Advancing diagnostic model evaluation to better understand water shortage mechanisms in institutionally complex river basins

Antonia Hadjimichael[1], Julianne Quinn[2], and Patrick Reed[1]

[1]Cornell University
[2]University of Virginia

November 24, 2022

Abstract

Water resources systems models enable valuable inferences on consequential system stressors by representing both the geophysical processes determining the movement of water, and the human elements distributing it to its various competing uses. This study contributes a diagnostic evaluation framework that pairs exploratory modeling with global sensitivity analysis to enhance our ability to make inferences on water scarcity vulnerabilities in institutionally complex river basins. Diagnostic evaluation of models representing institutionally complex river basins with many stakeholders poses significant challenges. First, it needs to exploit a large and diverse suite of simulations to capture important human-natural system interactions as well as institutionally-aware behavioral mechanisms. Second, it needs to have performance metrics that are consequential and draw on decision-relevant model outputs that adequately capture the multi-sector concerns that emerge from diverse basin stakeholders. We demonstrate the proposed model diagnostic framework by evaluating how potential interactions between changing hydrologic conditions and human demands influence the frequencies and durations of water shortages of varying magnitudes experienced by hundreds of users in a sub-basin of the Colorado river. We show that the dominant factors shaping these effects vary both across users and, for an individual user, across percentiles of shortage magnitude. These differences hold even for users sharing diversion locations, demand levels or water right seniority. Our findings underline the importance of detailed institutional representation for such basins, as institutions strongly shape how dominant factors of stakeholder vulnerabilities propagate through the complex network of users.

Advancing diagnostic model evaluation to better understand water shortage mechanisms in institutionally complex river basins

Antonia Hadjimichael[1], Julianne Quinn[2], Patrick Reed[1]

[1]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[2]Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA

Key Points:

- We contribute a novel diagnostic evaluation framework for understanding water scarcity in institutionally complex river basins
- Magnitude-varying sensitivity analysis of the frequency and duration of water shortages improves inferences on stakeholder-specific controls
- The most influential stressors vary notably across users and alternative measures of shortage, even for users of similar characteristics

Abstract

Water resources systems models enable valuable inferences on consequential system stressors by representing both the geophysical processes determining the movement of water, and the human elements distributing it to its various competing uses. This study contributes a diagnostic evaluation framework that pairs exploratory modeling with global sensitivity analysis to enhance our ability to make inferences on water scarcity vulnerabilities in institutionally complex river basins. Diagnostic evaluation of models representing institutionally complex river basins with many stakeholders poses significant challenges. First, it needs to exploit a large and diverse suite of simulations to capture important human-natural system interactions as well as institutionally-aware behavioral mechanisms. Second, it needs to have performance metrics that are consequential and draw on decision-relevant model outputs that adequately capture the multi-sector concerns that emerge from diverse basin stakeholders. We demonstrate the proposed model diagnostic framework by evaluating how potential interactions between changing hydrologic conditions and human demands influence the frequencies and durations of water shortages of varying magnitudes experienced by hundreds of users in a sub-basin of the Colorado river. We show that the dominant factors shaping these effects vary both across users and, for an individual user, across percentiles of shortage magnitude. These differences hold even for users sharing diversion locations, demand levels or water right seniority. Our findings underline the importance of detailed institutional representation for such basins, as institutions strongly shape how dominant factors of stakeholder vulnerabilities propagate through the complex network of users.

1 Introduction

1.1 Need for water systems model diagnostics

Water resources systems models are important boundary objects for supporting decision making and facilitating inferences on consequential risks (Moallemi et al., 2020; Star & Griesemer, 1989; Star, 2010; White et al., 2010). Their use can inform the comparison and assessment of alternative management policies, the evaluation of impacts of extreme events, and the allocation of limited resources among competing uses. Water systems models typically aim to represent the dynamics of water as a result of geophysical processes, as well as the elements of human systems used to manage it: infrastructure, institutions, and governance (Loucks, 1992). The inherent complexity of human behavior and institutions poses a non-trivial challenge for adequately capturing the unique human and natural system traits of regions. As such, they have been criticized for not producing insights that are generalizable to other contexts (Brown et al., 2015). A weakness and potential danger in this critique emerges if it is used to justify regional model representations that remain overly generic and limited in their decision relevance. We argue that the inclusion of human processes in our models can indeed inform our fundamental understanding of coupled human and natural processes as well as their interactions, if paired with formalized diagnostic frameworks that allow us to reconcile model behavior with the real-world decision-making contexts that shape consequential insights.

Like all models, water resources systems models are simplified abstractions of the real world that allow us to reason and make testable predictions about the modeled system. Their utility hinges on their ability to represent the real system with some "acceptable" fidelity, which can be quantified by several metrics. In the case of water resources systems, the multiple interdependent human and natural processes taking place translate into modeled representations that are highly complex, non-linear, and present strong interactions and threshold behaviors (Hornberger & Spear, 1981). Further, model complexity and detail has been increasing based on our ever-improving understanding of ecologic, geologic, atmospheric, and hydrologic processes, the availability of data, and the wider acceptance of their interactions, which in turn need to be represented (Saltelli et al., 2019). As a result, performing diagnostic analyses of such models becomes more difficult, both with regards to pinpointing the model components that most influence errors (Tang et al., 2007), but also with regards to broadening the scope of metrics used to evaluate model performance (Gupta et al., 1998). Recognition of this fact motivates the need for enhanced diagnostic tools and methods that move beyond error-driven evaluations of hydrologic state predictions (Gupta et al., 2008).

A centerpiece to modern model diagnostics, sensitivity analysis has had a long history of application in hydrologic, environmental, and Earth systems modeling (Pianosi et al., 2016; Razavi & Gupta, 2015; Saltelli et al., 2019; Shin et al., 2013; Song et al., 2015), with several domains of application (as reviewed by the aforementioned studies): identifying regions of sensitivity or uncertainty in model output, apportioning them to uncertain factors, identifying factors' function and importance, and, the focus of this study, diagnostic model evaluation. In diagnostic evaluations, the results of sensitivity analysis can be used to identify model components to be prioritized during calibration, their degrees of interaction, and to compare the behavior of the model and its components with what is expected in reality (Gupta et al., 2008). In the latter application, the aim is to check whether the processes and factors appearing to control model behaviour are indeed corroborated by our observations, so as to reject or support candidate formulation hypotheses, improve the model, and advance our fundamental understanding of such systems (Clark et al., 2011; Oreskes et al., 1994; Saltelli et al., 2004; Wagener et al., 2003). There is however a point of departure here between natural-systems-focused hydrologic modeling and water resources systems modeling. Many agencies and institutions are actively using water systems models as boundary objects that have met the conditions needed for establishing credibility (e.g., acceptable representational fidelity) but face broader tests on their salience and legitimacy in informing negotiated decisions (Cash et al., 2003; White et al., 2010). This consequence-oriented context is the lens through which we perform our diagnostic analysis.

Applying diagnostic model evaluation using consequence-oriented or decision-relevant sensitivity analysis (Herman et al., 2015) faces an additional challenge: selecting the parts of the model behavior space that are reflective of the stakeholders' viewpoints so that the most consequential uncertainties are identified and addressed (Saltelli & Funtowicz, 2014). Water systems models in particular are an amalgam of geophysical, hydrological, and infrastructure-constrained institutional processes that yield a model behavior space as a multitude of outputs that go beyond hydrologic states. These outputs may have very diverse levels of salience to the real system's stakeholders and to their goals being achieved. This is further complicated when such systems are also institutionally and dynamically complex, with multiple interacting domains and stakeholders. Consider the irrigation sector: access to several water sources and storage, differences in farming systems, the presence of contemporaneous risks and other factors affect the magnitude and duration of shortage that could be withstood by a farm in the case of drought (Komarek et al., 2020; Wallander et al., 2017).

Further, it has been recognised that the importance of various model components may vary in time and space (Pianosi et al., 2016). This gave rise to a slew of time-varying and space-varying sensitivity analyses (e.g., see Pianosi and Wagener (2017) and Rougé et al. (2019) and references therein, as well as the earlier review by Song et al. (2015)). With a few exceptions (discussed below), these studies have focused on either conceptual or case-specific hydrologic models describing rainfall-runoff processes in catchments, in which the human elements of the system are not a significant focus. This reflects a general challenge in hydrologic modeling, not limited to diagnostic studies: few models appropriately and adequately account for or represent the human activities that largely shape the flows of water (Wada et al., 2017; Wagener et al., 2010).

1.2 A framework for decision-relevant diagnostic evaluation

In recognition of the importance of human institutional coordination and control in large reservoir cascades, Quinn et al. (2019) and Rougé et al. (2019) applied time-varying sensitivity analyses as a diagnostic evaluation of reservoir release rules. Both studies highlight the dominating effect of operational coordination in achieving the systems' objectives. The present study expands on this growing body of work by contributing a diagnostic evaluation of a fine-scale model of an institutionally complex water resources system. The basin under study is the Upper Basin of the Colorado River within the state of Colorado (henceforth abbreviated to UCRB). The UCRB stretches from the headwaters of the Colorado River to the Colorado-Utah state line, from where it continues to deliver water downstream to Lake Powell. Water allocation in this basin, like in most other basins in the western United States, is determined by the doctrine of "prior appropriation", which allocates water to users based on right seniority. Seniority is determined based on the date each right was decreed and is associated with an amount of water that the user should put into a "beneficial use" (Kenney, 2005). In this manner, prior appropriation creates a hierarchical network of water allocation, where each of the users affects and is affected by water availability in the basin. This multiplex system of allocation is naturally accompanied by the infrastructure and conduits necessary to make the transfers possible, including large exports from the basin to other uses in the state of Colorado (further detailed in the Study area section). Given the important role that human systems play in regulating and distributing streamflow in such basins, one should question whether it is sensible to ignore them in model-based assessments, and, by extension, neglect them in our broader views on the diagnostic analyses of models. As of the time of writing, the authors are not aware of a diagnostic evaluation of a water resources model that attempted to assess dominant model controls using user-specific, decision-relevant metrics for institutionally complex multi-stakeholder systems.

Our proposed diagnostic framework is demonstrated using the State of Colorado's Stream Simulation Model (StateMod), a generic network-based water system model for water accounting and allocation. StateMod is a component of Colorado's Decision Support System (CDSS), jointly developed by the Colorado Water Conservation Board (CWCB) and the Division of Water Resources (DWR), which includes databases and data management tools, as well as several models for water resources planning for several basins in Colorado (Malers et al., 2001). Using detailed historic demand and operation records, StateMod replicates the UCRB's application of the prior appropriation doctrine, accounting for the entirety of the basin's consumptive water use. This allows us to represent the monthly allocation of water to each individual user in the basin, as well as their unmet demand (shortages). The explicit representation of human systems and institutions in such fine detail establishes a direct connection between the processes abstracted by the model and the many stakeholders affected by those processes in reality.

This study broadens the scope of traditional diagnostic evaluation of hydrologic models by contributing a diagnostic framework for institutionally complex river basins with a multitude of stakeholders. The framework brings together exploratory modeling (Bankes, 1993; Bankes et al., 2001; Lempert et al., 2003), global sensitivity analysis methods (Saltelli et al., 2008), and visual analytics (Keim et al., 2008; Thomas & Cook, 2005; von Landesberger et al., 2012). Exploratory modeling literature also views models as hypothetical computational experiments that give us a picture of how a system would behave if the various assumptions composing the model were correct. To be effective in producing a rich enough picture of a complex model's behavior space, exploratory modeling must examine a very large and diverse suite of model simulation runs that capture important interactions and mechanisms leading to consequences of interest (Goodwell et al., 2020; Gupta et al., 2008; Lamontagne et al., 2018; Raso et al., 2019). The model needs to therefore be evaluated under a large ensemble of potential states of the world (SOWs) which represent changes in "deeply uncertain" factors (Knight, 1921; Polasky et al., 2011; Walker & Marchau, 2003). These are factors that could potentially significantly affect a system, but that are so highly uncertain that experts either cannot know, or cannot agree on, statistical descriptions of the entire set of outcomes and their likelihoods (Kwakkel et al., 2010; Lempert, 2002; Lempert et al., 2003). Such deep uncertainties are typically investigated for their implications for stakeholders through exploratory modeling approaches (most recently reviewed by Moallemi et al. (2020)).

As we focus on better understanding how deeply uncertain factors affect each user as a function of the degree of water shortage being confronted, it is important to avoid myopic definitions of water scarcity extremes. Consequently, a core contribution of this study is to demonstrate the value of magnitude-varying water shortage diagnostics, akin to the time-varying approaches mentioned above. The reasoning behind our approach is similar in that different factors are likely to dominate different system states (Pianosi et al., 2016), but the critical states for different users may occur at different times, making magnitude-varying sensitivity analysis more decision-relevant to each user. Consequently, our magnitude-varying sensitivity analysis of the frequency and duration of each user's water shortages identifies how dominant factors might vary not only across different system modes, but also across the basin's water users. Visual analytics allow us to present this complex information across scales and users, and to derive insight about the dominant controls of the modelled shortages across the UCRB. The diagnostic framework presented in this study

2 Study area

Our study area spans 25,682 km^2 (9,915 mi^2) in Western Colorado, from the headwaters of the Colorado River at the Continental Divide to the Colorado-Utah state line (Fig. 1). The primary consumptive use of water in the UCRB is irrigation, with several thousand diversions drawing from the river and its tributaries to irrigate approximately 1,012 km^2 (391 mi^2). The basin is moderately populated, as most of Colorado's population lives east of the Continental Divide. As a result, major diversions of water need to cross the divide to deliver 567,400,000 m^3 (460,000 acre-feet) of water to northern and eastern Colorado for municipal, industrial, and agricultural uses (State of Colorado, 2015). These transbasin diversions are served through several tunnels present in the basin, the largest of which are indicated in Fig. 1.

Even though water for power generation is largely non-consumptive and does not deplete resources in the basin, the Shoshone Power Plant (indicated in Fig. 1) is a notable feature in this basin. Owing to its water right being one of the oldest and largest in the basin (dating to 1902 with a decree of 39.40 m^3/s—approximately 68% of the median river flow at the location), it significantly affects how many other users, both downstream and upstream, receive their allocation (Yates et al., 2015; USGS, 2019). When the Shoshone Power Plant requests their allotted water, junior (i.e., lower priority) transbasin and irrigation users upstream, as well as junior-right reservoirs, must cease or offset their diversion so the Shoshone call can be fully met. Conversely, users downstream from the plant benefit from its presence and senior call on the river, as almost all of the water Shoshone requests is immediately returned to the stream. This has led to the establishment of several recreation services along the Colorado River in towns downstream from the plant (from Glenwood Springs down to De Beque), worth an estimated $32 million/year to the local economy (Armistead & Mojica, 2018). Perhaps most crucially, water allocated to Shoshone, as with all rights, needs to be put into a beneficial use. This means that in the case that the 100-year-old plant shuts down, their right no longer needs to be honored by the junior users upstream, whereas the junior downstream users can no longer benefit from the availability of that water in the stream. In the past 15 years the plant had to shut down for repairs and maintenance twice Gardner-Smith (2019); Proctor (2008).

Further downstream from the Shoshone Power Plant is the so-called "15-mile reach", a segment of the river extending from the towns of Palisade to Grand Junction, the confluence of the Gunnison and Colorado rivers (indicated in Fig. 1). The US Fish and Wildlife Service has made several recommendations on maintaining critical flows in this part of the river (USFWS, 1999), as it is considered critical to the recovery of several endangered species of fish: the razorback sucker (*Xyrauchen texanus*), the Colorado pikeminnow (*Ptychocheilus lucius*), humpback chub (*Gila cypha*), and bonytail (*Gila elegans*) (IUCN, 2012b, 2012a, 2012c, 2012d; USFWS, 2020). The seniority of in-stream flow demands along the 15-mile reach, as well as the demands of irrigation users and transbasin diversions, and the functioning of the Shoshone Power Plant are all included in our ensemble of uncertain factors due to their immediate relevance to local agencies and stakeholders. This is further elaborated in the Methods section.

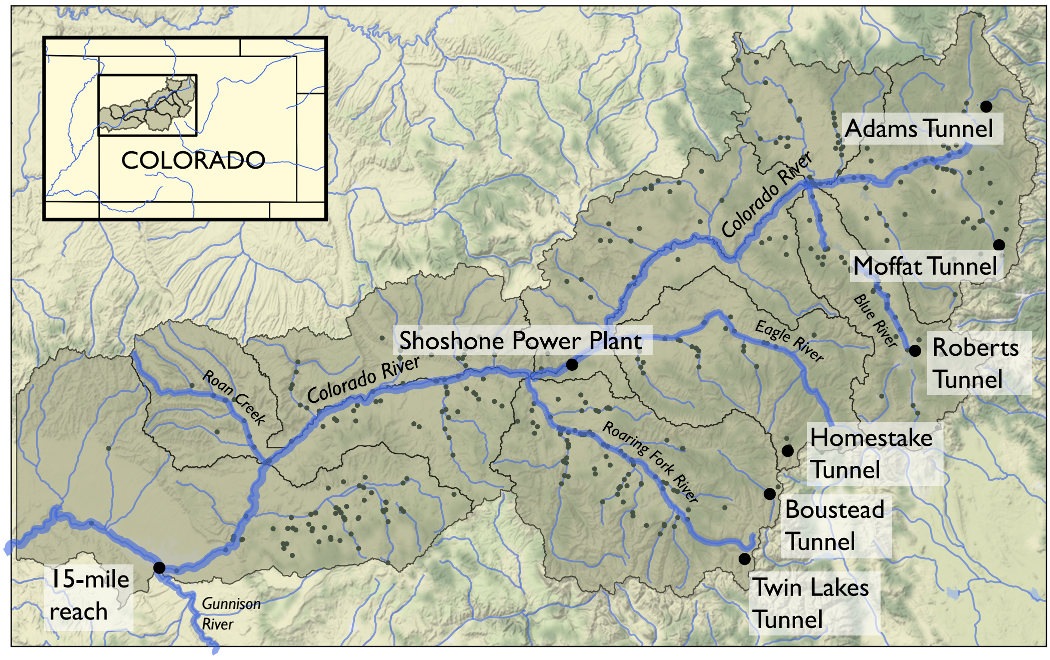

Figure 1. The Upper Colorado River Basin within the state of Colorado (UCRB), with structures of interest indicated. The Shoshone Power Plant owns one of the largest and oldest water rights in the basin. The 15-mile reach is critical for the recovery of endangered species. The highlighted tunnels are the largest transbasin diversions, exporting water to the eastern plains. The smaller points indicate all other modeled diversion points in StateMod (primarily irrigation). Figure from Hadjimichael et al. (2020).

3 Methods

3.1 Model

StateMod, the State of Colorado's water planning model, is part of a wider set of decision support tools (CDSS) developed by Colorado state agencies to facilitate comprehensive assessments of water allocation and use, as well as reservoir operations in all of the major sub-basins of the Colorado River within the state: White, Yampa, Gunnison, Dolores, San Juan, San Miguel, and Upper Colorado, modeled here (CWCB, 2012; Parsons & Bennett, 2006). StateMod explicitly represents all the aforementioned structures in the UCRB, as well as over 300 other diversions, indicated by the points in Fig. 1.

Each diversion point carries detailed operational records of demand and supply, stored in the central database of the CDSS, HydroBase. The water demand data in HydroBase reflect best-estimates of population, irrigation levels, and reservoir capacities up to 2010. Demands for irrigation diversions are produced by StateCU, the consumptive use model available within CDSS. StateCU calculates water consumption for each irrigation unit based on soil moisture, crop type, irrigated acreage and irrigation efficiencies, and also calculates return flows from each diversion. Monthly demands for municipal diversions are given by the average diversions at each month of the year between 1998-2005. Modeled monthly demands for transbasin diversions reflect their historical diversions whenever monthly data is available, or average estimates for dry, average, and wet conditions for the months without diversion data available. Reservoir filling demands are represented using minimum and maximum reservoir storage targets. The UCRB model's manual contains additional information on how historical diversion demands were estimated for all consumptive use diversions (CWCB & CDWR, 2016). StateMod represents water years 1909-2013, a period that includes extended periods of wet and dry flows, and years of extreme drought and high runoff. The current water year is defined as the period starting last October 1st, through the upcoming September 30th.

To estimate the effect of diversions and operations on stream flow and water availability, StateMod needs to first represent naturalized flow. To do so, historical diversion data, monthly reservoir storage, and return flows are superimposed on historical streamflow observations from USGS gauges. However, many of the thousands of diversions that take place in the basin do not occur near the USGS gauges, so ungauged river nodes also need to be modelled. StateMod distributes flow to these ungauged locations by using proration factors accounting for how much drainage area contributes to each gauged location. The model then applies demand and operational information to represent reservoir operations and diversions by each water right to reconstruct the remaining flow in the Colorado river and its tributaries within the UCRB.

Each water right is associated with its location on the stream, an administration number that represents its allocation seniority, and the decreed water flow it is allowed to divert. At every monthly timestep the model resolves all diversions and other transfer operations in order of seniority and estimates remaining river flow. Even though the basin's consumptive use of water is accounted for in its entirety, only 75% of the thousands of diversion points are represented in strictly correct locations, with the remaining grouped into aggregated representations based on size of diversion, location of water use, and tributary boundaries. In a similar manner, reservoirs and stock ponds with decreed capacities of less than 4,934,000 m^3 (4,000 acre-feet) of water are modeled in aggregate. The remaining 18 reservoirs are explicitly represented at their strict locations and make up 94% of the total storage capacity in the UCRB. The manner with which structures in the basin have been aggregated is described in great detail in the UCRB model's manual (CWCB & CDWR, 2016). Using this fine-scale set of data, StateMod is able to account for the effect of all users and their water rights on water availability in the UCRB. State agencies have in fact been actively using this model to assess impacts by proposed operations or other hypothetical scenarios for the past 30 years (Parsons & Bennett, 2006).

3.2 Experimental design

This study contributes a diagnostic framework for institutionally complex river basins by combining exploratory modeling with magnitude-varying sensitivity analyses of the frequencies and durations of shortages experienced by multiple users. StateMod is used to illustrate this diagnostic framework for the hundreds of users represented in the model.

For this exploratory assessment, we create a large ensemble of possible conditions for this system, representing changes in hydrology (drier and wetter conditions), human water demands, water rights, and physical storage. Our methodology then forces these changed conditions through the model to identify the factors dominating decision-relevant model outputs, specifically, the frequency and duration of different magnitudes of shortage for each user.

Fig. 2 presents the diagnostic framework contributed by this study, following the notation by McPhail et al. (2018) and Hadjimichael et al. (2020). Panel I describes the generation of the ensemble of uncertain factors to be propagated through the model. A set (Ψ) of 1,000 samples of uncertain factors is generated, each representing a potential state of the world (SOW). For each SOW, ten streamflow realizations (s) are also generated, for a total of 10,000 model evaluations. Model output is then produced as a result of each realization ($f(U, s)$) for all users in the basin (U), as shown in Panel II. Each $f(u, s)$ is therefore the model output related to user u for realization s, with the entire set of these ($f(u, S)$) being the model behavior for each user across all realizations in the ensemble (Panel III). Each $f(u, S)$ can then be used to perform diagnostic analysis on the model, using outputs that are of consequence to each user. In this particular case, the decision-relevant model outputs are chosen to be the unmet demands (shortages) experienced by each user, but this framework could be applied to other metrics in similar multi-actor systems. Panel IV shows how this model output is subsequently classified to percentiles of shortage magnitude for each user. Sensitivity analyses are then applied to the frequency and maximum duration of each shortage magnitude (Panel V), using three different methods: the Delta method, Sobol variance decomposition, and linear regression, detailed in section 3.4. Applying sensitivity analysis to the shortage magnitude corresponding to each discrete percentile of each individual user's shortage distribution allows us to identify which of 14 uncertain factors (and potentially, their interaction) control model behavior as it relates to different users, as well as identify how factor importance might vary at different shortage percentiles. The three sensitivity analysis methods each reveal different valuable information about the relationships between each important factor and the output.

3.3 Ensemble of uncertain factors

To generate the ensemble of uncertain factors used in the experiment, we explore parameterized representations of human demands and institutions included in StateMod, and of streamflows fed as inputs to the model. Listed in Table 1 are the 14 uncertain factors considered in this experiment, with the first six being parameters of a synthetic streamflow generator and the remainder being human-system StateMod parameters. The set Ψ of 1,000 parameter combinations is generated using a Latin hypercube sample (McKay et al., 1979) across the parameter ranges shown in Table 1, assuming parametric independence and uniform distributions. These ranges have been informed by the related literature (detailed below), but are intentionally expanded so as to capture important interactions and mechanisms with consequential effects (e.g., extreme multi-year droughts) (Bankes, 1993).

3.3.1 Changing hydrologic conditions

The synthetic streamflow generator used is a two-state Gaussian Hidden Markov Model (HMM), made up of two "hidden" climate states—one for dry and another for wet hydrologic conditions (Bracken et al., 2014). A HMM can be used to generate log-space flows from Gaussian distributions with different parameters; by changing the HMM's parameters we can represent changes in the frequency, severity, and persistence of droughts, as well as wet years. This particular basin has observed great historical hydrologic variability and persistence (Ault et al., 2013, 2014), something also reflected in the current bi-decadal drought being experienced in the region (Rhee et al., 2018; Schwartz, 2019).

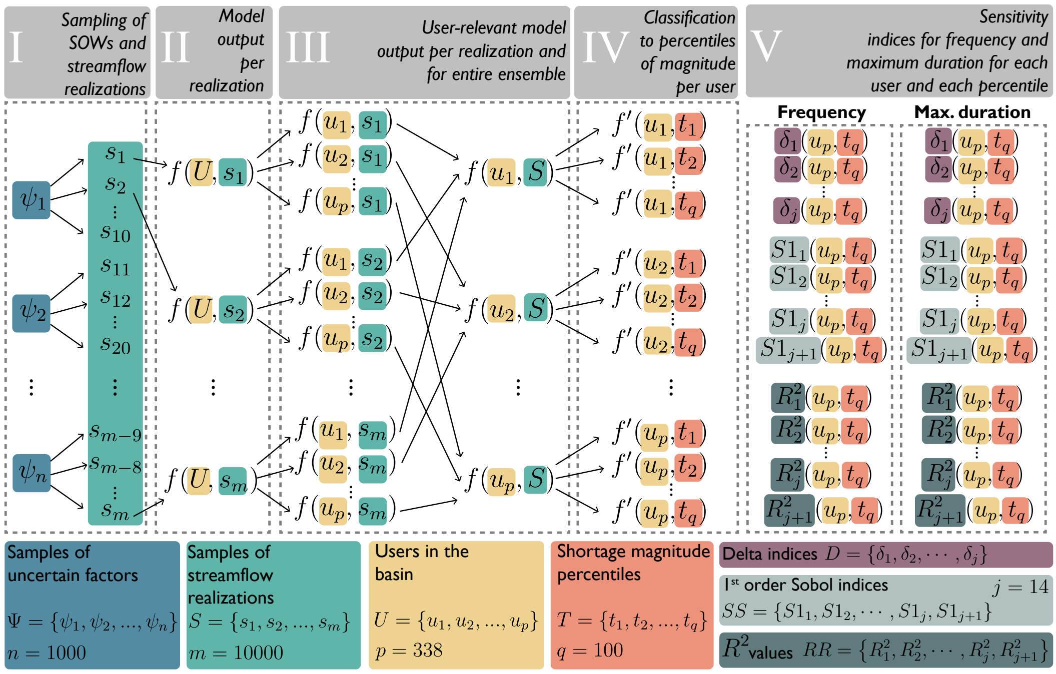

Figure 2. Experimental design of this study. Panel I shows how the samples of uncertain factors (Ψ) relate to the generation of streamflow realizations (s), i.e., ten streamflows for each sample Ψ. The performance of the system is evaluated for each realization using StateMod (panel II) producing shortages 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 to percentiles of magnitude for each user (t). Three sensitivity analysis methods are then applied to analyze the influence of 14 uncertain factors (and potentially, their interaction) on the frequency with which different annual shortage levels are experienced, and the maximum duration (in years) of annual shortages at that level. The chosen shortage levels correspond to discrete percentiles of the historical annual shortage distribution (panel V).

This drought has manifested through decreasing streamflows in the basin, despite greater precipitation, due to higher temperatures increasing evapotraspiration and causing earlier snowmelt (Xiao et al., 2018; Milly & Dunne, 2020). Such conditions are also consistent with regional projections (Christensen & Lettenmaier, 2007; Rasmussen et al., 2014).

The synthetic generator captures such dynamics by fitting the HMM to the naturalized log-space annual flows at the outflow node of the basin's model. This requires estimating the mean and standard deviation of the distributions of the two states, and the probabilities of transitioning between the two. To estimate the HMM parameters, we use the Expectation-Maximization algorithm available in the hmmlearn Python library (Lebedev, 2015) and fit the model to the last 70 years of the 105-year hydrologic record (due to non-stationarity in conditions over the whole record). To classify each annual flow into one of the two "hidden" model states (wet and dry) we use the Viterbi algorithm, also available in the package. The parameter estimates, Gaussian fits, as well as additional details on fitting and validating the generator can be found in the Supplementary Information (SI) of Hadjimichael et al. (2020).

Table 1. Uncertain factors and sampling ranges. Using a Latin hypercube sample, 1,000 parameter combinations (SOWs) are generated, under which the system performance is evaluated for every user.

Parameter	Current value	Lower bound	Upper bound
Hydrologic Factors			
Log-space dry flow mean (m^3) multiplier	1.0	0.98	1.02
Log-space dry flow standard deviation 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 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
Shift in timing of snowmelt (days earlier)	0	0	60
Change in evaporation (cm/month)	0.0	-15.24	30.48
Demand Factors			
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
Environmental and Institutional Factors			
Reservoir storage	1.0	0.8	1.0
Operation of Shoshone Power Plant	1	0	1
Seniority of environmental flows	0	0	1

To model changes in the mean and variance of the annual wet and dry flows in the record (the first four parameters listed in Table 1), as well as their persistence (the fifth and sixth parameters listed in Table 1), we modify the HMM parameter estimates using multipliers and delta operators, respectively. The ranges of these parameters were selected so the resulting flows span both the monthly and annual flows generated using synthetic stationary conditions and the Coupled Model Intercomparison Project 3 (CMIP3) and 5 (CMIP5) projections (CWCB, 2012). As a result, this HMM generator allows us to capture a range of possible hydrologic conditions: a broader range of extremes can be explored, not only with regards to the magnitude and variability of flow, but also its persistence (e.g., droughts longer than any historically observed). All these attributes of change can have consequential effects on the stakeholders—larger, longer or more frequent water shortages—making their integration critical to this exploratory modeling analysis.

Using the HMM, we synthetically generate log-space annual flows at the last model node under a range of HMM parameters. We then convert them to real-space, and then disaggregate them to monthly values, following an approach similar to Nowak et al. (2010): The monthly flow proportions are used from a historical year that is probabilistically selected based on how close its total annual flow is to that of the synthetically generated flow at the last node. A shift in the timing of snowmelt (controlled by the equivalent parameter in Table 1) is applied to the daily hydrograph of this year, thereby dissipating its peak and moving it earlier in the year. The application of this parameter has been included in this diagnostic analysis to reflect reduced snow cover durations observed in this region, due to increasing temperatures and dust on snow (Livneh et al., 2015; Skiles & Painter, 2019). We then spatially downscale the monthly flows at the last model node to all other upstream nodes, by proportionally scaling them using the monthly ratios of upstream node to last node in the selected historical year. Additional details about how this downscaling is performed, as well as figures of the generator's ability to capture the spatial streamflow correlations are given in the SI of Hadjimichael et al. (2020).

Lastly, the SOW ensemble includes a parameter to change the evaporation rate of all 18 reservoirs included in the model, as this is expected to increase with higher temperatures. The model accounts for reservoir surface evaporation by applying monthly evaporation rates (feet/month) to each reservoir. Annual rates are calculated by subtracting the weighted average effective precipitation from the estimated gross water surface evaporation. These are then scaled to monthly equivalents based on each reservoir's elevation (CWCB & CDWR, 2016). The equivalent parameter in Table 1 is applied as an additive delta operator to these monthly evaporation rates.

3.3.2 Changing water demands

The state of Colorado's population has grown significantly over the past century (DOLA, 2015), on a trend that is expected to continue, with the population doubling by 2050 by some estimates. This rise is typically accompanied by growing municipal and industrial (MI) water demands. However, even though population rise is expected, local governments and the state itself can influence where the population grows and how much water is needed to support this growth through conservation and efficiency measures implemented across the state (CWCB, 2010; State of Colorado, 2015). For example, conservation and efficiency programs have reduced per capita water consumption by 5% across the state and up to 30% in some communities (CWCB, 2010, 2019). The largest water demand in the state, however, is related to the agricultural sector, making up 89% of the state's total consumptive use (CWCB, 2019). Crop irrigation requirements are expected to increase by up to 30% by 2040, due to the growing season of many crops getting longer, as a result of rising temperatures (CWCB, 2012). At the same time, irrigated area in the UCRB is estimated to decrease by 55 km^2 (13,600 acres) by the mid-21st century, as cities expand into irrigated land and real estate developers purchase farmlands (CWCB, 2019; State of Colorado, 2015). The use of emerging technologies for more efficient water application is also expected to decrease crop requirements and mitigate some of the increased demands due to the changing climate (CWCB, 2019).

Based on these conflicting estimates, our diagnostic experiment explores the implications of both positive and negative changes in each type of demand present in the basin (municipal and industrial, transbasin, irrigation) by up to 50%. Applying this exploratory broad range of values allows us to assess sensitivities and consequences in a broader context of deeply uncertain SOWs, and capture important mechanisms of failure that might come about as a result of system conditions not previously observed. These scaling factors have been applied uniformly across all diversions representing each sector. For the transbasin diversions the scaling factors have been applied to their maximum historical monthly demand, to reflect their ideal amount of supply.

Lastly, as evapotranspiration increases during dry years (when streamflow is low), irrigation demands should be anti-correlated with the synthetically generated streamflows. To ensure this, for each sampled SOW, the annual flow anomalies are calculated for the last model node. Using a regression of historical annual total irrigation anomalies versus historical annual flow anomalies at the last node, we determine an appropriate annual total irrigation anomaly for each year of the synthetic SOW, with added noise to preserve variance. This time series is then added to the mean irrigation demand for that SOW. Additional details of how this step is performed can be found in the SI of (Hadjimichael et al., 2020).

3.3.3 Environmental and institutional changes

The last three factors listed in Table 1, represent other potential environmental and institutional changes. We sample reservoir storage losses of up to 20% as a potential result of sedimentation (Graf et al., 2010), and apply them uniformly to all reservoirs in the UCRB model. For the operation of the Shoshone Power Plant we use a binary variable sampled to indicate whether or not the plant is operational. If the plant is not operational, then it can no longer put water into a "beneficial use" and upstream junior users are not obligated to honor the call, while downstream users may be impacted by decreased deliveries. Finally, another binary variable is sampled to indicate a potential legal change to the seniority of the environmental flow right at the 15-mile reach. This change assigns a first priority senior right to this location, with a decree of 22.94 m^3/s (810 cf/s, the minimum flow rate recommended during dry conditions; USFWS (1999)).

3.4 Magnitude-varying sensitivity analysis

3.4.1 Understanding sensitivities at different shortage magnitudes

As illustrated in Fig. 2, the exploratory ensemble of potential changes (Ψ) is simulated through StateMod, producing decision relevant outputs for each user ($f(u, S)$). The outputs in this case are water shortages experienced by each user, further classified to increasing percentiles of annual magnitude ($f'(u, t)$). Fig. 3 shows the water shortages experienced by six basin users across the entire ensemble, presented in this magnitude-varying fashion. The six users shown here are: (a) a senior-right irrigation user located upstream; (b) a median-right irrigation user located downstream; (c) a junior-right irrigation user located downstream; (d) a senior irrigation user with a large decree of water allocation located downstream; (e) a transbasin diversion located midstream; and (f) the 15-mile reach (downstream). These users were selected out of the hundreds present in the basin so as to represent a range of diversion patterns, locations, levels, and right seniorities. The decreed diversion flows for these users are: (a) 0.47 m^3/s, (b) 0.35 m^3/s, (c) 0.96 m^3/s, (d) 26.63 m^3/s, and (f) 24.95 m^3/s. User (e) is a transbasin diversion represented by a tunnel in the model, which has an average annual demand of 69 million m^3.

In each panel, the black line shows the shortages that were experienced historically by each user, with the magnitude of annual shortage indicated by the y axis and its non-exceedance probability by the x axis. The blue shaded areas show the frequency with which each magnitude of shortage was experienced across the ensemble, with lighter shades indicating increased cumulative frequency. For example, the 80th percentile shortage experienced historically by the user shown at Fig. 3 (b) was approximately 1 million m^3. Across the ensemble, shortages at the 80th percentile of each realization ranged between 0.5 and 3 million m^3.

As elaborated in Hadjimichael et al. (2020), the users experience significantly different changes in their shortage distributions as a result of the same ensemble of changes in the 14 deeply uncertain factors. Comparing the historical magnitudes at each percentile with those across the ensemble, some users see their shortages increase both in magnitude and frequency in most realizations (e.g., users (a) and (b)), while others only experience more severe shortages in about half of the realizations (e.g., users (c) and (e)). The variability of magnitudes for each percentile also varies even when looking at a single user (e.g., 20th percentile and 90th percentile of shortage magnitude for user (c)). Lastly, the two dashed vertical lines in pink and yellow indicate the years during which the basin as a whole experienced its median and worst shortages, 1943 and 2002, respectively (further discussed in section 4.3). The differences between the relative shortages of the six users during these two reference years should also be noted: the worst year basin-wide is not nearly the worst for users (a) and (f), nor is the basin-wide median the median for users (a), (c), (e), and (f).

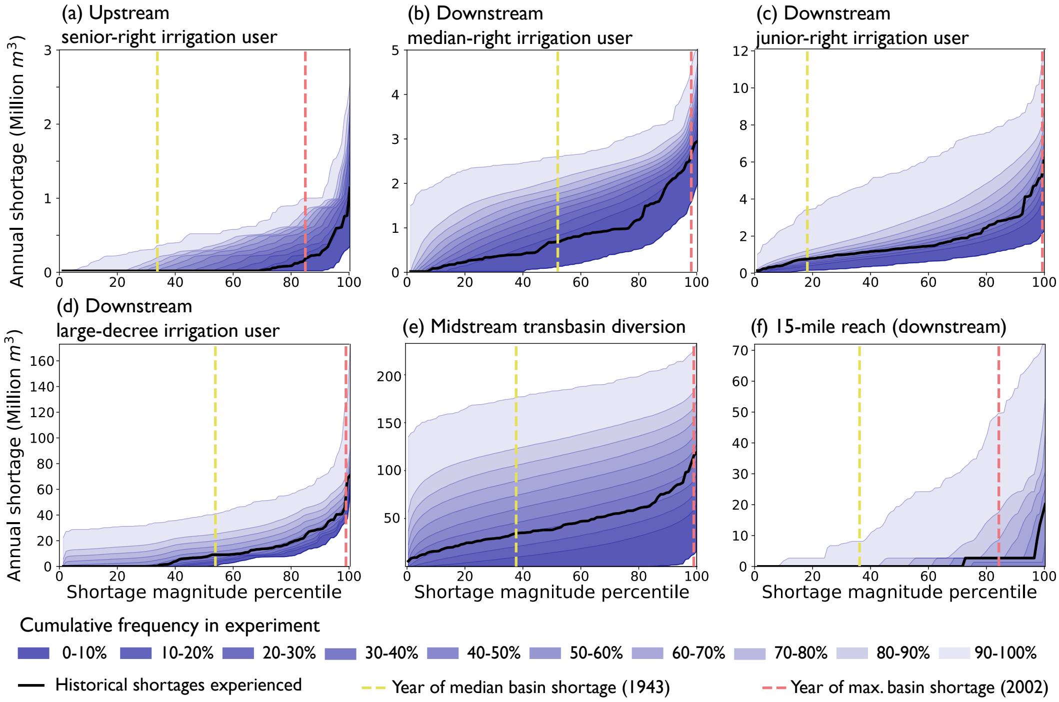

Figure 3. Percentile-varying impacts on shortage magnitude and frequency for six users in the basin. Each panel represents shortages experienced by a specific user in the basin: (a) a senior-right irrigation user; (b) a median-right irrigation user; (c) a junior-right irrigation user; (d) an irrigation user with a large decreed flow of water; (e) a transbasin diversion; (f) the 15-mile reach. The black line in every panel indicates the percent of time each annual shortage magnitude was experienced historically by each user. The shaded areas represent the frequency with which these magnitudes of shortage are experienced at each percentile across the simulated ensemble. Lighter shades indicate increased cumulative frequency. The dashed vertical lines in pink and yellow indicate the years during which the basin as a whole experienced its median and worst shortages, 1943 and 2002, respectively. Adapted from Hadjimichael et al. (2020).

One can draw two important inferences from this figure. First, selecting a single metric to be applied for the diagnostic performance of this model would be a very difficult (if not impossible) undertaking if one is indeed concerned with the metric capturing decision-relevant and consequential effects. Second, the differences seen across magnitude percentiles and users beg the question of whether different sets of factors are active at different levels of water scarcity extremes. These findings consequently motivate the application of magnitude-varying sensitivity analyses as a diagnostic tool for better understanding the state-consequence dynamics of this network of multi-sector stakeholders. For each user, the analyses are applied to both the frequency and the maximum duration of different annual shortages magnitudes corresponding to increasing percentiles of that user's historical annual shortage distribution.

3.4.2 Methods used for magnitude-varying sensitivity analysis

Three sensitivity analysis measures were applied to analyze the sensitivity of the frequencies and durations with which each user experiences different shortage magnitudes: the Delta moment-independent measure (Borgonovo, 2007; Plischke et al., 2013), Sobol variance decomposition (Sobol, 2001), and linear regression. Sobol sensitivity analysis is a widely applied method which decomposes the variance of a response variable (in this case, either the frequency or the maximum duration of a historical percentile of shortage) into the amounts contributed by each of the independent variables, both individually and by way of interactions. The first-order Sobol index represents the amount of variance in the response variable attributed to each parameter individually (i.e., without considering its interactions with the other parameters), whereas higher-order Sobol indices measure the additional variance caused by interactions between the parameters.

The Delta Method is a density-based measure that identifies model parameters that most influence the entire distribution of the response variable. The resulting Delta index for each parameter measures the normalized expected shift in the distribution of the response variable induced by the parameter (Borgonovo, 2007). The difference between the two lies in the fact that Sobol identifies parameters that achieve the greatest reduction in only the variance of the response variable, whereas the Delta index is a moment-independent measure. The application of the two methods is performed through the Python library SALib (Herman & Usher, 2017), which calculates both metrics using the method of Plischke et al. (2013), as it does not require a specific sampling scheme.

Lastly, we also apply ordinary least squares regression using the Python package statsmodels (Seabold & Perktold, 2010), as the third way to measure the parameters' influence on the dependent variables. For this step, a simple linear regression model is fit using each parameter alone, with the resulting R^2 indicating the proportion of variance in the response variable explained by that parameter. For all three methods we also perform the analysis for a control variable that has no bearing on the model. This is done so as to avoid attributing misplaced significance to any parameter that is in actuality an artifact of the bootstrap calculation of the indices. For all users, percentiles and methods, the sensitivity measure of each parameter is compared to that of the control variable and is set to zero if it does not exceed it.

The rationale for applying these three methods (Delta, Sobol, and ordinary least squares) is because they allow the diagnostic evaluation of different effects that the uncertain parameters might have on the response variables (the frequency and maximum duration of water shortages at different percentiles of historical shortage). The first-order Sobol index and R^2 attribute importance to each parameter according to its effect on output variance, with the difference that R^2 can only capture linear effects between the dependent and independent variables. The relative difference in parameter sensitivity resulting from the two methods suggests the presence of non-linear relationships between a parameter and the output of interest. The application of the Delta method allows us to further analyze potential effects the parameters might have on higher order moments of the distribution. This is particularly relevant to parameters changing the likelihoods of events in the tails of the output distribution—in this experiment, the hydrologic factors sampled are indeed expected to produce both drier and wetter conditions with increased durations in some realizations.

To summarize, our magnitude-varying sensitivity analysis is performed in the following steps. For every user, the percentiles of water shortage experienced historically are calculated based on the annual magnitudes of shortage (these are the black lines shown in Fig. 3). The shortage magnitude at each discrete percentile of this distribution ($f'(u,t)$) is experienced with different frequencies in each realization from our ensemble. It is also associated with a different maximum duration in each realization. For each user in the basin, the three sensitivity analyses methods described above are applied to understand which uncertain factors influence variability across realizations in both the frequency and the maximum duration of every historical percentile of shortage (Panel V in Fig. 2).

4 Results and discussion

The following sections present the diagnostic results from the three magnitude-varying sensitivity analyses applied to the frequency (section 4.1) and maximum duration (section 4.2) of shortage at each historical percentile. Although the analysis was performed for the over 300 users represented in StateMod, our initial detailed diagnostic results focus on the six users shown in Fig. 3, who span a range of sectors, basin locations and levels of seniority. Our final section of results provides broader diagnostic insights across the full suite of multi-sector users in the basin (section 4.3).

4.1 Factors controlling the frequency of shortages

Fig. 4 shows the results of the three sensitivity methods applied to the senior-, median- and junior-right irrigation users (panels (a-c) in Fig. 3). The figure illustrates interesting differences in which factors dominate the frequencies with which each user experiences their historical percentiles of shortage across realizations. The frequencies of water shortages for the senior-right user (a, d, e) are largely controlled by the snowmelt timing parameter (in gold color), the relative effect of which is reduced when looking at the other two users. The other two most significant factors for this user are the mean wet and dry flow parameters (in blue and red, respectively). This finding suggests that this user is not considerably (if at all) susceptible to changes in the water demands of the basin's irrigation sector, even though this includes their own demands. Rather, they are more susceptible to changes in water supply, and particularly its seasonality, as the snowmelt timing parameter simulates a shift in peak runoff to earlier in the year.

On the other hand, the median- and junior-right users' shortage frequencies are controlled in large part by the irrigation demand (in orange). The effect of this factor also increases for shortages of larger magnitudes. The relative importance of the mean wet and dry flow also switches between low and high percentiles of shortage, which can be attributed to different system states being active when different levels of shortage occur. In other words, the magnitude-varying sensitivity analysis allows us to directly link the magnitude of shortage experienced by these users to the dry and wet conditions in the basin: their largest shortages occur when streamflow is low (determined by the mean dry flow parameter) and their smallest shortages occur when streamflow is high (determined by the mean wet flow parameter). Furthermore, for the junior-right user (Fig. 4 (c, f, i)) snowmelt timing is not as significant as for the other two users. This could be attributed to the timing of their shortages not coinciding with the timing of the shift in snowmelt.

Examining this figure top to bottom (i.e., across methods) we see that the relative factor significance attributed by the first-order Sobol indices and the R^2 values are largely the same. This suggests that the relationship between these factors and the corresponding water shortage frequencies is mostly linear. The differences observed between these figures (Fig. 4 (d-i)) and those showing the Delta method indices (Fig. 4 (a-c)) can be attributed to the effects these parameters have on higher-order moments of those distributions. In particular, factors such as the operation of the Shoshone plant (in dark purple) and the seniority of the environmental flow right (in green) are more apparent in these panels. Note that white areas in these panels represent magnitude percentiles where indices could not be computed as historical shortages were zero (see equivalent panels in Fig. 3).

The equivalent results for the large-decree irrigation user, the transbasin diversion and the 15-mile reach are presented in Fig. 5. Here, the frequencies of shortages at different historical percentiles are controlled primarily by a single factor: the irrigation demand for the irrigation user (in orange), the transbasin demand for the transbasin diversion (in cyan), and the environmental flow right seniority (in green) for the 15-mile reach. Similar to the previous findings, comparing between the first-order Sobol and the

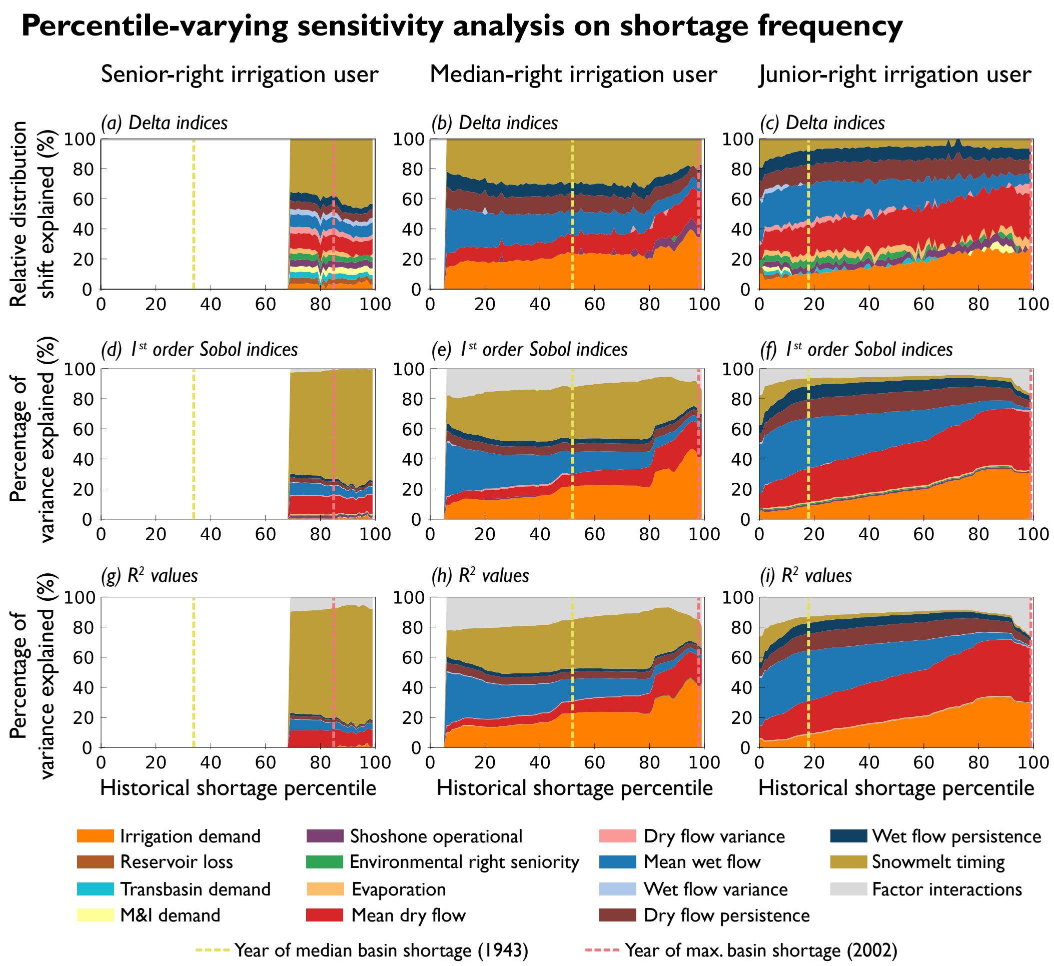

Figure 4. Percentile-varying sensitivity indices of shortage frequency for three irrigation users. Each panel presents the magnitude-varying sensitivity indices attributed to each factor, for the frequency of shortages experienced by: (a, d, g) the senior-right irrigation user; (b, e, h) the median-right irrigation user; and (c, f, i) the junior-right irrigation user. The results are ordered by method, with the first row showing factor significance as estimated using the Delta indices, the second row using the first-order Sobol indices, and the third row using the R^2 values. The dashed vertical lines in pink and yellow indicate the percentile of shortage experienced by each user in the years during which the basin as a whole experienced its median and worst shortages, 1943 and 2002, respectively. The colors in the legend are listed in the order that they are plotted, from bottom, up.

R^2 panels (Fig. 5 (d-i)) suggests that the nonlinear effects of the parameters do not change the rank order of factor importance. However, the effects are more nonlinear for these users, as the magnitude of the first order Sobol indices for irrigation demands (d), transbasin diversion demands (e) and environmental right seniority (f) are larger than their corresponding linear effects captured by the R^2 metric (g, h and i, respectively).

Several other factors are identified by the Delta method as affecting the frequency of shortages at moments beyond its variance for all three users (Fig. 5 (a-c)). Looking specifically at the 15-mile reach, the seniority of the environmental flow right explains

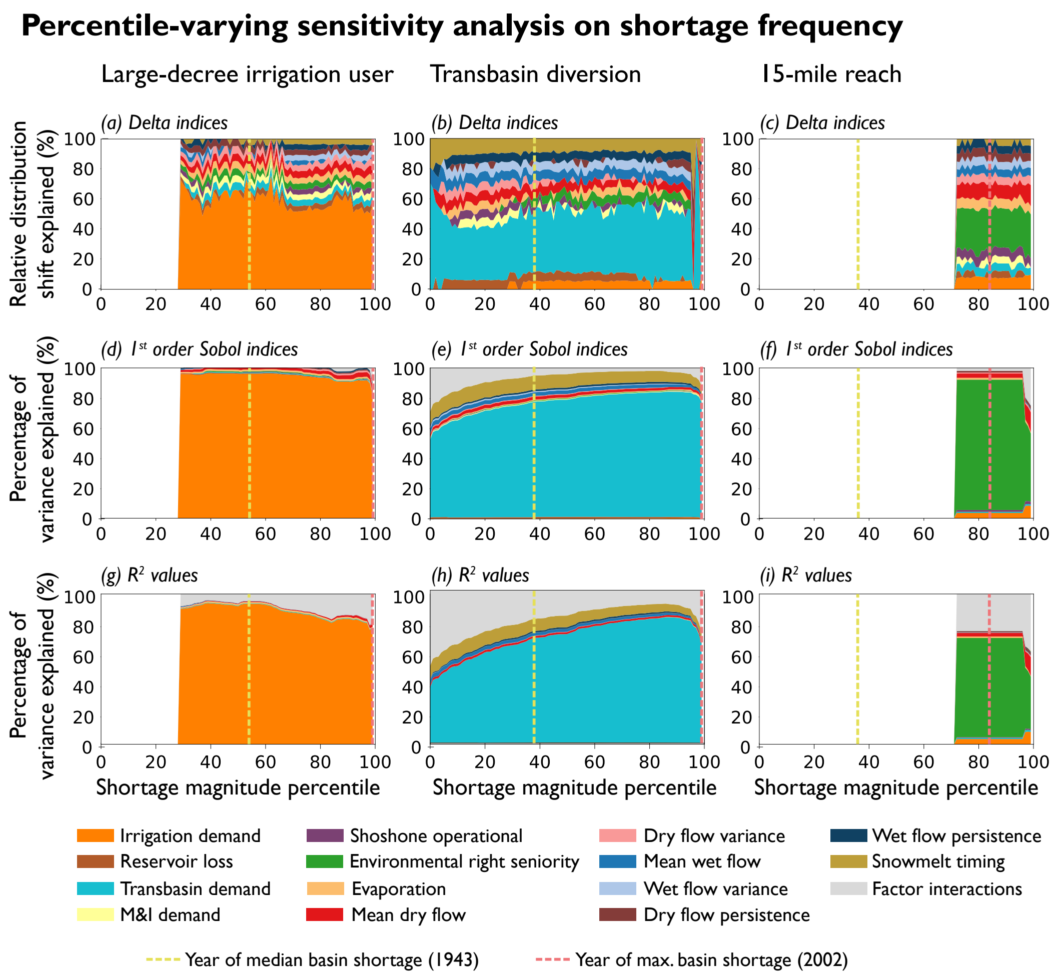

Figure 5. Percentile-varying sensitivity indices of shortage frequency for three basin users. Each panel presents the magnitude varying sensitivity indices attributed to each factor, for the frequency of shortages experienced by: (a, d, g) the large-decree irrigation user; (b, e, h) the transbasin diversion; and (c, f, i) the 15-mile reach. The results are ordered by method, with the first row showing factor significance as estimated using the Delta indices, the second row using the first-order Sobol indices, and the third row using the R^2 values. The dashed vertical lines in pink and yellow indicate the percentiles of shortage experienced by each user in the years during which the basin as a whole experienced its median and worst shortages, 1943 and 2002, respectively. The colors in the legend are listed in the order that they are plotted, from bottom, up.

more than 80% of the variance in shortage frequency. Looking at its equivalent Delta index (Fig. 5 (c)), the distributions of shortage frequency are also significantly affected by several other factors. This suggests that shortage frequencies at the tails of the distribution (i.e., SOWs that have extremely frequent large shortages) occur as a result of a combination of many factors changing together: increased demands (in orange, cyan, and yellow), drier and more variant flows (in red and light red), more evaporation (in light orange). Lastly, we compare the results from the Delta method applied to the transbasin diversion (Fig. 5 (b)) with those produced by the other two methods (Fig. 5 (e and h)). The findings suggest that reservoir loss (in brown) and irrigation demand (in orange) also affect the tail events in the distribution of frequencies, in that they contribute to the occurrence of either very frequent or very infrequent shortages of all magnitudes. Further, the first-order Sobol and R^2 indices (Fig. 5 (e and h)) attribute an increasing significance to the transbasin demand (in cyan) as we move to higher percentiles of shortage, something not seen in the equivalent Delta method panel (Fig. 5 (b)). This difference has implications regarding the effect of changing transbasin diversion demands on the frequency of shortages experienced. It appears that changing demands have a stronger relative effect on high-magnitude shortages with regard to the variance of their frequency, but affect equally significantly the asymmetry and tail thickness of the frequency of shortages across all percentiles.

4.2 Factors controlling the maximum duration of shortages

In addition to the severity and frequency of water shortages, their duration is a third dimension of concern, especially when considering large supply deficits caused by droughts (Lal et al., 2012; McKee et al., 2000; Timilsena et al., 2007). The maximum durations of different shortage magnitudes corresponding to historical percentiles of shortage are also examined for each user. Fig. 6 mirrors Fig. 3 but instead shows the maximum number of consecutive years each level of shortage was experienced historically and across the experiment. Let us use the median-right irrigation user (b) to illustrate how Fig. 6 should be interpreted and compared to Fig. 3. Historically, the magnitude of this user's 20th percentile annual shortage was approximately 0.2 million m^3 (Fig. 3 (b)). Fig. 6 shows that consecutive shortages of this magnitude and larger have historically been at most 25 years long (y position of black line when x=20). The maximum durations of shortages equal or in excess of the shortages in the two reference years, 1943 and 2002, are also shown in yellow and pink, respectively.

Reflecting the impacts on shortage frequency seen in Fig. 3, we again see varying implications for the presented users. Approximately half of the realizations in our exploratory ensemble see the large-decree irrigation user (Fig. 6 (d)) and the transbasin diversion (Fig. 6 (e)) having significantly longer durations of shortages at all levels. In many of the SOWs of the ensemble, the transbasin diversion (Fig. 6 (e)) experiences some level of shortage (albeit small) at all times. The senior-right irrigation user (Fig. 6 (a)) sees all their shortage durations increase in length across the majority of ensemble realizations. The junior-right irrigation user (Fig. 6 (c)) and the 15-mile reach (Fig. 6 (f)) see their shortage durations decrease in most of the SOWs in the exploratory ensemble. One should note that the cumulative frequencies with which these shortage durations are observed across the ensemble depend on the ranges of parameters sampled in the experimental design and should not be considered predictive for this system. They are rather used in a diagnostic manner, as explained in the introduction. In other words, the intent here is to gauge the response of the model output that is relevant to each user specifically (i.e., the duration of their water shortages) as a result of an intentionally broad range of perturbations (the deeply uncertain factors). The approach is used to demonstrate how performing model diagnostics at this decision-relevant scale indeed illuminates how the impacts to each user would differ should the hypothetical perturbations come to be in the system.

We further assess the relative importance of these changing factors on shortage duration, by following the same sensitivity analysis procedure for the maximum durations of all historical shortage percentiles. The results are presented for the six users in Figs. 7 and 8. Looking at the three irrigation users and comparing with the equivalent results for shortage frequency (Fig. 4), the factors controlling the maximum duration (Fig. 7) do show some differences. The persistence of dry flow (in dark red) has a larger relative significance for this aspect of drought behavior for all three users. Change in this factor reflects a change in the likelihood of a dry water year being followed by another dry

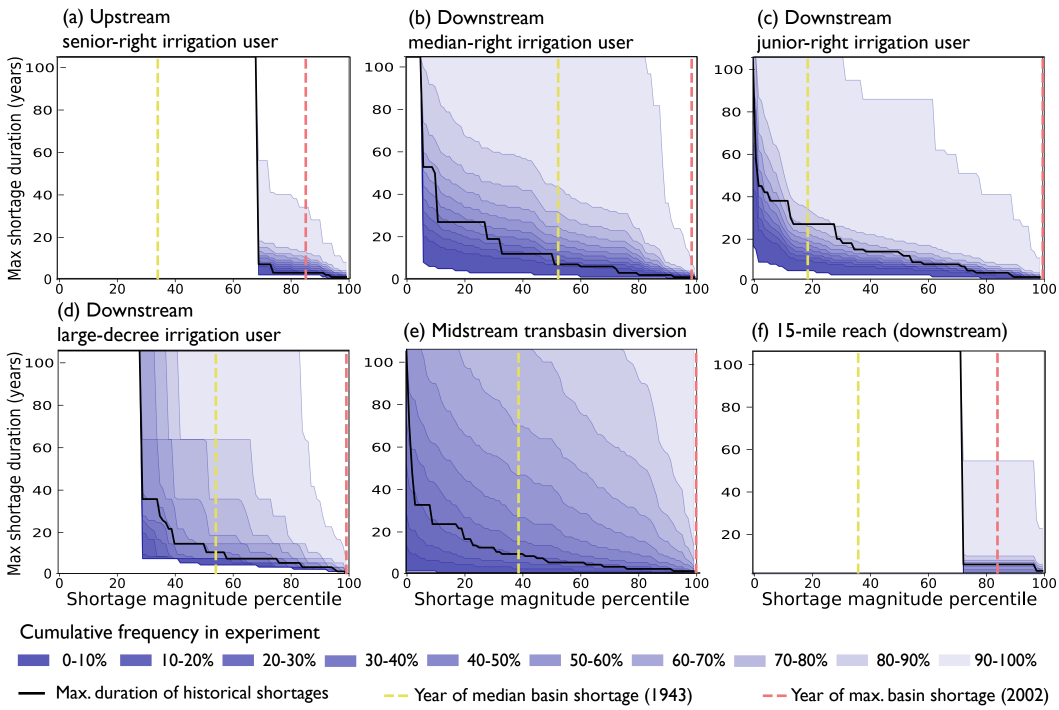

Figure 6. Percentile-varying impacts on the maximum duration of different levels of shortage for six users in the basin. Each panel represents the maximum duration of shortages in excess of different historical annual percentiles experienced by a specific user in the basin: (a) a senior-right irrigation 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; (f) the 15-mile reach. The black line in every panel indicates the maximum duration (y-axis) of consecutive annual shortages in excess of different historical shortage percentiles (x-axis) that was experienced historically by each user. The shaded areas represent the percent of realizations across the ensemble in which different maximum durations (y-axis) of shortage in excess of different historical shortage percentiles (x-axis) are experienced. Lighter shades indicate increased cumulative frequency within the ensemble. The dashed vertical lines in pink and yellow indicate the percentiles of shortage experienced by each user in the years during which the basin as a whole experienced its median and worst shortages, 1943 and 2002, respectively.

water year, thereby increasing the duration of experiencing water shortages at all levels.

Another notable difference between these results (Fig. 7) and the equivalent for the frequency of shortage (Fig. 4) is the increased relative effect of interactions between factors (in grey), especially for the junior-right user (Fig. 7 (f and i)). The implication of this difference is that there is increased non-linearity and complexity in how these changing factors are propagated through the networked system to affect the duration of shortages experienced by this user. This is also reflected in the diminished first-order and linear importance attributed to the wet-flow persistence (in dark blue) in these two panels, which is still captured by the Delta method (Fig. 7 (c)) as it affects the distribution of durations in a way that does not significantly change the variance.

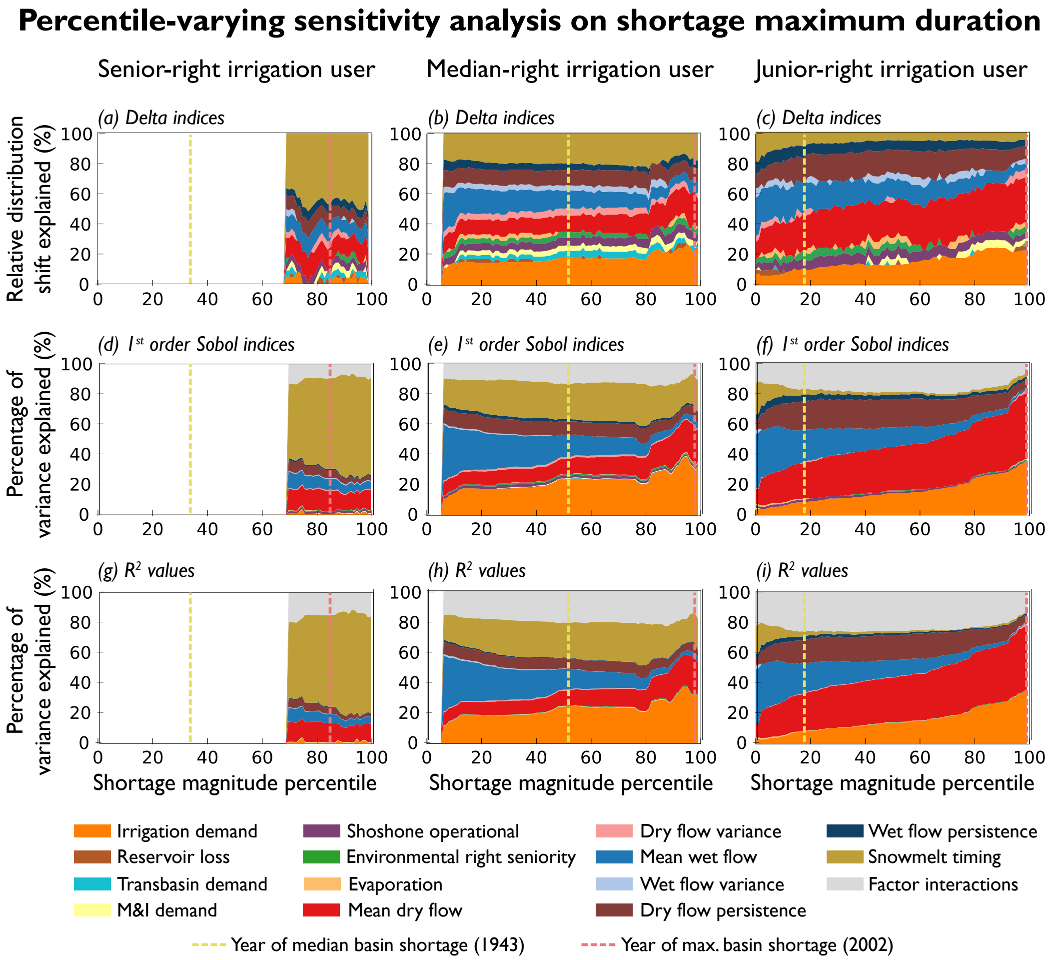

Figure 7. Percentile-varying sensitivity indices of the maximum duration of shortages in excess of different historical annual percentiles for three irrigation users. Each panel presents the magnitude varying sensitivity indices attributed to each factor, for: (a,d,g) the senior-right irrigation user; (b,e,h) the median-right irrigation user; and (c,f,i) the junior-right irrigation user. The results are ordered by method, with the first row showing factor significance as estimated using the Delta indices, the second row using the first-order Sobol indices, and the third row using the R^2 values. The dashed vertical lines in pink and yellow indicate the percentiles of shortage experienced by each user in the years during which the basin as a whole experienced its median and worst shortages, 1943 and 2002, respectively.

The largest differences in this regard are seen in the Delta indices for the median-right user: several factors that were not attributed any importance with regards to their effect on shortage frequency, do influence the maximum duration to at least some extent (Fig. 7 (b)). Specifically, municipal and industrial, and transbasin water demands (in yellow and cyan, respectively), the operation of Shoshone (in dark purple), the seniority of the environmental flow right at the 15-mile reach (in green), and dry and wet flow variance (in light red and light blue, respectively) all affect the distribution of shortage durations experienced by this user. This is not the case for the equivalent frequencies of these shortages (Fig. 4 (b)).

The results for the other three users (the large-decree irrigation user, transbasin diversion, and 15-mile reach) are largely consistent when comparing the sensitivity indices for frequency (Fig. 5) and those for maximum duration (Fig. 8). However, factor interactions (in grey) are relatively more important for the maximum shortage durations experienced by these users, as compared to their importance for shortage frequencies. Their importance also increases for durations of shortages of larger magnitudes (Fig. 8 (d-e) and (g-h)). These observations hold especially for the transbasin diversion: compare Fig. 5 (e) with Fig. 8 (e) and 5 (h) with Fig. 8 (h). This result could be attributed to the fact that there is less variation across the ensemble in the frequency of low magnitude shortages (Fig. 3 (e)) and in the duration of high magnitude shortages (Fig. 6 (e)), making it harder to decompose the variance to the various factors. Consequently, this complicates the identification of parameter changes leading to high-magnitude (and therefore high impact) drought durations, such as repeated years of 2002 shortage levels (the red dashed line).

The 2002 drought event, one of the most severe droughts ever recorded in the state of Colorado (Pielke et al., 2005), was indeed one of the worst years in terms of shortages experienced by the six users highlighted in this study (indicated by a red dashed line in Fig. 3 and others), but some users experienced worse shortages in other years. There is naturally more concern with these high-magnitude events, especially when their occurrence is sustained for several years, but if different users experience severe shortages in different years, diagnosing their influential factors using time-varying sensitivity analysis would not be appropriate for everyone. This is where performing sensitivity analysis in a magnitude-varying manner is most valuable, as it allows us to diagnose their manifestation in the model even if they occur at different times for different users.

Consider, for instance, the shortages for the three irrigation users presented in Figs. 4 and 7. Moving to higher percentiles (i.e., considering their largest experienced shortages) we generally see increasing first-order effects from irrigation demands and mean dry flow for the median- and junior-right users, and increasing effects from snowmelt timing for the senior-right user. However, this increasing importance is not always monotonic (sometimes there are abrupt shifts) which complicates the identification of consequential scenarios, especially when also looking at the results of the Delta method. If we just focus on the frequency and duration of higher-magnitude shortages (for instance, those above the 80th percentile), there are several factors contributing at least to some extent to these events. In fact, all 14 factors considered appear at least once in panels (a-c) of Figs. 4 and 7. As a reminder, we have also included an inconsequential control variable in all analyses to ensure that all identified factors indeed matter to the outcome more than the control.

Performing the sensitivity analysis on increasing percentiles of shortage allows us to illuminate the varying importance of the considered factors, as well as pinpoint the ones related with outcomes most relevant to each user. Performing the analysis on both the frequency and maximum duration of those shortages informs the identification of consequential scenarios pertaining to each. It also makes apparent that the duration of shortages, especially high-magnitude shortages, is a result of more interactive and non-linear relationships between the factors, making it therefore more difficult to develop triggers for adaptive management strategies (e.g., for water conservation) to reduce such impacts. Finally, performing the analysis on the many stakeholders of the basin further illuminates where such control strategies would be most effective, or how those triggers should vary across users.

4.3 Summarizing important effects across the basin

Even though the six users highlighted were selected to reflect stakeholders of different characteristics, demand levels, diversion locations and right seniorities, the con-

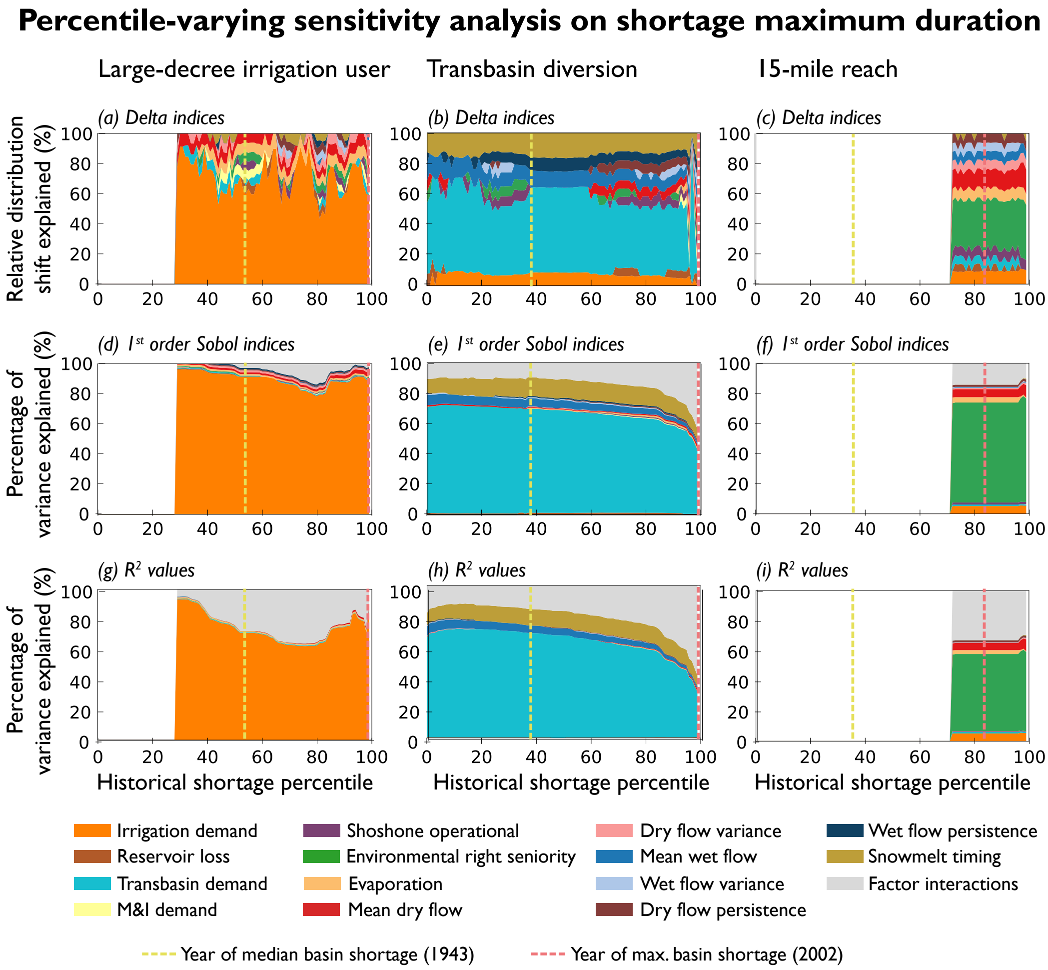

Figure 8. Percentile-varying sensitivity indices of the maximum duration of shortages in excess of different historical annual percentiles for three basin users. Each panel presents the magnitude varying sensitivity indices attributed to each factor, for: (a,d,g) the large-decree irrigation user; (b,e,h) the transbasin diversion; and (c,f,i) the 15-mile reach. The results are ordered by method, with the first row showing factor significance as estimated using the Delta indices, the second row using the first-order Sobol indices, and the third row using the R^2 values. The dashed vertical lines in pink and yellow indicate the percentiles of shortage experienced by each user in the years during which the basin as a whole experienced its median and worst shortages, 1943 and 2002, respectively.

clusions drawn with regards to the factors most strongly affecting the frequency and duration of their shortages do not necessarily extend to all users in the basin. The following section summarizes these results for all simulated users by examining the two water years in the basin's record during which the median and worst total basin shortages across users were experienced: 1943 and 2002, respectively. Water year 2002 is one of the most severe drought years in the record with effects felt throughout the state (Pielke et al., 2005). Yet the effects of the 2002 drought were not uniformly worst across users, nor was 1943 uniformly the median year. The shortages during water years 1943 and 2002 therefore correspond to different historical percentiles of shortage for every user (indicated by yellow and pink dashed lines in Fig. 3).

Fig. 9 presents the single most important factor affecting the frequency and maximum duration of shortages at the corresponding historical percentiles of shortage for years 1943 and 2002, for all users in the basin. The factors shown here are those identified using the Delta method, but as can be seen in Figs. 4 - 8, the three methods generally agree on the most dominant factor they identify. In every panel of this figure each user is represented by a radius. Radii lengths denote the ratio of shortage to demand experienced by every user in water year 1943 (Fig. 9 (a) and (c)), and water year 2002 (Fig. 9 (b) and (d)). The users are plotted in decreasing water right seniority, starting from the right-hand side of the circle and moving anti-clockwise. Since some users own multiple rights, seniority was determined using a volume-weighted rank. Radii colors indicate the single most important factor affecting the frequency (top row) or maximum duration (bottom row) of experiencing shortages in excess of historical 1943 (left column) and 2002 (right column) magnitudes for every user. To facilitate interpretation, the factors that do appear in the radial plots are indicated in bold text and a black border in the legend. For instance, reservoir loss (in light brown) is never the most important factor affecting either the frequency or maximum duration of these shortage levels for any user. There are users that did not experience a shortage in one or both of the years (e.g., in Fig. 3 we see that users (a) and (f) did not experience any shortage during 1943), in which case the dominant factors of frequency and shortage could not be identified.

Several model insights can be drawn from this visual summary. Looking at the lengths of the radii, it appears that water right seniority alone is surprisingly not predictive of impacts to the users in the basin. Moving down the seniority rank (anti-clockwise from the right-hand side) does not produce an equivalent sorting of shortage ratios for either of the two years. This conclusion is also supported by findings in Hadjimichael et al. (2020) with regards to the robustness of the basin's users to uncertainties in future basin characteristics. Comparing between the two water years, significantly larger shortages are seen in 2002, with several users being fully short of their demanded water. In contrast, none of the 1943 shortages are above 90% of demand. The majority of modeled users (approximately 65%) experienced no shortages during water year 1943; the majority of them (approximately 78%) also experienced at least some level of shortage in the 2002 water year.

Looking at radii colors, irrigation demand (in orange) and mean dry flow (in red) are the most commonly identified single most important factors. Other factors identified as most important for users are the transbasin demand (in cyan), the operation of Shoshone (in dark purple), the seniority of the environmental flow right (in green), the mean wet flow (in blue), and the change in snowmelt timing (in gold). As evident in Figs. 4 - 8, these are not the only important parameters for each user, but are shown here to illustrate the variety of factors that are most consequential to different users when such decision-relevant metrics are considered. Lastly, one notes differences in the color of several radii when comparing between Fig. 9 (a) and (c), and when comparing between Fig. 9 (b) and (d). This suggests that for the same level of shortage for the same user, the most dominant factor controlling the shortage's frequency differs from the equivalent factor controlling its maximum duration.

Relating these results to their spatial context, Fig. 10 places the important factors presented in radial form for each user in Fig. 9 (b) at their diversion location. The size of every point reflects the size of shortage during water year 2002. Black color is used to indicate diversions that did not have a shortage during that water year and therefore no dominant factor could be identified ("N/A"). The points indicated as having transbasin demands as their most important factor (in cyan) do in fact correspond to the transbasin diversion locations represented in the model (diversion tunnels). Shortage frequency for users diverting from Roan Creek and Roaring Fork is largely (but not entirely) driven by changes in mean dry flow (in red). This finding suggests that users in these tributaries might be more sensitive to streamflow (un)availability rather than demands depleting

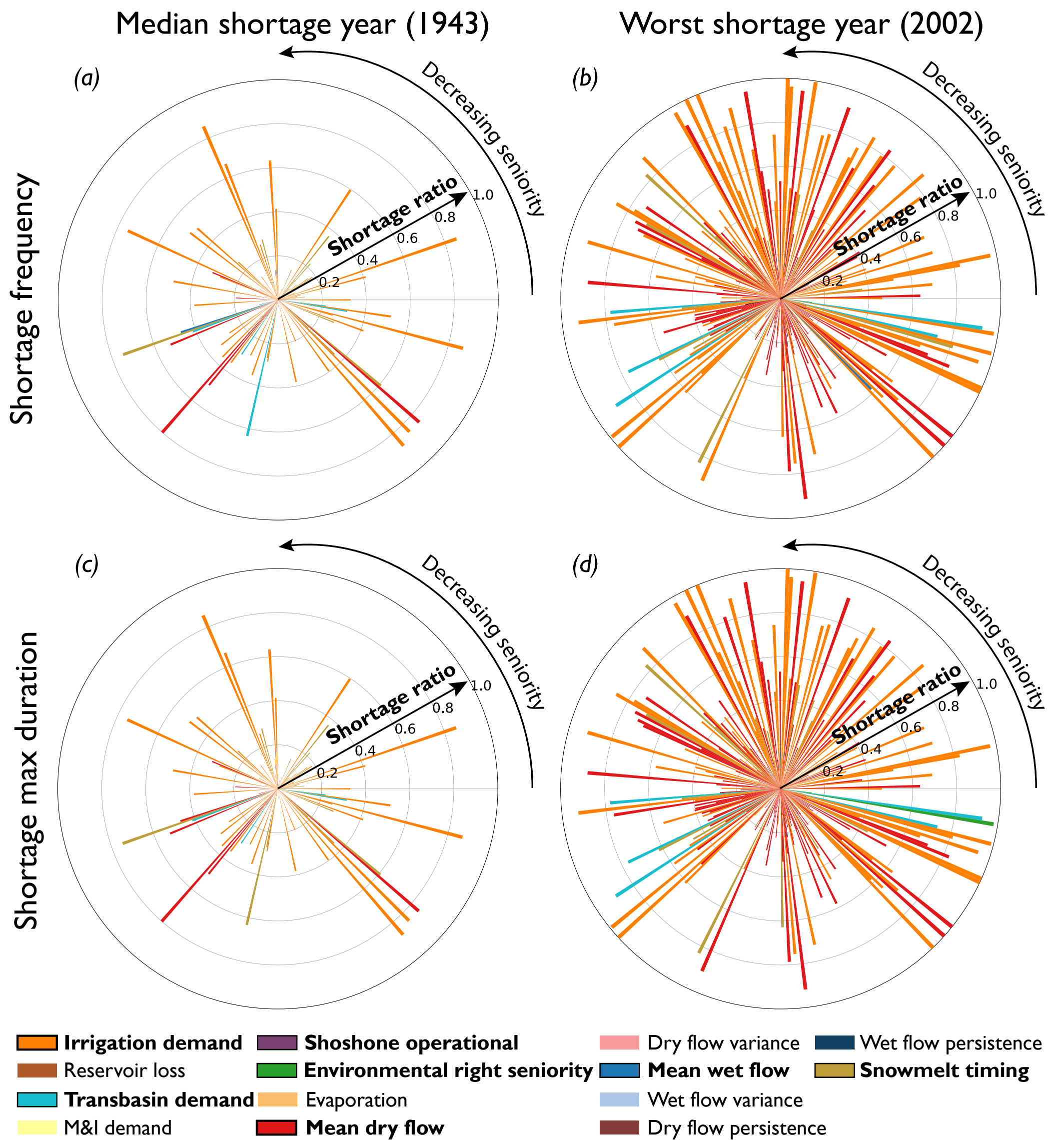

Figure 9. Shortages as a fraction of demand for all users in the basin, colored by the single most important factor influencing their frequency and maximum duration of 1943 and 2002-level shortages. Each radius represents a UCRB user, sorted anti-clockwise in decreasing seniority from 0 degrees. Radii lengths denote the shortage to demand ratio experienced by each user in water years 1943 (a and c) and 2002 (b and d). The color of each radius indicates the dominant factor identified by the Delta method as controlling the frequency (a-b) or the maximum duration (c-d) of experiencing shortages in excess of that year's shortage for each user. The factors indicated in bold text and a black border in the legend are factors identified to be most important for at least one user in at least one of the panels.

flow. Another notable insight from this figure is that beyond the aforementioned, there are no other clear spatial clusters of co-located users with the same dominant factor. This result underlines that water shortages as well as the factors most significantly shaping this consequential model output vary as a result of water right seniority and other user-specific characteristics. This contribution of our paper has only been possible through the detailed representation of institutional information on water supply for this basin, enabled by StateMod.

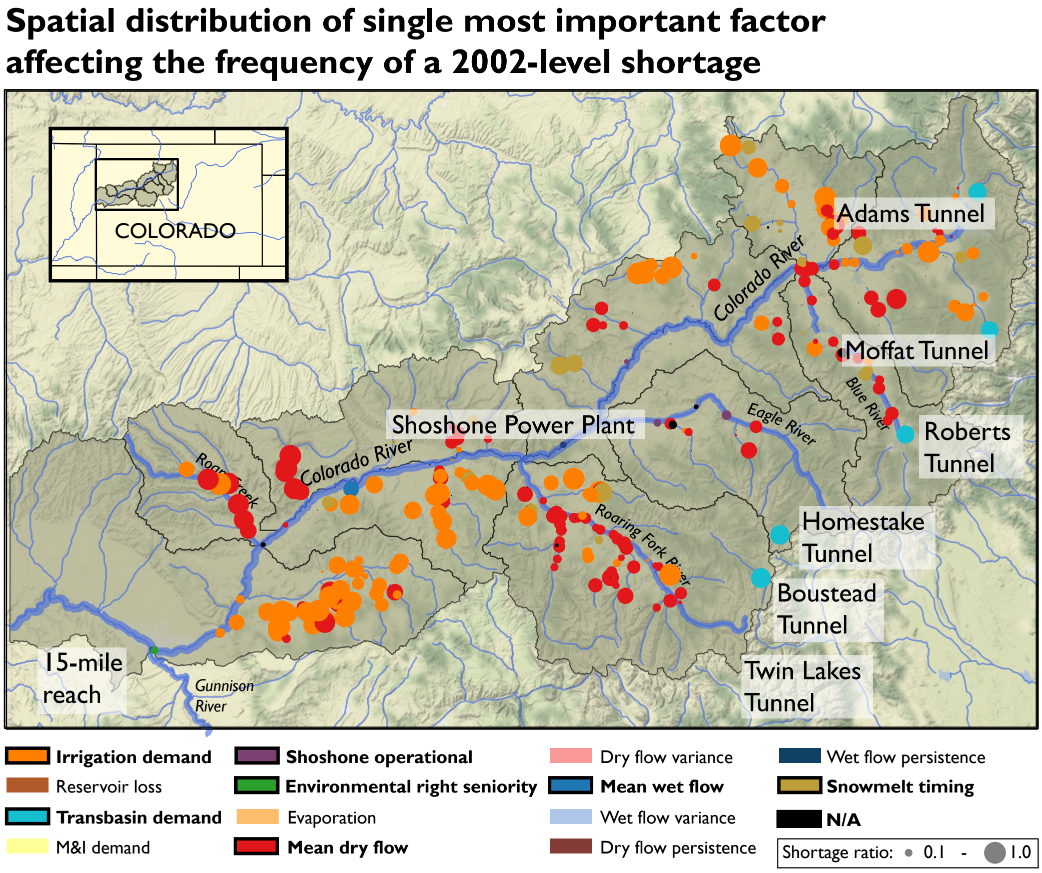

Figure 10. Shortages for all users in the basin during water year 2002, colored by the single most important factor identified as contributing to the frequency of an equivalent shortage. Each point represents a UCRB user and the size of each point denotes the shortage to demand ratio experienced by that user during the 2002 water year. The color of each point indicates the dominant factor identified as controlling the frequency of that level of shortage.

5 Conclusions and future work

This study contributes a diagnostic framework for water resources systems models representing institutionally complex river basins with many stakeholders. The framework pairs exploratory modeling with global sensitivity analysis methods and visual analytic techniques to evaluate a fine-scale water supply and allocation model, StateMod. The sensitivity analysis is applied in a novel, magnitude-varying manner that estimates factor contributions to the frequency and maximum duration of different levels of shortage. The reasoning behind this approach is primarily rooted in the hypothesis that different factors are likely to dominate the system states related to different shortage magnitudes (e.g., small shortages happening under wet conditions versus large shortages happening during droughts), articulated by several previous studies and also demonstrated herein (Herman et al., 2013; Pianosi et al., 2016; Quinn et al., 2019; Rougé et al., 2019). These prior studies have advocated for the use of time-varying sensitivity analysis in model diagnostics to understand how system sensitivities vary under flood versus drought events. However, different water users in a basin may experience varying degrees of impacts from such events, particularly in complex multi-user basins that are heavily influenced by institutional as well as hydrologic factors. In such cases, we illustrate the benefits of instead performing magnitude-varying sensitivity analysis for model diagnostics, so that the factors influencing drought severity for particular users can be assessed independent of the time they occurred. As seen here, while the *basin-wide* impacts of the 2002 drought were most severe, this was not true for all users. By performing magnitude-varying sensitivity analysis, we were able to find which factors are most important for each user's most extreme shortages, a more decision-relevant metric for everyone than the basin-level event. Applying this approach, we find that the dominant model parameters shaping the frequency and duration of shortages indeed vary among users and when transitioning across percentiles of shortage magnitude. Hydrologic factors in particular switch in relative importance when moving from smaller to larger shortage magnitudes. Across users, we see several different dominant shortage controls even for users of the same demand levels, water use, water right seniority, and basin location.

Future work will further examine how these characteristics relate and potentially shape how the dominant factors propagate through such institutionally complex river basins to affect their users. For example, clustering could be applied to discover groups of users with similar institutional and other characteristics that also have common factors influencing their water shortages. Even though the extent of factor interactions varies across users and percentiles, it appears to be more relevant to the durations of shortages, as opposed to their frequencies. Additional and more comprehensive sensitivity analyses should also be performed to more rigorously diagnose how uncertain factors interact to shape these decision-relevant model outputs. Lastly, to create the ensemble of uncertain factors we employ an experimental design that samples plausible, but wide and uniform ranges. Future analyses should investigate the implications of such design choices on the conclusions drawn by, for instance, fitting the synthetic streamflow generator to different hydrologic data or using more detailed water demand projections across the different sectors.

Acknowledgments

This research was supported by the U.S. Department of Energy, Office of Science, as part of research in Multi-Sector Dynamics, Earth and Environmental System Modeling Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding entities. StateMod is available at `https://github.com/OpenCDSS`. The input files to run StateMod for the UCRB can be found here `https://www.colorado.gov/pacific/cdss/surface-water-statemod`. All the scripts to replicate the analysis performed in this paper and regerate the figures can be found here `https://github.com/antonia-had/cdss-app-statemod-fortran/tree/2063d13/UCRB_analysis`.

References

Armistead, C., & Mojica, J. (2018). *Baseline Assessment: The Economic Contribution of Recreation in the Middle Colorado Watershed* (Tech. Rep.). Tacoma, WA: Earth Economics. Retrieved 2020-04-22, from https://static1.squarespace.com/static/582a05f020099e61757374f5/t/5afdfcc3758d466fdfc6a26c/1526594924982/EconomicContributionofRecreationMiddleColoradov1_2+w%3Apreface_title.pdf

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

Ault, T. R., Cole, J. E., Overpeck, J. T., Pederson, G. T., St. George, S., Otto-Bliesner, B., … Deser, C. (2013, March). The Continuum of Hydroclimate Variability in Western North America during the Last Millennium. *Journal of Climate*, *26*(16), 5863–5878. Retrieved 2020-04-28, from https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-11-00732.1 (Publisher: American Meteorological Society) doi: 10.1175/JCLI-D-11-00732.1

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

Bankes, S., Lempert, R., & Popper, S. (2001, March). Computer-Assisted Reasoning. *Computing in Science & Engineering*, *3*(2), 71. Retrieved 2020-04-15, from https://aip.scitation.org/doi/abs/10.1109/5992.909006 (Publisher: IEEE Computer Society) doi: 10.1109/5992.909006

Borgonovo, E. (2007, June). A new uncertainty importance measure. *Reliability Engineering & System Safety*, *92*(6), 771–784. Retrieved 2018-11-30, from http://www.sciencedirect.com/science/article/pii/S0951832006000883 doi: 10.1016/j.ress.2006.04.015

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

Brown, C. M., Lund, J. R., Cai, X., Reed, P. M., Zagona, E. A., Ostfeld, A., … Brekke, L. (2015). The future of water resources systems analysis: Toward a scientific framework for sustainable water management. *Water Resources Research*, *51*(8), 6110–6124. Retrieved 2019-09-30, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017114 doi: 10.1002/2015WR017114

Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., … Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the national academy of sciences*, *100*(14), 8086–8091. (Publisher: National Acad Sciences)

Christensen, N. S., & Lettenmaier, D. P. (2007, July). 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*, *11*(4), 1417–1434. Retrieved 2020-04-28, from https://www.hydrol-earth-syst-sci.net/11/1417/2007/ (Publisher: Copernicus GmbH) doi: https://doi.org/10.5194/hess-11-1417-2007

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9). Retrieved 2020-04-15, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR009827 (_eprint:

https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010WR009827) doi: 10.1029/2010WR009827

CWCB. (2010). *Statewide Water Supply Initiative* (Tech. Rep.). Colorado Water Conservation Board.

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

CWCB. (2019). *Analysis and Technical Update to the Water Plan* (Tech. Rep.). Colorado Water Conservation Board.

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

DOLA. (2015). *Population Totals Colorado Substate.* Retrieved 2019-10-10, from `https://demography.dola.colorado.gov/population/population-totals-colorado-substate/`

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/`

Goodwell, A. E., Jiang, P., Ruddell, B. L., & Kumar, P. (2020). Debates—Does Information Theory Provide a New Paradigm for Earth Science? Causality, Interaction, and Feedback. *Water Resources Research*, *56*(2), e2019WR024940. Retrieved 2020-04-30, from `http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019WR024940` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019WR024940) doi: 10.1029/2019WR024940

Graf, W. L., Wohl, E., Sinha, T., & Sabo, J. L. (2010). Sedimentation and sustainability of western American reservoirs. *Water Resources Research*, *46*(12).

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, *34*(4), 751–763. Retrieved 2020-04-07, from `http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97WR03495` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/97WR03495) doi: 10.1029/97WR03495

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes: An International Journal*, *22*(18), 3802–3813. (Publisher: Wiley Online Library)

Hadjimichael, A., Quinn, J., Wilson, E., Reed, P., Basdekas, L., Yates, D., & Garrison, M. (2020). Defining robustness, vulnerabilities, and consequential scenarios for diverse stakeholder interests within the upper colorado river basin. *Earth's Future*.

Herman, J. D., Reed, P. M., & Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research*, *49*(3), 1400–1414. Retrieved 2020-01-02, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/wrcr.20124` doi: 10.1002/wrcr.20124

Herman, J. D., Reed, P. M., Zeff, H. B., & Characklis, G. W. (2015). How Should Robustness Be Defined for Water Systems Planning under Change? *Journal of Water Resources Planning and Management*, *141*(10), 04015012. doi: 10.1061/(ASCE)WR.1943-5452.0000509

Herman, J. D., & Usher, W. (2017). SALib: An open-source Python library for Sensitivity Analysis. *J. Open Source Software*, *2*(9), 97. doi: http://dx.doi.org/10.21105/joss.00097

Hornberger, G. M., & Spear, R. C. (1981). Approach to the preliminary analysis of environmental systems. *J. Environ. Mgmt.*, *12*(1), 7–18.

IUCN. (2012a, April). *Gila cypha: NatureServe: The IUCN Red List of Threatened Species 2014: e.T9184A54036093.* (Tech. Rep.). International Union for Conservation of Nature. Retrieved 2020-01-30, from https://www.iucnredlist.org/species/9184/54036093 (type: dataset) doi: 10.2305/IUCN.UK.2014-1.RLTS.T9184A54036093.en.

IUCN. (2012b, April). *Gila elegans: NatureServe: The IUCN Red List of Threatened Species 2013: e.T9186A18232150.* (Tech. Rep.). International Union for Conservation of Nature. Retrieved 2020-01-30, from https://www.iucnredlist.org/species/9186/18232150 (type: dataset) doi: 10.2305/IUCN.UK.2013-1.RLTS.T9186A18232150.en.

IUCN. (2012c, April). *Ptychocheilus lucius: NatureServe: The IUCN Red List of Threatened Species 2013: e.T18829A19034371* (Tech. Rep.). International Union for Conservation of Nature. Retrieved 2019-10-02, from http://www.iucnredlist.org/details/18829/0 (type: dataset) doi: 10.2305/IUCN.UK.2013-1.RLTS.T18829A19034371.en

IUCN. (2012d, April). *Xyrauchen texanus: NatureServe: The IUCN Red List of Threatened Species 2013: e.T23162A19032625* (Tech. Rep.). International Union for Conservation of Nature. Retrieved 2019-10-02, from http://www.iucnredlist.org/details/23162/0 (type: dataset) doi: 10.2305/IUCN.UK.2013-1.RLTS.T23162A19032625.en

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual Analytics: Definition, Process, and Challenges. In A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North (Eds.), *Information Visualization: Human-Centered Issues and Perspectives* (pp. 154–175). Berlin, Heidelberg: Springer. Retrieved 2020-04-21, from https://doi.org/10.1007/978-3-540-70956-5_7 doi: 10.1007/978-3-540-70956-5_7

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

Knight, F. H. (1921). *Risk, Uncertainty and Profit.* Courier Dover Publications.

Komarek, A. M., De Pinto, A., & Smith, V. H. (2020, February). A review of types of risks in agriculture: What we know and what we need to know. *Agricultural Systems*, *178*, 102738. Retrieved 2020-05-18, from http://www.sciencedirect.com/science/article/pii/S0308521X18312034 doi: 10.1016/j.agsy.2019.102738

Kwakkel, J. H., Walker, W. E., & Marchau, V. A. (2010). Classifying and communicating uncertainties in model-based policy analysis. *International journal of technology, policy and management*, *10*(4), 299–315.

Lal, R., Delgado, J. A., Gulliford, J., Nielsen, D., Rice, C. W., & Pelt, R. S. V. (2012, November). Adapting agriculture to drought and extreme events. *Journal of Soil and Water Conservation*, *67*(6), 162A–166A. Retrieved 2020-05-28, from https://www.jswconline.org/content/67/6/162A (Publisher: Soil and Water Conservation Society Section: A Section) doi: 10.2489/jswc.67.6.162A

Lamontagne, J. R., Reed, P. M., Link, R., Calvin, K. V., Clarke, L. E., & Edmonds, J. A. (2018). Large Ensemble Analytic Framework for Consequence-Driven Discovery of Climate Change Scenarios. *Earth's Future*, *6*(3), 488–504. Retrieved 2020-03-06, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017EF000701 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017EF000701) doi: 10.1002/2017EF000701

Lebedev, S. (2015). *hmmlearn.* Retrieved from https://hmmlearn.readthedocs.io/

Lempert, R. J. (2002, May). A new decision sciences for complex systems. *Proceedings of the National Academy of Sciences*, *99*(suppl 3), 7309–7313. Retrieved 2019-02-22, from https://www.pnas.org/content/99/suppl_3/7309 doi: 10.1073/pnas.082081699

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

Livneh, B., Deems, J. S., Buma, B., Barsugli, J. J., Schneider, D., Molotch, N. P., ... Wessman, C. A. (2015, April). Catchment response to bark beetle outbreak and dust-on-snow in the Colorado Rocky Mountains. *Journal of Hydrology*, *523*, 196–210. Retrieved 2019-12-06, from http://www.sciencedirect.com/science/article/pii/S0022169415000566 doi: 10.1016/j.jhydrol.2015.01.039

Loucks, D. P. (1992, May). Water Resource Systems Models: Their Role in Planning. *Journal of Water Resources Planning and Management*, *118*(3), 214–223. Retrieved 2020-04-02, from http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%281992%29118%3A3%28214%29 doi: 10.1061/(ASCE)0733-9496(1992)118:3(214)

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

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

McKee, T. B., Doesken, N. J., Kleist, J., Shrier, C. J., & Stanton, W. P. (2000). *A history of drought in Colorado: Lessons learned and what lies ahead* (Tech. Rep.). SColorado Water Resources Research Institute. (Publisher: Colorado Water Institute (CWI))

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

Milly, P. C. D., & Dunne, K. A. (2020, March). Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. *Science*, *367*(6483), 1252–1255. Retrieved 2020-05-12, from https://science.sciencemag.org/content/367/6483/1252 (Publisher: American Association for the Advancement of Science Section: Report) doi: 10.1126/science.aay9187

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

Nowak, K., Prairie, J., Rajagopalan, B., & Lall, U. (2010). A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow. *Water Resources Research*, *46*(8).

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994, February). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, *263*(5147), 641–646. Retrieved 2020-04-15, from https://science.sciencemag.org/content/263/5147/641 (Publisher:

American Association for the Advancement of Science Section: Articles) doi: 10.1126/science.263.5147.641
Parsons, R., & Bennett, R. (2006). Reservoir Operations Management Using a Water Resources Model. *Operating Reservoirs in Changing Conditions*, 304–311. Retrieved 2019-07-08, from https://ascelibrary.org/doi/abs/10.1061/40875(212)30 doi: 10.1061/40875(212)30
Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016, May). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. Retrieved 2020-02-19, from http://www.sciencedirect.com/science/article/pii/S1364815216300287 doi: 10.1016/j.envsoft.2016.02.008
Pianosi, F., & Wagener, T. (2017, November). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrological Processes*, 3991–4003. Retrieved 2020-01-02, from https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.10968%4010.1111/%28ISSN%291099-1085.Kieth-Beven doi: 10.1002/hyp.10968@10.1111/(ISSN)1099-1085.Kieth-Beven
Pielke, R. A., Doesken, N., Bliss, O., Green, T., Chaffin, C., Salas, J. D., … Wolter, K. (2005, August). Drought 2002 in Colorado: An Unprecedented Drought or a Routine Drought? *Pure and Applied Geophysics*, *162*(8), 1455–1479. Retrieved 2018-09-24, from https://doi.org/10.1007/s00024-005-2679-6 doi: 10.1007/s00024-005-2679-6
Plischke, E., Borgonovo, E., & Smith, C. L. (2013, May). Global sensitivity measures from given data. *European Journal of Operational Research*, *226*(3), 536–550. Retrieved 2019-01-07, from http://www.sciencedirect.com/science/article/pii/S0377221712008995 doi: 10.1016/j.ejor.2012.11.047
Polasky, S., Carpenter, S. R., Folke, C., & Keeler, B. (2011, August). Decision-making under great uncertainty: environmental management in an era of global change. *Trends in Ecology & Evolution*, *26*(8), 398–404. Retrieved 2014-05-28, from http://www.sciencedirect.com/science/article/pii/S0169534711001157 doi: 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. (2019). What Is Controlling Our Control Rules? Opening the Black Box of Multireservoir Operating Policies Using Time-Varying Sensitivity Analysis. *Water Resources Research*, *55*(7), 5962–5984. Retrieved 2020-04-16, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024177 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR024177) doi: 10.1029/2018WR024177
Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M., Dai, A., … Zhang, G. (2014, March). Climate Change Impacts on the Water Balance of the Colorado Headwaters: High-Resolution Regional Climate Model Simulations. *Journal of Hydrometeorology*, *15*(3), 1091–1116. Retrieved 2020-04-28, from https://journals.ametsoc.org/doi/full/10.1175/JHM-D-13-0118.1 (Publisher: American Meteorological Society) doi: 10.1175/JHM-D-13-0118.1
Raso, L., Kwakkel, J., Timmermans, J., & Panthou, G. (2019, March). How to evaluate a monitoring system for adaptive policies: criteria for signposts selection and their model-based evaluation. *Climatic Change*, *153*(1), 267–283. Retrieved 2020-05-12, from https://doi.org/10.1007/s10584-018-2355-3 doi: 10.1007/s10584-018-2355-3
Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitiv-

ity in Earth and Environmental systems models. *Water Resources Research*, *51*(5), 3070–3092. Retrieved 2020-04-14, from `http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR016527` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR016527) doi: 10.1002/2014WR016527

Rhee, G., Salazar, J., & Grigg, C. (2018, December). How Long Does a 15-Year Drought Last? On the Correlation of Rare Events. *Journal of Climate*, *32*(5), 1345–1359. Retrieved 2019-12-04, from `https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-18-0326.1` doi: 10.1175/JCLI-D-18-0326.1

Rougé, C., Reed, P. M., Grogan, D. S., Zuidema, S., Prusevich, A., Glidden, S., ... Lammers, R. B. (2019, November). Coordination and Control: Limits in Standard Representations of Multi-Reservoir Operations in Hydrological Modeling. *Hydrology and Earth System Sciences Discussions*, 1–37. Retrieved 2020-04-01, from `https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-589/` (Publisher: Copernicus GmbH) doi: https://doi.org/10.5194/hess-2019-589

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., ... Wu, Q. (2019, April). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29–39. Retrieved 2020-02-20, from `http://www.sciencedirect.com/science/article/pii/S1364815218302822` doi: 10.1016/j.envsoft.2019.01.012

Saltelli, A., & Funtowicz, S. (2014). When all models are wrong. *Issues in Science and Technology*, *30*(2), 79–85. (Publisher: JSTOR)

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer* (1edition ed.). Chichester, England ; Hoboken, NJ: Wiley-Interscience.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models.* John Wiley & Sons.

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`

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In *9th python in science conference.*

Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W., & Jakeman, A. J. (2013, October). Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. *Journal of Hydrology*, *503*, 135–152. Retrieved 2020-04-15, from `http://www.sciencedirect.com/science/article/pii/S0022169413006422` doi: 10.1016/j.jhydrol.2013.08.047

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*(8), 7362–7378. Retrieved 2019-12-06, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024573` doi: 10.1029/2018WR024573

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation*, *55*(1-3), 271–280. doi: https://doi.org/10.1016/S0378-4754(00)00270-6

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015, April). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, *523*, 739–757. Retrieved 2020-02-19, from `http://www.sciencedirect.com/science/article/pii/S0022169415001249` doi: 10.1016/j.jhydrol.2015.02.013

Star, S. L. (2010). This is not a boundary object: Reflections on the origin of a concept. *Science, Technology, & Human Values*, *35*(5), 601–617. (Publisher: Sage

Publications Sage CA: Los Angeles, CA)

Star, S. L., & Griesemer, J. R. (1989, August). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420. Retrieved 2020-05-16, from `https://doi.org/10.1177/030631289019003001` (Publisher: SAGE Publications Ltd) doi: 10.1177/030631289019003001

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

Tang, Y., Reed, P., Wagener, T., & Werkhoven, K. v. (2007, February). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrology and Earth System Sciences*, *11*(2), 793–817. Retrieved 2020-02-17, from `https://www.hydrol-earth-syst-sci.net/11/793/2007/` doi: https://doi.org/10.5194/hess-11-793-2007

Thomas, J. J., & Cook, K. A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* IEEE Computer Society, Los Alamitos, CA, United States(US). Retrieved 2020-04-21, from `https://www.osti.gov/biblio/912515`

Timilsena, J., Piechota, T. C., Hidalgo, H., & Tootle, G. (2007). Five Hundred Years of Hydrological Drought in the Upper Colorado River Basin. *Journal of the American Water Resources Association*, *43*(3), 798–812. Retrieved 2020-05-28, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1752-1688.2007.00064.x` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1752-1688.2007.00064.x) doi: 10.1111/j.1752-1688.2007.00064.x

USFWS. (1999, December). *Final programmatic biological opinion for Bureau of Reclamation's 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* (Tech. Rep.). United States Fish and Wildlife Service.

USFWS. (2020). *Upper Colorado River Endangered Fish Recovery Program.* Retrieved 2020-01-30, from `https://coloradoriverrecovery.org/`

USGS. (2019). *USGS Current Conditions for USGS 09085100 COLORADO RIVER BELOW GLENWOOD SPRINGS, CO* (Tech. Rep.). United States Geological Survey. Retrieved 2019-11-06, from `https://waterdata.usgs.gov/co/nwis/uv?site_no=09085100`

von Landesberger, T., Schreck, T., Fellner, D. W., & Kohlhammer, J. (2012). Visual Search and Analysis in Complex Information Spaces—Approaches and Research Challenges. In J. Dill, R. Earnshaw, D. Kasik, J. Vince, & P. C. Wong (Eds.), *Expanding the Frontiers of Visual Analytics and Visualization* (pp. 45–67). London: Springer. Retrieved 2020-04-21, from `https://doi.org/10.1007/978-1-4471-2804-5_4` doi: 10.1007/978-1-4471-2804-5_4

Wada, Y., Bierkens, M. F. P., Roo, A. d., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., … Wheater, H. (2017, August). Human–water interface in hydrological modelling: current status and future directions. *Hydrology and Earth System Sciences*, *21*(8), 4169–4193. Retrieved 2020-04-01, from `https://www.hydrol-earth-syst-sci.net/21/4169/2017/` (Publisher: Copernicus GmbH) doi: https://doi.org/10.5194/hess-21-4169-2017

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological processes*, *17*(2), 455–476. (Publisher: Wiley Online Library)

Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., … Wilson, J. S. (2010). The future of hydrology: An evolving science for a changing world. *Water Resources Research*, *46*(5). Retrieved 2020-05-12, from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008906` (_eprint:

https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009WR008906) doi: 10.1029/2009WR008906

Walker, W. E., & Marchau, V. (2003). Dealing with uncertainty in policy analysis and policymaking. *Integrated Assessment*, *4*(1), 1–4. Retrieved 2016-05-10, from http://www.tandfonline.com/doi/pdf/10.1076/iaij.4.1.1.16462

Wallander, S., Marshall, E., & Aillery, M. (2017, June). Farmers Employ Strategies To Reduce Risk of Drought Damages. *Amber Waves*. Retrieved 2020-05-18, from https://www.ers.usda.gov/amber-waves/2017/june/farmers-employ-strategies-to-reduce-risk-of-drought-damages/

White, D. D., Wutich, A., Larson, K. L., Gober, P., Lant, T., & Senneville, C. (2010, April). Credibility, salience, and legitimacy of boundary objects: water managers' assessment of a simulation model in an immersive decision theater. *Science and Public Policy*, *37*(3), 219–232. Retrieved 2020-05-12, from https://academic.oup.com/spp/article/37/3/219/1626552 (Publisher: Oxford Academic) doi: 10.3152/030234210X497726

Xiao, M., Udall, B., & Lettenmaier, D. P. (2018). On the Causes of Declining Colorado River Streamflows. *Water Resources Research*, *54*(9), 6739–6756. Retrieved 2020-04-28, from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023153 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR023153) doi: 10.1029/2018WR023153

Yates, D. N., Miller, K. A., Wilby, R. L., & Kaatz, L. (2015, January). Decision-centric adaptation appraisal for water management across Colorado's Continental Divide. *Climate Risk Management*, *10*, 35–50. Retrieved 2019-10-01, from http://www.sciencedirect.com/science/article/pii/S2212096315000194 doi: 10.1016/j.crm.2015.06.001