ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

METHOD

10.1029/2022WR032349

Uncertainty Decomposition to Understand the Influence of Water Systems Model Error in Climate Vulnerability Assessments

Scott Steinschneider[1], Jonathan D. Herman[2], John Kucharski[3], Marriah Abellera[4], and Peter Ruggiero[5]

[1]Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY, USA, [2]Department of Civil and Environmental Engineering, University of California, Davis, Davis, CA, USA, [3]Engineering Research and Development Center, U.S. Army Corps of Engineers, Washington, DC, USA, [4]Institute for Water Resources, U.S. Army Corps of Engineers, Washington, DC, USA, [5]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA

Key Points:

- We present a method to determine if a water resources systems model is sufficiently accurate to assess climate vulnerability
- The method couples systems model uncertainty quantification with variance decomposition in ensemble scenario experiments
- The method is demonstrated on a simulation model of reservoirs and pumping operations in the Sacramento-San Joaquin River Basin and Delta

Correspondence to:

S. Steinschneider,
ss3378@cornell.edu

Citation:

Steinschneider, S., Herman, J. D., Kucharski, J., Abellera, M., & Ruggiero, P. (2023). Uncertainty decomposition to understand the influence of water systems model error in climate vulnerability assessments. *Water Resources Research*, *59*, e2022WR032349. https://doi.org/10.1029/2022WR032349

Received 10 MAR 2022
Accepted 8 JAN 2023

Author Contributions:

Conceptualization: Scott Steinschneider, Jonathan D. Herman, John Kucharski, Marriah Abellera, Peter Ruggiero
Data curation: Scott Steinschneider, Jonathan D. Herman, John Kucharski
Formal analysis: Scott Steinschneider, Jonathan D. Herman
Funding acquisition: John Kucharski, Marriah Abellera
Investigation: Scott Steinschneider, Jonathan D. Herman
Methodology: Scott Steinschneider, Jonathan D. Herman, Peter Ruggiero
Project Administration: John Kucharski, Marriah Abellera
Resources: Scott Steinschneider, Jonathan D. Herman
Software: Scott Steinschneider, Jonathan D. Herman, John Kucharski
Supervision: John Kucharski, Marriah Abellera, Peter Ruggiero
Validation: Scott Steinschneider, Jonathan D. Herman

Abstract Climate vulnerability assessments rely on water infrastructure system models that imperfectly predict performance metrics under ensembles of future scenarios. There is a benefit to reduced complexity system representations to support these assessments, especially when large ensembles are used to better characterize future uncertainties. An important question is whether the total uncertainty in the output metrics is primarily attributable to the climate ensemble or to the systems model itself. Here we develop a method to address this question by combining time series error models of performance metrics with time-varying Sobol sensitivity analysis. The method is applied to a reduced complexity multi-reservoir systems model of the Sacramento-San Joaquin River Basin in California to demonstrate the decomposition of flood risk and water supply uncertainties under an ensemble of climate change scenarios. The results show that the contribution of systems model error to total uncertainty is small (∼5%–15%) relative to climate based uncertainties. This indicates that the reduced complexity systems model is sufficiently accurate for use in the context of the vulnerability assessment. We also observe that climate uncertainty is dominated by the choice of general circulation model and its interactive effects with the representative concentration pathway (RCP), rather than the RCP alone. This observation has implications for how climate vulnerabilities should be interpreted.

1. Introduction

Climate vulnerability assessments have become a common feature of water resources systems planning studies (Arnell, 2011; Plummer et al., 2012; US Bureau of Reclamation, 2012; Weaver et al., 2013). These assessments generally require ensemble simulations of future climate scenarios that are passed through a combination of hydrologic models and water resources systems models to measure the vulnerability of the water system to properties of future climate. Once these vulnerabilities are identified, additional simulation or optimization experiments are used to determine how well different adaptation actions mitigate these vulnerabilities (J. D. Herman et al., 2015, 2020).

The literature on the uncertainties that underlie future climate scenarios (Knutti et al., 2008; Lehner et al., 2020; Northrop & Chandler, 2014), associated hydrologic responses (Kundzewicz et al., 2018; Mendoza et al., 2016; Steinschneider, Wi, & Brown, 2015; Wilby & Harris, 2006), and water system vulnerability under climate change (Steinschneider, McCrary, Mearns, & Brown, 2015; Steinschneider, McCrary, Wi, et al., 2015) is extensive. However, water resources systems model uncertainties are usually neglected in these climate impact assessments, presumably under the assumption that they are negligible in comparison to other uncertainties. This assumption is likely valid for systems models underpinned by years to decades of development. However, many of these high-fidelity models are computationally expensive and ill-suited for ensemble experiments required by climate vulnerability and adaptation assessments. More parsimonious system models and emulators of complex systems models have become a popular means of reducing the computational cost of systems model simulation in ensemble experiments (Badham et al., 2019; Basco-Carrera & Mendoza, 2017; Gijsbers et al., 2017; Haasnoot et al., 2014; Helgeson et al., 2021; Voinov et al., 2018). These models provide faster runtimes at the expense of some accuracy in system representation. Their use raises the question of whether such reduced complexity models are suitable for use in climate vulnerability assessments and how this should be assessed.

Visualization: Scott Steinschneider, Jonathan D. Herman
Writing – original draft: Scott Steinschneider, Jonathan D. Herman
Writing – review & editing: Scott Steinschneider, Jonathan D. Herman, John Kucharski, Marriah Abellera, Peter Ruggiero

Past work has considered the question of whether a systems model is fit-for-purpose (Haasnoot et al., 2014; Hamilton et al., 2022). For example, Haasnoot et al. (2014) emphasize the ability of the model to produce "credible outcomes with sufficient accuracy for the screening and ranking of promising actions and pathways in order to support… strategic adaptive planning decisions." They describe a simplified systems model as fit-for-purpose if it produces decisions that are consistent with a more complex model. In the context of climate vulnerability assessments, we investigate a related concept: whether prediction errors arising from the systems model are negligible compared to the uncertainty in forcing, particularly around key output metrics that are most relevant to decision-making. This emphasizes the relative accuracy of the systems model against the background of other exogenous uncertainties, and thus contributes a complementary viewpoint by assessing if the model is fit-for-purpose in the context of planning under uncertain future conditions. This viewpoint is consistent with recent recommendations to ensure greater transparency and more robustness in climate change impact assessments (Wagener, 2022).

This technical note advances variance decomposition as an approach to assess the suitability of water systems models in climate vulnerability studies. Uncertainty decomposition is widely used to assess sources of uncertainty in climate models and their influence on key variables of interest (Hawkins & Sutton, 2009; Lehner et al., 2020). It is also used to identify factors that drive uncertainty in water systems performance metrics (Greve et al., 2018; Schlef et al., 2018) and the broader human-Earth system (Lamontagne et al., 2019). In this study we extend this technique to assess whether a water systems model is sufficiently accurate for its intended purpose in a climate vulnerability assessment, providing a diagnostic method to propagate and decompose systems model error in the context of an ensemble of climate scenarios.

2. Data and Methods

2.1. Case Study and Simulation Model

The proposed method is demonstrated using a new, daily time step simulation model of the eight largest reservoirs in the Sacramento-San Joaquin River Basin (SSJRB), California, and the water supply pumping operations near the system outlet in the Sacramento-San Joaquin Delta (Figure 1). The model structure and data requirements were simplified from recently developed simulation models of this system (ORCA—Cohen et al., 2020, 2021; CALFEWS—Zeff et al., 2021) for the purpose of efficiently estimating water supply and flooding metrics in the Delta in large ensemble climate vulnerability assessments.

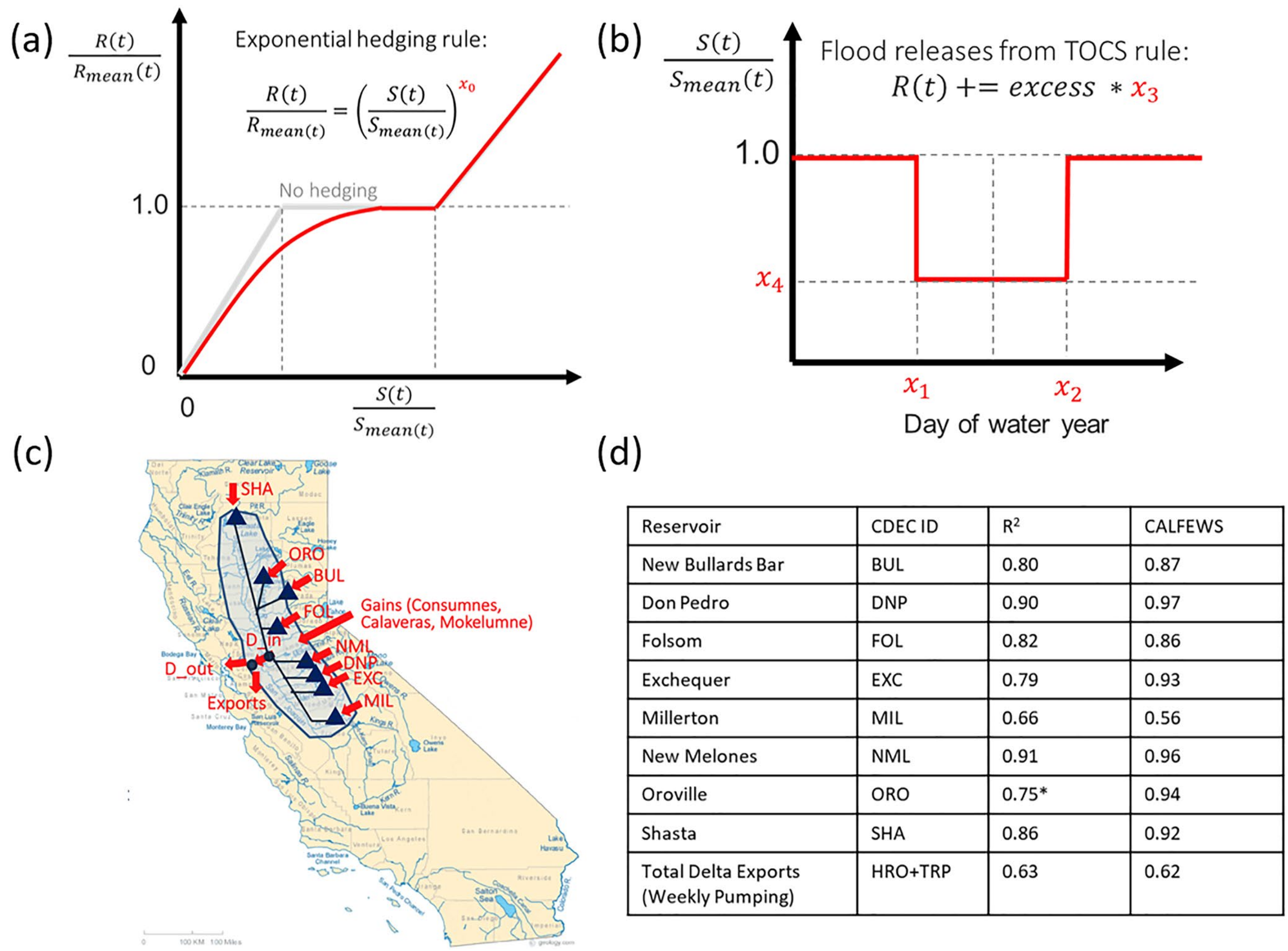
The SSJRB reduced complexity systems model (referenced hereafter as the SSJRB model) consists of three components: reservoir release policies, gains, and Delta pumping. The model contains 43 parameters, including: five release policy parameters for each of the eight reservoirs, two parameters for gains, and one parameter for Delta pumping. The model uses the historical observed median operating pattern (storage and release for each day of the water year) over the period 1997–2021, and adjusts this pattern based on current hydrologic conditions. Initially, reservoir operations are described by 5-parameter $(x_0, \ldots x_4)$ exponential water supply and linear flood hedging rules (Figures 1a and 1b). The water supply rule is given by:

$$\frac{R_i(t)}{R_{i,m}(t)} = \left(\frac{S_i(t)}{S_{i,m}(t)}\right)^{x_0} \tag{1}$$

where $R_i(t)$ is the release for the ith reservoir, $S_i(t)$ is the storage, and $R_{i,m}(t)$ and $S_{i,m}(t)$ are the median release and storage for that day of the water year, respectively. The water supply release determined from Equation 1 is then increased to model flood control operations. Specifically, if the day of the water year falls between $[x_1, x_2]$ and $S_i(t) > x_4 S_{i,m}(t)$, then $R_i(t)$ is increased by the amount $x_3(S_i(t) - x_4 S_{i,m}(t))$.

Next, the hydrologic gains into the Delta, G (t), are defined as the Delta inflow $D_{\text{in}}(t)$ minus the sum of reservoir outflows. This term represents the tributaries for which reservoirs are not modeled (see Figure 1c) as well as additional inflows downstream of the reservoirs. Gains can be either positive or negative. Positive gains represent winter inflows, while negative gains represent consumptive withdrawals in the summer. These gains are estimated from historical patterns using two parameters (x_5, x_6):

$$G(t) = G_m(t)\left(\sum_i \frac{S_i(t)}{K_i}\right)^{x_5} + x_6 \sum_i Q_i(t) \tag{2}$$



Reservoir	CDEC ID	R^2	CALFEWS
New Bullards Bar	BUL	0.80	0.87
Don Pedro	DNP	0.90	0.97
Folsom	FOL	0.82	0.86
Exchequer	EXC	0.79	0.93
Millerton	MIL	0.66	0.56
New Melones	NML	0.91	0.96
Oroville	ORO	0.75*	0.94
Shasta	SHA	0.86	0.92
Total Delta Exports (Weekly Pumping)	HRO+TRP	0.63	0.62

Figure 1. Overview of the Sacramento-San Joaquin River Basin (SSJRB) simulation model. (a) Reservoir operating policy relating releases R to storage S via an exponential hedging rule; (b) adjustments to water supply releases to model flood control operations; (c) inflow and pumping locations; (d) model accuracy (R^2) for storage at the eight reservoirs in the system and total delta exports compared to the CALFEWS systems model (Zeff et al., 2021). The time periods for the comparison are October 1997–September 2021 (SSJRB) and October 1996–September 2016 (CALFEWS). *The Oroville result for the SSJRB model is impacted by the operational response to the February 2017 spillway failure over the subsequent year.

where $G_m(t)$ are the median gains for that day of the water year, K_i is the storage capacity of reservoir $i \in [1, 8]$, and $Q_i(t)$ is the inflow into reservoir i. The first term in Equation 2 covers the broader seasonal patterns of withdrawals, while the second term represents additional Delta inflows that are assumed to be correlated with reservoir inflows included in the model.

The Delta pumping policy is represented by the following equation with one parameter, x_7:

$$P(t) = D_{\text{in}}(t)\, p_m(t) \left(\sum_i \frac{S_i(t)}{K_i} \right)^{x_7} \tag{3}$$

where $P(t)$ is the total pumping volume, $p_m(t)$ is the median pumping for that day of the water year (percent of inflow), and the storage fraction term is the same as in Equation 2. We impose an upper bound on pumping to approximate a combination of the infrastructure capacity and environmental guidelines, though this amount can be exceeded if needed.

The parameters for all of the model components are fit with differential evolution (Storn & Price, 1997). The historical data used to find the median daily patterns and to fit the parameters are taken from the California Data Exchange Center (cdec.water.ca.gov). These operating rules are empirical simplifications based on the observed

10.1029/2022WR032349

data, and do not exactly match those published in water control manuals. However, they ensure that all reservoirs follow the same model structure, and that the model is parsimonious enough to calibrate and modify. Figure 1d shows the ability of the systems model to replicate historical storage for each of the eight reservoirs. Overall, the systems model adequately represents the operations of these facilities, with R^2 values for daily simulated and observed storage ranging between 0.66 and 0.91, with an average of 0.81. This performance is slightly worse than that of a recently published, more detailed, state-of-the-art simulation model of the California water system (CALFEWS; Zeff et al., 2021), which has an average storage R^2 of 0.88 for the same reservoirs, and also contains a more detailed system representation south of the Delta to describe deliveries to irrigation districts. However, the SSJRB model is significantly faster (3–4 orders of magnitude) due to a combination of simplified structure and Numba just-in-time compilation (Lam et al., 2015). In the analysis that follows, we investigate whether the decrease in accuracy significantly influences our understanding of system performance in the context of broader climate uncertainties.

2.2. Error Models for Key System Model Outputs

Two output variables from the SSJRB systems model are of interest in this study: a water supply metric—delta pumping exports, $P(t)$—and a flood control metric—delta outflows, $D_{\text{out}}(t) = D_{\text{in}}(t) - P(t)$. We develop error models for these two metrics using the historical simulation from the SSJRB systems model, which enables stochastic simulation of these errors under a wide range of future scenarios at a daily time step. Through stochastic simulation, we are able to thoroughly explore how systems model error interacts with future scenarios of climate and hydrology to determine plausible future ranges of these variables.

We follow the general approach in McInerney et al. (2017) and define a residual term, ε_t, equal to the difference between the daily observations and simulations after transformation:

$$\varepsilon_t = f(O_t|\lambda) - f(M_t|\lambda) \tag{4}$$

Here, O_t is the observed data associated with the decision-relevant variable of interest (e.g., delta outflows or delta pumping exports), M_t is the systems model simulation of that variable, t is a time step (daily in this case) within the historical record of length T, and $f(\cdot|\lambda)$ is a transformation with parameter λ. The transformation is used to simplify the probabilistic behavior of the observed and simulated time series before calculating the residuals. Here, we employ the Box-Cox transformation (Box & Cox, 1964), which becomes the identity transformation for $\lambda = 1$ and approaches a logarithmic transformation as λ approaches 0. However, other transformations (e.g., log-sinh) could also be applied.

To remove any systematic bias between the simulations and observations, the residuals are regressed against the transformed simulation:

$$\varepsilon_t = \beta_0 + \beta_1 f(M_t|\lambda) + \eta_t \tag{5}$$

Two assumptions are made: (a) systems model bias has a constant component (β_0) and a component i.e., correlated with the magnitude of the simulated response, for example, the systems model tends to underestimate the observations when it predicts large flows and overestimate the observations when it predicts low flows; (b) this bias is a linear function of the magnitude of the simulation itself (via the coefficient β_1). Importantly, the bias correction in Equation 5 adjusts the error distribution as the systems model simulation changes under a future climate scenario. That is, if certain values of M_t are simulated more often under future climate conditions, then the frequency of certain types of error will change. We note that this bias correction could be made more general using a non-linear function (e.g., a local weighted regression or generalized additive model), but initial analysis (not shown) suggested this was unnecessary for the case study used in this work.

The bias corrected residuals η_t are then decorrelated in time using an autoregressive (AR) model:

$$\eta_t = \theta_0 + \theta_1 \eta_{t-1} + \xi_t \tag{6}$$

Here, θ_0 and θ_1 are the coefficients of the AR model and ξ_t is a white noise term. An AR(1) model was found to be sufficient to remove autocorrelation in the delta outflow and export residual time series. However, any higher-order autoregressive moving average model can be selected based on the behavior of the residual series η_t.

10.1029/2022WR032349

Once the model above is fit to the historical series of data, stochastic traces $\widetilde{O_{1:T^*}}$ are simulated for new periods of interest ($t^* = 1, \ldots, T^*$, e.g., future decades under climate change). These simulations are produced with the following steps:

1. Bootstrap a value of ξ_t from the historical record.
2. For a new time t^*, estimate $\widetilde{\eta_{t^*}}$ using Equation 6 (i.e., the AR model), the resampled value of ξ_t, and the previous value $\widetilde{\eta_{t^*-1}}$.
3. Estimate $\widetilde{\varepsilon_{t^*}}$ using Equation 5 (i.e., the bias correction model), the value $\widetilde{\eta_{t^*}}$, and the systems model simulation $\widetilde{M_{t^*}}$.
4. Estimate $\widetilde{O_{t^*}} = f\left(\widetilde{\varepsilon_{t^*}} + f\left(\widetilde{M_{t^*}}|\lambda\right)|\lambda\right)^{-1}$

Steps 1–4 are then repeated for all time steps $t^* = 1, \ldots, T^*$. $\tilde{\eta}_0$ is initialized as 0.

In short, steps 1–4 above resample white noise terms ξ_t from the historical period, embed autocorrelation into those noise terms using the AR model, reinsert any bias into the errors that was present historically, add those autocorrelated and biased errors to the transformed, system model simulation, and then back-transform the result into an estimate of the variable of interest.

2.3. Climate Scenarios

In this study, we assess whether the SSJRB systems model error in key variables of interest (e.g., delta outflows and exports) is sufficiently small in the broader context of climate uncertainty. This is tested by forcing the systems model with an ensemble of projected flows between 2020 and 2099 for each of the eight reservoir inflow points of the system (see Figure 1). The ensemble, developed in Brekke et al. (2014), is derived from CMIP5 general circulation model (GCM) simulations (Taylor et al., 2012), downscaled to a daily timescale and 1/8° spatial resolution using the updated Bias-Correction and Spatial Disaggregation technique (NCAR, 2014) followed by daily disaggregation (Wood et al., 2004). The ensemble is composed of four different representative concentration pathways (RCPs; 2.6, 4.5, 6.0, 8.5) and 31 different GCMs, with 97 scenarios altogether (not every RCP is used with every GCM). The downscaled, daily climate data force the Variable Infiltration Capacity (VIC) hydrologic model, previously calibrated for watersheds across the US West (see Brekke et al., 2014).

To generate a balanced ensemble, we filter the full ensemble described above to include only those GCMs with simulations under the RCP 4.5 and 8.5 emission scenarios (i.e., the most common emission scenarios across all GCMs). This leads to 29 GCMs under these two RCPs, for a total of 58 scenarios. For each scenario and its associated trace of simulated daily delta outflows and delta exports from the SSJRB systems model, we develop a stochastic ensemble of 100 traces of delta outflow and exports using the error modeling procedure in Section 2.2. This results in a total of 5,800 80-year traces of daily delta outflows and exports, which are then analyzed using sensitivity analysis (described next) to partition variance among the various uncertainty sources: GCMs, RCPs, and systems model error. We also consider whether the variance partitioning changes considerably if a smaller set of GCMs is used, for instance based on their ability to capture aspects of regional climate (Gershunov et al., 2017; Pierce et al., 2018).

2.4. Sensitivity Analysis

We use Sobol sensitivity analysis to attribute variance within the ensemble of 5,800 traces of delta outflows and exports to the GCMs, the RCPs, systems model error, and interactive effects between these different sources of uncertainty. These three inputs are sampled as integer factors in the sensitivity analysis (i.e., the choice of GCM, RCP, and systems model error realization). A major benefit of using Sobol sensitivity analysis is its ability to attribute variance not only to individual factors (GCMs, RCPs, systems model error), but also to the interactions between these factors. This is important, because the effect of systems model error on the overall variance of output variables of interest can depend on the underlying hydroclimate regime (e.g., larger system model errors during wet periods could become more influential if wet periods occur more often). Ultimately, the model is classified as fit-for-purpose for large ensemble experiments in a climate vulnerability assessment if the variance attributed to the systems model uncertainty and its interactive effects is small compared to the other sources of uncertainty.

10.1029/2022WR032349

Sobol sensitivity analysis is described in detail elsewhere (J. Herman & Usher, 2017; Pianosi et al., 2016; Saltelli et al., 2010; Sobol, 2001) and therefore only briefly reviewed here. Let Y_k be the output metric of interest from the SSJRB systems model, which we define separately for each year of simulation ($k = 2020, \ldots, 2099$). For delta outflows, we define Y_k as the annual maximum outflow in year k; for delta exports, Y_k is defined as the annual sum of exports. While other metrics could be chosen, these metrics are representative of annual flood and drought risk in the Sacramento-San Joaquin system. The Sobol method is used to attribute variance in Y_k to individual uncertainty factors and their interactions, and can be written as follows:

$$D(Y_k) = \sum_i D_i + \sum_{i<j} D_{ij} + D_{12\ldots h} \tag{7}$$

Here, $D(Y_k)$ represents the total variance in Y_k, D_i is the first-order variance contribution of the ith factor, D_{ij} is the second-order variance contribution of the interaction between the ith and jth factors, and $D_{12\ldots h}$ represents the variance contribution of all higher-order interactions greater than second-order. Sensitivity indices are then defined as the fraction of individual variance contribution terms to the total variance (e.g., $\frac{D_i}{D}$ and $\frac{D_{ij}}{D}$ represent the first-order sensitivity index for factor i and second-order sensitivity index for factors i and j, respectively). Similarly, the total-order sensitivity for a given factor $(1 - \frac{D_{\sim i}}{D})$ uses the variance associated with all factors besides factor i ($D_{\sim i}$) to define the variance attributed to all first-order and higher-order interactions associated with factor i. See Saltelli (2002) for additional detail on the numerical estimation of terms (D_i, D_{ij}, $D_{\sim i}$).

When making a determination of whether the systems model is sufficiently accurate for use in a climate impact analysis, we avoid setting a distinct threshold for the variance attributed to the systems model. This is ultimately a subjective choice based on user preference, and we believe that forwarding an (arbitrary) threshold here would discourage critical evaluation and collaborative decision-making on a case-by-case basis that is central to effective water resources planning. Similar logic supports recent efforts in the statistical literature to discourage the use of arbitrary p-values when assessing the statistical significance of relationships (Wasserstein & Lazar, 2016).

3. Results and Discussion

3.1. Systems Error Models

Error accumulates throughout the system to influence the key metrics of interest: delta outflows and exports. Figures 2a and 2d show the observed and simulated time series of daily delta outflows and exports, respectively. The systems model captures the observed outflows well, with a Nash-Sutcliffe efficiency (NSE) over the entire period of 0.74 and a percent bias of 3.2%. Simulations of daily exports are less skillful, with an NSE of 0.32 and percent bias of 12.1%, although when aggregating to a weekly time step the simulations are better (NSE of 0.63). Notably, the systems model tends to underestimate the largest delta outflows, albeit with one exception in 2017. The SSJRB systems model imposes a cap on delta exports, but observed exports occasionally exceed this limit. The systems model also overestimates the smallest delta outflows (Figure 2a), particularly during dry years, and exhibits fewer and less extreme declines in daily exports (Figure 2d).

We fit error models for both the delta outflow and export series in order to generate stochastic traces of these variables. A Box-Cox transformation with $\lambda = 0.3$ is applied to the delta outflows, while the delta exports are not transformed (i.e., $\lambda = 1$). Also, delta outflow residuals ε_t did not vary significantly with model simulations, so $\beta_1 = 0$ in Equation 5. Figures 2a and 2d show 95% bounds generated using the stochastic ensemble. These uncertainty bounds are developed by simulating delta outflow and export residuals and adding them to the SSJRB simulation. Several important features emerge from the ensemble. The delta outflows ensemble clearly captures some of the lowest outflow values, but also captures several of the peak outflow events. Similarly, the stochastic delta exports ensemble better captures many of the lowest observed exports and observed exports that extend above the cap imposed in the SSJRB model.

This is clear in Figures 2b and 2e, which show coverage probabilities for the stochastic ensemble. Coverage probabilities quantify how often the observations fall between the $\frac{\alpha}{2}$ and $\left(1 - \frac{\alpha}{2}\right)$ percentiles of the stochastic ensemble, with α ranging from 0.01 to 0.99 in Figures 2b and 2e. The target coverage probabilities (i.e., $\alpha \times 100\%$) are compared against the observed coverage probabilities to assess the reliability of the stochastic ensemble. The results show that the ensemble is very reliable, with the largest deviations between target and observed coverage probabilities only reaching ~5.5%.

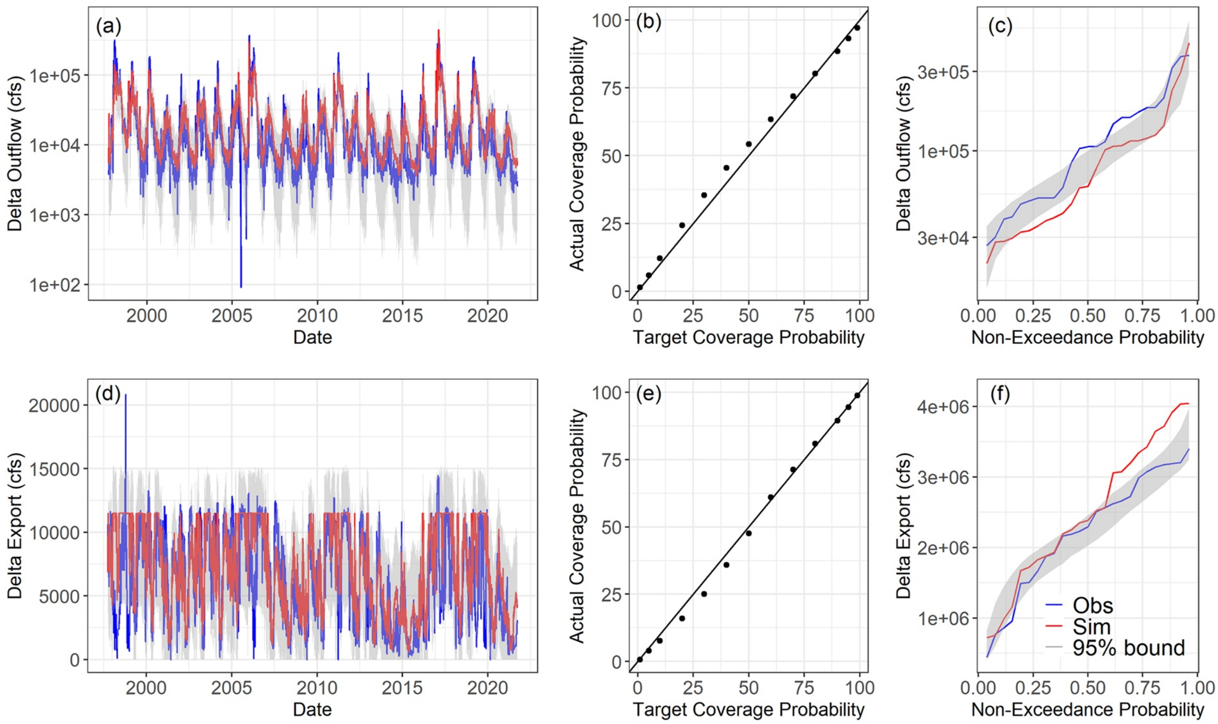


Figure 2. (a) Observed (blue) and simulated (red) daily delta outflows. The gray area shows 95% bounds of the stochastic delta outflow simulations. (b) Target versus actual coverage probabilities (i.e., the probability that observations fall within the *p* percent bounds of the simulated ensemble), with *p* varied from 1% to 99%. (c) The CDF of observed (blue) and simulated (red) annual maximum delta outflows, as well as the 95% bounds for the CDF of annual maxima from the stochastic traces. (d) and (e) Same as (a) and (b) but for daily delta exports. (f) Same as (c) but for the annual sum of delta exports.

The assessments above focus on daily data, but we are often interested in specific aggregated metrics of model output. Figures 2c and 2f show the distribution of observed and simulated annual maximum delta outflows and annual total delta exports, respectively. These are the focus of the uncertainty decomposition shown in Section 3.2 below. Also shown in Figures 2c and 2f are the same distributions from the stochastic traces. For delta outflows, the simulated distribution tends to underestimate that of the observed, except for the very largest annual maxima. However, the stochastic ensemble captures the observed distribution reasonably well, albeit with an underestimation between the 40th and 75th percentiles of the distribution. The opposite issue is apparent in the delta exports data, as the SSJRB model tends to overestimate the observed annual totals. The stochastic ensemble, however, adequately bounds the observed distribution.

3.2. Sensitivity Analysis

The SSJRB model was used to generate a 5,800-member ensemble of future simulations across 2 RCPs, 29 GCMs, and with 100 stochastic traces of model error (as described above). The Sobol analysis resamples from these discrete factors to decompose the variance in the two output metrics across the entire ensemble for each year, as shown in Figures 3a and 3b. Here we show the first-order sensitivity indices for RCP, GCM, and systems model error, as well as second-order sensitivity indices that quantify the variance associated with interactive effects (see Section 2.4 for the definition of first- and second-order sensitivity indices). Higher order sensitivity is not shown, and so the cumulative variance shown in Figure 3 is always slightly less than 100% of the total variance. All sensitivity indices are smoothed with a 10-year rolling average.

Several insights emerge from Figures 3a and 3b. First, the largest source of uncertainty in both metrics stems from the first-order effects of the GCMs. Averaged across all years, first-order GCM uncertainty accounts for 49% and 46% of the total variance in delta outflows and exports, respectively. This result demonstrates that the GCM used to simulate system inflows has the single largest impact on the variance in decision-relevant metrics of interest, regardless of emission scenario or other factors. We speculate that climate model uncertainty dominates the total uncertainty due to the range of regional precipitation responses (from both internal variability and change) across the GCMs (Schlef et al., 2018), but we do not verify that here.

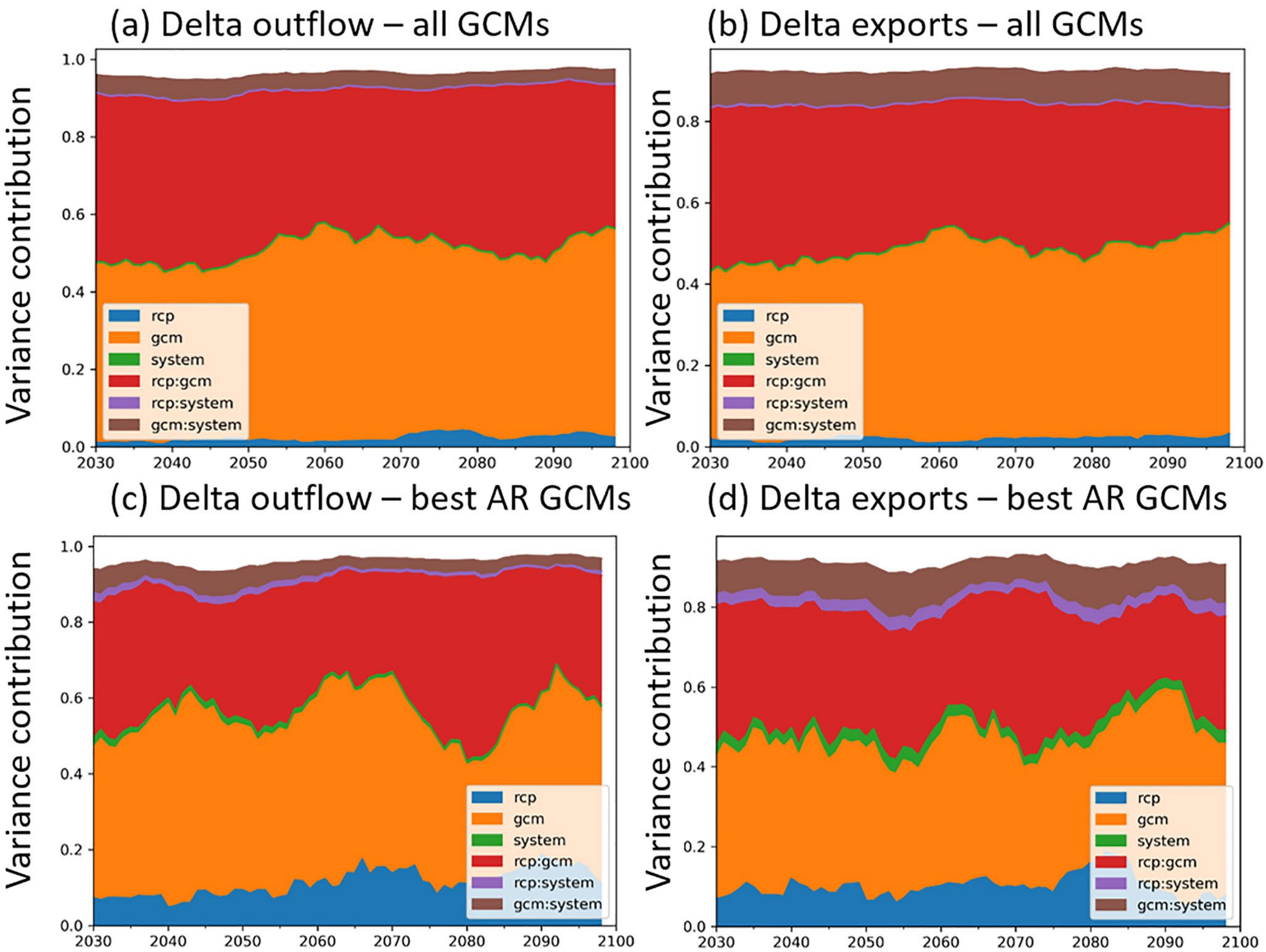


Figure 3. Variance contribution of general circulation models (GCMs), representative concentration pathways, and system model uncertainty to the total variance of (a), (c) annual maximum delta outflows and (b), (d) annual total delta exports, shown by year and for first-order effects and second-order interactions. Results are smoothed using a lagged 10-year rolling average and shown when using (a), (b) all GCMs and (c), (d) only four GCMs selected for their accurate representation of atmospheric rivers.

The second largest source of uncertainty stems from the interactive effect of GCMs with RCPs. The magnitude of this uncertainty rivals that of the first-order GCM effect, and when averaged across years, accounts for 41% and 35% of the variance in delta outflows and exports, respectively. In contrast, the first-order effect of the RCPs is small, on average less than 2%. This result highlights that neither annual maximum delta outflow nor annual total delta exports are explained by consistent differences that emerge across emission scenarios. This is true even by the end of the century when temperature differences between RCP 4.5 and RCP 8.5 are greatest. Instead, it is the variable response of the GCMs to a given RCP that explains a large portion of variance in the delta outflows and exports. That is, if one GCM projects drier conditions as emissions rise, but another projects wetter conditions with greater emissions, only the combination of both the emissions scenario *and* the specific GCM can explain the response of the water system.

Finally, we consider the variance explained by factors related to the SSJRB systems model uncertainty. Figures 3a and 3b show that for both delta outflows and delta pumping, uncertainties related to the system model are small in comparison to the other sources. The uncertainty attributed to the first-order systems model effect is under 1% across time, as is the interactive effect between the system model error and RCP. The interactive effect between systems model error and GCM is larger, with time-averaged values of 4% for delta outflows and 7.9% for delta exports. However, the cumulative uncertainty for all first and higher-order terms related to systems model error (i.e., total-order sensitivity, not shown in Figure 3) is 8.2% for delta outflows and 16.5% for delta exports, far below the uncertainties related to the GCMs and their interactive effects with emission scenarios. From this perspective, we deem the SSJRB systems model sufficiently accurate to support climate change vulnerability studies of the system, given the set of 29 GCMs considered in the ensemble.

It is reasonable to question if this result is consistent when a smaller set of GCMs with less variability and a more consistent representation of the climate system is selected for analysis. To answer this question, Figures 3c and 3d show the same results as in Figures 3a and 3b but based on a subset of the original ensemble composed of simulations from the two RCPs and only four GCMs (access1-0, canesm2, cnrm-cm5, gfdl-cm3). These GCMs have been shown to capture atmospheric river dynamics along the US West Coast that are important to water supply and flood risk in the SSJRB system (Gershunov et al., 2017). Note that access1-3 is also highlighted as skillful in Gershunov et al., 2017 but is not available in the ensemble of VIC simulations (see Section 2.3).

In the restricted set of GCMs, the uncertainty attributable to the first-order effect from the GCMs declines, though it is still substantial. A similar result is seen for the interactive effect between the GCMs and RCPs. Importantly, we only observe a small increase in the variance attributed to the systems model. Instead, the variance contribution associated with the first-order effect of the RCPs increases the most, from less than 2% in the full ensemble to 11% and 10% in the limited ensemble for delta outflows and exports, respectively. This is consistent with the expectation that the effects of emission scenario on hydrologic response are more apparent in GCMs which provide a more consistent representation of regional climate. Still, the main result from the full ensemble persists: systems model uncertainty is small compared to other uncertainties, therefore the reduced complexity model is suitable for climate vulnerability analysis. We note that we repeated this same experiment with four different GCMs that were shown to produce realistic simulations of precipitation and temperature in California (Pierce et al., 2018), with very similar results (not shown).

4. Conclusions

This technical note contributes a novel approach to determine whether a water resources systems model is sufficiently accurate for ensemble experiments in climate vulnerability assessments. The approach decomposes the variance in decision-relevant metrics (delta outflows and exports) to compare systems model error to other uncertainty sources (e.g., future climate scenarios). Model error in the output metrics is represented by a time series error model. To our knowledge, this is the first time error models and variance decomposition have been used to assess model suitability for climate impact experiments.

This technique is demonstrated with a new, computationally efficient, daily systems model of reservoirs within the Sacramento—San Joaquin River Basin California. The analysis shows that uncertainties from the systems model are small in comparison to those associated with future climate scenarios, especially in the key metric of annual maximum delta outflow. Systems model uncertainty contributes more to the total variance in delta exports, although the future climate scenarios continue to dominate variance in the delta export metric. GCMs and their interactive effects with RCPs contribute the most uncertainty in future delta outflows and exports. In comparison, the first-order contributions from the RCPs themselves are small. The conclusions above hold even if smaller ensembles of GCMs are selected based on their ability to represent regional climate, although with a somewhat larger direct contribution from emission scenario.

The approach presented here is widely applicable and can be extended to consider other uncertainty sources not accounted for in this experiment. For instance, the effects of climate model error and natural climate variability, which were combined in the experiment presented here, could be separated using single-model initial condition large ensembles (Lehner et al., 2020). Similarly, hydrologic model uncertainty can significantly affect water resource impact assessments (Malek et al., 2022) and could be accounted for by using multiple model structures and behavioral parameter sets. Importantly, hydrologic model uncertainty could interact with systems model error to further increase the influence of systems model error on the overall variance of key variables of interest. This possibility, which was not explored in this work, deserved further scrutiny. As additional uncertainty sources are considered and the ensemble size grows, the computational costs of the approach (in particular the Sobol method) may become infeasible. Therefore, future work is needed to consider alternative methods for sensitivity analysis that are more efficient, such as ANOVA or the method of Morris (J. D. Herman et al., 2013; Morris, 1991). Further consideration of the tradeoff between model accuracy and parsimony is also warranted.

A few caveats of the approach deserve mention. The systems model uncertainty is quantified based on historical errors between observed and modeled outcomes, but this approach will not capture any nonstationarity in the error distribution that emerges under new boundary conditions. Similarly, the current approach does not capture structural uncertainties related to changes in the system, for example, infrastructure and institutional adaptation or land-use change in response to exogenous forcing. The uncertainties around such structural system changes are large, and the approach taken in this work (i.e., the combined use of time series error models and sensitivity analysis) should not be used to conclude that future systems uncertainty is small compared to other future uncertainties. Rather, the method advanced here only considers the uncertainty in the system as it has existed and operated historically, and therefore can only be used to infer whether the systems model is fit-for-purpose in large ensemble climate vulnerability assessments of the water system if that system did not adapt to the changes it experiences. More broadly, sensitivity analyses such as the variance decomposition used in this work could be used to compare the relative importance of future hydroclimate and institutional uncertainties, if such institutional uncertainties could be quantified. This effort is an important avenue of future research.

Data Availability Statement

The code, inputs, and final outputs for all analyses conducted in this study are available at Kucharski (2022).

Acknowledgments
This work was supported by the United States Army Corps of Engineers, Institute for Water Resources. Jon Herman also received partial support from NSF Grant EnvSus-2041826. We further acknowledge the World Climate Research Program's Working Group on Coupled Modeling and the climate modeling groups for producing and making available their model output.

References

Arnell, N. W. (2011). Incorporating climate change into water resources planning in England and Wales. *JAWRA Journal of the American Water Resources Association*, *47*(3), 541–549. https://doi.org/10.1111/j.1752-1688.2011.00548.x

Badham, J., Elsawah, S., Guillaume, J. H., Hamilton, S. H., Hunt, R. J., Jakeman, A. J., et al. (2019). Effective modeling for integrated water resource management: A guide to contextual practices by phases and steps and future opportunities. *Environmental Modelling & Software*, *116*, 40–56. https://doi.org/10.1016/j.envsoft.2019.02.013

Basco-Carrera, L., & Mendoza, G. (2017). Collaborative modelling. Engaging stakeholders in solving complex problems of water management, global water partnership, perspective paper No. 10.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Brekke, L., Wood, A., & Pruitt, T. (2014). *Downscaled CMIP3 and CMIP5 hydrology projections: Release of hydrology projections, comparison with preceding information, and summary of user needs*. National Center for Atmospheric Research.

Cohen, J., Zeff, H., & Herman, J. (2020). Adaptation of multi-objective reservoir operations to snowpack decline in the Western United States. *Journal of Water Resources Planning and Management*, *146*(12). https://doi.org/10.1061/(ASCE)WR.1943-5452.0001300

Cohen, J., Zeff, H., & Herman, J. (2021). How do the properties of training scenarios influence the robustness of reservoir operating policies to climate uncertainty? *Environmental Modelling & Software*, *141*, 105047. https://doi.org/10.1016/j.envsoft.2021.105047

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A., & Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophysical Research Letters*, *44*(15), 7900–7908. https://doi.org/10.1002/2017GL074175

Gijsbers, P. J. A., Baayen, J. H., & ter Maat, G. J. (2017). Quick scan tool for water allocation in The Netherlands. In J. Hřebíček, R. Denzer, G. Schimak, & T. Pitner (Eds.), *Environmental software systems. Computer science for environmental protection. ISESS 2017*, *IFIP advances in information and communication technology* (Vol. 507). Springer. https://doi.org/10.1007/978-3-319-89935-0_9

Greve, P., Kahil, T., Mochizuki, J., Schinko, T., Satoh, Y., Burek, P., et al. (2018). Global assessment of water challenges under uncertainty in water scarcity projections. *Nature Sustainability*, *1*(9), 486–494. https://doi.org/10.1038/s41893-018-0134-9

Haasnoot, M., van Deursen, W. P. A., GuillaumeKwakkel, J. H. A. J. H., van Beek, E., & Middelkoop, H. (2014). Fit for purpose? Building and evaluating a fast, integrated model for exploring water policy pathways. *Environmental Modelling & Software*, *60*, 99–120. https://doi.org/10.1016/j.envsoft.2014.05.020

Hamilton, S. H., Pollino, C. A., Stratford, D. S., Fu, B., & Jakeman, A. J. (2022). Fit-for-purpose environmental modeling: Targeting the intersection of usability, reliability and feasibility. *Environmental Modelling & Software*, *148*, 105278. https://doi.org/10.1016/j.envsoft.2021.105278

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin America Meteorology Social*, *90*(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Helgeson, C., Srikrishnan, V., Keller, K., & Tuana, N. (2021). Why simpler computer simulation models can be epistemically better for informing decisions. *Philosophy of Science*, *88*(2), 213–233. https://doi.org/10.1086/711501

Herman, J., & Usher, W. (2017). SALib: An open-source Python library for sensitivity analysis. *Journal of Open Source Software*, *2*(9), 97. https://doi.org/10.21105/joss.00097

Herman, J. D., Kollat, J. B., Reed, P. M., & Wagener, T. (2013). Technical Note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. *Hydrology and Earth System Sciences*, *17*(7), 2893–2903. https://doi.org/10.5194/hess-17-2893-2013

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M., & Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty. *Water Resources Research*, *56*(2), e24389. https://doi.org/10.1029/2019WR025502

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

Knutti, R., Allen, M. R., Friedlingstein, P., Gregory, J. M., Hegerl, G. C., Meehl, G. A., et al. (2008). A review of uncertainties in global temperature projections over the twenty-first century. *Journal of Climate*, *21*(11), 2651–2663. https://doi.org/10.1175/2007jcli2119.1

Kucharski, J. (2022). JohnRushKucharski/systemuncertainty: Submission (v1.0.0-research) [Dataset and Software]. Zenodo. https://doi.org/10.5281/zenodo.6331306

Kundzewicz, Z. W., Krysanova, V., Benestad, R. E., Hov, Ø., Piniewski, M., & Otto, I. M. (2018). Uncertainty in climate change impacts on water resources. *Environmental Science & Policy*, *79*, 1–8. https://doi.org/10.1016/J.ENVSCI.2017.10.008

Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: A LLVM-based python JIT compiler. In *Proceedings of the second workshop on the LLVM compiler infrastructure in HPC*.

10.1029/2022WR032349

Lamontagne, J. R., Reed, P. M., Marangoni, G., Keller, K., & Garner, G. G. (2019). Robust abatement pathways to tolerable climate futures require immediate global action. *Nature Climate Change*, *9*(4), 290–294. https://doi.org/10.1038/s41558-019-0426-8

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., et al. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, *11*(2), 491–508. https://doi.org/10.5194/esd-11-491-2020

Malek, K., Reed, P. M., Zeff, H., Hamilton, A., Wrzesien, M., Holtzman, N., et al. (2022). Bias correction of hydrologic projections strongly impacts inferred climate vulnerabilities in institutionally complex water systems. *Journal of Water Resources Planning and Management*, *148*(1). https://doi.org/10.1061/(ASCE)WR.1943-5452.0001493

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, *53*(3), 2199–2239. https://doi.org/10.1002/2016WR019168

Mendoza, P. A., Clark, M. P., Mizukami, N., Gutmann, E. D., Arnold, J. R., Brekke, L. D., & Rajagopalan, B. (2016). How do hydrologic modeling decisions affect the portrayal of climate change impacts? *Hydrological Processes*, *30*(7), 1071–1095. https://doi.org/10.1002/hyp.10684

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometric*s, *33*(2), 161–174. https://doi.org/10.1080/00401706.1991.10484804

National Center for Atmospheric Research (NCAR). (2014). *NCAR final project report to USACE responses to climate change, project W26HM423495778*. National Center for Atmospheric Research. Retrieved from http://corpsclimate.us/rcchsa.cfm

Northrop, P. J., & Chandler, R. E. (2014). Quantifying sources of uncertainty in projections of future climate. *Journal of Climate*, *27*(23), 8793–8808. https://doi.org/10.1175/jcli-d-14-00265.1

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Pierce, D. W., Kalansky, J. F., & Cayan, D. R. (2018). *Climate, drought, and sea level rise scenarios for the fourth California climate assessment. California's fourth climate change assessment, California energy commission*. Scripps Institution of Oceanography. Publication Number: CNRA-CEC-2018-006.

Plummer, R., de Loë, R., & Armitage, D. A. (2012). Systematic review of water vulnerability assessment tools. *Water Resources Management*, *26*(15), 4327–4346. https://doi.org/10.1007/s11269-012-0147-5

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, *145*(2), 280–297. https://doi.org/10.1016/S0010-4655(02)00280-1

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Schlef, K. E., Steinschneider, S., & Brown, C. M. (2018). Spatiotemporal impacts of climate and demand on water supply in the Apalachicola-Chattahoochee-Flint Basin. *Journal of Water Resources Planning and Management*, *144*(2). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000865

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. https://doi.org/10.1016/S0378-4754(00)00270-6

Steinschneider, S., McCrary, R., Mearns, L. O., & Brown, C. (2015). The effects of climate model similarity on probabilistic climate projections and the implications for local, risk-based adaptation planning. *Geophysical Research Letters*, *42*(12), 5014–5044. https://doi.org/10.1002/2015GL064529

Steinschneider, S., McCrary, R., Wi, S., Mulligan, K., Mearns, L., & Brown, C. (2015). Expanded decision-scaling framework to select robust long-term water-system plans under hydroclimatic uncertainties. *Journal of Water Resources Planning and Management*, *141*(11), 04015023. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000536

Steinschneider, S., Wi, S., & Brown, C. (2015). The integrated effects of climate and hydrologic uncertainty on future flood risk assessments. *Hydrological Processes*, *29*(12), 2823–2839. https://doi.org/10.1002/hyp.10409

Storn, R., & Price, K. (1997). Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341–359. https://doi.org/10.1023/A:1008202821328

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1

U.S. Bureau of Reclamation. (2012). Colorado River Basin water supply and demand study (p. 85).

Voinov, A., Jenni, K., Gray, S., Kolagani, N., Glynn, P. D., Bommel, P., et al. (2018). Tools and methods in participatory modeling: Selecting the right tool for the job. *Environmental Modelling & Software*, *109*, 232–255. https://doi.org/10.1016/j.envsoft.2018.08.028

Wagener, T. (2022). On the evaluation of climate change impact models for adaptation decisions, In C. Kondrup, P Mercogliano, F Bosello, J Mysiak, E Scoccimarro, A Rizzo, et al. (Eds.). In *Climate adaptation modelling*. Springer Climate. Springer. https://doi.org/10.1007/978-3-030-86211-4_5

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Weaver, C. P., Lempert, R. J., Brown, C. M., Hall, J. A., Revell, D., & Sarewitz, D. (2013). Improving the contribution of climate model information to decision making: The value and demands of robust decision frameworks. *Wiley Interdisciplinary Reviews: Climate Change*, *4*(1), 39–60. https://doi.org/10.1002/wcc.202

Wilby, R. L., & Harris, I. (2006). A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water Resources Research*, *42*(2), W02419. https://doi.org/10.1029/2005WR004065

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, *15*(1–3), 189–216. https://doi.org/10.1023/b:clim.0000013685.99609.9e

Zeff, H. B., Hamilton, A. L., Malek, K., Herman, J. D., Cohen, J. S., Medellin-Azuara, J., et al. (2021). California's food-energy-water system: An open source simulation model of adaptive surface and groundwater management in the Central Valley. *Environmental Modelling & Software*, *141*, 105052. https://doi.org/10.1016/j.envsoft.2021.105052