range of robustness values experienced by different users in meeting their history-informed magnitude-frequency combinations (squares with black borders). These combinations represent users' previous experiences and could potentially be decision-relevant candidates as robustness performance metrics: In Figure 6a the robustness values at the historical thresholds range from 7% to 24% robustness, in Figure 6b they range from 3% to 99% robustness, in Figure 6c from 39% to 99%, in Figure 6e from 13% to 63%, and in Figure 6f from 35% to 88%. For the single history-informed threshold in Figure 6d, the robustness value is 62% (indicated by the light blue color).

Most importantly, the relative robustness ranking of these six users is not always stable across performance thresholds. For the sake of illustration, if one were to gauge the robustness of this system using “Robust if shortages of 10% of demand do not occur more frequently than 80% of the time” as the metric definition, the relative robustness of the four irrigation and the transbasin diversion would be the following (from most robust to least robust): $d > a > c > b > e$. If, instead, the robustness metric was selected to be “Robust if shortages of 30% of demand do not occur more frequently than 30% of the time,” their relative ranking would become: $d > c > a > b = e$ (notice relative rankings for (a) and (c) switching). Lastly, if robustness was assessed using “Robust if shortages of 80% of demand do not occur more frequently than 10% of the time,” their relative ranking would again switch to: $d = a > c > e > b$ (notice relative rankings of (b) and (e) changing). This finding further highlights how different robustness metrics, applied for all users in the basin, can yield different and unpredictable multiactor robustness inferences.

Table 2
History-Informed Robustness Metrics Used for Scenario Discovery

Performance metrics					
Median-right irrigation user					
(a) Robust if	10%	shortage occurs	(b) Robust if	80%	shortage occurs
no more than	80%	of the time	no more than	20%	of the time
Junior-right irrigation user					
(c) Robust if	10%	shortage occurs	(d) Robust if	40%	shortage occurs
no more than	80%	of the time	no more than	10%	of the time
Transbasin diversion					
(e) Robust if	70%	shortage occurs	(f) Robust if	80%	shortage occurs
no more than	30%	of the time	no more than	10%	of the time

Note. Two performance thresholds for the median-right, junior-right, and transbasin diversion user were selected to perform scenario discovery. The respective factor maps are presented in Figure 7.

Alternative robustness metrics might also yield different conclusions about which deeply uncertain factors most influence success and failure at those satisficing thresholds. We investigate these implications by applying scenario discovery to six of the history-informed robustness performance metrics, highlighted with a yellow border in Figure 6. There are two performance thresholds each for the median-right, junior-right, and transbasin diversion users, and they are listed in Table 2.

Exploiting logistic regression, each panel in Figure 7 plots the top two most important factors controlling robustness performance for a given magnitude-frequency threshold, along with their interaction (*most important factor* × *second most important factor*). In our factor mapping, the values of the most important factor are always plotted on the *x* axis and the values of the second most important factor are always plotted on the *y* axis in each panel. The most important factors were determined based on their predictive ability in the regression model. Their predictive ability is indicated by their McFadden's pseudo- R^2 values, which are provided in the SI for each panel in Figure 7. Points in each panel represent the sampled SOWs, in blue if at least 50% of the realizations for that SOW met the respective criterion in each panel (successes) and in red if less than 50% of the realizations met the criterion (failures). Using this method, we can map parametric regions of the uncertain factors to conditional probabilities of failure and success and use the model to predict the likelihood that a criterion will be met given a change in that factor. As changes in mean wet and dry flows are sampled independently, in approximately 6% of the sampled SOWs the values for the mean dry flows exceed those for the sampled mean wet flows. These samples have their parameter labels swapped during the scenario discovery part of the analysis, to avoid misclassifying the SOWs. The parametric area corresponding to those samples is not assigned conditional probabilities of success or failure (blank area in Figure 6c). Note that the mean wet and dry flow multipliers have also been converted to real space in the figure to aid interpretability.

The two alternative robustness performance metrics for the median-right irrigation user (Figures 7a and 7b) result in different robustness values (Figure 6b) and factor maps. Using the “10% demand-80% of the time” performance metric (Figure 7a), robustness for this user is 16% of the sampled SOWs (derived from square color in Figure 6b). Success and failure are first controlled by the shift in the timing of snowmelt and then by changes in the mean wet flow. The failure region for this metric is a large portion of the sampled space, with failures occurring along the entire range of snowmelt shifts (0–60 days earlier) and successes only occurring when combined with wetter worlds (higher mean wet flows). Using the “80% demand-20% of the time” performance metric (Figure 7b) the most important factor change is to the mean dry flow, followed by the shift in snowmelt. Higher magnitude shortages typically occur in drier years, the flow of which is in part controlled by changes in the distribution of mean dry flows. Use of this metric results in a 52% robustness value (derived from square color in Figure 6b), reflected in the now smaller failure region (red areas), characterized by drier low flows and snowmelt shift. Both these panels also show that interactions between shifts in snowmelt timing and changes in the mean wet and dry flows appear to further affect success in Figures 6a and 6b, respectively.

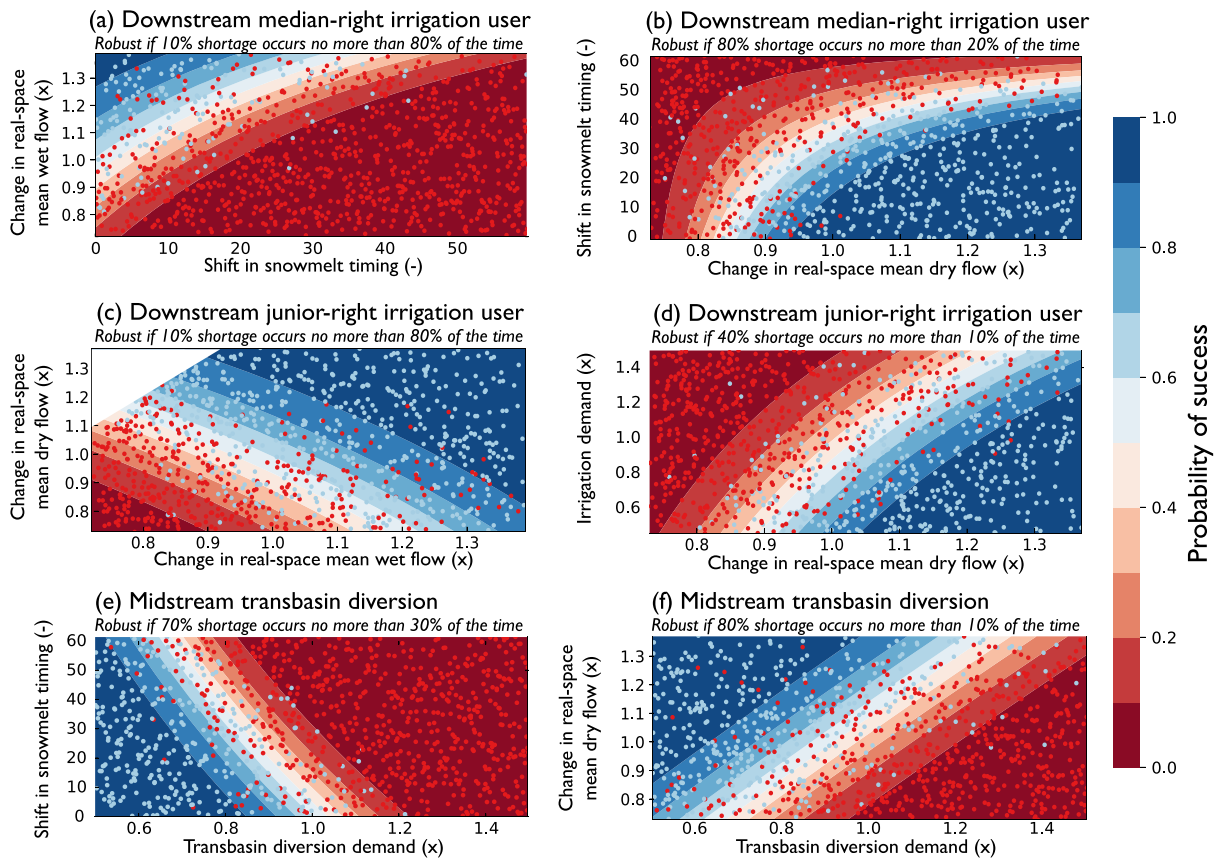


Figure 7. Scenario discovery for two robustness performance thresholds for three users in the basin. (a, b) Two alternative metrics for the median-right irrigation user, (c, d) two alternative metrics for the junior-right irrigation user, and (e, f) two alternative metrics for the transbasin diversion. All the performance metrics are listed in Table 2. Points in each panel represent the sampled SOWs, in blue if at least 50% of the realizations for that SOW met the criterion (successes) and in red if less than 50% met the criterion (failures). For each factor map the parameter with the highest predictive ability is plotted on the x axis, and the parameter with the second highest predictive ability is plotted on the y axis. In brackets next to each parameter name, we indicate whether the parameter had an additive or a multiplicative effect on the system. Note that the blank area in the top left corner of panel (c) is not assigned conditional probabilities of success as it represents sampled SOWs in which the mean dry flow exceeds the mean wet flow. Mean wet and dry flow parameters have been converted to real space to aid interpretability.

Looking at the two alternative performance metrics for the junior-right irrigation user (Figures 7c and 7d), similar changes can be seen. Using the “10% demand-80% of the time” performance metric (Figure 7c), robustness for this user is 60% of the realizations (derived from square color in Figure 6c). Success and failure are determined primarily by the distributions of wet and dry years. This means that applying the exact same robustness metric to two different irrigation users in this basin results not only in different robustness values (16% for the median- and 60% for the junior-right user) but also different drivers of their failures. Even though prior studies have highlighted that robustness definitions can yield unintended multiactor consequences (Herman et al., 2014; Trindade et al., 2017), we believe this to be the first study to evaluate and quantify how the highly diverse users in an institutionally complex river basin are faced not only with different levels of vulnerabilities but also and how changes in their preferred measures of robustness can fundamentally reshape which deeply uncertain factors control their performance.

The two alternative performance metrics for the transbasin diversion (Figures 7e and 7f) also result in different robustness values (Figure 6d) and factor maps. In both instances the factor most significantly controlling success in the performance metrics is transbasin diversion demands, followed by the shift in snowmelt timing for the “70% demand-30% of the time” metric, and by changes in the dry flow distribution for the “80% demand-10% of the time” metric. The “70% demand-30% of the time” metric is met by 34% of the simulated realizations (derived from square color in Figure 6e) and can almost never be met in realizations with transbasin diversion demands increasing by more than 20%. For this metric to always be met, transbasin diversion

demands would actually need to decrease, either by conservation measures or through other means. This is also true for the “80% demand-10% of the time” metric.

Finally, the failures highlighted in the panels of Figure 7 (the red points) show failure realizations have streamflows across the sampled range, both in drier and wetter conditions. This indicates that failure across users and preference thresholds is determined by factor interactions beyond those determining streamflow, as even in the most favorable SOWs (with wetter conditions) there are multiple realizations failing to meet the preference thresholds. These findings underline the importance of exploring the impacts of a wide number of uncertain factors on a variety of performance metrics of interest, as (i) there might exist several failure modes resulting from the interactions between the uncertain factors and (ii) the relative effects of these factors and interactions might vary for different impacts of concern. This appears to be the case when comparing across users and across types of impacts for a single user. The scenario discovery also suggests that the probabilities of failure and success in all performance metrics can be variably sensitive to the identified factors. In some instances, minor shifts in a factor's value may result in movement to the failure region (for example, changes of 10% in the mean dry and wet flows). Moving across different combinations of magnitude and frequency of shortage, the relative effect of different factors varies (even when only looking at the effects on a single user).

5. Conclusions and Future Work

The UCRB is an institutionally complex basin supporting a multitude of diverse stakeholders and agricultural, municipal, and industrial activities. Water allocation in the basin happens through an intricate network of hierarchical water rights shaping both supply and shortage dynamics for all users. In this study, we use StateMod, a model of the basin's water flow, accounting, and allocation, to resolve the dynamics of supply for hundreds of the basin's stakeholders, and investigate their changing vulnerabilities under an ensemble of potential (deeply uncertain) futures. We pair StateMod with a bottom-up vulnerability assessment framework, advancing previous work in this area by formalizing an a posteriori exploration of a spectrum of performance preference thresholds, defined by combinations of shortage magnitude and frequency. We investigate the consequences of alternative preference thresholds for robustness conclusions, as well as for the identification of consequential scenarios for the different users.

Our analysis demonstrates large differences in how various users in the basin are affected by the same ensemble of uncertain futures, shaping both the magnitude and frequency of shortages in diverse ways across users. Furthermore, their relative vulnerabilities to changing hydrologic and socioeconomic conditions differ from their historical experience, stressing that water resources management concerned with addressing shortage vulnerabilities for individual users cannot only depend on their past performance to infer future impacts. Our findings also show a complex mapping between performance thresholds and robustness values that is shaped by the interactions between the considered uncertain factors and the water allocation hierarchy. This highlights the complexity of selecting a single performance threshold a priori, that is equally relevant across the diverse stakeholders in the basin, even within a single sector (say, agriculture). Lastly, when discovering consequential scenarios for UCRB users, we also show for the first time that the most important factors driving shortages can change significantly across users and preferred performance thresholds (i.e., inferences of what matters and for whom can change significantly).

These results imply that, for systems of this type, predicting impacts on, say, agriculture using a single functional relationship or metric would dramatically misrepresent the impacts for the different users. Management policies designed based on these simplified estimations and key factors could potentially be blindsided by uncertainties affecting ignored types of shortage or users. While it is beyond the scope of this paper to quantify the sources of heterogeneity in effects across the users, it is apparent that the human elements of this system, the water rights in particular, are fundamental components shaping the experienced impacts, by translating hydrologic and demand changes to supply and allocation. StateMod allowed us to resolve these effects at a high resolution and demonstrate their role in distributing vulnerabilities for the users. We believe this finding to be strong evidence that the modeling representation of such complex systems affects the conclusions drawn, and consequently, it is important to use highly resolved models of multiuser systems when assessing impacts specific to their stakeholders. Simply aggregating either shortage quantification or robustness valuation per sector would collapse their diversity to a single, poorly representative, value.

Future work will explore the mechanisms by which these rights influence the sensitivities of different metrics for different users through cluster analysis. We would like to discover if sensitivities are more informed by right seniority, magnitude of decree, basin geography, sectoral use, or other factors, as well as the extent of their interactions. Moreover, we plan to investigate the stability of robustness values achieved using different robustness metrics (e.g., regret-based metrics as opposed to satisficing) or using alternative performance threshold definitions. These definitions could, for example, take into account shortage duration. This is a relevant shortage characteristic that has not been included in these results but that could be influenced by changing hydrologic persistence, an especially pertinent concern given the current bidecadal drought of the 21st century.

Data Availability Statement

StateMod is available at GitHub (<https://github.com/OpenCDSS>). The input files to run StateMod for the UCRB can be found at the website (<https://www.colorado.gov/pacific/cdss/surface-water-statemod>). All the scripts to replicate the analysis performed in this paper and regenerateft Hadjimichael and Quinn (2020).

Acknowledgments

This study was partially supported by the National Oceanic and Atmospheric Administration (NOAA) through the Sectoral Applications Research Program (SARP). 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

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., & Meko, D. M. (2014). Assessing the risk of persistent drought using climate model simulations and paleoclimate data. *Journal of Climate*, *27*(20), 7529–7549. <https://doi.org/10.1175/JCLI-D-12-00282.1>

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., St. George, S., Otto-Bliessner, B., et al. (2013). The continuum of hydroclimate variability in western North America during the last millennium. *Journal of Climate*, *26*(16), 5863–5878. <https://doi.org/10.1175/JCLI-D-11-00732.1>

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Research*, *41*(3), 435–449. <https://doi.org/10.1287/opre.41.3.435>

Beh, E. H., Zheng, F., Dandy, G. C., Maier, H. R., & Kapelan, Z. (2017). Robust optimization of water infrastructure planning under deep uncertainty using metamodels. *Environmental Modelling & Software*, *93*, 92–105. <https://doi.org/10.1016/j.envsoft.2017.03.013>

Bracken, C., Rajagopalan, B., & Zagona, E. (2014). A hidden Markov model combined with climate indices for multidecadal streamflow simulation. *Water Resources Research*, *50*, 7836–7846. <https://doi.org/10.1002/2014WR015567>

Breiman, L. (1984). *Classification and regression trees*. New York: Routledge. <https://doi.org/10.1201/9781315139470>

CWCB (2010). Statewide water supply initiative: Colorado Water Conservation Board.

CWCB (2012). Colorado River water availability study Phase I Report: Colorado Water Conservation Board.

CWCB, & CDWR (2016). Upper Colorado River basin water resources planning model user's manual: Colorado Water Conservation Board and Colorado Division of Water Resources.

Chavarria, S. B., & Gutzler, D. S. (2018). Observed changes in climate and streamflow in the pper Rio Grande Basin. *Journal of the American Water Resources Association*, *54*(3), 644–659. <https://doi.org/10.1111/1752-1688.12640>

Christensen, N., & Lettenmaier, D. (2006). A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River basin. *Hydrology and Earth System Sciences Discussions*, *3*(6), 3727–3770. <https://doi.org/10.5194/hess-11-1417-2007>

Christensen, N., Wood, A., Voisin, N., Lettenmaier, D., & Palmer, R. (2004). The effects of climate change on the hydrology and water resources of the Colorado River basin. *Climatic Change*, *62*(1-3), 337–363. <https://doi.org/10.1023/B:CLIM.0000013684.13621.1f>

Clarke, L., Nichols, L., Vallario, R., Hejazi, M., Horing, J., Janetos, A. C., et al. (2018). Sector interactions, multiple stressors, and complex systems. In Reidmiller, D. R., Avery, C. W., Easterling, D. R., Kunkel, K. E., Lewis, K. L. M., Maycock, T. K., & Stewart, B. C. (Eds.), *Impacts, risks, and adaptation in the United States: Fourth National Climate Assessment, Volume II* (vol. II, pp. 638–668). Washington, DC, USA: U.S. Global Change Research Program.

DOLA (2015). Population totals Colorado substate. Retrieved 2019–10–10, from <https://demography.dola.colorado.gov/population/population-totals-colorado-substate/>

Dessai, S., Hulme, M., Lempert, R., & Pielke, R. (2009). Do we need better predictions to adapt to a changing climate? *Eos, Transactions American Geophysical Union*, *90*(13), 111–112. <https://doi.org/10.1029/2009EO130003>

Dittrich, R., Wreford, A., & Moran, D. (2016). A survey of decision-making approaches for climate change adaptation: Are robust methods the way forward? *Ecological Economics*, *122*, 79–89. <https://doi.org/10.1016/j.ecolecon.2015.12.006>

Elias, E. H., Rango, A., Steele, C. M., Mejia, J. F., & Smith, R. (2015). Assessing climate change impacts on water availability of snowmelt-dominated basins of the Upper Rio Grande basin. *Journal of Hydrology: Regional Studies*, *3*, 525–546. <https://doi.org/10.1016/j.ejrh.2015.04.004>

Franssen, M. (2005). Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. *Research in Engineering Design*, *16*(1), 42–56. <https://doi.org/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. <https://doi.org/10.1023/A:1008894516817>

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., et al. (2017). Large near-term projected snowpack loss over the western United States. *Nature Communications*, *8*(1), 1–7. <https://doi.org/10.1038/ncomms14996>

Gardner-Smith, B. (2019, April). Shoshone hydro plant in Glenwood Canyon went down last week, but not the river. The Aspen Times. Retrieved 2019–10–02, from <https://www.aspentimes.com/news/local/shoshone-hydro-plant-in-glenwood-canyon-went-down-last-week-but-not-the-river/>

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*, *55*, 9024–9050. <https://doi.org/10.1029/2019WR025462>

Graf, W. L., Wohl, E., Sinha, T., & Sabo, J. L. (2010). Sedimentation and sustainability of western American reservoirs. *Water Resources Research*, *46*, W12535. <https://doi.org/10.1029/2009WR008836>

Hadjimichael, A., & Quinn, J. (2020). *antonia-had/UCRB analysis: Robustness analysis*. Zenodo. <https://doi.org/10.5281/zenodo.3742375>

- Hall, J., & Borgomeo, E. (2013). Risk-based principles for defining and managing water security. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(2002), 20120407. <https://doi.org/10.1098/rsta.2012.0407>
- Hall, J., Grey, D., Garrick, D., Fung, F., Brown, C., Dadson, S. J., & Sadoff, C. W. (2014). Coping with the curse of freshwater variability. *Science*, 346(6208), 429–430. <https://doi.org/10.1126/science.1257890>
- Harding, B., Wood, A., & Prairie, J. (2012). The implications of climate change scenario selection for future streamflow projection in the Upper Colorado River basin. *Hydrology and Earth System Sciences*, 16(11), 3989–4007. <https://doi.org/10.5194/hess-16-3989-2012>
- 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. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000509](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000509)
- Herman, J. D., Zeff, H. B., Reed, P. M., & Characklis, G. W. (2014). Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, 50, 7692–7713. <https://doi.org/10.1002/2014WR015338>
- IUCN (2012). Gila cypha: NatureServe: The IUCN red list of threatened species 2014: e.T9184A54036093.: International Union for Conservation of Nature.
- IUCN (2012). Gila elegans: NatureServe: The IUCN red list of threatened species 2013: e.T9186A18232150.: International Union for Conservation of Nature.
- IUCN (2012). Ptychocheilus lucius: NatureServe: The IUCN red list of threatened species 2013: e.T18829a19034371: International Union for Conservation of Nature.
- IUCN (2012). Xyrauchen texanus: NatureServe: The IUCN red list of threatened species 2013: e.T23162a19032625: International Union for Conservation of Nature.
- Kasprzyk, J. R., Reed, P. M., & Hadka, D. M. (2016). Battling arrow's paradox to discover robust water management alternatives. *Journal of Water Resources Planning and Management*, 142(2), 04015053. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000572](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000572)
- Kenney, D. S. (2005). Prior appropriation and water rights reform in the western United States. In Bruns, Bryan Randolph, Ringler, Claudia, & Meinzen-Dick, Ruth Suseela (Eds.), *Water rights reform: Lessons for institutional design* (pp. 336). International Food Policy Research Institute.
- Knight, F. H. (1921). *Risk, uncertainty and profit*: Courier Dover Publications.
- Kwakkel, J. (2019). A generalized many-objective optimization approach for scenario discovery. *Futures & Foresight Science*, 1(2), e8. <https://doi.org/10.1002/ffo2.8>
- Kwakkel, J., & Haasnoot, M. (2019). Supporting DMDU: A taxonomy of approaches and tools. In Marchau, V., Walker, W., Bloemen, P., & Popper, S. (Eds.), *Decision making under deep uncertainty*: Springer.
- Kwakkel, J., Walker, W., & Marchau, V. (2010). Classifying and communicating uncertainties in model-based policy analysis. *International Journal of Technology, Policy and Management*, 10(4), 299–315. <https://doi.org/10.1504/IJTPM.2010.036918>
- Lamontagne, J. R., Reed, P. M., Marangoni, G., Keller, K., & Garner, G. G. (2019). Robust abatement pathways to tolerable climate futures require immediate global action. *Nature Climate Change*, 9(4), 290–294. <https://doi.org/10.1038/s41558-019-0426-8>
- Lebedev, S. (2015). hmmlern. Retrieved from <https://hmmlern.readthedocs.io/>
- Lehner, F., Wood, A. W., Vano, J. A., Lawrence, D. M., Clark, M. P., & Mankin, J. S. (2019). The potential to reduce uncertainty in regional runoff projections from climate models. *Nature Climate Change*, 9(12), 926–933. <https://doi.org/10.1038/s41558-019-0639-x>
- Lempert, R. J. (2002). A new decision systems for complex systems. *Proceedings of the National Academy of Sciences*, 99(suppl 3), 7309–7313. <https://doi.org/10.1073/pnas.082081699>
- Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A general, analytic method for generating robust strategies and narrative scenarios. *Management science*, 52(4), 514–528. <https://doi.org/10.1287/mnsc.1050.0472>
- Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the next one hundred years*: RAND Corporation.
- Lempert, R. J., Popper, S. W., & Bankes, S. C. (2010). Robust decision making: Coping with uncertainty. *The Futurist*, 44(1), 47.
- Livneh, B., Deems, J. S., Buma, B., Barsugli, J. J., Schneider, D., Molotch, N. P., et al. (2015). Catchment response to bark beetle outbreak and dust-on-snow in the Colorado Rocky Mountains. *Journal of Hydrology*, 523, 196–210. <https://doi.org/10.1016/j.jhydrol.2015.01.039>
- 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. <https://doi.org/10.1016/j.envsoft.2016.03.014>
- Malers, S. A., Ray, R. B., & Catherine, N.-L. (2001). Colorado's decision support systems: Data-centered water resources planning and administration. *Watershed Management and Operations Management* 2000, 1–9. [https://doi.org/10.1061/40499\(2000\)153](https://doi.org/10.1061/40499(2000)153)
- 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. <https://doi.org/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. [https://doi.org/10.1002/2017EF000649@10.1002/\(ISSN\)2328-4277.RESDEC1](https://doi.org/10.1002/2017EF000649@10.1002/(ISSN)2328-4277.RESDEC1)
- Moallemi, E. A., Zare, F., Reed, P. M., Elsayah, S., Ryan, M. J., & Bryan, B. A. (2020). Structuring and evaluating decision support processes to enhance the robustness of complex humannatural systems. *Environmental Modelling & Software*, 123, 104551. <https://doi.org/10.1016/j.envsoft.2019.104551>
- Moss, R., Fisher-Vanden, K., Delgado, A., Backhaus, S., Barrett, C. L., Bhaduri, B., et al. (2016). Understanding dynamics and resilience in complex interdependent systems: U.S. Global Change Research Program.
- 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, W08529. <https://doi.org/10.1029/2009WR008530>
- Painter, T. H., Skiles, S. M., Deems, J. S., Bryant, A. C., & Landry, C. C. (2012). Dust radiative forcing in snow of the Upper Colorado River Basin: 1. A 6 year record of energy balance, radiation, and dust concentrations. *Water Resources Research*, 48, W07521. <https://doi.org/10.1029/2012WR011985>
- Parsons, R., & Bennett, R. (2006). Reservoir operations management using a water resources model. *Operating Reservoirs in Changing Conditions*, 304–311. [https://doi.org/10.1061/40875\(212\)30](https://doi.org/10.1061/40875(212)30)
- Pielke, R. A., Doesken, N., Bliss, O., Green, T., Chaffin, C., Salas, J. D., et al. (2005). Drought 2002 in Colorado: An unprecedented drought or a routine drought? *Pure and Applied Geophysics*, 162(8), 1455–1479. <https://doi.org/10.1007/s00024-005-2679-6>
- Polasky, S., Carpenter, S. R., Folke, C., & Keeler, B. (2011). Decision-making under great uncertainty: Environmental management in an era of global change. *Trends in Ecology & Evolution*, 26(8), 398–404. <https://doi.org/10.1016/j.tree.2011.04.007>
- Proctor, C. (2008, May). Xcel's Shoshone power plant back in business. Denver Business Journal. Retrieved 2019–10–02, from <https://www.bizjournals.com/denver/stories/2008/04/28/daily37.html>

- Quinn, J. D., Reed, P. M., Giuliani, M., Castelletti, A., Oyler, J. W., & Nicholas, R. E. (2018). Exploring how changing monsoonal dynamics and human pressures challenge multireservoir management for flood protection, hydropower production, and agricultural water supply. *Water Resources Research*, *54*, 4638–4662. <https://doi.org/10.1029/2018WR022743>
- Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M., Dai, A., et al. (2014). Climate change impacts on the water balance of the Colorado headwaters: High-resolution regional climate model simulations. *Journal of Hydrometeorology*, *15*(3), 1091–1116. <https://doi.org/10.1175/JHM-D-13-0118.1>
- Rhee, G., Salazar, J., & Grigg, C. (2018). How long does a 15-year drought last? On the correlation of rare events. *Journal of Climate*, *32*(5), 1345–1359. <https://doi.org/10.1175/JCLI-D-18-0326.1>
- Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical Association*, *46*(253), 55–67. <https://doi.org/10.1080/01621459.1951.10500768>
- Schwartz, J. (2019, March). Amid 19-year drought, states sign deal to conserve Colorado River water. The New York Times. Retrieved 2019–12–04, from <https://www.nytimes.com/2019/03/19/climate/colorado-river-water.html>
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129. <https://doi.org/10.1037/h0042769>
- Skiles, S. M., & Painter, T. H. (2019). Toward understanding direct absorption and grain size feedbacks by dust radiative forcing in snow with coupled snow physical and radiative transfer modeling. *Water Resources Research*, *55*, 7362–7378. <https://doi.org/10.1029/2018WR024573>
- Solander, K. C., Bennett, K. E., & Middleton, R. S. (2017). Shifts in historical streamflow extremes in the Colorado River Basin. *Journal of Hydrology: Regional Studies*, *12*, 363–377. <https://doi.org/10.1016/j.ejrh.2017.05.004>
- Starr, M. K. (1962). *Product design and decision theory*, Prentice-Hall series in engineering design. Englewood Cliffs, NJ: Prentice-Hall.
- State of Colorado (2015). Colorado's water plan. Denver, Colorado.
- Trindade, B. C., Reed, P. M., & Characklis, G. W. (2019). Deeply uncertain pathways: Integrated multi-city regional water supply infrastructure investment and portfolio management. *Advances in Water Resources*, *134*, 103,442. <https://doi.org/10.1016/j.advwatres.2019.103442>
- Trindade, B. C., Reed, P. M., Herman, J. D., Zeff, H. B., & Characklis, G. W. (2017). 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. <https://doi.org/10.1016/j.advwatres.2017.03.023>
- USFWS (1999). Final programmatic biological opinion for Bureau of Reclamations operations and depletions, other depletions, and funding and implementation of recovery program actions in the Upper Colorado River above the confluence with the Gunnison river: United States Fish and Wildlife Service.
- USFWS (2020). Upper Colorado River Endangered Fish Recovery Program.
- USGS (2019). USGS current conditions for USGS 09085100 Colorado River below GLenwood Springs, CO: United States Geological Survey.
- Udall, B., & Overpeck, J. (2017). The twenty-first century Colorado River hot drought and implications for the future. *Water Resources Research*, *53*, 2404–2418. <https://doi.org/10.1002/2016WR019638>
- Vano, J. A., Udall, B., Cayan, D. R., Overpeck, J. T., Brekke, L. D., Das, T., et al. (2014). Understanding uncertainties in future Colorado River streamflow. *Bulletin of the American Meteorological Society*, *95*(1), 59–78. <https://doi.org/10.1175/BAMS-D-12-00228.1>
- Wald, A. (1950). *Statistical decision functions*: Wiley.
- Walker, W. E., Harremos, P., Rotmans, J., Sluijs, J. P. v. d., Asselt, M. B. A. v., Janssen, P., & Krauss, M. P. Krayevon (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, *4*(1), 5–17. <https://doi.org/10.1076/iaij.4.1.5.16466>
- Woodhouse, C. A., Gray, S. T., & Meko, D. M. (2006). Updated streamflow reconstructions for the Upper Colorado River basin. *Water Resources Research*, *42*, W05415. <https://doi.org/10.1029/2005WR004455>
- Xiao, M., Udall, B., & Lettenmaier, D. P. (2018). On the causes of declining Colorado River streamflows. *Water Resources Research*, *54*, 6739–6756. <https://doi.org/10.1029/2018WR023153>
- Yates, D. N., Miller, K. A., Wilby, R. L., & Kaatz, L. (2015). Decision-centric adaptation appraisal for water management across Colorado Continental Divide. *Climate Risk Management*, *10*, 35–50. <https://doi.org/10.1016/j.crm.2015.06.001>

References From the Supporting Information

- Akintug, B., & Rasmussen, P. (2005). A Markov switching model for annual hydrologic time series. *Water Resources Research*, *41*, W09424. <https://doi.org/10.1029/2004WR003605>
- Ault, T. R., Mankin, J. S., Cook, B. I., & Smerdon, J. E. (2016). Relative impacts of mitigation, temperature, and precipitation on 21st-century megadrought risk in the American Southwest. *Science Advances*, *2*(10), e1600873. <https://doi.org/10.1126/sciadv.1600873>