Influence of Subseasonal-to-Annual Water Supply Forecasts on Many-Objective Water System Robustness under Long-Term Change

Kyla Semmendinger, Ph.D.[1]; and Scott Steinschneider, Ph.D.[2]

Abstract: The sensitivity of forecast-informed reservoir operating policies to forecast attributes (lead-time and skill) in many-objective water systems has been well-established. However, the viability of forecast-informed operations as a climate change adaptation strategy remains underexplored, especially in many-objective systems with complex trade-offs across interests. Little is known about the relationships between forecast attribute and policy robustness under deep uncertainty in future conditions and the relationships between forecast-informed performance and future hydrologic state. This study explores the sensitivity of forecast-informed policy robustness to forecast lead-time and skill in the outflow management plan of the Lake Ontario basin. We create water supply forecasts at four different subseasonal-to-annual lead-times and two levels of skill and further employ a many-objective evolutionary algorithm to discover policies tailored for each forecast case, historical supply conditions, and six objectives. We also leverage a partnership with decision-makers to identify a subset of candidate policies, which are reevaluated under a large set of plausible hydrologic conditions that reflect stationary and nonstationary climates. Scenario discovery techniques are used to map attributes of future hydrology to forecast-informed policy performance. Results show policy robustness is directly related to forecast lead-time, where policies conditioned on 12-month forecasts were more robust under future hydrology. Policies tailored for noisier long-lead forecasts were more robust under a wide range of plausible futures compared with policies trained to perfect forecasts, which highlights the potential to overfit control policies to historical information, even for a forecast-informed policy with perfect foresight. The relationship between performance and the hydrologic regime is dependent on the complexity of the interactions between control decisions and objectives. A threshold of objective performance as a function of supply conditions can support adaptive management of the system. However, more complex interactions make it difficult to identify simple hydrologic indicators that can serve as triggers for dynamic management. **DOI: 10.1061/JWRMD5.WRENG-6205.** *© 2024 American Society of Civil Engineers.*

Introduction

Water supply forecasts have a long history of use in reservoir operations. Early efforts utilized simple statistical forecast models of water supply based on hydrologic persistence, which were valuable for short-term (hourly-to-daily) operations and monthly allocation decisions under near-average conditions (Van Der Beken et al. 1980; Russell and Caselton 1971; Stedinger et al. 1984; Wilson and Kirdar 1970). Over time, the sophistication of forecast-informed reservoir operating policies grew with improvements in the forecasts themselves (Faber and Stedinger 2001), but adoption of forecast use in practice remained limited due to concerns around skill and the risk attitudes of water managers, among other factors (Rayner et al. 2005). Recently, though, there has been renewed interest in forecast-informed policies, particularly in locations where water stress is growing and forecast skill has improved significantly at longer lead times (weeks to months) and for more extreme conditions using state-of-the-art forecasting systems (Alexander et al. 2021; Delaney et al. 2020; Woodside et al. 2022; Zarei et al. 2021). This has made water managers more comfortable using forecasts to inform longer-term release decisions meant to balance water supply, flood control, and other system objectives. The promise of these forecast-based operating policies is poised to grow further as artificial intelligence accelerates forecasting system improvements (Khatun et al. 2023; Lam et al. 2022).

However, the value of forecasts to decision-making does not necessarily increase proportionally with forecast skill. There is a long literature exploring the sensitivity of forecast value to forecast attributes (e.g., forecast skill and lead-time), most often in single objective contexts (Anghileri et al. 2016; Faber and Stedinger 2001; Hamlet et al. 2002; Turner et al. 2017) but also in multiobjective settings (Denaro et al. 2017; Doering et al. 2021; Nayak et al. 2018; Yang et al. 2021; Yao and Georgakakos 2001). This work has highlighted that forecast value is dependent on multiple factors, including forecast uncertainty, lead-time, management regime, operating objective(s), and characteristics of the basin and reservoirs themselves. While more accurate forecasts typically improve forecast value for decision-making, uncertain forecasts can still prove useful (Zhao et al. 2011; Zhao and Zhao 2014), and the relationship between improved skill and value is not guaranteed to be linear or even monotonic.

In a separate line of work, there is a growing focus on identifying water resource planning and management strategies that can effectively navigate long-term (i.e., decadal) uncertainty in future conditions (Christensen et al. 2004; Culley et al. 2016; Fayaz et al. 2020; Herman and Giuliani 2018; Mereu et al. 2016). One approach is to identify static adaptation strategies that are robust (i.e., provide satisfactory performance) across a range of future

[1]Dept. of Biological and Environmental Engineering, Cornell Univ., Ithaca, NY 14853 (corresponding author). ORCID: https://orcid.org/0000-0002-0976-0140. Email: kts48@cornell.edu

[2]Associate Professor, Dept. of Biological and Environmental Engineering, Cornell Univ., Ithaca, NY 14853.

Note. This manuscript was submitted on March 31, 2023; approved on December 3, 2023; published online on February 23, 2024. Discussion period open until July 23, 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.

conditions (sometimes referred to as future states-of-the-world; Herman et al. 2015). However, finding robust, static policies is often difficult, particularly as the range of plausible futures grows (Culley et al. 2016). In response, an alternative approach is to identify dynamic adaptation strategies that trigger changes in the system (either operational or structural) based on the particular trajectory of future conditions that unfolds over time (Herman et al. 2020). This approach has been shown to hold significant promise (Arango-Aramburo et al. 2019; Cradock-Henry et al. 2020; Haasnoot et al. 2020; Hall et al. 2019; Trindade et al. 2019), although there are limits to the types of future change that dynamic operational adjustments can address (Fayaz et al. 2020; Herman and Giuliani 2018).

Despite the recent proliferation of work in both forecast-informed reservoir operations and long-term water resources planning under uncertainty, there has to date been only limited work exploring how forecast informed operations can itself serve as a long-term adaptation strategy to change, despite recent advocacy for this approach (Wilson et al. 2021). Even a static operating policy that is forecast-informed can in some sense be considered dynamic, as the forecasts (and therefore operating decisions) will change in character over time in response to long-term changes in the underlying hydroclimate system. For example, if a system experiences significantly wetter conditions over several decades, and short-term forecasts are able to reflect that change via more frequent forecasts of above-average water supply conditions, the frequency of flood control-oriented operations will increase even without changing the underlying structure of the policy.

The concept of forecasts as a climate change adaptation strategy was explored in Steinschneider and Brown (2012), who showed that dynamic reservoir operations based on seasonal water supply forecasts, coupled with real-options risk hedging, was able to significantly improve water supply reservoir operations under uncertain nonstationary hydrology. More recently, in an unpublished thesis, Goulart (2019) used synthetic short-term (three-day) streamflow forecasts and policy optimization to show that a single policy tailored for forecast use significantly improved water supply operations regardless of the trajectory of future climate, although the degree of improvement over a baseline, no-forecast policy did depend on the nature of the future climate scenario. The authors also showed that forecast-informed policies only trained to historical data could improve water supply performance in the future, albeit with a potential increase in flood risk. Cohen et al. (2020) explored the use of probabilistic seasonal forecasts (including perfect forecasts) of water year type and snowpack-to-streamflow estimates as an adaptation strategy in California. The authors found that perfect forecasts significantly improved water supply performance under a wide range of future climate scenarios, and some of these benefits could be realized using probabilistic forecasts dynamically trained through time to each scenario.

These studies show the potential for forecast-informed operations as an adaptation strategy to climate change, including the robustness that forecast-based operations afford water systems under deep uncertainty in future conditions. The studies also show how the viability of forecast-based operations as a climate change adaptation strategy depends on the future states of the world to which they are applied and the skill of the forecasts. However, in all cases, the lead time of the forecasts were assumed a priori rather than selected based on their contribution to performance in uncertain future conditions. In addition, most of the previous work has focused on systems with one or two objectives (Steinschneider and Brown 2012; Goulart 2019), and work that considered more than two objectives used preselected adaptations without policy optimization (Cohen et al. 2020). The value of forecast information as an adaptation to future hydroclimate conditions in systems with many competing objectives remains underexplored, particularly when complex policies must be optimized to accommodate different forecast attributes. Similarly, while past work has mapped future hydrologic states (long-term wetting or drying) to forecast value (Goulart 2019; Cohen et al. 2020), these analyses have not considered how this mapping depends on system objectives or the spatiotemporal characteristics of those future hydrologic states (i.e., where and in which seasons are hydrologic trends strongest).

The present study seeks to build on the previously cited work to improve our understanding of how forecast-informed operating policies are able to navigate shifting hydrologic regimes in multiobjective water systems and how the viability of this approach varies with forecast attribute (lead-time and skill) and type of future condition. We explore these questions in a case study of the Lake Ontario–St. Lawrence River (LOSLR) system, the levels and flows of which are controlled by operations at the Moses Saunders Dam on the St. Lawrence River. Semmendinger et al. (2022) found that the current control policy of the Moses Saunders Dam is limited in its ability to leverage forecast improvements to enhance multiobjective performance. Here, we build off the work in Semmendinger et al. (2022) by employing many-objective optimization to adjust the control policy using forecasts of different lead-times and skill levels. We seek to identify alternative policies that can better leverage forecast information and balance trade-offs between six different system objectives. We also identify a subset of historically optimal control policies, informed by forecasts of varying lead-time and skill, and reevaluate how these control policies meet minimum thresholds of performance under alternative stationary and nonstationary hydrologic traces (states-of-the-world) to quantify policy robustness under future uncertainty. We then employ scenario discovery (Groves and Lempert 2007) to identify complex relationships between system performance for different objectives, forecast attribute, and future hydrologic conditions across seasons and locations in the system. The results of this work, developed in a researcher–practitioner partnership (Badham et al. 2019; Syme and Sadler 1994) with decision-makers in the LOSLR system, are designed to directly support an ongoing review of the current operating policy and the formulation and evaluation of new candidate operating policies for the Moses Saunders Dam.

Lake Ontario–St. Lawrence River System

Outflows from Lake Ontario are regulated at the Moses Saunders Dam on the St. Lawrence River (Fig. 1) by the International Lake Ontario–St. Lawrence River Board (Board of Control) under the purview of the International Joint Commission (IJC). Lake Ontario is the last lake in the series of hydrologically connected Great Lakes, and the St. Lawrence River serves as the outlet of the Great Lakes to the Atlantic Ocean. Outflows from the Upper Great Lakes (via Lake Erie) are the primary inflow into Lake Ontario, in addition to inputs from local tributaries and over-lake precipitation and losses from over-lake evaporation. Lake Ontario has a surface area of approximately 19,010 km^2, a watershed area of 64,025 km^2, and a volume of approximately 1,646 km^3. Water levels vary seasonally, rising in the spring and summer with increased snowmelt and runoff and reduced over-lake evaporation and falling in the autumn as over-lake evaporation rises due to temperature differences between colder air and warmer waters.

The IJC was established under the Boundary Waters Treaty of 1909 to manage the shared waterways between Canada and the United States. Flow regulation on Lake Ontario began in the 1950s when the Moses Saunders Dam was constructed as part of the St. Lawrence Seaway project. The first control policies at the

Fig. 1. Map of the Lake Ontario–St. Lawrence River (LOSLR) water system, with key locations within the baseline highlighted. The Moses Saunders Dam is shown on the St. Lawrence River as the triangle. (Sources: Esri, TomTom, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors and the GIS User Community.)

dam, including Plan 1958-A, Plan 1958-C, and Plan 1958-D, aimed to minimize navigation costs while maximizing hydropower production and providing reasonable protection for coastal riparians as required by the Boundary Waters Treaty of 1909, 1952 Order of Approval, and 1956 Supplemental Order of Approval. Following the implementation of Plan 1958-D in 1963, the basin experienced prolonged periods of low water supplies and water levels. To maintain performance for system interests, like commercial navigation, the Board of Control discretionarily deviated from Plan 1958-D, later referred to as Plan 1958-DD (Plan 1958D with deviations).

The Plan 1958-DD regime continued until 2017 when a new regulation plan, Plan 2014, was implemented along with the 2016 Supplemental Order of Approval (International Joint Commission 2016). Plan 2014 was developed with the goal of reintroducing more variable water levels to rehabilitate wetland health and services while still providing the same level of benefits for commercial navigation and hydropower interests (International Joint Commission 2014). Despite efforts to balance impacts to system stakeholders, a trade-off remains between more variable lake levels for wetland health and more confined lake levels for coastal riparians. A few months after the implementation of Plan 2014, the Upper Great Lakes and LOSLR basins experienced prolonged periods of unprecedented water supplies. Lake levels on Lake Ontario rose rapidly, resulting in the worst flood event on record (International Lake Ontario–St. Lawrence River Board 2018). The 2017 flood record was broken just two years later in 2019. The recent floods prompted the review of Plan 2014 and other candidate plan alternatives. We leverage a partnership with the basin's scientific advisory board [the Great Lakes Adaptative Management (GLAM) Committee] to develop findings that will support the expedited review of Plan 2014.

Data and Methods

This analysis has four major components: forecast generation; policy optimization; candidate policy selection; and scenario discovery (Fig. 2). Water supply forecasts are created at four lead times (1, 3, 6, and 12 month) and for two levels of skill (baseline and perfect) over the historic water supply sequence [Fig. 2(a)]. These forecasts are fed into a many-objective optimization algorithm to discover policies that approximate the Pareto frontier for each combination of forecast lead-time and skill [Fig. 2(b)]. The resulting policies are subset using a set of satisficing criteria (i.e., minimal thresholds for performance across objectives) and stakeholder input on policy preferences [Fig. 2(c)]. This results in a small set of candidate policies, which are reevaluated on out-of-sample water supply conditions and associated forecast information that reflect plausible future hydrologic scenarios. Policy performance across these scenarios is used to define policy robustness, and scenario discovery is performed to diagnose what exogeneous hydrologic variables most often cause system failures for different cases of forecast attribute [Fig. 2(d)]. These four components define a generalizable framework to assess the impact of forecasts on water system robustness under long-term hydrologic variability and change and can be applied to any water system for selected forecast lead times and skill levels. In the following sections, we provide details on each of these components in the context of the LOSLR system, with specific methodological decisions tailored to support the ongoing expedited review of the system's current operational policy.

Forecast Generation

Regulation plans for the Moses Saunders Dam have traditionally been developed and tested using simulation analysis over the

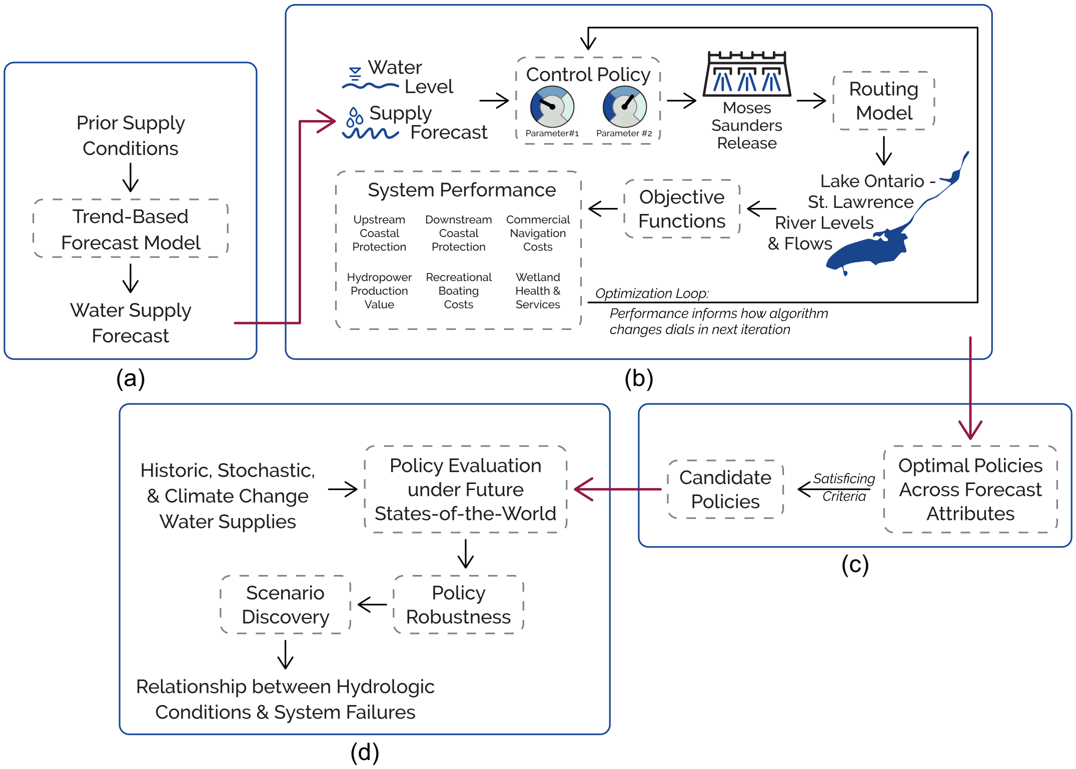

Fig. 2. Workflow of analysis: (a) forecast generation; (b) policy optimization; (c) candidate policy selection; and (d) scenario discovery.

historical record of water supplies, which in this study runs from 1900 through 2020. All data are simulated at a quarter-monthly (near-weekly) timescale, whereby there are 48 quarter-months in a year. All water supply data (described here) are gathered from the GLAM Committee.

The current management policy, Plan 2014, utilizes a trend-based net total supply (NTS) forecast into Lake Ontario at a 12-month lead-time to inform release decisions (D. Lee, *Deterministic forecasts for Lake Ontario plan formulation*, unpublished report). The NTS is the sum of flows into Lake Ontario from the Upper Great Lakes via the Niagara River and the Welland Canal and the net basin supply (NBS) of Lake Ontario, the latter which is defined as the sum of over-lake precipitation and runoff into the lake minus over-lake evaporation. For Lake Ontario, outflows from Lake Erie drive the NTS term. Under Plan 2014, the annual average NTS (i.e., the average NTS over the next year) is forecasted using a first-order autoregressive (AR1) time-series model [Eq. (1)]. For any given quarter-month, the previous 48 quarter-months (i.e., 12 months) of NTS are averaged [Eq. (1*a*)] and a Box-Cox transformation is applied to the average [Eq. (1*b*)]. The transformed average is input into an AR1 forecast model [Eq. (1*c*)], and the output is backtransformed [Eq. (1*d*)] to obtain the forecasted annual average NTS for the next 48 quarter-months

$$\mathrm{NTS}_p = \mathrm{avg}(\mathrm{NTS}_{t-49}\mathrm{:NTS}_{t-1}) \qquad (1a)$$

$$\mathrm{NTS}_{p,bc} = \mathrm{Box-Cox\ transform}\,(\mathrm{NTS}_p) \qquad (1b)$$

$$\mathrm{NTS}_{f,bc} = \mathrm{AR}_1(\mathrm{NTS}_{p,bc}) \qquad (1c)$$

$$\mathrm{NTS}_f = \mathrm{Box-Cox\ backtransform}\,(\mathrm{NTS}_{f,bc}) \qquad (1d)$$

The parameters of the Box-Cox transformation (lambda) and AR1 forecast model (first-order autocorrelation coefficient and shift) in Plan 2014 were calibrated using the historical data from 1900 through 2000. In this study, we use a similar trend-based forecast structure to create new forecasts at the 1-, 3-, 6-, and 12-month lead-times. For example, rather than averaging the previous 48 quarter-months of water supplies [Eq. (1*a*)] for the 12-month lead time, we average the previous 24 quarter-months of water supplies for the 6-month lead time. For each forecast lead-time, we calculate the rolling-average and then average the quarter-monthly values across all years. To remove the effects of seasonality, we standardize the rolling average by subtracting off this average and dividing by the standard deviation of the rolling average across all years. The standardized rolling averages are normally distributed, which negates the need to impose a Box-Cox transformation. We refit the parameters in the AR1 forecast model for each lead-time using updated data from 1900 through 2020, and all forecasts are unstandardized before use. The updated AR1 parameters result in slightly different forecasts at the 12-month lead-time when compared with the embedded Plan 2014 12-month forecast model. The results from forecast model training and testing as well as a comparison with the Plan 2014 forecast model are shown in Supporting Information S1. For each forecast lead-time, we create forecasts using the baseline trend-model skill (hereafter status quo skill) as well as perfect skill (i.e., perfect insight into future conditions). This results in a total of eight unique forecast combinations across four lead-times and two levels of skill developed for the historical record.

Policy Optimization

Plan 2014 conditions release decisions on estimated preproject outflows (i.e., outflows that would have occurred if the dam were not there) with adjustments made for forecasted water supplies via a sliding rule curve function (International Joint Commission 2014).

J. Water Resour. Plann. Manage., 2024, 150(5): 04024009

Releases are determined on a quarter-monthly (i.e., near weekly) time scale. Eq. (2) shows the preproject release term as a function of lake level:

$$\text{preproject release} = 555.823 \times (\text{level}_{\text{ontario}} - \text{adj} - 69.474)^{1.5} \tag{2}$$

where $\text{level}_{\text{ontario}}$ = the water level on Lake Ontario at the end of the previous quarter-month; and adj = an adjustment factor for differential crustal movements. Releases are then prescribed as a function of the preproject term and the water supply forecast in a sliding rule curve composed of two release regimes for above and below average supply conditions [Eq. (3)]

$$\text{release} = \begin{cases} \text{pre project release} + \left[\dfrac{\text{NTS}_f - \text{NTS}_{h,\text{avg}}}{\text{NTS}_{h,\max} - \text{NTS}_{h,\text{avg}}}\right]^{P_1} \times C_1, & \text{NTS}_f \geq \text{NTS}_{h,\text{avg}} \\ \text{pre project release} - \left[\dfrac{\text{NTS}_{h,\text{avg}} - \text{NTS}_f}{\text{NTS}_{h,\text{avg}} - \text{NTS}_{h,\min}}\right]^{P_2} \times C_2, & \text{NTS}_f < \text{NTS}_{h,\text{avg}} \end{cases} \tag{3}$$

where NTS_f = the forecasted annual average NTS [calculated in Eq. (1)]; $\text{NTS}_{h,\max}$ = the historical maximum NTS; $\text{NTS}_{h,\min}$ = the historical minimum NTS; and $\text{NTS}_{h,\text{avg}}$ = the historical average NTS and the threshold that designates which regime to follow. The historical values in the control policy are calculated from the period record from 1900 through 2000.

The multipliers (C_1 and C_2) and exponents (P_1 and P_2) in Eq. (3) are sets of constants, some of which are determined by comparing the forecasted supply to a threshold of wet conditions (T_w) and forecast confidence (CI_{99}):

$$C_1 = \begin{cases} C_{1w}, & (\text{NTS}_f - CI_{99}) \geq T_w \\ C_{1m}, & \text{otherwise} \end{cases} \tag{4}$$

Eq. (4) shows that, when there is high confidence in wet basin conditions, releases increase by setting C_1 to C_{1w}. Additionally, during extremely dry conditions (designated when water levels fall below some threshold, L_d), the rule curve release in Eq. (3) is further reduced by F_d

$$\text{release} = \text{release} - F_d \quad \text{if level}_{\text{ontario}} < L_d \tag{5}$$

When necessary, the rule curve release is adjusted to ensure it does not violate flow limits for various system constraints. For example, during ice formation, releases are limited to prevent an ice jam, and, during the navigation season, flows are limited to maintain safe velocities for transiting ships. If the rule curve release exceeds or falls below any of these limits, the release is set to whichever limit is most constraining. Therefore, at any given time, either releases follow the rule curve or the flow prescribed under the most limiting system constraint. In addition to these operational limits, Plan 2014 includes quarter-monthly thresholds for extreme high and low water levels, known as the H14 criteria (International Joint Commission 2016). When these thresholds are exceeded, the Board of Control has the authority to deviate from the Plan 2014 release regime (the rule curve and operational limit flows) to maintain system integrity. A more detailed description of the Plan 2014 operational limits can be found in the management plan documentation (International Joint Commission 2014, 2016).

All parameters of this control policy [coefficient and power parameters in Eq. (3)] were calibrated during the initial development of Plan 2014 in the 2000s to maximize multiobjective system performance (International Lake Ontario–St. Lawrence River Study Board 2006). Performance was measured using a large number of performance indicators that represent seven major objectives: coastal flood control upstream of the Moses Saunders Dam; flood control downstream of the dam; commercial navigation; hydropower production; municipal and industrial water supply; recreational boating; and wetland health and services. The parameters governing supply adjustments were calibrated using trial and error (as opposed to a formal optimization algorithm).

In this analysis, key parameters within the Plan 2014 control policy are formally optimized using the Borg multiobjective evolutionary algorithm (Hadka and Reed 2013). We optimize a total of 13 decision variables: the supply adjustments applied to the forecasted NTS [$\text{NTS}_{h,\max}$, $\text{NTS}_{h,\text{avg}}$, and $\text{NTS}_{h,\min}$ in Eq. (3)]; the threshold that determines what regime to follow [$\text{NTS}_{h,\text{avg}}$ in Eq. (3) but set as a separate parameter in the optimization], the rule curve coefficients and powers [C_{1m}, C_{1w}, C_2, P_1, and P_2 in Eq. (3)]; thresholds of basin conditions [T_w in Eq. (4)]; forecast confidence interval [CI_{99} in Eq. (4)]; and dry condition adjustments [L_d and F_d in Eq. (5)]. Importantly, we only explore alternative values for the parameters of the current rule curve but do not attempt to identify an alternative structure for the release rule. We also do not adjust the operational flow constraints in the optimization procedure. These design choices were made to (1) help communicate study results to system stakeholders, who are familiar with the current rule curve structure; and (2) ensure legal requirements that govern system management and are enforced in the current constraint set are respected.

For each combination of forecast lead-time ($n = 4$) and skill level ($n = 2$), we replace the NTS_f term in Eq. (3) with the respective forecast trace and allow Borg to optimize the 13 decision variables over the historic supply record (1900–2020). Based on initial convergence tests, we allow Borg to iterate over 100,000 function evaluations. At the ith iteration of the algorithm, Borg will identify N_i sets of policy parameters, each set defining a unique control policy for a particular forecast scenario. Each unique control policy can be used to simulate a quarter-monthly time-series of flows over the historical period (1900–2020), which are then routed through the system to calculate water levels along Lake Ontario and the St. Lawrence River. Objective functions, described in more detail here, relate these flows and levels to system performance for different stakeholder groups. Borg modifies the decision variables in the next function evaluation based on the objective values from the current evaluation, with the goal of identifying new policies that expand the Pareto front of possible solutions at each iteration. For each of the eight forecast scenarios, we allow Borg to optimize policies across five random seeds and then pool nondominated polices across those seeds to represent the final, Pareto-approximate set of policies for a given forecast lead time and skill level.

We focus on six objectives in this analysis: flood risk upstream of the Moses Saunders Dam; flood risk downstream of the dam; commercial navigation; hydropower production; recreational boating; and wetland health and services. We do not consider the objective of municipal and industrial water supply, as this objective is largely insensitive to system operations. Models for the six selected objectives were adapted from those developed by the GLAM scientific advisory board. Flood risks upstream and downstream of the dam are defined based on the number of homes inundated at different water levels on Lake Ontario and the St. Lawrence River, which were compiled in a recently developed decision-support tool produced by the GLAM Committee (Great Lakes–St. Lawrence River Adaptive Management Committee 2021). Models for commercial navigation costs, value of hydropower production, and recreational boating costs are based on the impact models developed in the evaluation of Plan 2014 before implementation, and which were updated in 2017–2018 with stakeholder feedback. Commercial navigation and recreational boating costs are primarily a function of time of year and water levels on Lake Ontario and the St. Lawrence River, while hydropower production is primarily a function of Lake Ontario water levels and releases from the Moses Saunders Dam. Ecosystem health is based on the area of meadow marsh, which is calculated following the methodology described in Wilcox and Xie (2007) and Wilcox and Bateman (2018). The presence of meadow marsh area is dependent on water levels fluctuating at certain times of the year (e.g., the growing season) to flood and dewater certain areas of shoreline to promote biodiversity. More details on the models used for each objective are presented in Supporting Information S2.

Candidate Policy Selection

Results from the policy optimization provide a large set of mathematically optimal policies that maximize performance for each system objective or provide a nondominated trade-off in performance across objectives. To help select a small set of candidate policies for further analysis, we solicited feedback from decision-makers in the LOSLR region to identify minimal thresholds of performance across the six system objectives. The minimal thresholds of performance, or satisficing criteria, indicate whether a plan is realistically implementable. These minimal thresholds of performance are meant to only serve as a starting point to filter policies and discover trade-offs among system objectives in new candidate plans. The satisficing criteria are informed by the Boundary Waters Treaty of 1909, 2016 Supplemental Order of Approval (International Joint Commission 2016), and discussions with Board of Control members, representing various system interests. Performance under Plan 2014 is considered as a baseline, and the satisficing criteria are expressed as percent improvements or reductions from Plan 2014 performance. The following criteria were decided upon for a control policy to be considered a candidate plan:

1. No worse than Plan 2014 performance for upstream riparian protection (0% reduction);
2. No worse than Plan 2014 performance for downstream riparian protection (0% reduction);
3. No worse than Plan 2014 performance for commercial navigation (0% reduction);
4. Slight reduction from Plan 2014 performance for hydropower production (0.5% reduction);
5. Slight reduction from Plan 2014 performance for meadow marsh (5% reduction); and
6. Moderate reduction in from Plan 2014 performance for recreational boating (20% reduction).

We note that Plan 2014 increased the value of hydropower from the previous management plan (Plan 1958DD), so a 0.5% reduction in the value of hydropower from Plan 2014 is equivalent to a return to Plan 1958DD in this objective (and hence viewed as satisficing).

The Pareto approximate policies for all eight forecasting scenarios were pooled and screened using these satisficing criteria, thereby determining which combinations of forecast lead-time and skill contribute the most satisficing policies under the historical record. Policies from each forecast scenario that meet the satisficing criteria were further screened in deliberation with our partners in the GLAM Committee based on stakeholder preference in order to select a handful of final candidate policies for further in-depth analysis.

Robustness and Scenario Discovery

The final candidate policies are reevaluated using the historic supply trace from 1900 to 2020 (for which they were optimized), a stochastically generated set of 500 centuries of water supply data based on historical statistics from 1900 to 2000 (Fagherazzi et al. 2007), and a climate-change-driven supply data set composed of 159 centuries of plausible future supply traces (Steinschneider 2022). The stochastic traces were generated using a multivariate contemporaneous mix of shifting mean and ARMA processes fit to NBS for the 1900–2000 reference period (Sveinsson and Salas 2006), coupled with space-time disaggregation procedures (Mejia and Rousselle 1976; Stedinger et al. 1984). These stochastic traces provide plausible realizations of natural hydroclimate variability assuming stationarity, which include wetter and drier periods that were not experienced during the 1900–2000 reference period. We note that the period between 2001 and 2020 was relatively wet compared with the 1900–2000 reference period, so the stochastic climate traces are drier on average compared with the historic supply trace from 1900 to 2020. The climate-change driven supply traces were developed using long short-term memory (LSTM) artificial recurrent neural networks that use input time-series of quarter-monthly precipitation and average temperature to predict NBS into each of the five Great Lakes, Ottawa River flows, and ice conditions on the St. Lawrence River using midcentury (2036–2065) projections of precipitation and temperature across the entire Great Lakes basin. These data were derived from the recently released CMIP6 climate model projection database (Eyring et al. 2016). The projections, taken from 46 separate GCMs and four different emission scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5), were downscaled to the Great Lakes Basin using the delta method (note that not all GCMs have projections for each emission scenario). The historical precipitation between 1952 and 2019 were adjusted on a monthly basis to reflect change factors derived for each combination of GCM and emission scenario. Each century, or trace, from the three data sets is referred to as a distinct state-of-the-world, for a total of 660 states-of-the-world in the reevaluation (one historic trace, 500 stochastic traces, and 159 climate-change traces). In each state-of-the-world, forecasts at a selected lead-time and skill level are generated, flows and water levels are simulated using the candidate policy, and all objectives are reestimated to quantify system performance under that particular century of supply conditions.

For each candidate policy, we identify the future states-of-the-world for which the satisficing criteria are met, which are based on percent changes in objective function values relative to the baseline performance of Plan 2014 (see Candidate Policy Selection). Importantly, we define the satisficing criteria conditional on Plan 2014 performance separately for each state-of-the-world, rather than Plan 2014 performance under the historical record. Therefore, a candidate policy can be satisficing for a particular state-of-the-world even if its absolute objective scores are poor, as long as they compare favorably with the score of Plan 2014 under that state-of-the-world. We then define the robustness of a given policy as the fraction of states-of-the-world in which the satisficing criteria are met (Hadjimichael et al. 2020).

Scenario discovery (Bryant and Lempert 2010; Groves and Lempert 2007) is then performed to help explain why a given candidate policy does or does not meet the satisficing criteria for different states-of-the-world. Eleven variables are used to represent distinct hydrologic features for each state-of-the-world, as shown in Table S1. We use cumulative annual NTS and Ottawa River flows into the St. Lawrence River to represent average water supply conditions upstream and downstream of the dam, respectively. We also consider seasonal versions of these variables (December–February, March–May, June–August, September–November). Last, we consider the average ice condition on the St. Lawrence River, where higher values indicate a higher frequency of unstable ice conditions (or ice formation periods) that are highly sensitive to air temperature and can constrain releases from the Moses Saunders Dam. We use gradient boosted trees (Schapire 2009) to develop mapping among these 11 hydrologic variables and the occurrence/nonoccurrence of satisficing performance for each state-of-the-world. The parameters of the gradient boosted trees were calibrated using a grid search with average precision scoring as the evaluation metric. The gradient boosted trees return relative feature importance (Pedregosa et al. 2011) to identify which hydrologic variables are most useful in predicting whether a candidate policy will provide satisficing performance under a given future state-of-the-world. Ultimately, this analysis helps to determine which hydrologic features of a particular future cause a candidate plan to underperform the current status-quo regulation plan.

Results

A total of 4,034 policies were identified across the eight separate optimization runs for each forecast lead-time and skill level, with 734, 472, 265, and 435 (697, 578, 466, and 387) policies identified for the 12-, 6-, 3-, and 1-month status quo (perfect) forecast scenarios. Among this larger set, 257 control policies met the satisficing criteria under historical water supply conditions. Fig. 3 shows the breakdown of the 257 satisficing policies by forecast lead-time and skill level as well as the entire set of nondominated policies pooled across all forecasting scenarios (gray lines). Within each lead-time panel, policies that meet the satisficing criteria are colored, and the line type is drawn to the forecast skill, where light blue lines correspond to status quo skill and dark blue lines correspond to a perfect forecast. Policies that are part of the Pareto front but do not meet the satisficing criteria are colored in gray. Plan 2014 performance is shown as the solid red line. The 12-month lead-time contributes the most satisficing policies (205), followed by the 6-month (25), 3-month (23), and 1-month (4) lead-times, respectively. At longer lead-times, both levels of forecast skill contribute satisficing policies, with 100 (105) satisficing policies using perfect (status-quo) forecasts for the 12-month lead time and 13 (12) policies for the 6-month lead. However, at shorter lead-times (1- and 3-month), the ability of policies to meet the satisficing criteria is heavily dependent on forecast skill. At the 3-month lead-time, the optimization algorithm discovers 22 satisficing policies using perfect forecasts and only one satisficing policy using status quo forecasts. At the 1-month lead-time, the optimization algorithm discovers four satisficing policies using perfect forecasts and no satisficing policies using status quo forecasts.

Over the historic hydrological trace, similar trade-offs among system objectives persist across forecast lead-time and skill level, although to varying degrees. The largest trade-offs that emerge are among the benefits for coastal riparians, commercial navigation, and recreational boating as well as among wetland health, commercial navigation, and recreational boating. The optimization algorithm was not able to discover any policies that outperformed Plan 2014 across all six system objectives. The coastal riparian objectives, both upstream and downstream, appear more sensitive to forecast skill. On average, policies using perfect forecasts outperform those using status quo forecasts across forecast lead-time for most objectives, although this is not the case for hydropower production and recreational boating. Policies conditioned on perfect forecasts do not result in improved performance for every objective due to the structure of the Plan 2014 control policy that we optimize. The Plan 2014 control policy was developed for a status quo forecast, and even when policy parameters are optimized for a new forecast with greater accuracy, the current structure of the control policy is not able to leverage these improved forecasts to improve outcomes across all objectives. This point is revisited in more detail later in this text.

Before selecting individual policies for additional analysis, we first assess whether the larger set of satisficing policies under historical supply conditions remains satisficing under alternative water supply conditions. Here, each policy that meets the satisficing criteria for the historic supply trace (one state-of-the-world) is reevaluated for the stochastically generated and climate-change supply traces (659 additional states-of-the-world). By conducting this analysis for all satisficing policies identified for the historical trace (shown in Fig. 3), we ensure that insights remain generalizable and are not specific to individually selected policies. Fig. 4 shows the total number of historically satisficing policies that meet the satisficing criteria across each forecast lead-time and skill level in each state-of-the-world, where redder states-of-the-world are drier and bluer states-of-the-world are wetter. The number of satisficing policies is greatest for the historical supply trace (triangles in Fig. 4), as this was the trace used to initially define satisficing policies. The number of policies that remain satisficing for alternative states-of-the-world decreases below the number for the historical trace and can range from near the historical number of policies to zero policies. This suggests that the identification of satisficing policies is dependent on the water supply trace used to define satisficing performance, and the policies explored here are somewhat overfit to the historical trace to which they were optimized. However, this effect is less stark for policies informed by a 12-month lead forecast. For these policies, for which there were 205 that were satisficing under the historical water supply trace, between 20 and 50 policies remain satisficing for approximately 75% of the out-of-sample states-of-the-world tested, depending on the level of forecast skill. This decreases significantly for policies operated with shorter lead-time forecasts, such that the few satisficing policies under the historical trace are satisficing under most alternative states-of-the-world. Finally, we note that the number of satisficing policies across forecast lead-times and skill levels is generally independent of the long-term average hydrologic conditions (dry or wet) of a given state-of-the-world, as seen in Fig. 4, by the lack of stratification in the coloration of points for any scenario of forecast lead-time and skill level. The one exception is for the 3-month perfect forecast, where policies tend to perform better under wetter conditions.

The comparison of policies using status quo and perfect forecasts is helpful to establish the lower and upper bounds of

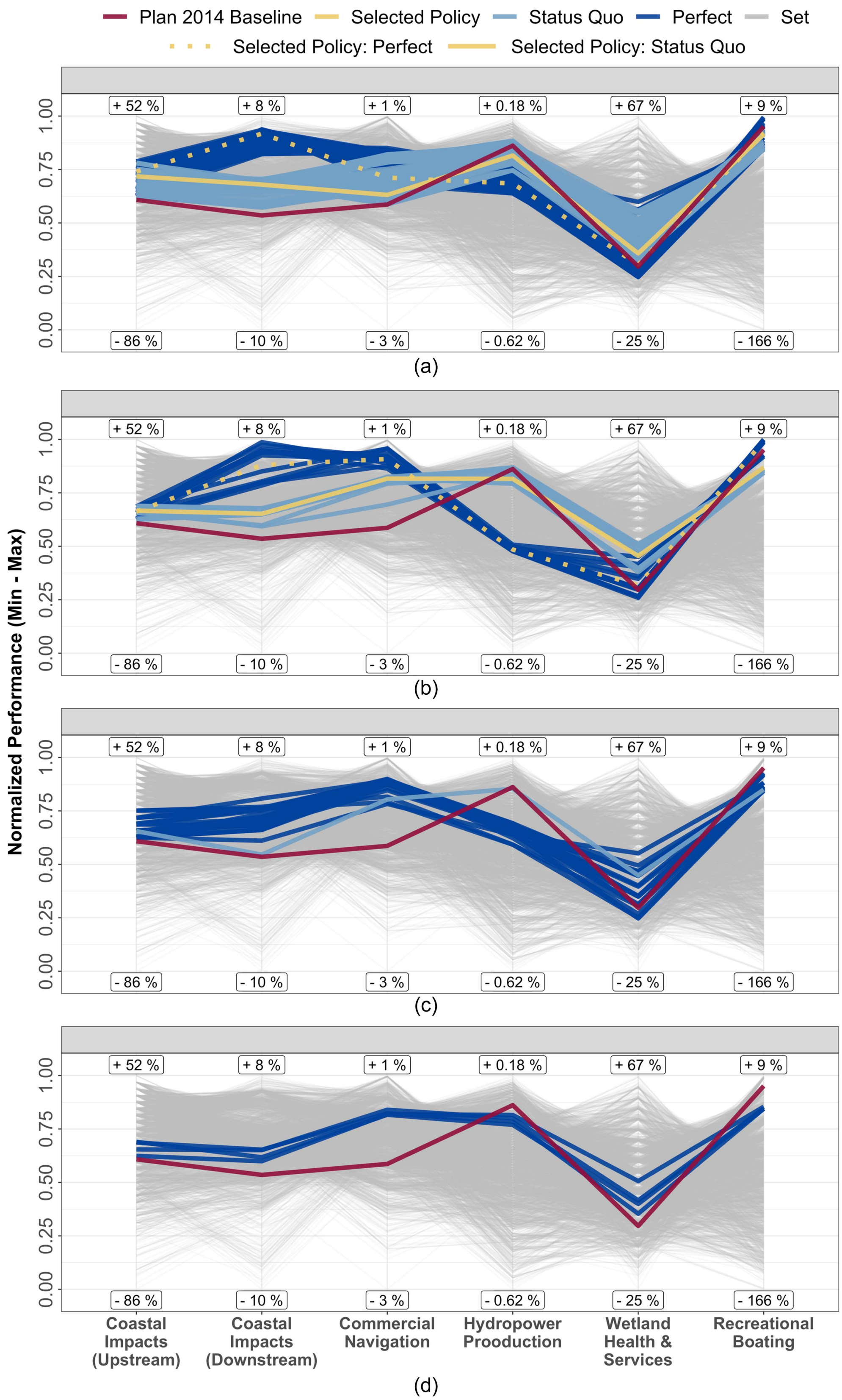

Fig. 3. Parallel axis plot of nondominated policies aggregated across policies optimized for: (a) 12-month; (b) 6-month; (c) 3-month; and (d) 1-month lead-time. System objectives shown on the x and y axes are normalized performance from 0 (worst performance) to 1 (best performance) for a given objective. A single line across all objectives represents a control policy, i.e., a specific combination of decision variables. Within each lead-time panel, policies that meet the satisficing criteria are highlighted, and the line type is drawn to the forecast skill. The minimum and maximum policy performance across all optimal policies for each objective are displayed as the percent difference from Plan 2014 performance, where positive values are equivalent to better performance and improvements over Plan 2014. Some forecast lead-times do not have satisficing policies across both forecast skill levels and, therefore, do not have both skill levels highlighted.

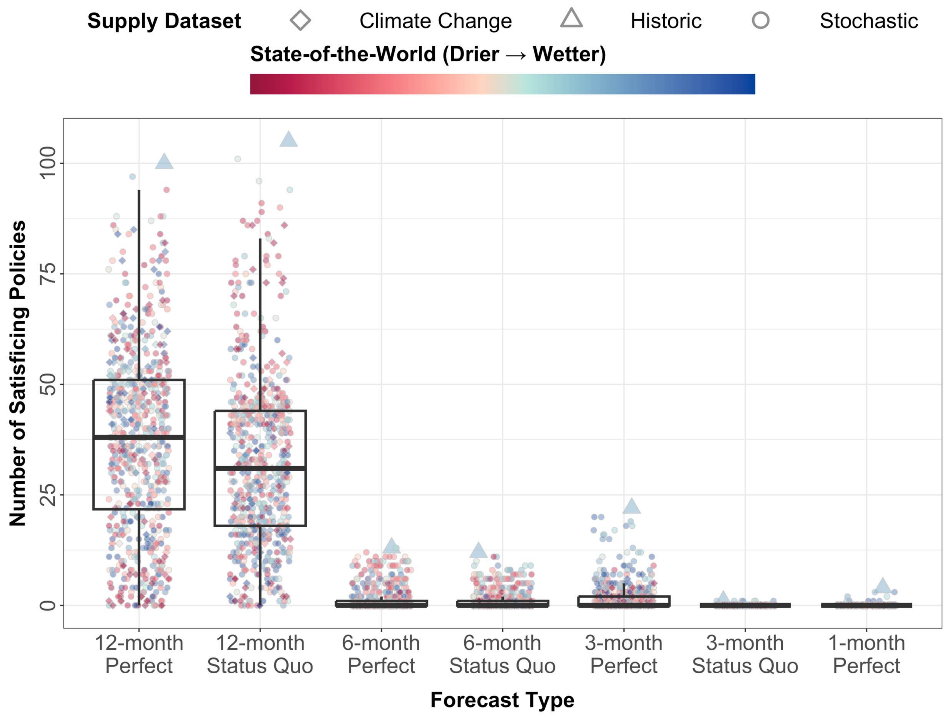

Fig. 4. The number of candidate policies that meet the satisficing criteria (y axis) across each combination of forecast type (x axis) for the 660 states-of-the-world. Individual states-of-the-world are shown within the boxplots and shaded according to their relative dryness/wetness (based on their long-term average NTS). The points are shaped according to the supply data set (historic supplies, climate-change supplies, stochastic supplies).

forecast-informed policy performance. In addition, results from Figs. 3 and 4 suggest that only policies using 12- and 6-month lead time forecasts are competitive with Plan 2014 for cases of perfect and status-quo forecast skill. Therefore, we proceed by selecting four final candidate policies from the 6- and 12-month lead-times for more in-depth analysis (one from each combination of forecast lead-time and skill level). Based on stakeholder feedback, we selected as our final candidates those policies that exhibited the largest improvements in coastal riparian protection (particularly upstream on Lake Ontario) under the historical water supply trace, while still minimizing reductions in performance for other objectives. In particular, we select final candidate policies that balance improvements for upstream riparians without major reductions in recreational boating, which is one of the starkest trade-offs that emerge from the full set of satisficing policies. These selected policies are colored yellow in Fig. 3.

Fig. 5 shows whether each of the four selected candidate policies are satisficing for all states-of-the-world, which have been ordered based on their average water supply conditions. The policies that are satisficing for a particular state-of-the-world and objective are colored blue, while the policies that fail to meet the satisficing criteria are colored red. Policies that are not satisficing can fail to meet the satisficing criteria in multiple ways and to varying degrees. Therefore, for each policy, we show in varying shades of red the degree to which satisficing criteria are not met for individual objectives. Failures are normalized from zero to one across each system objective, where a failure score of one is the most severe failure. The number of objectives that fail to meet the satisficing criteria for each state-of-the-world are colored white, gray, yellow, orange, and red and correspond to 0, 1, 2, 3, and 4 objective failures, respectively.

Fig. 5 shows that satisficing behavior varies by objective and candidate policy. The selected 12-month status quo policy meets the satisficing criteria in 536 out of 660 states-of-the-world, i.e., a robustness score of ∼80% [Fig. 5(a)]. Of the remaining 124 states-of-the world where the satisficing criteria are not met, a large majority (112) of them exhibit failures in only one objective (the wetland index). These failures tend to occur in drier states-of-the-world. The 12-month status quo policy fails in 10 states-of-the-world for two objectives and two states-of-the-world for three objectives. These failures are spread across upstream coastal flooding, downstream coastal flooding, commercial navigation, and recreational boating. There are no failures in hydropower production.

The selected 12-month perfect policy meets the satisficing criteria in 267 out of 660 states-of-the-world, i.e., a robustness score of ∼40% [Fig. 5(b)]. Notably, robustness for this perfect-forecast policy is less than that for the policy designed with noisier (status-quo) forecasts. Similar to the 12-month status-quo policy, the 12-month perfect policy tends to fail in only the wetland objective (373 states-of-the world), although failures to meet minimum performance for the wetland indicator do not appear more frequent in dry states-of-the-world. There are an additional 17 states-of-the world that fail across two objectives, which is primarily due to failures in upstream coastal flooding under dry conditions. In the driest states-of-the-world, failures can occur across three objectives (wetland indicator, upstream flooding, and commercial navigation).

The candidate policies that utilize forecasts at the 6-month lead-time are significantly less robust than 12-month forecast policies, with robustness scores of ∼11% and ∼12% for the status-quo and perfect forecast cases, respectively [Figs. 5(c and d)]. For most states-of-the-world, there are at least one or two system objective failures. Across both levels of forecast skill, there are pronounced failures for upstream coastal riparians. The 6-month status quo policy is unable to meet minimum levels of performance for downstream coastal riparians across most states-of-the-world and upstream flooding and recreational boating in wetter states-of-the-world.

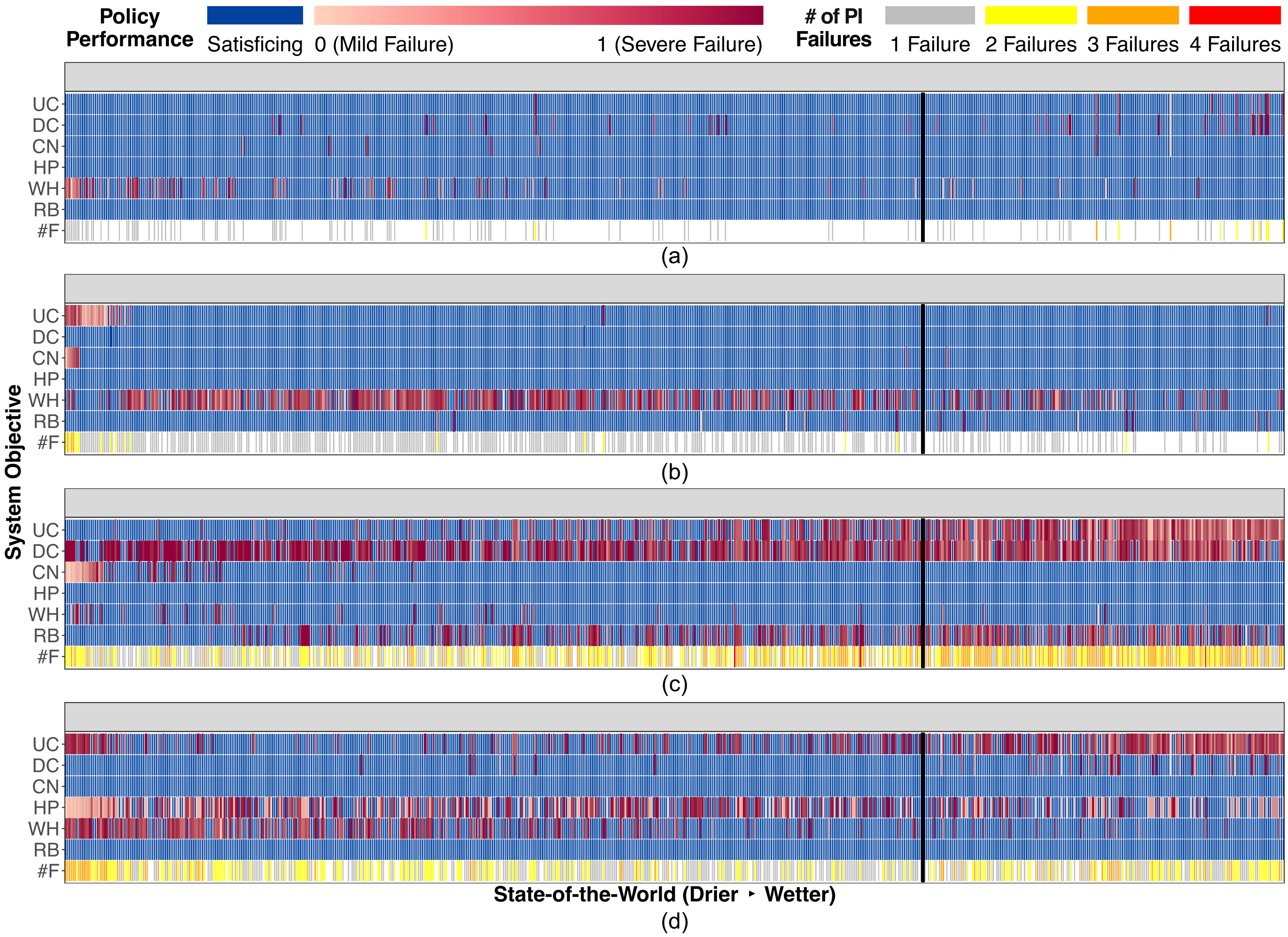

Fig. 5. System performance for individual objectives in plausible states-of-the-world across the selected (a) 12-month status quo; (b) 12-month perfect; (c) 6-month status quo; and (d) 6-month perfect candidate policies. Each tile shows whether or not the selected candidate policy meets the satisficing criteria for every combination of plausible state-of-the-world (x axis) and performance indicator (y axis). Tiles are filled according to the satisficing criteria. Failures to meet the satisficing criteria are shaded by the normalized severity of the failure. The y axis is arranged by system objective, where UC is upstream coastal impact, DC is downstream coastal impact, CN is commercial navigation, HP is hydropower production, WH is wetland health and services, and RB is recreational boating. The last element in the y axis, #F, is the number of objective failures for the associated state-of-the-world. The x axis is arranged from drier to wetter states-of-the-world. The historic state-of-the-world is outlined in black.

The 6-month perfect policy fails in hydropower production and upstream flooding across states-of-the-world and the wetland indicator in drier states-of-the-world.

The percentage of states-of-the-world that are satisficing for each policy varies depending on whether supply traces are derived from the stochastic or climate-change data sets and which objectives are being evaluated (Table 1). For some objectives (e.g., downstream coastal impacts, commercial navigation, recreational boating), the percentage of satisfying policies is relatively insensitive to the particular supply data set used across all four candidate policies. For other objectives, like hydropower production, the percentage of satisfying policies is insensitive to supply data set for three of the four candidate policies, but there are large differences for the 6-month perfect forecast policy. Differences between the stochastic and climate-change driven traces are most consistently observed across multiple policies for upstream coastal impacts and wetland health. For upstream coastal impacts, performance tends to be better under the stochastic supply traces for all policies; for wetland health, performance is better under the climate-change traces for all policies except the 6-month perfect forecast policy. Similar to Fig. 5, the percentage of satisfying policies for wetland health drops considerably for stochastic and climate-change based supply data sets when policies use perfect forecasts instead of status quo forecasts.

The equal or lower robustness of perfect forecast policies compared with status-quo forecast policies at both lead-times is somewhat unintuitive. However, these results are consistent with the findings in Semmendinger et al. (2022) and are likely related to complex interactions between forecast information and policy structure that can lead to overfitting. One example of this occurs during flood events on Lake Ontario. With perfect insight into future conditions, water levels are often drawn down well in advance of a flood event. However, as water levels on Lake Ontario decrease, the hydraulic head of the system also decreases and so does the preproject term in the rule curve. Thus, if the drawdown occurs too early before a flood, outflows will decline right as water supplies increase, and the drawdown will reverse before the flood event peaks. This can eliminate the flood control benefits of the original drawdown. During the historical trace (for which the system was optimized), this undesirable progression can be avoided through

Table 1. Percentage of satisficing traces for each selected policy, objective, and supply data set

Policy	Supply Data Set	UC	DC	CN	HP	WH	RB
12-month status quo	Stochastic	98.8	94.0	99.6	100	87.2	99.8
	Climate-change	96.2	94.3	96.9	100	90.6	100
12-month perfect	Stochastic	99.2	99.8	100	100	43.0	96.6
	Climate-change	82.4	98.1	93.7	100	57.2	100
6-month status quo	Stochastic	59.0	21.2	94.8	100	93.6	65.4
	Climate-change	41.5	19.5	88.1	100	95.6	44.7
6-month perfect	Stochastic	61.8	94.6	100	55.0	70.8	100
	Climate-change	47.8	90.6	100	0.0	57.9	100

Note: The stochastic and climate-change driven data sets contain 500 and 159 traces, respectively. The columns show the percentage of satisficing traces for each objective, including upstream coastal impacts (UC), downstream coastal impacts (DC), commercial navigation (CN), hydropower production (HP), wetland health and services (WH), and recreational boating (RB).

overfitting of the policy parameters. However, this may not be the case for out-of-sample hydrologic states-of-the-world to which the policy was not trained. For the policy trained to noisier (status-quo) forecasts, policy parameters cannot be as overfit for specific floods because the historical forecasts are wrong in different ways prior to different floods in the historical record. This then produced a policy that generalized better to new hydrologic traces.

To better understand why different candidate polices fail to meet the satisficing criteria under different states of the world, we employ scenario discovery on the four selected policies. Fig. 6 shows the results from the gradient boosted trees classification for two of these policies (6-month status quo and 12-month perfect). The points shown in Plot (ii) are shaded by if minimum performance thresholds are or are not met for the objective of interest, with darker shades indicating the severity of failure (as defined by a normalized deficit to the satisficing criterion). For the 6-month status quo policy, the flooding objective for upstream coastal riparians failed to meet minimum performance levels in 298 states-of-the-world [Fig. 6(a)]. The gradient boosted trees found downstream winter water supplies into the St. Lawrence River as the most influential hydrologic variable for this type of failure, followed by autumn downstream supplies [Fig. 6(a)]. Fig. 6(a)(ii) shows how failures in the upstream coastal flooding objective vary with these two hydrologic variables. In the stochastic and climate-change-based water supply data sets, there is a clear relationship between above-average combinations of autumn and winter downstream supplies and failures to meet minimum performance for upstream coastal riparians. This result highlights the complex interactions that occur spatially through the system, where, in this case, high water supplies downstream on the St. Lawrence River require lower

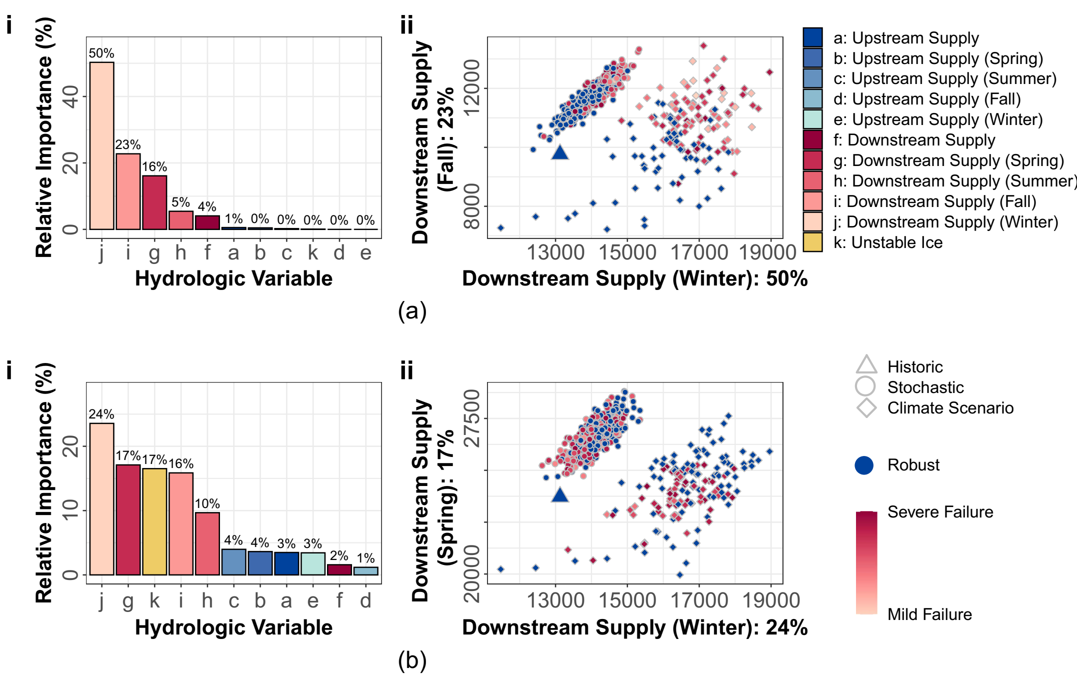

Fig. 6. Results from the gradient boosted tree classification mapping hydrologic state variables to satisficing or nonsatisficing performance for the (a) status quo 6-month policy and upstream coastal riparian objective; and (b) perfect 12-month policy and meadow marsh ecosystem objective. Factor maps for all combinations of forecast attributes and objectives are in the supplemental material. Plot (i) in each panel shows the influence (as percent importance) of each hydrologic variable. Plot (ii) shows each state-of-the-world plotted with its two most influential factors as the axes (most influential on x axis). Points are shaped according to the respective supply data set, with the historic record shown as the larger triangle. States-of-the-world are shaded if minimum performance thresholds are or are not met for the objective of interest.

releases from the Moses Saunders Dam to control downstream flooding. This can then lead to flooding challenges on Lake Ontario. The 6-month status quo candidate policy appears to struggle with this dynamic more than the original Plan 2014, leading to many states-of-the-world that fail the satisficing criteria for the upstream flooding objective.

A similar result is seen for the meadow marsh objective and the 12-month perfect forecast policy, although the relationship between hydrologic variables and objective failures is slightly more complex [Fig. 6(b)]. The meadow marsh objective, which is a proxy for wetland health, failed to meet minimum performance levels in 353 states-of-the-world. Here, lower downstream winter and spring water supply conditions tend to cause the 12-month perfect forecast policy to underperform Plan 2014 in the meadow marsh objective [Fig. 6(b)(i)]. However, there is not a clear threshold for failures as a function of supplies [i.e., red and blue points are not well separated; Fig. 6(b)(ii)], likely due to the complicated and nonlinear dynamics that drive the formation and preservation of meadow marsh. This finding demonstrates the complexity of predicting system objective failures based on a limited number of hydrologic measures. For some objectives, one might be able to identify simple mappings between hydrologic regimes and system vulnerabilities. However, that is not the case for every objective.

Conclusion

This study contributes an understanding of how forecast-informed reservoir operations can serve as a robust adaptation strategy to uncertain, future hydrologic conditions in many-objective water systems. We focused on the viability of this approach as a function of forecast attribute (lead-time and skill), type of future condition, and the interaction between these two factors. These issues were explored in a case study of the LOSLR basin, where water levels and flows in Lake Ontario and the St. Lawrence River are regulated by the operating policy of the Moses Saunders Dam. We used many-objective optimization to identify Pareto approximate control policies tailored for forecasts at four different subseasonal to seasonal lead times and two levels of skill, and then used a set of satisficing criteria developed with basin stakeholders and decision-makers to select a subset of promising control policies for further analysis. These promising candidates were then reevaluated under hundreds of plausible hydrologic regimes, including those that reflect stationary and nonstationary climate. Finally, we explored how features of future hydrologic conditions across space and season influence the satisficing nature of forecast-informed policies, as compared with a baseline policy.

Results showed that, for the LOSLR basin, policies conditioned on 12-month forecasts were more robust under the future states-of-the-worlds tested here than policies tailored for shorter lead-time forecasts, especially the 1- and 3-month lead-times. While forecast skill significantly affected the overall robustness of forecast-informed operations based on shorter lead-times, this was not the case for the 6- and 12-month forecast-based policies. In fact, policies tailored for noisier long-lead forecasts were more robust under a wide range of plausible futures as compared with those policies trained to perfect forecasts. This result highlights the potential to overfit control policies to historical information, even in the case of a forecast-informed policy with perfect foresight.

We note that the robustness of 12-month policies compared with other lead-times may be related to the mathematical formulation of the release rule, which was kept the same as in the baseline Plan 2014 control policy, which was originally formulated to use 12-month lead forecasts. This choice was made to help communicate the results of policy optimization to stakeholders who were familiar with the existing control policy but likely limited the identification of alternative policies better adapted to shorter-term (and more variable) forecast information. These findings support the need to explore alternative release functions through approaches like direct policy search (DPS; Quinn et al. 2017b), which can use global approximating functions (e.g., neural networks, radial basis functions) to more flexibly map exogenous and system state information to high-performing control decisions. We leave this effort for future work.

Findings from the scenario discovery analysis showed that, for some system objectives and policies (e.g., upstream coastal flood control in the 6-month status quo policy), there are clear relationships between performance and the hydrologic regime. A clear threshold of system performance as a function of supply conditions can directly support adaptive management of the system. If hydrologic conditions that trigger failures for specific objectives and control policies can be identified, system managers can carefully monitor those specific hydrological variables over time to determine if the system is approaching a state of vulnerability. This can then be used to dynamically trigger changes in the control policy to better manage emerging hydrologic conditions, although such action requires confidence that recent trends toward vulnerable climate states will continue and not revert to the historical mean state. This strategy is further complicated if failures in some system objectives cannot be easily predicted based on the hydrologic regime, as was the case for meadow marsh under the 12-month perfect forecast-based policy. Our results showed that complex relationships and dynamics between control decisions and system objectives can make it difficult to identify simple hydrologic indicators and thresholds that can serve as triggers for dynamic management. In these cases, alternative strategies may be needed for risk mitigation that extend beyond what large-scale water infrastructure can provide (e.g., land use management decisions that support more robust coastal wetlands regardless of water level regulation policy).

Results of this work are being used to support the expedited review of the current control policy of the Moses Saunders Dam, Plan 2014, and the exploration of alternatives by the GLAM Committee. Through the researcher–practitioner partnership that underscored the work presented in this study, we found disagreement around appropriate system objectives to be among the largest hurdles in formulating the problem explored in this work. As shown by Quinn et al. (2017a), uncertainty in the underlying formulation of system objectives can dramatically alter the interpretation of policy performance. Therefore, a major avenue of future work will be to work with system stakeholders to identify satisficing criteria and objective functions that best represent evolving interests across the LOSLR basin and, more broadly, to explore the sensitivity of forecast-informed operations as a climate change adaptation strategy to the underlying formulation of the many-objective control problem.

Data Availability Statement

All models and code generated or used during the study are available in an online repository at github.com/ksemmendinger/Plan-2014-Python.

Acknowledgments

We acknowledge the Great Lakes–St. Lawrence River Adaptive Management Committee for providing feedback on this work to align study outcomes with decision-making needs. This work was supported by the U.S. Geological Survey Northeast Climate Adaptation Science Center, which is managed by the USGS National Climate Adaptation Science Center, under Grant/Cooperative Agreement No. G21AC10601-00, and also by the National Science Foundation under Grant No. CBET-2144332 and through the NSF Graduate Research Fellowship under Grant No. DGE-1650441. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the Northeast Climate Adaptation Science Center or the USGS.

Supplemental Materials

Eqs. S1–S5, Figs. S1–S4, and Table S1 are available online in the ASCE Library (www.ascelibrary.org).

References

Alexander, S., G. Yang, G. Addisu, and P. Block. 2021. "Forecast-informed reservoir operations to guide hydropower and agriculture allocations in the Blue Nile Basin, Ethiopia." *Int. J. Water Resour. Dev.* 37 (2): 208–233. https://doi.org/10.1080/07900627.2020.1745159.

Anghileri, D., N. Voisin, A. Castelletti, F. Pianosi, B. Nijssen, and D. P. Lettenmaier. 2016. "Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments." *Water Resour. Res.* 52 (6): 4209–4225. https://doi.org/10.1002/2015WR017864.

Arango-Aramburo, S., S. W. D. Turner, K. Daenzer, J. P. Ríos-Ocampo, M. I. Hejazi, T. Kober, A. C. Álvarez-Espinosa, G. D. Romero-Otalora, and B. van der Zwaan. 2019. "Climate impacts on hydropower in Colombia: A multi-model assessment of power sector adaptation pathways." *Energy Policy* 128 (May): 179–188. https://doi.org/10.1016/j.enpol.2018.12.057.

Badham, J., et al. 2019. "Effective modeling for integrated water resource management: A guide to contextual practices by phases and steps and future opportunities." *Environ. Modell. Software* 116 (Jun): 40–56. https://doi.org/10.1016/j.envsoft.2019.02.013.

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

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

Cohen, J. S., H. B. Zeff, and J. D. Herman. 2020. "Adaptation of multiobjective reservoir operations to snowpack decline in the Western United States." *J. Water Resour. Plann. Manage.* 146 (12): 04020091. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001300.

Cradock-Henry, N. A., P. Blackett, M. Hall, P. Johnstone, E. Teixeira, and A. Wreford. 2020. "Climate adaptation pathways for agriculture: Insights from a participatory process." *Environ. Sci. Policy* 107 (May): 66–79. https://doi.org/10.1016/j.envsci.2020.02.020.

Culley, S., S. Noble, A. Yates, M. Timbs, S. Westra, H. R. Maier, M. Giuliani, and A. Castelletti. 2016. "A bottom-up approach to identifying the maximum operational adaptive capacity of water resource systems to a changing climate." *Water Resour. Res.* 52 (9): 6751–6768. https://doi.org/10.1002/2015WR018253.

Delaney, C. J., et al. 2020. "Forecast informed reservoir operations using ensemble streamflow predictions for a multipurpose reservoir in Northern California." *Water Resour. Res.* 56 (9): e2019WR026604. https://doi.org/10.1029/2019WR026604.

Denaro, S., D. Anghileri, M. Giuliani, and A. Castelletti. 2017. "Informing the operations of water reservoirs over multiple temporal scales by direct use of hydro-meteorological data." *Adv. Water Resour.* 103 (May): 51–63. https://doi.org/10.1016/j.advwatres.2017.02.012.

Doering, K., J. Quinn, P. M. Reed, and S. Steinschneider. 2021. "Diagnosing the time-varying value of forecasts in multiobjective reservoir control." *J. Water Resour. Plann. Manage.* 147 (7): 04021031. https://doi.org/10.1061/(asce)wr.1943-5452.0001386.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. 2016. "Overview of the coupled model intercomparison project Phase 6 (CMIP6) experimental design and organization." *Geosci. Model Dev.* 9 (5): 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Faber, B. A., and J. R. Stedinger. 2001. "Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts." *J. Hydrol.* 249 (1–4): 113–133. https://doi.org/10.1016/S0022-1694(01)00419-X.

Fagherazzi, L., D. Fay, and J. Salas. 2007. "Synthetic hydrology and climate change scenarios to improve multi-purpose complex water resource systems management. The Lake Ontario–St Lawrence River study of the International Canada and US joint commission." *WIT Trans. Ecol. Environ.* 103 (May): 171. https://doi.org/10.2495/WRM070171.

Fayaz, N., L. E. Condon, and D. G. Chandler. 2020. "Evaluating the sensitivity of projected reservoir reliability to the choice of climate projection: A case study of bull run watershed, Portland, Oregon." *Water Resour. Manage.* 34 (6): 1991–2009. https://doi.org/10.1007/s11269-020-02542-3.

Goulart, H. M. D. 2019. *Assessing the operational value of forecast information under climate change*. Milan, Italy: Politecnico di Milano.

Great Lakes–St. Lawrence River Adaptive Management Committee. 2021. "Expedited review of plan 2014, Phase 1: Informing plan 2014 deviation decisions under extreme conditions." *Report to the international joint commission*. Accessed January 31, 2023. https://www.ijc.org/sites/default/files/GLAM_ExpeditedReview_Phase1Report_2021-11-19.pdf.

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

Haasnoot, M., J. Kwadijk, J. Van Alphen, D. Le Bars, B. Van Den Hurk, F. Diermanse, A. Van Der Spek, G. Oude Essink, J. Delsman, and M. Mens. 2020. "Adaptation to uncertain sea-level rise: How uncertainty in Antarctic mass-loss impacts the coastal adaptation strategy of the Netherlands." *Environ. Res. Lett.* 15 (3): 034007. https://doi.org/10.1088/1748-9326/ab666c.

Hadjimichael, A., J. Quinn, E. Wilson, P. Reed, L. Basdekas, D. Yates, and M. Garrison. 2020. "Defining robustness, vulnerabilities, and consequential scenarios for diverse stakeholder interests in institutionally complex river basins." *Earth's Future* 8 (7): e2020EF001503. https://doi.org/10.1029/2020EF001503.

Hadka, D., and P. Reed. 2013. "Borg: An auto-adaptive many-objective evolutionary computing framework." *Evol. Comput.* 21 (2): 231–259. https://doi.org/10.1162/EVCO_a_00075.

Hall, J. W., H. Harvey, and L. J. Manning. 2019. "Adaptation thresholds and pathways for tidal flood risk management in London." *Clim. Risk Manage.* 24 (Jan): 42–58. https://doi.org/10.1016/j.crm.2019.04.001.

Hamlet, A. F., D. Huppert, and D. P. Lettenmaier. 2002. "Economic value of long-lead streamflow forecasts for Columbia River hydropower." *J. Water Resour. Plann. Manage.* 128 (2): 91–101. https://doi.org/10.1061/(asce)0733-9496(2002)128:2(91).

Herman, J. D., and M. Giuliani. 2018. "Policy tree optimization for threshold-based water resources management over multiple timescales." *Environ. Modell. Software* 99 (Jan): 39–51. https://doi.org/10.1016/j.envsoft.2017.09.016.

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.

Herman, J. D., P. M. Reed, H. B. Zeff, and G. W. Characklis. 2015. "How should robustness be defined for water systems planning under change?" *J. Water Resour. Plann. Manage.* 141 (10): 04015012. https://doi.org/10.1061/(asce)wr.1943-5452.0000509.

International Joint Commission. 2014. *Lake Ontario St. Lawrence River Plan 2014: Protecting against extreme water levels, restoring wetlands* *and preparing for climate change*. Washington, DC: International Joint Commission.

International Joint Commission. 2016. *Regulation Plan 2014 for Lake Ontario and the St. Lawrence River: Compendium document*. Washington, DC: International Joint Commission.

International Lake Ontario–St. Lawrence River Board. 2018. *Observed conditions regulated outflows in 2017*. Washington, DC: International Joint Commission.

International Lake Ontario–St. Lawrence River Study Board. 2006. "Options for Managing Lake Ontario and St. Lawrence River water levels and flows." *Report to the International Joint Commission*. Accessed January 31, 2023. https://ijc.org/sites/default/files/L40.pdf.

Khatun, A., B. Sahoo, and C. Chatterjee. 2023. "Two novel error-updating model frameworks for short-to-medium range streamflow forecasting using bias-corrected rainfall inputs: Development and comparative assessment." *J. Hydrol.* 618 (Mar): 129199. https://doi.org/10.1016/j.jhydrol.2023.129199.

Lam, R., et al. 2022. "GraphCast: Learning skillful medium-range global weather forecasting." Preprint, submitted December 24, 2022. http://arxiv.org/abs/2212.12794.

Mejia, J. M., and J. Rousselle. 1976. "Disaggregation models in hydrology revisited." *Water Resour. Res.* 12 (2): 185–186. https://doi.org/10.1029/WR012i002p00185.

Mereu, S., J. Sušnik, A. Trabucco, A. Daccache, L. Vamvakeridou-Lyroudia, S. Renoldi, A. Virdis, D. Savić, and D. Assimacopoulos. 2016. "Operational resilience of reservoirs to climate change, agricultural demand, and tourism: A case study from Sardinia." *Sci. Total Environ.* 543 (Feb): 1028–1038. https://doi.org/10.1016/j.scitotenv.2015.04.066.

Nayak, M. A., J. D. Herman, and S. Steinschneider. 2018. "Balancing flood risk and water supply in California: Policy search integrating short-term forecast ensembles with conjunctive use." *Water Resour. Res.* 54 (10): 7557–7576. https://doi.org/10.1029/2018WR023177.

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

Quinn, J. D., P. M. Reed, M. Giuliani, and A. Castelletti. 2017a. "Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems." *Water Resour. Res.* 53 (8): 7208–7233. https://doi.org/10.1002/2017WR020524.

Quinn, J. D., P. M. Reed, and K. Keller. 2017b. "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.

Rayner, S., D. Lach, and H. Ingram. 2005. "Weather forecasts are for wimps: Why water resource managers do not use climate forecasts." *Clim. Change* 69 (2–3): 197–227. https://doi.org/10.1007/s10584-005-3148-z.

Russell, S. O., and W. F. Caselton. 1971. "Reservoir operation with imperfect flow forecasts." *J. Hydraul. Div.* 97 (2): 323–331. https://doi.org/10.1061/jyceaj.0002873.

Schapire, R. E. 2009. "A short introduction to boosting." *Society* 14 (5): 1612. https://doi.org/10.1.1.112.5912.

Semmendinger, K., D. Lee, L. Fry, and S. Steinschneider. 2022. "Establishing opportunities and limitations of forecast use in the operational management of highly constrained multiobjective water systems." *J. Water Resour. Plann. Manage.* 148 (8): 04022044. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001585.

Stedinger, J. R., B. F. Sule, and D. P. Loucks. 1984. "Stochastic dynamic programming models for reservoir operation optimization." *Water Resour. Res.* 20 (11): 1499–1505. https://doi.org/10.1029/WR020i011p01499.

Steinschneider, S. 2022. *Using hydroclimate modeling and social science to enhance flood resilience on Lake Ontario through the climate smart communities program (Version 1.0) [Data set]*. Honolulu, HI: Zenodo. https://doi.org/10.5281/zenodo.7186904.

Steinschneider, S., and C. Brown. 2012. "Dynamic reservoir management with real-option risk hedging as a robust adaptation to nonstationary climate." *Water Resour. Res.* 48 (5): 12. https://doi.org/10.1029/2011WR011540.

Sveinsson, Ó. G. B., and J. D. Salas. 2006. "Multivariate shifting mean plus persistence model for simulating the Great Lakes net basin supplies." In *Proc., 2006 Annual AGU Hydrology Days*, 173–184. Fort Collins, CO: Colorado State Univ.

Syme, G. J., and B. S. Sadler. 1994. "Evaluation of public involvement in water resources planning: A researcher-practitioner dialogue." *Eval. Rev.* 18 (5): 523–542. https://doi.org/10.1177/0193841X9401800501.

Trindade, B. C., P. M. Reed, and G. W. Characklis. 2019. "Deeply uncertain pathways: Integrated multi-city regional water supply infrastructure investment and portfolio management." *Adv. Water Resour.* 134 (Dec): 103442. https://doi.org/10.1016/j.advwatres.2019.103442.

Turner, S. W. D., J. C. Bennett, D. E. Robertson, and S. Galelli. 2017. "Complex relationship between seasonal streamflow forecast skill and value in reservoir operations." *Hydrol. Earth Syst. Sci.* 21 (9): 4841–4859. https://doi.org/10.5194/hess-21-4841-2017.

Van Der Beken, A., G. L. Vandewiele, J. Tiberghien, J. Marien, J. Marivort, and M. Meersseman. 1980. "On-line flow forecasting for automatic operation of a flood reservoir." In *Proc., Oxford Symp. Hydrological forecasting, April 1980, (Int. Association of Hydrological Sciences*. Washington, DC: IAHS-AISH Publication.

Wilcox, D. A., and J. A. Bateman. 2018. "Photointerpretation analysis of plant communities in Lake Ontario wetlands following 65 years of lake-level regulation." *J. Great Lakes Res.* 44 (6): 1306–1313. https://doi.org/10.1016/j.jglr.2018.08.007.

Wilcox, D. A., and Y. Xie. 2007. "Predicting wetland plant community responses to proposed water-level-regulation plans for Lake Ontario: GIS-based modeling." *J. Great Lakes Res.* 33 (4): 751–773. https://doi.org/10.3394/0380-1330(2007)33[751:PWPCRT]2.0.CO;2.

Wilson, A., R. Cifelli, F. Munoz-Arriola, J. Giovannettone, J. Vano, T. Parzybok, A. Dufour, J. Jasperse, K. Mahoney, and B. McCormick. 2021. "Efforts to build infrastructure resiliency to future hydroclimate extremes." In *Geo-extreme 2021: Climatic extremes and earthquake modeling*, edited by C. L. Meehan, M. A. Pando, B. A. Leshchinsky, and N. H. Jafari, 222–223. Reston, VA: ASCE.

Wilson, T. T., and E. Kirdar. 1970. "Use of runoff forecasting in reservoir operations." *J. Irrig. Drain. Div.* 96 (3): 299–308. https://doi.org/10.1061/jrcea4.0000732.

Woodside, G. D., A. S. Hutchinson, F. M. Ralph, C. Talbot, R. Hartman, and C. Delaney. 2022. "Increasing stormwater capture and recharge using forecast informed reservoir operations, Prado Dam." *Ground Water* 60 (5): 634–640. https://doi.org/10.1111/gwat.13162.

Yang, G., S. Guo, P. Liu, and P. Block. 2021. "Sensitivity of forecast value in multiobjective reservoir operation to forecast lead time and reservoir characteristics." *J. Water Resour. Plann. Manage.* 147 (6): 04021027. https://doi.org/10.1061/(asce)wr.1943-5452.0001384.

Yao, H., and A. Georgakakos. 2001. "Assessment of Folsom Lake response to historical and potential future climate scenarios 2. Reservoir management." *J. Hydrol.* 249 (1–4): 176–196. https://doi.org/10.1016/S0022-1694(01)00418-8.

Zarei, M., O. Bozorg-Haddad, S. Baghban, M. Delpasand, E. Goharian, and H. A. Loáiciga. 2021. "Machine-learning algorithms for forecast-informed reservoir operation (FIRO) to reduce flood damages." *Sci. Rep.* 11 (1): 24295. https://doi.org/10.1038/s41598-021-03699-6.

Zhao, T., X. Cai, and D. Yang. 2011. "Effect of streamflow forecast uncertainty on real-time reservoir operation." *Adv. Water Resour.* 34 (4): 495–504. https://doi.org/10.1016/j.advwatres.2011.01.004.

Zhao, T., and J. Zhao. 2014. "Joint and respective effects of long- and short-term forecast uncertainties on reservoir operations." *J. Hydrol.* 517 (Sep): 83–94. https://doi.org/10.1016/j.jhydrol.2014.04.063.