Detection Time for Nonstationary Reservoir System Performance Driven by Climate and Land-Use Change

Matthew Chen[1] and Jonathan D. Herman[2]

Abstract: Dynamic planning of water infrastructure requires identifying signals for adaptation, including measures of system performance linked to vulnerabilities. However, it remains a challenge to detect projected changes in performance outside the envelope of natural variability, and to identify whether such detections can be attributed to one or more uncertain drivers. This study investigated these questions using a combination of ensemble simulation, nonparametric tests, and variance decomposition, which were demonstrated for a case study of the Sacramento–San Joaquin River Basin, California. We trained a logistic regression classifier to predict future detections given observed trends in performance over time. The scenario ensemble includes coupled climate and land-use change through the end of the century, evaluated using a multireservoir simulation model to determine changes in water supply reliability and flooding metrics relative to the historical period (1951–2000). The results show that the reliability metric is far more likely to exhibit a significant change within the century, with the most severe scenarios tending to be detected earlier, reflecting long-term trends. Changes in flooding often are not detected due to natural variability despite severe events in some scenarios. We found that the variance in detection times is attributable largely to the choice of climate model, and also to the emissions scenario and its interaction with the choice of climate model. Finally, in the prediction model for both cases, reliability and flooding, the model learns to associate more-recent observations of system performance with nonstationarity detection. These findings underscore the importance of differentiating between long-term change and natural variability in identifying signals for adaptation. **DOI: 10.1061/JWRMD5.WRENG-6184.** © *2024 American Society of Civil Engineers.*

Introduction

Water resources planning is challenged by cascading uncertainty in multiple factors, including future greenhouse gas emissions, climate model structure, downscaling, hydrologic modeling, and land use (Wilby and Dessai 2010). In response, there are two main approaches to long-term water resources decision support: robust planning, and dynamic planning. The goal of robust planning is to identify vulnerabilities under uncertainty and select actions with adequate performance across a wide range of plausible scenarios (Hinkel and Bisaro 2016; Walker et al. 2013; Weaver et al. 2013). However, the choice of static planning alternatives may lead to overinvestment, especially regarding irreversible infrastructure decisions (Borgomeo et al. 2018). In contrast, dynamic adaptation planning involves gradual investments in which decisions are made in response to new information over time (Fletcher et al. 2019; Haasnoot et al. 2013; Hui et al. 2018). The design of dynamic policies requires mapping indicators to actions, which can involve optimizing the sequence, timing, and threshold values assigned to observations and projections (Herman et al. 2020). Planning frameworks utilizing adaptive planning have been applied successfully to a variety of applications, including dynamic adaptive policy pathways (Haasnoot et al. 2013) and engineering options analysis (de Neufville and Smet 2019). Such approaches are appropriate for many applications that involve planning under uncertainty; some examples include planning flood mitigation infrastructure using indicators of sea level rise (Woodward et al. 2014) or event levels (Kwakkel et al. 2015), or planning for water supply infrastructure adaptation and operations with reservoir storage as an indicator (Mortazavi-Naeini et al. 2015). The present study focused on the detection of potential indicators as statistical changepoints, rather than on the mapping of these indicators to actions.

The choice and design of indicators is critical to an adaptive management strategy. Multiple strategies exist for identifying and designing indicators, including scenario discovery (Bryant and Lempert 2010; Groves et al. 2015; Lempert and Groves 2010) and multiobjective optimization of fixed or flexible policy structures (Cohen and Herman 2021; Hamarat et al. 2014; Kwakkel et al. 2015; Quinn et al. 2017; Zeff et al. 2016). Robinson and Herman (2019) proposed indicators based on a threshold classification method for future water supply vulnerabilities, which becomes more accurate at a multidecadal timescale. Ideally, indicators are linked to tipping points in system performance or predictions of future vulnerability, and are able to separate signals from noise and natural variability (Haasnoot et al. 2013; Hegerl and Zwiers 2011). However, this is a difficult problem for climate adaptation, because noise and natural variability may mask trends of key indicator variables (Bass et al. 2022; Siler et al. 2019).

The adaptation signals identified in these studies generally have not considered whether their timing is caused by nonstationary system performance reflecting long-term change rather than natural variability. This has been investigated indirectly through robustness testing, i.e., ensuring that adaptive policies continue to perform well in other samples of scenarios outside the training set (Cohen and Herman 2021). It also is possible to optimize policies directly to search for plans that are both dynamic and robust (Kwakkel et al. 2015). However, these approaches have not yet investigated whether and when nonstationary performance will be detected in projections of future climate and land use. Numerous studies

[1]Ph.D. Student, Dept. of Civil and Environmental Engineering, Univ. of California, Davis, Davis, CA 95616 (corresponding author). ORCID: https://orcid.org/0009-0005-9158-6231. Email: mtwchen@ucdavis.edu

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of California, Davis, Davis, CA 95616. ORCID: https://orcid.org/0000-0002-4081-3175

Note. This manuscript was submitted on April 15, 2023; approved on September 21, 2023; published online on January 29, 2024. Discussion period open until June 29, 2024; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, © ASCE, ISSN 0733-9496.

have addressed nonstationary trends in hydroclimatic variables; Slater et al. (2021) provided a review. Detection methods typically include regression methods such as ordinary least squares (Hecht and Vogel 2020; Papalexiou and Montanari 2019), nonparametric methods such as the Mann–Kendall test (Westra et al. 2013; Wilby 2006; Ziegler et al. 2005), and parametric methods in which distribution parameters vary with time (Katz 2013; Prosdocimi et al. 2015). Some studies using parametric methods have found that detection times depend on structural uncertainty in the choice of distribution (Ceres et al. 2017; Lee et al. 2017). However, the detection of nonstationary change in engineering system objectives to inform adaptation signals remains an open question. Some nonstationarity in hydroclimate can be buffered by existing infrastructure and operations, and would not necessarily result in a significant change to system performance.

The detection of nonstationary system performance must be attributed to one or more uncertain drivers to gain further insight into appropriate adaptation planning. For example, if uncertainty is attributable to natural variability, adaptation should be flexible and focused on the short term (e.g., operational changes such as hedging water supply); if uncertainty is attributable to climate uncertainty, adaptation should be robust to long-term change (Whateley and Brown 2016). Hawkins and Sutton (2009) developed an ANOVA approach to decompose variance in mean global temperature due to internal variability, model variability, and radiative forcing uncertainty. This methodology has been adapted to other hydroclimatic variables, such as precipitation (Greve et al. 2018) and drought risk (Orlowsky and Seneviratne 2013). In the water resources field, Whateley and Brown (2016) analyzed the sensitivity of uncertainties in decision-relevant water resources objectives such as water supply reliability and vulnerability using a similar approach. Additionally, the sampling of natural variability may be strengthened by large ensemble climate projections (Lehner et al. 2020). For example, Steinschneider et al. (2023) used ensemble projections to conduct variance decomposition via global sensitivity analysis to understand model suitability in climate risk assessments.

Finally, dynamic adaptation benefits from signals that identify changes in system performance before they occur. These signals may be drawn from a combination of observed hydroclimate and system performance, and can be used to reduce uncertainty in future scenarios and trigger adaptation based on this updated projection. Multiple studies have used Bayesian statistics of observed hydroclimatic variables to improve estimates of future trends (Smith et al. 2009; Tebaldi et al. 2005). Furthermore, in the context of dynamic adaptation planning, studies have shown the importance of proactive monitoring and timely action in response (van Ginkel et al. 2022; Haasnoot et al. 2018). This emphasizes the need to detect signals ahead of time, especially for cases in which adaptations that occur too early or too late are highly costly (Raso et al. 2019). It has been found that there is potential to detect future water supply vulnerability in advance using machine learning (Robinson et al. 2020), although without distinguishing between outcomes that occur within the distribution of historical variability and those that reflect nonstationary trends.

This study addressed indicators of nonstationarity as statistical changepoints—specifically, identifying tipping points as times in which current statistical properties of objective time series have shifted compared to the historical distribution. We approached this using a nonparametric statistical testing procedure that evolves with time, i.e., over a sliding window of a fixed size. This study analyzed whether and when detections of nonstationary reservoir system performance will occur in ensembles of climate and land-use projections, and decomposed the uncertainty in detection time to these drivers of change. We additionally investigated the ability to predict future detections based on recently observed performance objectives. Two metrics for water supply reliability and flood volume were tested using a series of nonparametric significance tests against the historical baseline. Sources of uncertainty in the detection time then were attributed to three uncertain drivers: the emissions scenario, land-use scenario, and climate model uncertainty using global sensitivity analysis. Finally, a logistic regression model was trained to predict future detections, in order to support dynamic adaptation planning through the design of early warning signals for nonstationary system performance. As mentioned previously, the optimization of actions conditioned on the detection of nonstationary performance was outside the scope of this study, and is left for future work. Instead, we focused on the behavior of detecting nonstationary changes on the projection data set, including analyzing how frequently detections occur, decomposing the sources of uncertainty in detections, and evaluating the possibility of predicting detections in advance.

Methods

Case Study and Scenario Data

Historical and future water supply and flooding objectives were simulated in a daily-timestep model of eight reservoirs in the Sacramento–San Joaquin River Basin (SSJRB) system in Northern California (Fig. 1). This case study is well suited to studying flood risk and water supply objectives because the system aims to balance both flood releases downstream of the reservoirs and water supply, in particular, exports pumped from the Sacramento–San Joaquin River Delta. Reservoir operating policies are fit empirically based on median observed data and adjusted for current conditions. This structure promotes parsimony and consistency among how each reservoir is operated within the model.

In all, the model consists of 43 parameters (5 parameters describe release policies for each of the 8 reservoirs, 2 parameters describe hydrological gains, and 1 parameter describes water

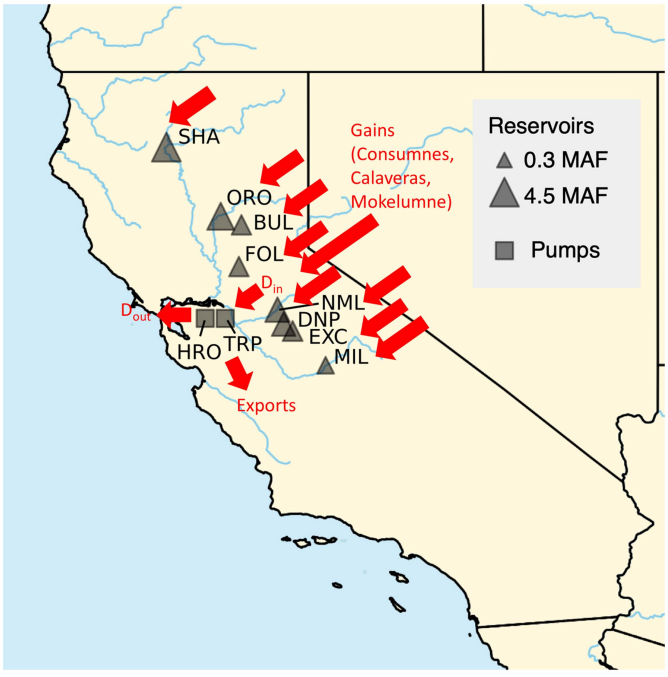

Fig. 1. Map of the SSJRB system.

supply exports pumped from the Sacramento–San Joaquin Delta) that are calibrated to historical observations using differential evolution (Storn and Price 1997). A previous version of the model served as the case study for uncertainty decomposition methods and was found to have sufficient accuracy in the broader context of climate uncertainty (Steinschneider et al. 2023). Water supply releases are given by Eq. (1), which represents a hedging rule for reservoir releases

$$\frac{R_i(t)}{R_{i,m}(t)} = \left(\frac{S_i(t)}{S_{i,m}(t)}\right)^{x_0} \quad \forall\ i \in \{0, 1, \ldots, 8\} \tag{1}$$

where i denotes one of the eight reservoirs; R = release; S = storage; x_0 is a fitted parameter; and subscript m denotes median observed value for that day of the water year. The release from Eq. (1) then is modified to include flood control releases if the day of the water year is between parameters $[x_1, x_2]$ and the storage exceeds a threshold, $S_i(t) > x_4 S_{i,m}(t)$. This obtains Eq. (2), which consists of a simplified seasonal flood pool rule

$$R_i(t)' = R_i(t) + x_3(S_i(t) - x_4 S_{i,m}(t)) \quad \forall\ i \in \{0, 1, \ldots, 8\} \tag{2}$$

where $\{x_1, \ldots x_4\}$ are fitted parameters.

Hydrologic gains into the Delta, $G_{\text{obs}}(t)$, are defined by Eq. (3), which captures any additional tributary inflows into the system that were not modeled explicitly. Eq. (4) describes how $G_{\text{obs}}(t)$ is estimated using observed data, which we denote $G(t)$

$$G_{\text{obs}}(t) \coloneqq D_{\text{in}}(t) - \sum_i R_i(t) \tag{3}$$

$$G(t) = G_m(t)\left(\sum_i \frac{S_i(t)}{K_i}\right)^{x_5} + x_6 \sum_i Q_i(t) \tag{4}$$

where D_{in} = Delta inflows; K = reservoir capacity; Q = reservoir inflows that are modeled explicitly; and x_5 and x_6 are fitted parameters. The first term in Eq. (4) adjusts the median gains for general hydrologic conditions, and the second term assumes that additional reservoir inflows are correlated with modeled inflows.

Finally, water supply exports pumped from the Delta are given by Eq. (5), which also consists of a hedging rule similar to Eq. (1)

$$P(t) = D_{\text{in}}(t) p_m(t)\left(\sum_i \frac{S_i(t)}{K_i}\right)^{x_7} \tag{5}$$

where p_m = median percentage of inflow pumped on that day of the water year, and x_7 is a fitted parameter. Except for Eq. (2), which is formally defined in USACE control manuals, the other operating rules are empirical and are fit to historical data with reasonable accuracy, as shown by Steinschneider et al. (2023). Steinschneider et al. (2023) also provided more detail about the model definition.

Streamflow inputs from the CMIP5 Climate and Hydrology Projections (Brekke et al. 2014) are input into the reservoir system model, containing a total of 97 climate scenarios generated from 4 Representative Concentration Pathways (RCPs) (2.6, 4.5, 6.0, and 8.5) and an ensemble of 31 General Circulation Models (GCMs). Furthermore, 36 different land-use scenarios (LULCs), including those from the USGS LUCAS model (Sleeter and Wilson 2017), the US Department of Energy GCAM model (West and Le Page 2014), and the USGS FORE-SCE model (Sohl et al. 2014), are used to estimate changes in water demand using estimates of crop water intensity (Mall and Herman 2019) and urban water demand (Christian-Smith et al. 2012). Reservoir releases then are adjusted using a demand multiplier estimated from the water demand projections relative to the historical baseline (Mall and Herman 2019; Robinson et al. 2020). Together these form an ensemble of 3,492 climate and land-use scenarios. Finally, system objectives are calculated on an annual timestep through the end of the century using these simulation results. Water supply reliability is calculated as the ratio of Delta pumping to the median historical pumping, which is a proxy for export demand from the SSJRB reservoirs, and is capped at 100%. The flooding objective is the summed total release exceeding the safe downstream release capacity for each reservoir. These models were used to determine the historical (1951–2000) and projected future (through 2100) distribution of system performance. The projections here reflect future performance predictions if the system is operated in the same way after the policies in Eqs. (1)–(5) are fit to historical releases. Specifically, the reservoir operating policy is not dynamic to changes in performance; if a detection of nonstationarity is a good indicator variable, it potentially could serve as a trigger to change the operating policy. This choice aligns with our focus on the detections of nonstationarity themselves, and not on dynamic reservoir operations.

Nonparametric Significance Detection

To detect significant trends in the projected future objectives versus the historical distribution (1951–2000), one-sided nonparametric Mann–Whitney U tests were conducted sequentially on a sliding window of fixed length through time on the annual projections, until the end of the simulation period. The one-sided tests capture decreasing system performance (i.e., an increase in the flood objective, or a decrease in the reliability objective). A window size of 30 years was chosen to obtain an adequate sample size for hypothesis testing while maintaining the ability to detect changes during the planning horizon. The threshold for detection was set at $p = 0.05$. Detection may occur either due to nonstationarity arising from processes that drive long-term change, or through natural variability that is sampled incompletely in the historical window. Differentiating a detection between these two causes is important to inform the design of adaptation signals.

The Mann–Whitney U tests make the assumptions of serial independence between annual observations. This assumption may hold for the inflow timeseries, but may be violated in this study by sequences of projected water supply reliability or flood volume because an autocorrelative structure is likely to arise due to year-to-year carryover storage. Yue and Wang (2002) found that a prewhitening procedure can effectively remove the effects of serial correlation from the Mann–Whitney U test. Thus, to account for the effect of serial correlation, we preprocess the reliability and flooding projections using a similar prewhitening process. In early experiments that were conducted without correcting for autocorrelation, we found that detections of nonstationarity for reliability scenarios generally occurred earlier than when a correction was applied. This confirms the findings of Yue and Wang, who concluded that the presence of positive autocorrelation can create false positives when using the Mann–Whitney U test. Figs. S2 and S3 in the Supplemental Materials show the autocorrelation functions of the reliability and flooding metrics before and after applying the correction.

For each year of each scenario in the ensemble, we record the p-value reflecting whether a detection occurs for each objective. Additionally, scenarios that never detect a significant change in the projection horizon are deemed no-detect scenarios, in which detection may occur after the end of the simulation period, given ongoing long-term changes in forcing. If a scenario is dominated by extreme events and high natural variability, a detection may occur but may return to the original distribution after the period of

high natural variability has passed through the detection window. The flood volume objective may be especially vulnerable to this; thus it is important to analyze how the p-values evolve with time in order to gauge such behavior.

We also compute the fraction of scenarios in the ensemble with a significant detection for each year in the timeseries. That is, if each plausible realization of the projected future is equally likely to occur, the detection rate represents an idealized measure of detection probability with time, although the model ensemble does not represent the full set of all possible futures.

Attribution and Uncertainty Decomposition

We aggregate the detection results using two metrics: the detection rate across the ensemble at the end of the century, and the year of first detection in each scenario. These metrics change based on the choice of GCM, RCP, and LULC scenario. To understand the relative influence of each, we first analyze the distribution of the metrics conditioned on the scenario factors. For the year of first detection in the 3,492 individual scenarios, descriptive statistics, including the median, standard deviation, and sample size are aggregated by GCM, RCP, and LULC. Additionally, scenarios incurring a significant detection are mapped to their projected objective severity at the end of the century. This step identifies whether a relationship exists between when a detection occurs and the severity of the scenario, particularly to assess cases in which severe scenarios are not detected.

Next, we conduct a global sensitivity analysis to attribute variance in the timing of first detection to the choice of GCM, RCP, and LULC scenario. The first-, second-, and total-order global sensitivity of first detection years to these input variables was computed using Sobol sensitivity analysis, providing a measure of how much variability in results is driven by each parameter both independently and through interactions with other parameters (Sobol 2001). An example decomposition of the variance in the year of first detection Y for $k = 3$ inputs is

$$\mathrm{Var}(Y) = \sum_{i=1}^{k} V_i + \sum_{i<j}^{k} V_{ij} + \cdots + V_{1,2,3,\ldots,k} \tag{6}$$

First- and total-order Sobol sensitivity indices, S, are expressed in Eqs. (7) and (8), respectively, where $V_{\sim i}$ indicates the variance contribution of all terms except input i. The Sobol analysis is performed using the open-source SALib software version 1.4.5 package (Herman and Usher 2017)

$$S_i = \frac{V_i}{\mathrm{Var}(Y)} \quad \forall\ i \in \{1, 2 \ldots k\} \tag{7}$$

$$S_{Ti} = 1 - \frac{V_{\sim i}}{\mathrm{Var}(Y)} \quad \forall\ i \in \{1, 2 \ldots k\} \tag{8}$$

Early Warning of Detection

Logistic regression classifiers were trained to predict the probability of a detection of significant change in system performance between the present year, t^*, and a lead time $t^* + L$—in other words, to predict if $p \leq 0.05$ for any time before L years into the future based on observed objective values up to t^*, i.e., the positive target class represents a detection occurring, and the negative target class represents no detections within the next L years. The input covariates include the standardized mean and standard deviation of objective values for each decade of available historical data up to t^*. The open-source library scikit-learn was used train the model parameters, and by default implemented L2-regularization (Pedregosa et al. 2011). To compare the behavior between models consistently, we assume a probability threshold of 0.5 (i.e., choosing the class with the higher predicted probability), although depending on the objectives a different sensitivity and specificity trade-off can be achieved in practice. A total of 75% of scenarios in the ensemble were selected randomly to train the model, and the remaining 25% were used to test model performance. Different models were trained for multiple combinations of L ranging from 10 to 40 years and t^* from 2001 to 2051. To assess model interpretability, trained model coefficients were studied to identify trends in which input features most strongly influence the probability of detection.

Results

Simulation Model Projections

Fig. 2 shows the 30-year moving average of modeled system objectives, water supply reliability, and upstream flood volume for all scenarios. For reliability, it is apparent that the mean of all scenarios

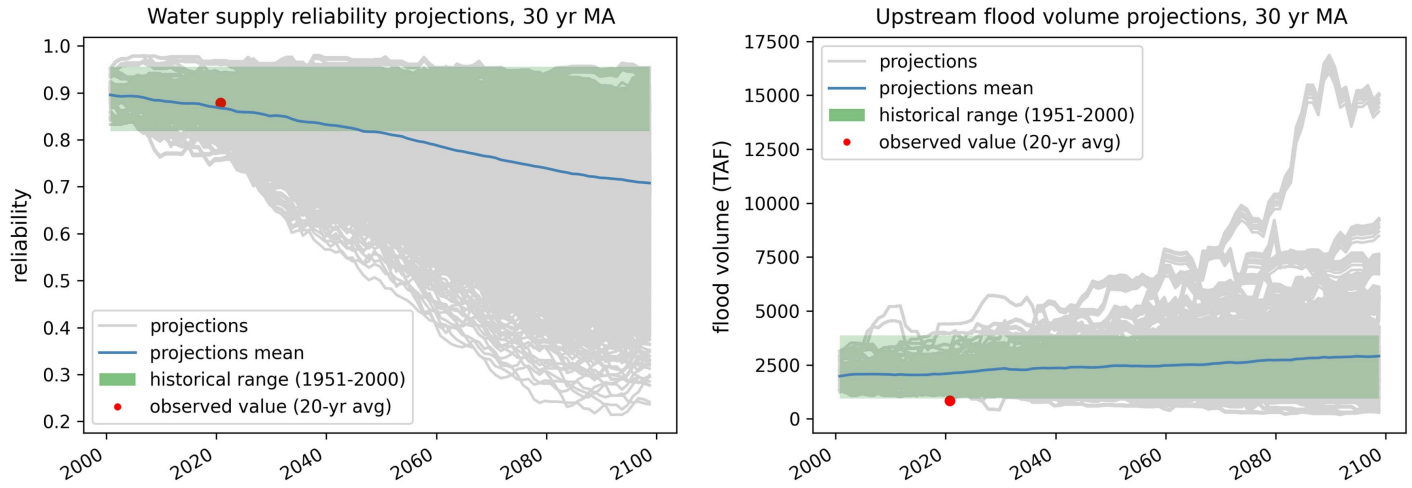

Fig. 2. 30-year moving averages of ensemble projections of water supply reliability and cumulative flood volume for 3,492 GCM, RCP, and LULC scenarios compared with the historical range.

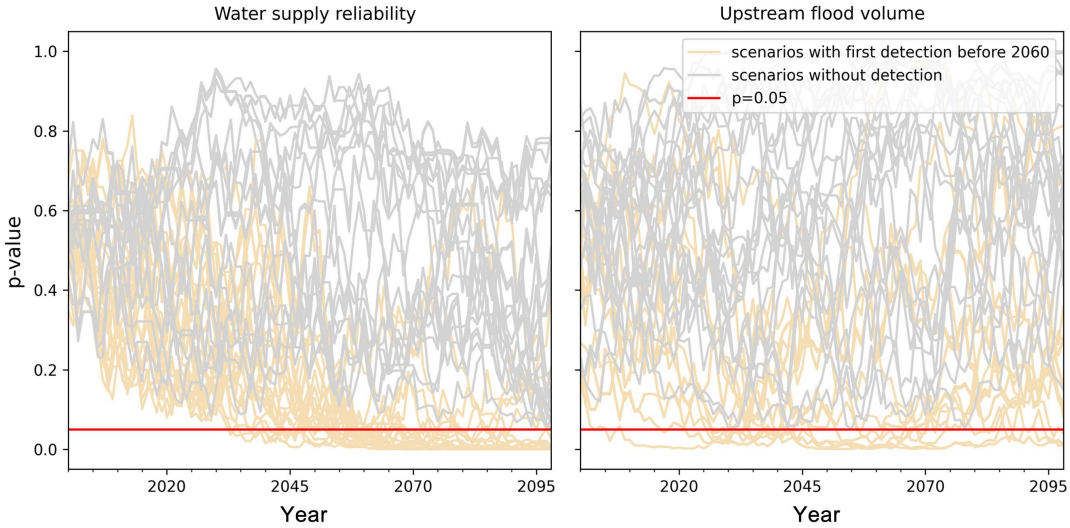

Fig. 3. Mann–Whitney U-test p-values. A total of 25 scenarios exhibiting detection and 25 no-detect scenarios were selected randomly. Scenarios exhibiting detection were chosen from those that first detected a change before 2060 in order to visualize behavior after the initial detection.

is decreasing, although variability increases substantially with time. For flood volume, the mean of all scenarios remains relatively constant, although variability also is increasing with time. Furthermore, it is clear that many scenarios fall well outside of the range of 30-year rolling means from the historical period of the projections, which indicates the potential for significant detections.

Fig. 2 also shows the mean of observed objectives for the 20 years of available data (i.e., simulation results using observed hydrology). Although observed water supply reliability is reasonably within the range of model projections over the historical period, observed upstream flood volumes are below the range. This indicates that the recent observed period was unusually dry, and/or that the scenario modeling chain contains a wet bias for the flood volume objective. Significance testing within each scenario is determined relative to its own historical period, and should not be affected by this bias. However, this result shows that the models and their coupled forcings may capture the observed distribution of objectives poorly despite high variability between scenarios, especially for extreme events.

Significance Detection

Fig. 3 shows the timeseries of p-values from the Mann–Whitney test for a subset of 50 scenarios—25 with and 25 without a detection. Of the scenarios with a detection, we visualized those that exhibited their first detection before 2060 to visualize detection behavior after the initial detection. For water supply reliability, most scenarios with a detection have homogenous detections that generally continue to occur after the first year. Additionally, despite some noise, p-values generally are decreasing before the first detection. This is indicative of detections of long-term change, rather than of short-term detections of natural variability. However, even if the trend for reliability is more monotonic than for flooding, there is still a component of detections due to natural variability. This is evident from the scenarios that do not remain significant after 2060, after the first detection has occurred. The selected no-detect scenarios in Fig. 3 do not have a clear trend in terms of p-values.

In contrast, p-values for the flooding objective do not follow a clear trend and have substantial noise. This is the case for the samples of scenarios both with and without a detection. This is evidence for detections driven by periods of natural variability, which then reverse as the detection window advances. False positive detections also may occur due to the Type 1 error of the statistical test, which was controlled to be under 5%. Additional experiments using expanding windows instead of rolling windows led to largely the same results: detections are infrequent and noisy for the flooding objective. Overall, these results highlight the impact of natural variability as the primary driver for detecting significant changes in the flooding objective. Conversely, reliability detections generally are more indicative of long-term change, although noisy detections due to natural variability also occur.

Fig. 4 shows the detection rates among the full simulation ensemble. The reliability objective had far greater detection rates than the flood volume objective. For reliability, 73.2% of scenarios had a detection at the end of the projection period, whereas only 12.9% of scenarios for flood volume had a significant change. Additionally, detections increased steadily throughout the projection for both reliability and flooding, even for the flooding objective in a small fraction of scenarios, although detections were rare due to natural variability.

Fig. 5 shows the distributions of the first year of detection. For reliability, the distribution of first detection years included 14.6% no-detects (not shown). The distribution was roughly symmetric, with the median detection occurring in 2058. For flood volume, the distribution included 65.1% no-detects, because the majority of flooding scenarios did not detect a significant change in the projection period. The distribution was more evenly distributed throughout the century, consistent with the influence of natural variability on this objective.

Fig. 6 shows the relationship between the first detection year and the projected objective severity at the end of the simulation period for the same scenario. For reliability, earlier detection years have a wide range of possible outcomes, but later detections occur predominantly for less severe scenarios. Thus, severe scenarios are more likely to be detected earlier, and are not likely to go undetected. This is confirmed by the distribution of reliability for no-detect scenarios, which are centered on high values. For flood volume, there is little relationship between detection year and severity, although the most severe scenarios in the ensemble were detected at some point during the century. As indicated previously, most scenarios have no detection, and these objectives are concentrated on lower values.

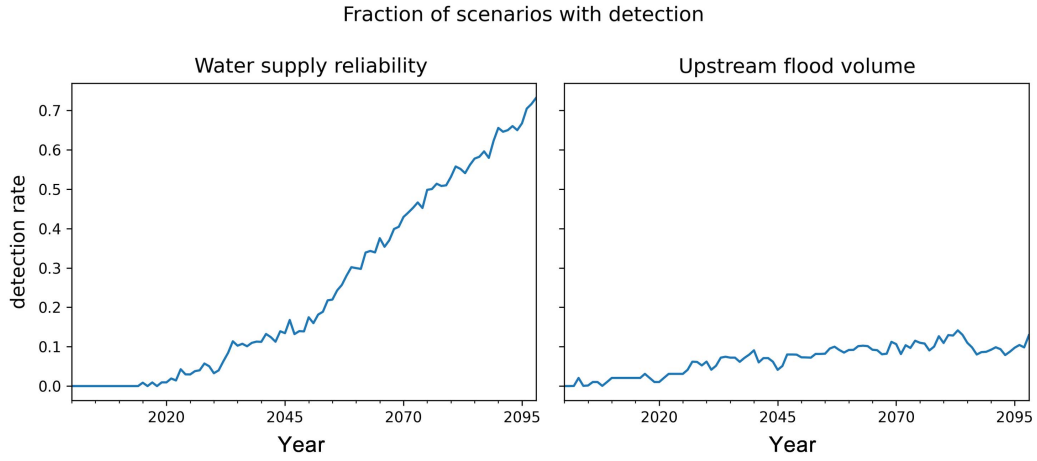

Fig. 4. Annual detection rates of significant detections out of the entire ensemble for the reliability objective and flood volume.

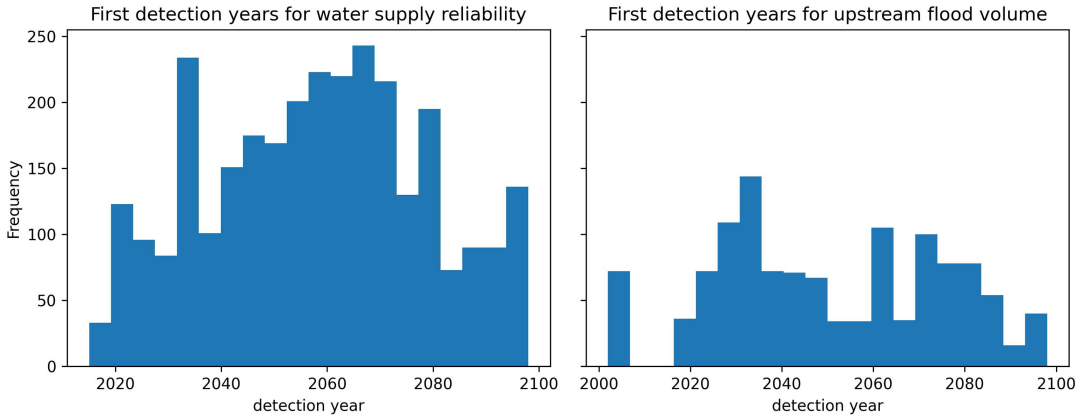

Fig. 5. Distributions of first detection years for reliability and flood volume for all scenarios. No-detect scenarios are not shown (509 for reliability and 2,275 for flood volume).

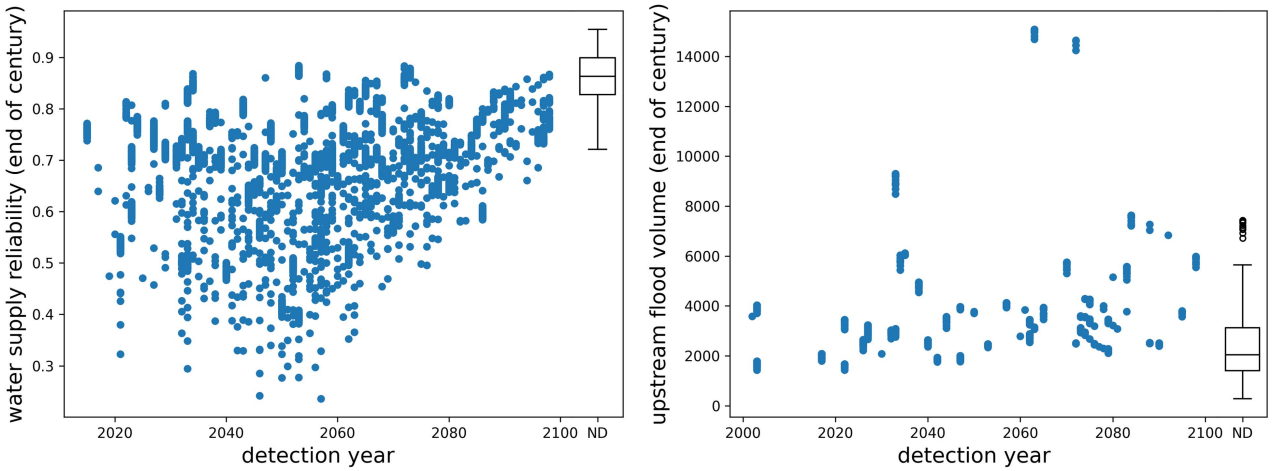

Fig. 6. Detection year versus reliability and flood volume at the end of the projection period. ND = no-detect scenarios.

Uncertainty Decomposition

Detection rates at the end of the projection period (2098) are sorted by choice of GCM, RCP, and LULC (Fig. 7). The results for GCM and LULC are shown as histograms, whereas for the four RCPs the individual detection rates are shown as a bar plot. For reliability, the GCM contributes the most to the variability in detection rates, because there are GCMs with both near-zero and near-100% detection rates; the majority are above 60%. Additionally, all RCPs and

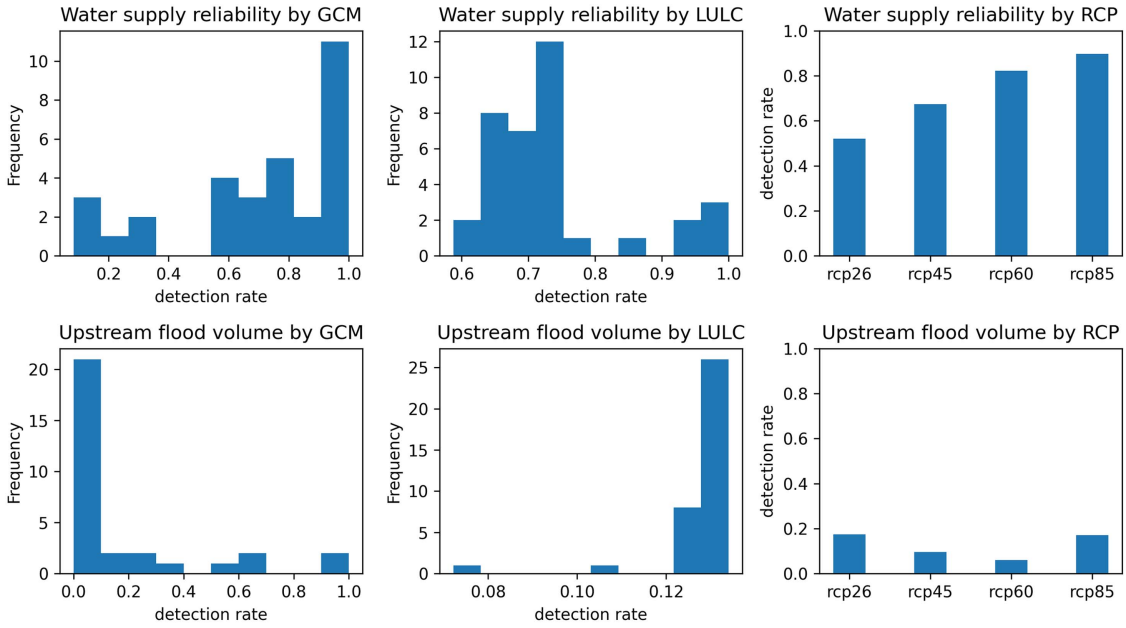

Fig. 7. Distribution of relative detection counts at the end of the projection period (2098) sorted by GCM, RCP, and LULC for reliability and flood volume. The results for GCM and LULC are shown as histograms, in which each observation represents the detection rate, or fraction of scenarios with detection, within one GCM or LULC. RCP results are shown individually in a bar plot with each RCP labelled.

LULCs have detection rates greater than 50%. In other words, at least 50% of scenarios containing any given RCP or LULC will detect a significant change at the end of the projection period. The variability is attributable mostly to GCM choice, for which detection rates generally are high but have much greater variance overall. Interestingly, detection rates also increase under more severe RCPs. For flood volume, detection rates are consistently low (20% or less) and do not vary much with the choice of RCP or LULC scenario. The variability in detection rates is driven primarily by GCM choice, for which detection rates mostly are low, but with several high outliers.

In addition, when first detection years are sorted by GCM, RCP, and LULC, it is apparent that GCM drives most of the variability in the median. The first detections occur throughout the century for both the water supply and flooding objectives. However, for the flooding objective 10 of 31 GCMs did not detect a significant change in the projection period. For LULC, the median exhibits a much smaller range, although standard deviations can be somewhat larger than from the GCM aggregates in some cases. This is centered on about the year 2059 for water supply and the year 2047 for flooding (Fig. S1 in the Supplemental Materials). Additional histograms are plotted separately in Fig. 8 for the RCP aggregates. For water supply, the median appears to be somewhat increasing with increasing RCP, and there appear to be more RCP8.5 detections later in the century than in the other RCPs. This suggests that there are cases of RCP8.5 scenarios that are less severe until later in the century, given the result that less severe scenarios tend to be detected later. This is consistent with previous results, because water supply reliability for RCP8.5 generally is less severe than for the other RCPs until midcentury (Fig. 9). This pattern does not occur in the flooding objective, which did not exhibit obvious trends.

The Sobol sensitivity analysis confirms the finding that GCM contributes to the greatest variability in the detection year, followed by the second-order interaction between the GCM and RCP (Tables 1–4). For total-order sensitivity, the choice of GCM contributes Sobol sensitivity indexes of 0.971 and 1.012 for reliability and flood volume, respectively. The RCP, which represents the effects of climate change, is next in importance, but only through its interaction with GCM choice. The total-order sensitivity for RCP is 0.537 and 0.304, respectively, for the reliability and flood volume objectives, whereas the second-order interaction scores with GCM are 0.497 and 0.297. The RCP has essentially no first-order contribution for reliability and flood volume (GCM is notably higher in both). Bootstrapped 95% confidence intervals for the sensitivity indices are reasonably small, suggesting a sufficient sample size for convergence.

Early Warning of Detection

Finally, we investigated whether the nonstationarity detections can be predicted in advance from observed system performance. Fig. 10 shows the true positive rate (sensitivity), true negative rate (specificity), and area under the receiver operating characteristic curve (AUC) scores for logistic regression classifiers trained on varying combinations of starting time t^* and lead time L. Overall, we found high AUC scores for both objectives, suggesting that the models have high predictive ability in general. For water supply reliability, sensitivity increases for larger t^* values, especially at longer lead times. This suggests that the classifiers are more accurate at discerning detections within a long time window, but are less accurate at determining the precise timing of a detection. For the flood volume objective, the sensitivity is noticeably lower than the specificity of the models, except at short lead times. Additionally, the specificity is close to 100%. This is consistent with the prior results indicating that detections in upstream flood volume are driven largely by natural variability and may not be predictable from recent observations. At short lead times, the model likely has learned

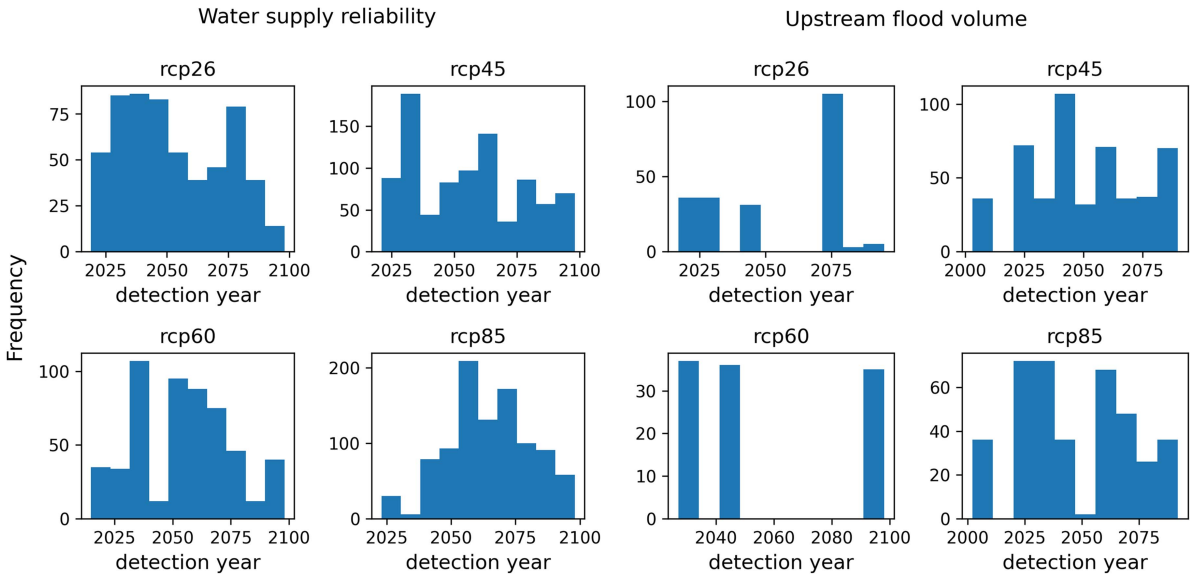

Fig. 8. Histograms of first detection years sorted by RCP for both the water supply reliability and flood volume objectives. Each observation in the histogram represents a scenario with a respective RCP.

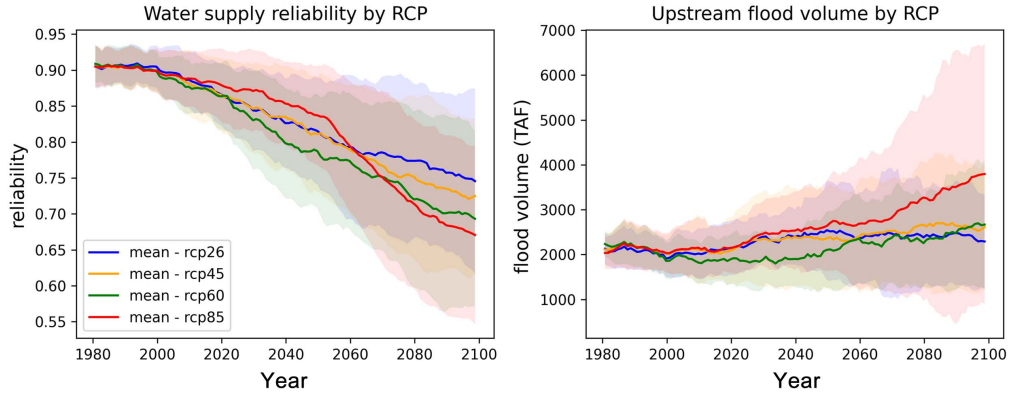

Fig. 9. Objective severity separated by RCP. Shaded bands indicate ±1 standard deviation.

Table 1. First- and total-order Sobol sensitivity analysis of year of first detection for water supply reliability

Variable	ST	ST_conf	S1	S1_conf
RCP	0.537	0.050	0.008	0.064
GCM	0.971	0.092	0.392	0.082
LULC	0.107	0.024	0.043	0.026

Note: ST = total-order sensitivity; ST_conf = half-width of 95% confidence interval (CI) for total order sensitivity; S1 = first-order sensitivity; and S1_conf = half-width of 95% CI for first order sensitivity. Confidence intervals are the half-width of the bootstrapped 95% CI.

Table 2. Second-order (interaction) Sobol sensitivity analysis of year of first detection for water supply reliability

Variable pairs	S2	S2_conf
RCP, GCM	0.497	0.087
RCP, LULC	−0.008	0.083
GCM, LULC	0.002	0.095

Note: S2 = second-order (interaction) sensitivity; and S2_conf = half-width of 95% CI for second order sensitivity. Confidence intervals are the half-width of the bootstrapped 95% CI. Negative values are statistically zero.

Table 3. First- and total-order Sobol sensitivity analysis of year of first detection for upstream flood volume

Variable	ST	ST_conf	S1	S1_conf
RCP	0.304	0.035	0.001	0.045
GCM	1.012	0.066	0.674	0.084
LULC	0.005	0.005	−0.003	0.004

Note: ST = total-order sensitivity; ST_conf = half-width of 95% confidence interval (CI) for total order sensitivity; S1 = first-order sensitivity; and S1_conf = half-width of 95% CI for first order sensitivity. Confidence intervals are the half-width of the bootstrapped 95% CI. Negative values are statistically zero.

Table 4. Second-order (interaction) Sobol sensitivity analysis of year of first detection for upstream flood volume

Variable pairs	S2	S2_conf
RCP, GCM	0.297	0.058
RCP, LULC	−0.035	0.054
GCM, LULC	0.002	0.086

Note: S2 = second-order (interaction) sensitivity; and S2_conf = half-width of 95% CI for second order sensitivity. Confidence intervals are the half-width of the bootstrapped 95% CI. Negative values are statistically zero.

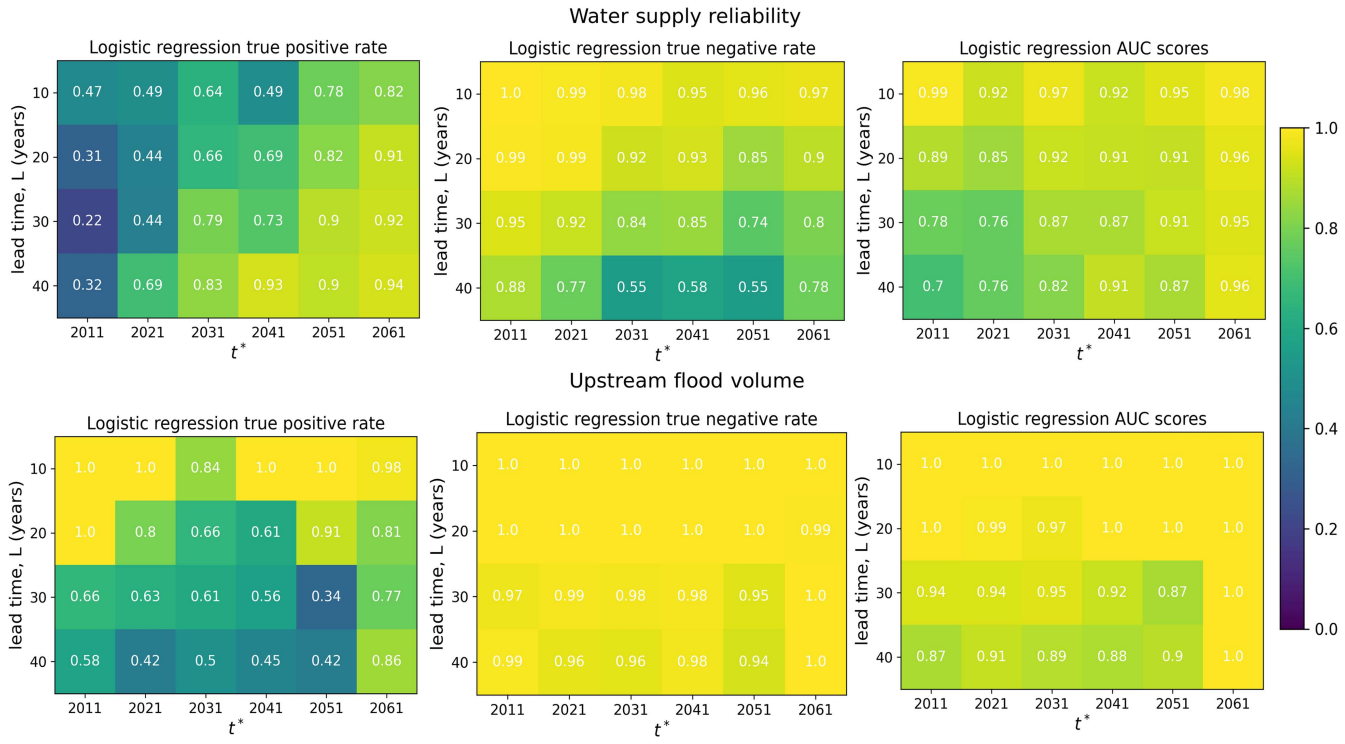

Fig. 10. True positive and true negative scores for logistic regression classifier according to the time of prediction, t^* and the lead time L.

to predict the negative class with near-perfect accuracy. In addition, flooding detections occur much less frequently in the data set, so true negatives are an easier prediction task in general.

Fig. 11 shows trends in trained model coefficients for a selected prediction year, $t^* = 2051$, in order to understand which input observations are driving the early warning signals. For water supply reliability, coefficients in the near term are notably negative for the last 2 decades, and generally are positive before 2030. This indicates that a scenario with high reliability later in the projection period is less likely to be classified as a detection. Conversely, high reliability earlier in the projection increases the probability of detection. The coefficients of the standard deviation features have less of a trend and are less interpretable. For flood volume, coefficients for mean features tend to trend upward, and tend to be consistently positive in the near term. As a result, a plausible hypothesis is that detection probability is higher when the historical floods are less extreme, such as when the historical period undersamples the true natural variability, and following a recent large flood event. This is consistent with the finding that detections for the flooding objective are dominated by natural variability instead of long-term change. The coefficients for the standard deviation features do not have a clear trend.

Discussion and Conclusion

This study contributes a method to analyze whether and when detections of nonstationary reservoir system performance will occur, to support adaptive planning in the context of uncertain ensemble projections. We propose three key findings

1. The ability to detect nonstationarity strongly depends on the performance objective. This indicates that multiple adaptation indicators may be needed for multiobjective reservoir management problems (Quinn et al. 2017). Specifically, we find that water supply reliability is far more likely to detect change within the century, driven by long-term change, compared with the flooding objective, for which detections are dominated by natural variability. This finding most likely holds beyond this case study, because the difficulty of detecting nonstationarity in extreme events has been recognized widely (e.g., Bass et al. 2022; Siler et al. 2019). In addition, the scenarios with the most severe changes in system performance are detectable earlier in the period.
2. The variance in detection time is attributable mostly to the choice of GCM, including its interaction with the emissions scenario. The GCM factor also embeds natural variability, because the scenario ensemble used in this experiment contained only one realization per GCM; this could be augmented in future work by large-ensemble scenarios containing many realizations (Lehner et al. 2020). Model uncertainty suggests a high potential for dynamic learning as new observations become available. The scenario ensemble also did not include hydrologic uncertainty or endogenous feedbacks in reservoir management, both of which could affect the scale of the climate uncertainty (e.g., Anghileri et al. 2018; Broderick et al. 2019).
3. A logistic regression model trained to predict future detections had mixed results. Early warning predictions for both water supply and flood objectives performed well overall, although for reliability it is difficult to predict the exact timing of the future detection within the designated window because the model true positive rate in the near term is comparatively low. In contrast, for the flooding objective, it is difficult to predict detections at longer lead times, because the model tends to have a lower true positive rate for longer lead times. This underscores natural variability as the main driver for flooding detection. This result expands prior work on early warning signals for adaptation (e.g., Raso et al. 2019; Robinson et al. 2020) by considering the detection of nonstationarity as the target variable in the classification.

These results are applicable directly to inform adaptive planning. Detection of nonstationary reliability, or early warning signals

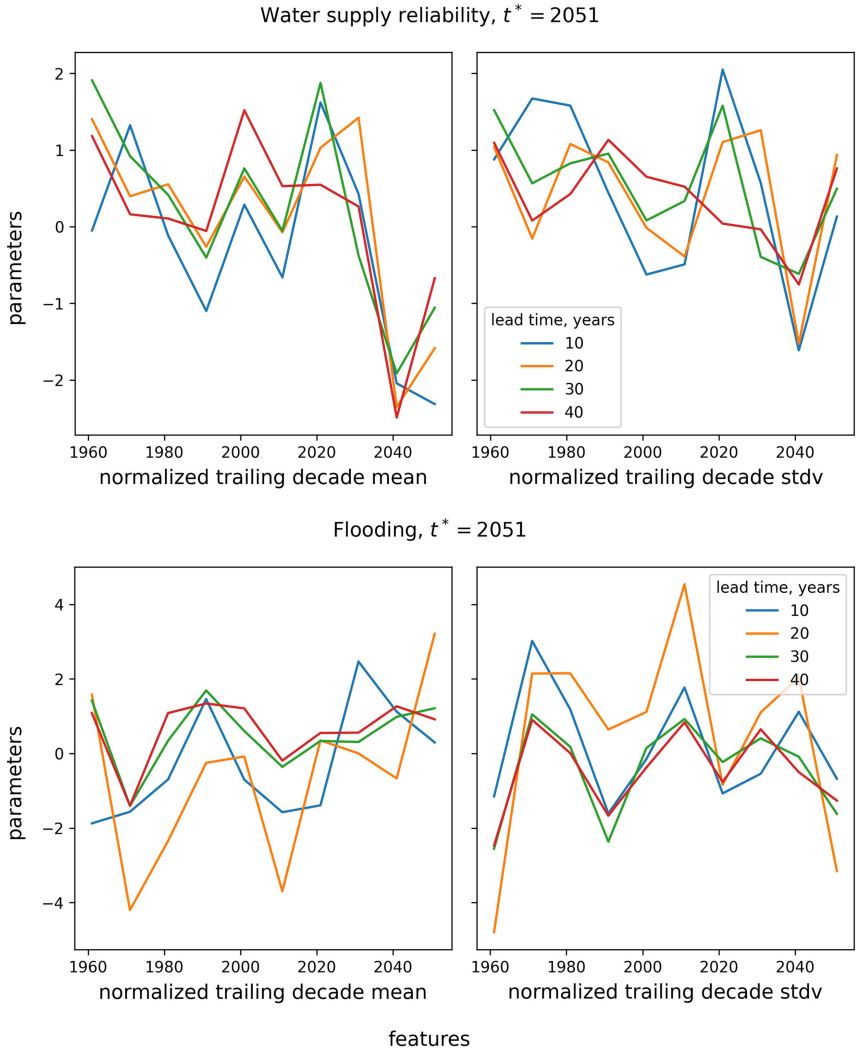

Fig. 11. Trained logistic regression coefficients for water supply reliability and flood volume at $t^* = 2051$.

thereof, would be more reliable indicators to trigger action because they imply long-term change and are less influenced by natural variability. In contrast, the flooding objective, and its lack of detection, indicates difficulty in separating long-term trends from natural variability. However, this does not preclude the success of dynamic adaptation, especially if adaptation focuses more on flexible decisions that are reversible when the future unfolds differently than expected (Fletcher et al. 2019). Similarly, these results highlight the importance of differentiating natural variability and long-term change when making irreversible infrastructure decisions, because such decisions may incur high regret (Borgomeo et al. 2018). There also may be cases in which successful adaptations are not restricted to long-term change; for example, when collecting more observations of natural variability over time leads to revised estimates of extreme event magnitudes.

Several opportunities for future work remain. First, indicators also could be developed by applying the same detection methodology to hydroclimate variables, for example as Haasnoot et al. (2018) proposed for precipitation. In this case, a difference in detection time may occur relative to the engineering system objectives. This also may help to map the occurrence of detections to their physical drivers. Second, nonstationary system behavior also will be driven by endogenous changes to reservoir operations, water demand, and the feedback between them, which could either mitigate or amplify the challenges caused by climate change (Jafino et al. 2019). Finally, nonstationarity detections and/or early warning signals can be used as indicator variables in policy search methods for dynamic adaptation to climate change (Cohen and Herman 2021). Their information content can be compared with other indicators that do not necessarily reflect nonstationarity; although statistical significance is not a requirement to trigger adaptation, there may be potential to limit false positives and improve performance for out-of-sample scenarios. Given the substantial uncertainty in climate and land-use scenarios, this study provides a method to understand the potential for learning about dynamic changes in reservoir system performance, and the timing of adaptations needed to maintain robustness in the long term.

Data Availability Statement

All code corresponding to methods and figure generation are available at the repository https://github.com/Matt2371/climate_detection.

Acknowledgments

This work was supported by the US National Science Foundation grant CBET-2041826. We further acknowledge the World Climate Research Program's Working Group on Coupled Modeling and the

climate modeling groups listed in the Supplemental Materials for producing and making available their model output.

Supplemental Materials

Figs. S1–S3 and Table S1 are available online in the ASCE Library (www.ascelibrary.org).

References

Anghileri, D., M. Botter, A. Castelletti, H. Weigt, and P. Burlando. 2018. "A comparative assessment of the impact of climate change and energy policies on Alpine hydropower." *Water Resour. Res.* 54 (11): 9144–9161. https://doi.org/10.1029/2017WR022289.

Bass, B., J. Norris, C. Thackeray, and A. Hall. 2022. "Natural variability has concealed increases in Western US flood hazard since the 1970s." *Geophys. Res. Lett.* 49 (7): 1–10. https://doi.org/10.1029/2021GL097706.

Borgomeo, E., M. Mortazavi-Naeini, J. W. Hall, and B. P. Guillod. 2018. "Risk, robustness and water resources planning under uncertainty." *Earth's Future* 6 (3): 468–487. https://doi.org/10.1002/2017EF000730.

Brekke, L., A. Wood, and T. Pruitt. 2014. "Downscaled CMIP3 and CMIP5 hydrology projections release of hydrology projections, comparison with preceding information, and summary of user needs." Accessed September 9, 2022. https://gdo-dcp.ucllnl.org/downscaled_cmip_projections/.

Broderick, C., C. Murphy, R. L. Wilby, T. Matthews, C. Prudhomme, and M. Adamson. 2019. "Using a scenario-neutral framework to avoid potential maladaptation to future flood risk." *Water Resour. Res.* 55 (2): 1079–1104. https://doi.org/10.1029/2018WR023623.

Bryant, B. P., and R. J. Lempert. 2010. "Thinking inside the box: A participatory, computer-assisted approach to scenario discovery." *Technol. Forecasting Social Change* 77 (1): 34–49. https://doi.org/10.1016/j.techfore.2009.08.002.

Ceres, R. L., C. E. Forest, and K. Keller. 2017. "Understanding the detectability of potential changes to the 100-year peak storm surge." *Clim. Change* 145 (1–2): 221–235. https://doi.org/10.1007/s10584-017-2075-0.

Christian-Smith, J., H. Matthew, and L. Allen. 2012. "Urban water demand in California to 2100: Incorporating climate change." Accessed September 14, 2022. https://www.pacinst.org.

Cohen, J. S., and J. D. Herman. 2021. "Dynamic adaptation of water resources systems under uncertainty by learning policy structure and indicators." *Water Resour. Res.* 57 (11): 1–24. https://doi.org/10.1029/2021WR030433.

de Neufville, R., and K. Smet. 2019. "Engineering options analysis (EOA)." In *Decision making under deep uncertainty*, 117–132. Berlin: Springer.

Fletcher, S., M. Lickley, and K. Strzepek. 2019. "Learning about climate change uncertainty enables flexible water infrastructure planning." *Nat. Commun.* 10 (1): 1782. https://doi.org/10.1038/s41467-019-09677-x.

Greve, P., L. Gudmundsson, and S. I. Seneviratne. 2018. "Regional scaling of annual mean precipitation and water availability with global temperature change." *Earth Syst. Dyn.* 9 (1): 227–240. https://doi.org/10.5194/esd-9-227-2018.

Groves, D. G., E. Bloom, R. J. Lempert, J. R. Fischbach, J. Nevills, and B. Goshi. 2015. "Developing key indicators for adaptive water planning." *J. Water Resour. Plann. Manage.* 141 (7): 05014008. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000471.

Haasnoot, M., J. H. Kwakkel, W. E. Walker, and J. ter Maat. 2013. "Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world." *Global Environ. Change* 23 (2): 485–498. https://doi.org/10.1016/j.gloenvcha.2012.12.006.

Haasnoot, M., S. van't Klooster, and J. van Alphen. 2018. "Designing a monitoring system to detect signals to adapt to uncertain climate change." *Global Environ. Change* 52 (Sep): 273–285. https://doi.org/10.1016/j.gloenvcha.2018.08.003.

Hamarat, C., J. H. Kwakkel, E. Pruyt, and E. T. Loonen. 2014. "An exploratory approach for adaptive policymaking by using multi-objective robust optimization." *Simul. Modell. Pract. Theory* 46 (Aug): 25–39. https://doi.org/10.1016/j.simpat.2014.02.008.

Hawkins, E., and R. Sutton. 2009. "The potential to narrow uncertainty in regional climate predictions." *Bull. Am. Meteorol. Soc.* 90 (8): 1095–1108. https://doi.org/10.1175/2009BAMS2607.1.

Hecht, J. S., and R. M. Vogel. 2020. "Updating urban design floods for changes in central tendency and variability using regression." *Adv. Water Resour.* 136 (Feb): 103484. https://doi.org/10.1016/j.advwatres.2019.103484.

Hegerl, G., and F. Zwiers. 2011. "Use of models in detection and attribution of climate change." In *Wiley interdisciplinary reviews: Climate change*. Hoboken, NJ: Wiley. https://doi.org/10.1002/wcc.121.

Herman, J., and W. Usher. 2017. "SALib: An open-source Python library for sensitivity analysis." *J. Open Source Software* 2 (9): 1–2. https://doi.org/10.21105/joss.00097.

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

Hinkel, J., and A. Bisaro. 2016. "Methodological choices in solution-oriented adaptation research: A diagnostic framework." *Reg. Environ. Change* 16 (1): 7–20. https://doi.org/10.1007/s10113-014-0682-0.

Hui, R., J. Herman, J. Lund, and K. Madani. 2018. "Adaptive water infrastructure planning for nonstationary hydrology." *Adv. Water Resour.* 118 (Aug): 83–94. https://doi.org/10.1016/j.advwatres.2018.05.009.

Jafino, B. A., M. Haasnoot, and J. H. Kwakkel. 2019. "What are the merits of endogenising land-use change dynamics into model-based climate adaptation planning?" *Socio-Environ. Syst. Modell.* 1 (Feb): 16126. https://doi.org/10.18174/sesmo.2019a16126.

Katz, R. W. 2013. "Statistical methods for nonstationary extremes." In *Extremes in a changing climate*, edited by A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, 15–37. Dordrecht, Netherlands: Springer.

Kwakkel, J. H., M. Haasnoot, and W. E. Walker. 2015. "Developing dynamic adaptive policy pathways: A computer-assisted approach for developing adaptive strategies for a deeply uncertain world." *Clim. Change* 132 (3): 373–386. https://doi.org/10.1007/s10584-014-1210-4.

Lee, B. S., M. Haran, and K. Keller. 2017. "Multidecadal scale detection time for potentially increasing Atlantic storm surges in a warming climate." *Geophys. Res. Lett.* 44 (20): 10–617. https://doi.org/10.1002/2017GL074606.

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

Lempert, R. J., and D. G. Groves. 2010. "Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the American west." *Technol. Forecasting Social Change* 77 (6): 960–974. https://doi.org/10.1016/j.techfore.2010.04.007.

Mall, N. K., and J. D. Herman. 2019. "Water shortage risks from perennial crop expansion in California's Central Valley." *Environ. Res. Lett.* 14 (10): 104014. https://doi.org/10.1088/1748-9326/ab4035.

Mortazavi-Naeini, M., G. Kuczera, A. S. Kiem, L. Cui, B. Henley, B. Berghout, and E. Turner. 2015. "Robust optimization to secure urban bulk water supply against extreme drought and uncertain climate change." *Environ. Modell. Software* 69 (Jul): 437–451. https://doi.org/10.1016/j.envsoft.2015.02.021.

Orlowsky, B., and S. I. Seneviratne. 2013. "Elusive drought: Uncertainty in observed trends and short- and long-term CMIP5 projections." *Hydrol. Earth Syst. Sci.* 17 (5): 1765–1781. https://doi.org/10.5194/hess-17-1765-2013.

Papalexiou, S. M., and A. Montanari. 2019. "Global and regional increase of precipitation extremes under global warming." *Water Resour. Res.* 55 (6): 4901–4914. https://doi.org/10.1029/2018WR024067.

Pedregosa, F., et al. 2011. "Scikit-learn: Machine learning in python." *J. Mach. Learn. Res.* 12 (85): 2825–2830.

Prosdocimi, I., T. R. Kjeldsen, and J. D. Miller. 2015. "Detection and attribution of urbanization effect on flood extremes using nonstationary

flood-frequency models." *Water Resour. Res.* 51 (6): 4244–4262. https://doi.org/10.1002/2015WR017065.
Quinn, J. D., P. M. Reed, and K. Keller. 2017. "Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points." *Environ. Modell. Software* 92 (Jun): 125–141. https://doi.org/10.1016/j.envsoft.2017.02.017.
Raso, L., J. Kwakkel, and J. Timmermans. 2019. "Assessing the capacity of adaptive policy pathways to adapt on time by mapping trigger values to their outcomes." *Sustainability* 11 (6): 1–16. https://doi.org/10.3390/su11061716.
Robinson, B., J. S. Cohen, and J. D. Herman. 2020. "Detecting early warning signals of long-term water supply vulnerability using machine learning." *Environ. Modell. Software* 131 (Sep): 104781. https://doi.org/10.1016/j.envsoft.2020.104781.
Robinson, B., and J. D. Herman. 2019. "A framework for testing dynamic classification of vulnerable scenarios in ensemble water supply projections." *Clim. Change* 152 (3–4): 431–448. https://doi.org/10.1007/s10584-018-2347-3.
Siler, N., C. Proistosescu, and S. Po-Chedley. 2019. "Natural variability has slowed the decline in western US snowpack since the 1980s." *Geophys. Res. Lett.* 46 (1): 346–355. https://doi.org/10.1029/2018GL081080.
Slater, L. J., et al. 2021. "Nonstationary weather and water extremes: A review of methods for their detection, attribution, and management." *Hydrol. Earth Syst. Sci.* 25 (7): 3897–3935. https://doi.org/10.5194/hess-25-3897-2021.
Sleeter, B. M., and T. T. Wilson. 2017. *Land-use and land-cover projections for California's 4th climate assessment*. Reston, VA: USGS.
Smith, R. L., C. Tebaldi, D. Nychka, and L. O. Mearns. 2009. "Bayesian modeling of uncertainty in ensembles of climate models." *J. Am. Stat. Assoc.* 104 (485): 97–116. https://doi.org/10.1198/jasa.2009.0007.
Sobol, I. M. 2001. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." *Math. Comput. Simul.* 55 (1–3): 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6.
Sohl, T. L., et al. 2014. "Spatially explicit modeling of 1992–2100 land cover and forest stand age for the conterminous United States." *Ecol. Appl.* 24 (5): 1015–1036. https://doi.org/10.1890/13-1245.1.
Steinschneider, S., J. D. Herman, J. Kucharski, M. Abellera, and P. Ruggiero. 2023. "Uncertainty decomposition to understand the influence of water systems model error in climate vulnerability assessments." *Water Resour. Res.* 59 (1): e2022WR032349. https://doi.org/10.1029/2022WR032349.
Storn, R., and K. Price. 1997. "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces." *J. Global Optim.* 11 (4): 341–359. https://doi.org/10.1023/A:1008202821328.
Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns. 2005. "Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles." *J. Clim.* 18 (10): 1524–1540. https://doi.org/10.1175/JCLI3363.1.
van Ginkel, K. C. H., M. Haasnoot, and W. J. Wouter Botzen. 2022. "A stepwise approach for identifying climate change induced socio-economic tipping points." *Clim. Risk Manage.* 37 (Apr): 100445. https://doi.org/10.1016/j.crm.2022.100445.
Walker, W. E., M. Haasnoot, and J. H. Kwakkel. 2013. "Adapt or perish: A review of planning approaches for adaptation under deep uncertainty." *Sustainability* 5 (3): 955–979. https://doi.org/10.3390/su5030955.
Weaver, C. P., R. J. Lempert, C. Brown, J. A. Hall, D. Revell, and D. Sarewitz. 2013. "Improving the contribution of climate model information to decision making: The value and demands of robust decision frameworks." *Wiley Interdiscip. Rev. Clim. Change* 4 (1): 39–60. https://doi.org/10.1002/wcc.202.
West, T. O., and Y. Le Page. 2014. *CMS: Land cover projections (5.6-km) from GCAM v3.1 for Conterminous USA, 2005–2095*. Washington, DC: US DOE.
Westra, S., L. Alexander, and F. W. Zwiers. 2013. "Global increasing trends in annual maximum daily precipitation." *J. Clim.* 26 (11): 3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1.
Whateley, S., and C. Brown. 2016. "Assessing the relative effects of emissions, climate means, and variability on large water supply systems." *Geophys. Res. Lett.* 43 (21): 11–338. https://doi.org/10.1002/2016GL070241.
Wilby, R. L. 2006. "When and where might climate change be detectable in UK river flows?" *Geophys. Res. Lett.* 33 (19): 1–5. https://doi.org/10.1029/2006GL027552.
Wilby, R. L., and S. Dessai. 2010. "Robust adaptation to climate change." *Weather* 65 (7): 180–185. https://doi.org/10.1002/wea.543.
Woodward, M., Z. Kapelan, and B. Gouldby. 2014. "Adaptive flood risk management under climate change uncertainty using real options and optimization." *Risk Anal.* 34 (1): 75–92. https://doi.org/10.1111/risa.12088.
Yue, S., and C. Wang. 2002. "The influence of serial correlation on the Mann–Whitney test for detecting a shift in median." *Adv. Water Resour.* 25 (3): 325–333. https://doi.org/10.1016/S0309-1708(01)00049-5.
Zeff, H. B., J. D. Herman, P. M. Reed, and G. W. Characklis. 2016. "Cooperative drought adaptation: Integrating infrastructure development, conservation, and water transfers into adaptive policy pathways." *Water Resour. Res.* 52 (9): 7327–7346. https://doi.org/10.1002/2016WR018771.
Ziegler, A. D., E. P. Maurer, J. Sheffield, B. Nijssen, E. F. Wood, and D. P. Lettenmaier. 2005. "Detection time for plausible changes in annual precipitation, evapotranspiration, and streamflow in three Mississippi River sub-basins." *Clim. Change* 72 (1–2): 17–36. https://doi.org/10.1007/s10584-005-5379-4.