Title
Dynamic and robust adaptation to climate uncertainty in water resources systems

Author
Cohen, Jonathan

Publication Date
2021

Peer reviewed|Thesis/dissertation

Dynamic and robust adaptation to climate uncertainty in water resources systems

By

JONATHAN SCOTT COHEN

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Jonathan D. Herman, Chair

Jay R. Lund

Samuel Sandoval Solis

Committee in Charge

2021

In loving memory of my beloved Grandfather, Sol Marenberg z"l.

Contents

List of Figures

List of Tables

Abstract

Adaptation to the multi-scale impacts of climate change in water resources systems is challenged by substantial uncertainty in future hydroclimatic projections, specifically regarding snowpack decline, flood and drought risks, and long-term transient trends in projections. This dissertation develops new approaches to address these challenges, specifically for the northern California water resources system. A new daily timestep simulation model of the system is introduced. Model simulations over an ensemble of climate scenarios provide a baseline system response for each chapter. In Chapter 1, a statistical analysis of these baseline simulations links several vulnerabilities directly to snowpack decline and the shift of snowmelt-fed streamflow earlier in the water year. To adapt seasonal reservoir management to snowpack decline in the region without the cost of additional infrastructure, the study proposes and tests adaptations that parameterize the structure of existing operating policies: a dynamic flood control rule curve, and revised snowpack-to-streamflow forecasting methods to improve seasonal runoff predictability given declining snowpack. These adaptations are shown to mitigate the majority of vulnerabilities caused by snowpack decline across the scenario ensemble. To address the issue of adapting to future flood and drought risk, Chapter 2 introduces a scenario selection framework. Scenarios are clustered by hydrologic properties, including full natural flow and snow water equivalent, along with a baseline-regret metric-the difference between the status quo and a perfect foresight adaptation. Findings suggest that reservoir operating policies should be trained to scenarios with a high baseline regret value in order to be most robust to climate uncertainty. Lastly, Chapter 3 develops a framework for dynamic climate adaptation based on multi-objective policy tree optimization, a heuristic search method that combines relevant indicators, actions, and thresholds in a flexible policy structure. Analysis of a robust set of policy trees identifies the most common indicators, actions, policy structure, and timing that produce robust policies. This represents a new and transferable problem framing for adaptation under uncertainty in which indicator variables, relevant actions, and policy structure are identified simultaneously. In total, this dissertation provides a synthe-

sized approach combining novel robust and dynamic planning methods to adapt water resources systems to uncertainty under climate change.

Acknowledgements

I first must thank my advisor, mentor, and friend Jon Herman, for whom I am deeply grateful. Jon taught me how to think independently, how to ask and answer meaningful scientific questions, and how to question and justify my own findings. He helped teach me how to educate myself to advance my knowledge in systems analysis, water resources engineering, and climate adaptation. Thanks also goes to Jay Lund for advancing my knowledge of water resources management in California, economics, and how to think probabilistically. Jay's courses were also crucial in improving my writing skills throughout graduate school. I want to thank my dissertation and qualifying exam committee members Sam Sandoval-Solis, Bryan Jenkins, and Josue Medellin-Azuara for their guidance and feedback on my research. In addition, I must acknowledge the whole UC Davis Department of Civil and Environmental Engineering for its support in this process.

I am also very grateful for the Department of Engineering at Swarthmore College, which gave me an exemplary education in preparation for graduate school. I am deeply grateful for my undergraduate advisor Arthur McGarity, who helped to give me a foundation in environmental systems analysis, operations research, and water resources engineering, while also guiding me in how to conduct research. Additionally, thank you to Carr Everbach for teaching me to consider the interactions of our role as engineers with the environment and communities our work effects. Thank you also to those at Politechnika Krakowska for the knowledge in environmental engineering and policy I received in Poland.

Several sources have funded this research over the past four years. They include U.S. National Science Foundation grants CBET-1803589 and CNS-1639268, and the 2020-2021 Summer GSR Award from the UC Davis Department of Civil Engineering. In addition, the opportunities from Jon Herman to be teaching assistant for ECI 115 and from Jay Lund to co-instruct ECI 153 helped to fund my time in graduate school.

Special thanks goes to my partner Emily Hain, not only for her companionship, but also for her exceptional support during the process of writing this dissertation

and finishing my degree. I owe a lot of gratitude to Francisco Bellido-Leiva, not only for his friendship and exchange of resourceful ideas, but also for being an exemplary classmate, colleague, and co-instructor. I also will thank and acknowledge Emily Snider, Drew Friedrichs, Keara Tuso, and Samantha Sharp, for our ongoing support of each other during the COVID-19 pandemic. Thank you to my lab-mates Marina Mautner, Bethany Robinson, Natalie Mall, Nusrat Molla, and Liam Ekblad, for our constant support of each other and sharing of ideas and knowledge. Many other classmates in the UD Davis College of Engineering helped to shape my ideas and friendships through this process. Just some of these include Federico Zabaleta, Sergio Valbuena, Aaron Alexander, Aisha Faruqi, Lauren Lynn, Andy Schmalenbach, Leo Pescador, and Quinn Horvath.

I owe so much thanks to my mom, Marjorie Marenberg, for encouraging me to think with curiosity and creativity. Thank you also to my dad, Chris Cohen, for helping and teaching me to overcome many obstacles in life. To my younger siblings, Daniel and Sarah, witnessing your many accomplishments over the past four years has only further motivated me on my academic journey.

Lastly, this dissertation is dedicated to the loving memory of my Grandfather, Sol Marenberg z"l. He grew up during the Great Depression with almost nothing, and was the first in his family of Jewish immigrants to attend college. A civil engineer himself, he served in the U.S. Army Corps of Engineers during World War II and then spent his career designing schools, hospitals, and concert halls. Without his dedication and passion for his work and family, I would not have had the foundation to complete this achievement today.

Chapter summaries

Notes on Chapter 1

Chapter 1 introduces a novel approach to analyze vulnerabilities and adaptations specifically focused on the challenge of snowpack decline under climate change. This work ties to the broader literature on climate adaptation in which a water resources system is tested under an ensemble of climate projections to determine potential vulnerabilities and appropriate operational changes. Introducing a new daily timestep simulation model of the northern California reservoir system, the region's response to an ensemble of climate scenarios is assessed. Through a statistical analysis, the impacts of a long-term transient snowpack decline on water supply, reservoir storage, and environmental flows is assessed. Given these isolated vulnerabilities, adaptations that parameterize the structure of existing operating policies are developed. These include a dynamic flood control rule curve and revised snowpack-to-streamflow forecasting methods. The coupled approach to vulnerability assessment and adaptation will be generalizable to other snowmelt-dominated water resources systems facing the loss of seasonal snowpack storage. Chapter 1 is the result of an article that is published in *Journal of Water Resources Planning and Management*, which was chosen as an addition to the Editor's Choice Collection for the December 2020 issue (https://doi.org/10.1061/(ASCE)WR.1943-5452.0001300)

Notes on Chapter 2

Chapter 2 explores how the properties of training scenarios for policy search impact policy robustness of optimized reservoir policies under climate change. Using the northern California reservoir system as a template, specifically scenario properties are investigated. These include annual runoff, snowpack, and baseline regret—the difference between baseline policy and perfect foresight performance in an individual scenario. Results indicate that policies trained to scenario subsets with high baseline regret outperform those generated with other training sets in both wetter and drier futures, largely by adopting an intra-annual hedging strategy. Additionally, policies trained to scenarios with higher annual flow and maximum snowpack will perform poorly in scenarios with lesser streamflow and snowpack values, and vice-verse. The addition of scenarios with high baseline regret values in a training set helps to mitigate this issue. The approach highlights the potential to improve the efficiency and robustness of policy training by considering both the hydrologic properties and adaptation potential of the training ensemble. Chapter 2 is the result of an article that is published in *Environmental Modeling and Software* (https://doi.org/10.1016/j.envsoft.2021.105047).

Notes on Chapter 3

Chapter 3 introduces a framework for dynamic adaptation to climate and land-use uncertainty based on multi-objective policy tree optimization, a heuristic search method that combines relevant indicators, actions, and thresholds in a flexible policy structure. The approach is demonstrated for a case study of northern California, where a mix of infrastructure, management, and operational adaptations are considered over time in response to an ensemble of nonstationary hydrology, water demand, and economic conditions. The study identifies a subset of non-dominated policies that are robust to held-out scenarios, and then analyzes their most common actions and indicators compared to the non-robust policies. Results show that the robust policies are not differentiated by the actions they select, but show

substantial differences in their indicator variables, which can be interpreted in the context of physical hydrologic trends. In particular, the statistical transformations of the indicator variables highlight the balance between adapting quickly versus correctly, which the non-robust policies fail to achieve. Additionally, the indicators most frequently associated with each action are determined, along with the distribution of action timing across the scenario ensemble. This study presents a new and transferable problem framing for adaptation under uncertainty in which indicator variables, relevant actions, and policy structure are identified simultaneously during the optimization. Chapter 3 is the result of a manuscript that will be submitted to *Water Resources Research*.

Chapter 1

Adaptation of multi-objective reservoir operations to snowpack decline in the western United States[1]

1.1 Abstract

Long-term snowpack decline is among the best-understood impacts of climate change on water resources systems. This trend has been observed for decades, and is projected to continue even in climate projections in which total runoff volumes do not change significantly. For basins in which snowpack has historically provided intra-annual water storage, snowpack decline creates several issues that may require adaptation to infrastructure, operations, or both. This study develops an approach to analyze vulnerabilities and adaptations specifically focused on the challenge of snowpack decline, using the Northern California reservoir system as a case study. First, an open-source daily timestep simulation model of this system, validated against historical observations of operations, is introduced. Multi-objective vulnerabilities to snowpack decline are then examined using a set of downscaled climate scenarios to capture the physically-based effects of rising temperatures. A statistical analysis shows that the primary impacts include water supply shortage and lower reservoir storage resulting from the seasonal shift in runoff timing. These challenges identified from the vul-

[1]This chapter has been published: Cohen, J. S., Zeff, H. B. and Herman, J. D. (2020). "Adaptation of multiobjective reservoir operations to snowpack decline in the western United States". *Journal of Water Resources Planning and Management* ***146*** (12): 04020091

nerability assessment inform proposed adaptations to operations to maintain multi-objective performance across the ensemble of plausible future scenarios, which include other uncertain hydrologic changes. To adapt seasonal reservoir management without the cost of additional infrastructure, we specifically propose and test adaptations that parameterize the structure of existing operating policies: a dynamic flood control rule curve, and revised snowpack-to-streamflow forecasting methods to improve seasonal runoff predictability given declining snowpack. These adaptations are shown to mitigate the majority of vulnerabilities caused by snowpack decline across the scenario ensemble, with remaining opportunities for improvement using formal policy search and dynamic adaptation techniques. The coupled approach to vulnerability assessment and adaptation is generalizable to other snowmelt-dominated water resources systems facing the loss of seasonal storage due to rising temperatures.

1.2 Introduction

In many mountainous regions, snowpack provides valuable intra-annual water storage to support summer irrigation, urban uses, and environmental flows (Sturm et al.; 2017; Rhoades et al.; 2018). Rising temperatures due to climate change have led to long-term declines in mountain snowpack, resulting from both changes in precipitation phase (Cayan et al.; 2001; Klos et al.; 2014) as well as earlier spring melt timing (Cayan; 1996). These trends have been observed in records dating back to the mid-20th century (Mote et al.; 2005; Stewart et al.; 2005; Barnett et al.; 2008; Donat et al.; 2013; Belmecheri et al.; 2016) and are projected to continue with high confidence (Hayhoe et al.; 2004; Leung et al.; 2004), making snowpack decline one of the best-predicted impacts of climate change (McCabe et al.; 2007; Huang et al.; 2018) (Figure 1.1b,d), despite substantial uncertainty in total runoff volumes in many river basins. The primary impact is an intra-annual shift in the hydrologic regime, moving streamflows earlier in the water year (Knowles et al.; 2006; Kapnick and Hall; 2010) (Figure 1.1a,c). As this shift continues, river basins historically reliant on snowpack storage

may require additional infrastructure; for those with existing downstream storage infrastructure, reservoir operating policies must be adapted to mitigate vulnerabilities associated with snowpack decline.

In the absence of adaptation, snowpack decline will lead to several potentially severe consequences for water resources management (Barnett et al.; 2005). First, inflows concentrated in the winter and early spring will likely lead to lower reservoir storage levels later in the water year, eliminating flexibility for water deliveries given fixed storage capacities (Christensen et al.; 2004). This lost storage would decrease system performance in other objectives, including environmental flows and hydropower generation (Vicuna and Dracup; 2007). Additionally, the loss of snowmelt and resulting seasonal streamflow shift will be detrimental to agriculture, given conflicting intra-annual timing with irrigation demands (Qin et al.; 2020). Second, an increase in rain-on-snow events, potentially combined with more intense precipitation events, may amplify flood risk (McCabe et al.; 2007; Surfleet and Tullos; 2013; Huang et al.; 2018). When coupled with the loss of snowpack storage, this will increase the tension between flood control and water supply operations (Knowles et al.; 2006; Lee et al.; 2009; Mateus and Tullos; 2017). Finally, snowpack decline will reduce or eliminate the seasonal hydrologic predictability offered by snowpack-to-streamflow forecasts, which historically have been quite accurate (Koster et al.; 2010; Mahanama et al.; 2012; Livneh and Badger; 2020). This loss of seasonal predictability may cause substantial economic losses in the future, particularly in the agricultural sector (Simpson et al.; 2004; Pederson et al.; 2011).

Vulnerability assessments are meant to evaluate the potential impacts of long-term hydrological changes, including snowpack decline, on water resources systems. These often involve the response of a simulation model to an ensemble of downscaled Global Circulation Model (GCM) scenarios that capture a portion of the future uncertainty in hydroclimate variability (Christensen and Lettenmaier; 2006; Knowles et al.; 2018). This type of top-down approach is particularly relevant when climate models show broad agreement on the direction of future

change, and where the impacts under consideration can be directly linked to physical processes, as is the case with snowpack decline. The ensemble simulation can be accompanied by a statistical analysis to quantify risk across the climate models and scenario definitions (Brekke et al.; 2009; Goharian et al.; 2016; Mateus and Tullos; 2017). Several studies have used top-down vulnerability assessments to explore the impacts of uncertainty across ensembles of climate models and emission scenarios (e.g., Hamlet and Lettenmaier; 1999; Minville et al.; 2010; Karamouz et al.; 2013). Conversely, bottom up approaches (Weaver et al.; 2013) generate synthetic sequences representing potential changes in variability and magnitude of regional hydrologic and atmospheric variables to bypass the inherent structural and parametric uncertainties that propagate through climate projections, also known as the cascade of uncertainty (Wilby and Dessai; 2010). While bottom-up scenario generation methods can parameterize the causes of change across a range of scenarios for vulnerability assessment, they may not capture long-term transient trends and complex physical processes that are present in climate projections (Herman et al.; 2020). Therefore, significant opportunity exists within top-down approaches to isolate the impacts of a particular long-term transient change, such as snowpack decline, on the dynamics of reservoir systems from the broader uncertainty in total runoff magnitude.

For river basins with downstream reservoir storage, adapting to the hydrologic impacts of climate change will require revised operating policies to increase efficiency before investing in new infrastructure (Gleick; 2002; Culley et al.; 2016). A subset of robust planning studies explore adaptations to system operations under climate change, combining simulation or optimization models with downscaled runoff projections or synthetic scenarios (Wilby and Dessai; 2010; Herman et al.; 2015). Unlike vulnerability assessments, these studies rely on designing new policies for adaptation - a change in operations that has the potential to mitigate climate vulnerabilities or the impacts of uncertainty. We group adaptation studies into three subsets: first, those that adapt to future exogenous hydrologic uncertainty associated with transient climate change in projection ensembles (e.g., Georgakakos et al.;

2012; Steinschneider et al.; 2015); second, studies that propose adaptations to climate change scenarios with specific isolated properties (i.e. drier scenarios, increases in floods) (Medellín-Azuara et al.; 2008; Wilby and Keenan; 2012, e.g.,); and lastly, studies that aim to mitigate impacts of thermodynamic climate change (i.e., due to rising temperatures) including sea level rise and seasonal streamflow shifts, (e.g., Willis et al.; 2011; Mateus and Tullos; 2017; Sterle et al.; 2020). An opportunity exists to synthesize these three approaches. Given the high confidence in predicting thermodynamic changes and their consequences, adaptations specific to those changes can be proposed after formally isolating their impacts on system objectives. This would be followed by an assessment of adaptation performance with respect to both exogenous dynamic uncertainty across an ensemble, as well as to specific transient properties of individual scenarios.

This study develops this synthesized adaptation approach in order to isolate and mitigate system vulnerabilities that result from a specific physical impact of climate change projected with confidence—in this case, snowpack decline. An ensemble of downscaled GCM scenarios represents uncertainty in transient streamflow and precipitation trends while also containing varying magnitudes of temperature rise and physical snowpack decline trajectories. We use a new daily time-step model of the northern California reservoir system to produce a top-down system response to these scenarios. By isolating the impacts of individual hydrologic variables on system vulnerabilities through a statistical analysis, this top-down assessment attains the focus on specific uncertainties defined a priori in bottom-up approaches, while also benefiting from the detailed physical properties and long-term transient trends in climate projection outputs. Noting the isolated effects of snowpack decline on specific system objectives found in the vulnerability assessment, we propose two adaptations to specifically mitigate the impacts of snowpack decline and its resulting intra-annual streamflow shifts. These include an alteration of the dynamic flood control curve and a revised snowpack-to-streamflow forecasting method. We analyze how these adaptations dynamically reduce vulnerabilities directly related to snowpack decline, considering the magnitude of these changes

across scenarios. To then synthesize the three adaptation study approaches, we consider how these adaptations hold up against more uncertain changes across the ensemble, such as those in total annual streamflow. This coupled vulnerability assessment-adaptation study approach will be broadly transferable to river basins facing snowpack decline, where operating policies for upstream and downstream infrastructure can be redesigned to compensate for the loss of this critical natural water storage.

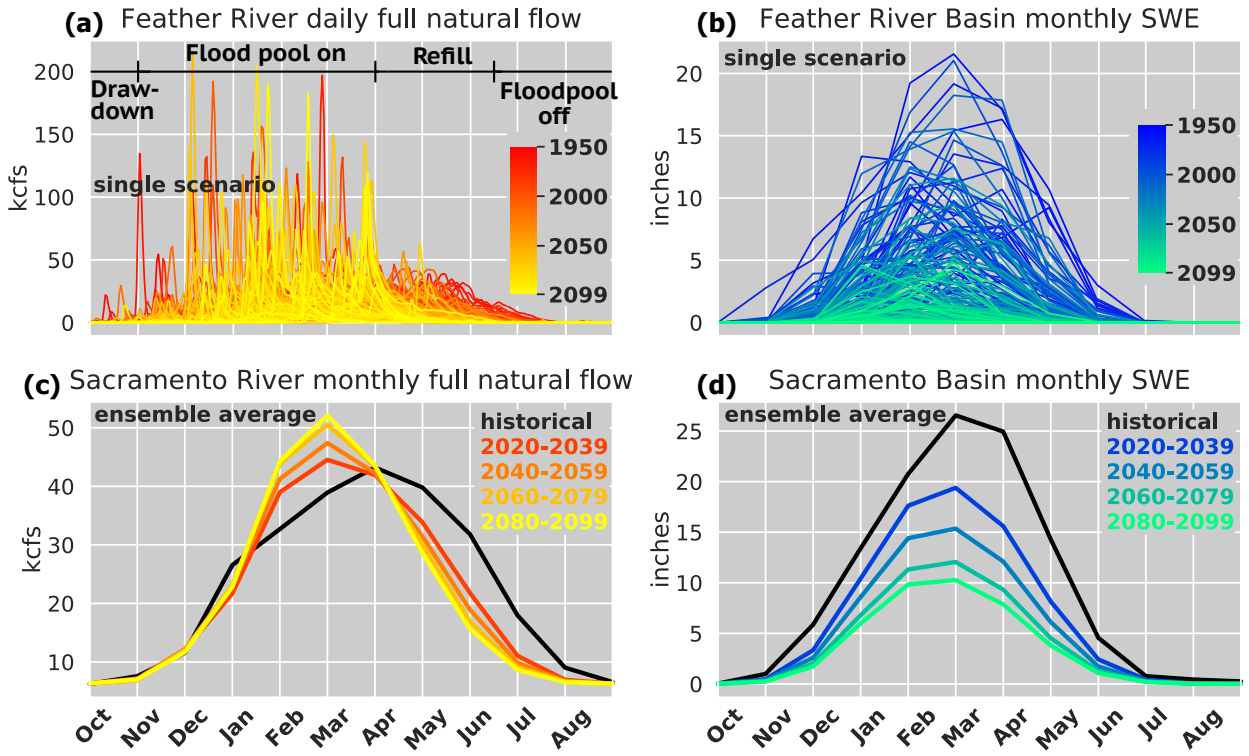

Figure 1.1: (a) Daily streamflow displaying an intra-annual streamflow shift and (b) monthly basin-averaged snow water equivalent (SWE) showing snowpack decline in a single downscaled hydrologic scenario (NOAA GFDL-CM3, RCP 8.5). Historical (1996-2018) and climate projection ensemble averages over time of (c) average monthly streamflow and (d) monthly basin-averaged (SWE). Data source: USBR (2014).

1.3 Model and study area

1.3.1 Northern California reservoir system

To accommodate its Mediterranean climate and high inter-annual variability, California has built a vast and complex system of water supply and flood control infrastructure. Reservoirs at the foothills of the Sierra Nevada Range store high flows during the winter and spring to be delivered for agriculture and municipal supply, while also managing flood events. Historically, reservoir inflows in the early irrigation season have been driven by snowmelt, suggesting that the management of downstream reservoir storage will become even more crucial under climate change (e.g. Figure 1.1). Storage and conveyance infrastructure include both the State Water Project (SWP) and federal Central Valley Project (CVP), which consist of a number of reservoirs and aqueducts throughout the Sacramento-San Joaquin river basin. The terminal Delta of this system is the site of pumped water exports from north to south to support agriculture and municipal supply, delivering annual averages of 2.7 and 3.2 MAF (million-acre feet) for the CVP and SWP, respectively (Table 1.1). These exports are constrained by critical environmental flow requirements related to the salinity of Delta outflows. Delta exports are a key metric for water supply reliability in the state and have been found vulnerable to climate change, due to combined changes in precipitation and seasonal runoff timing (Anderson et al.; 2008; Ray et al.; 2020).

In the northern Sacramento basin, three of the largest Sierra foothill reservoirs by volume (Shasta, Oroville, and Folsom) provide a combined 9 MAF (11.1 km^3) of storage (Table 1.2), and play a key role in balancing human and environmental water needs. Their releases satisfy demands for deliveries north of the Delta, Delta outflows, and south of Delta exports (see Table 1.2 for demand values), while also maintaining downstream environmental flow targets. Additionally, these reservoirs are crucial for flood control, while also providing an ancillary benefit of hydropower production. Carryover storage in these reservoirs, measured at the end of the water year on September 30th, is a strong indicator of system performance and

economic vulnerability (Draper and Lund; 2004). Potential warmer, drier climate change has the possibility of being most detrimental to these economic benefits (Medellín-Azuara et al.; 2008).

Pumping plants	Tracy (CVP)	Banks (SWP)
Average annual demand (MAF/year)	2.7	3.2
Max pumping capacity $Tmax, Pmax$ (cfs)	4300	8500
Max target TT, BT (cfs)	4300	6000
Min target TT, BT (cfs)	1000	1000
Max intake limit TP, BP (cfs)	4300	6680

Table 1.1: Characteristics of the Delta pumping plants modeled in ORCA. Attributes that are followed by notation correspond to parameters and constraints that are included in the simulation model.

The impacts of climate change on California water resources is a topic that has been studied extensively (Vicuna and Dracup; 2007), with several studies concluding that hydrologic changes have high potential to reduce Delta exports and reservoir carryover storage (e.g., Lettenmaier and Sheer; 1991; VanRheenen et al.; 2004; Vicuna et al.; 2007; Brekke et al.; 2009)). These studies have primarily used planning models on monthly time steps, inhibiting the ability to analyze vulnerabilities to flooding, while also focusing on vulnerability rather than adaptation. Those that have considered adaptive operations (e.g. Yao and Georgakakos; 2001; Tanaka et al.; 2006; Georgakakos et al.; 2012) have considered only a few climate scenarios, potentially not capturing the range of outcomes using an ensemble approach. Under the many projected changes to the hydrologic regime, operational adaptations are needed to maintain storage levels to support multiple objectives and yield adequate water supply, while continuing to provide flood control benefits. Additionally, adaptations can be targeted to specific hydrologic changes, such as snowpack decline, that are projected with higher certainty while remaining robust to other uncertain changes in total water availability.

Reservoir	Shasta	Oroville	Folsom
Storage capacity $S^k_{\max}$ (TAF)	4,552	3,537	975
Deadpool $S^k_{\min}$ (TAF)	550	852	90
Minimum winter flood control storage f^k_{tocs} (TAF)	3,252	2,837	375
Downstream levee capacity DQ^k (cfs)	79,000	150,000	115,000
Maximum environmental flow demands E^k_{min} (cfs)	3,250	1,700	1,750
Minimum environmental flow demands E^k_{min} (cfs)	2,000	1,000	250
Maximum south of Delta demands SOD^k (TAF/day)	6.82	16.86	1.71
Minimum south of Delta demands SOD^k (TAF/day)	1.19	1.98	0.3
Average annual south of Delta demand (MAF/year)	2.2	3.2	0.48
Maximum north of Delta demands NOD^k (TAF/day)	9.84	0	2.16
Minimum north of Delta demands NOD^k (TAF/day)	1.64	0	0.36
Average annual north of Delta demand (MAF/year)	1.5	0	0.3
Average Delta outflow demand (MAF/year)	6.0	2.0	2.0

Table 1.2: Characteristics of the reservoirs modeled in ORCA. Attributes that are followed by notation correspond to parameters and constraints that are included in the simulation model.

1.3.2 Simulation model (ORCA)

To analyze this problem, we construct an open source simulation model of the northern portion of the California water system, named Operation of Reservoirs in California (ORCA) (https://github.com/jscohen4/orca). ORCA simulates several major components of Northern California's water resource system, including the interaction of snowpack-to-streamflow forecasting, Shasta, Oroville, and Folsom Reservoirs, and management of the Sacramento-San Joaquin Delta (Figure 1.2a,b). While not as spatially comprehensive as other existing statewide models, ORCA is a pure simulation model that runs on a daily timestep, which allows flexible adjustments to operating rules and straightforward testing of alternate scenarios. The model demonstrates accuracy in simulating historical daily system operations (Figure 1.2c,d,e), including reservoir releases and Delta exports. See Section 1.7.4 in the Appendix for further model accuracy results.

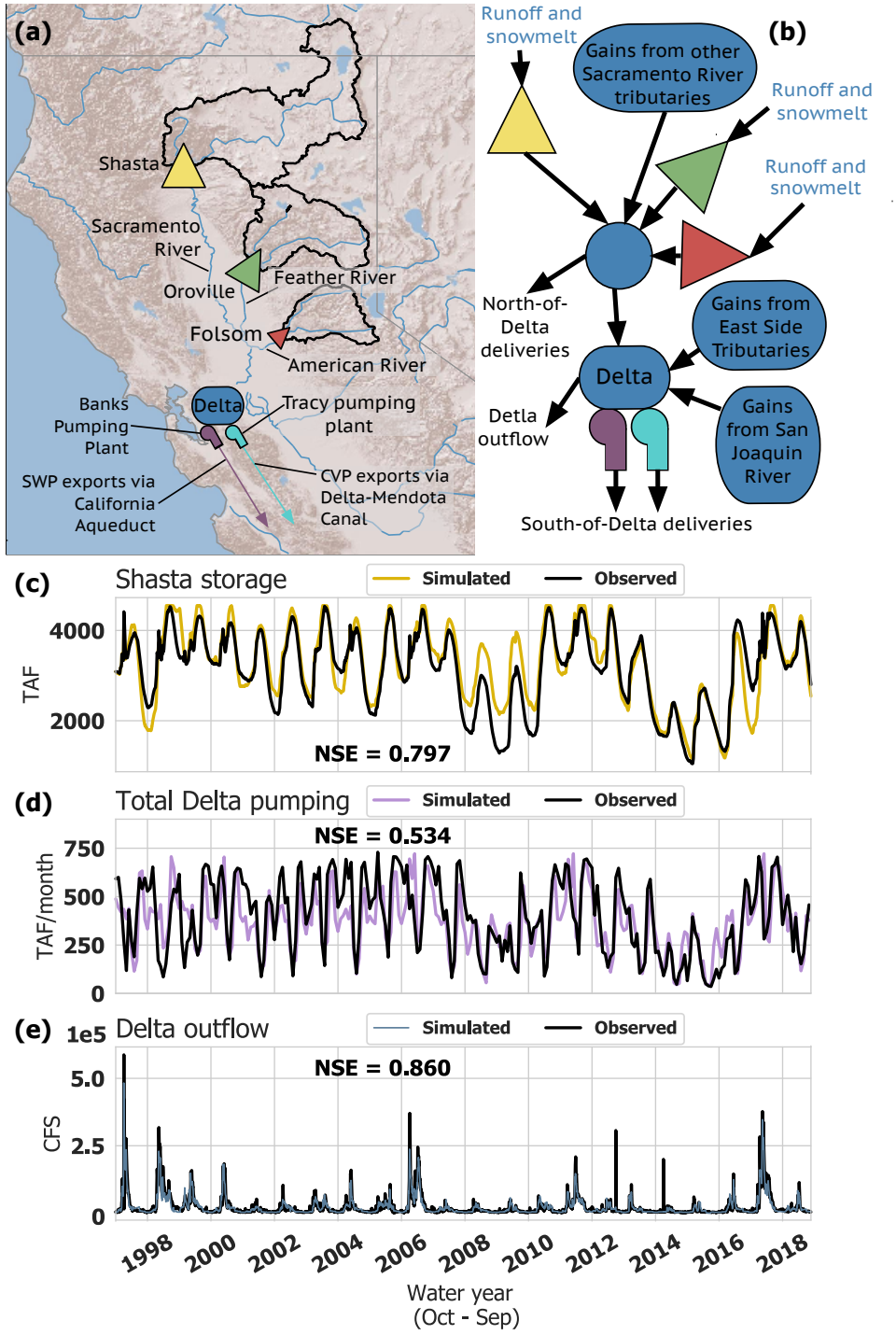

Figure 1.2: (a) Map of Northern California Water Resource System modeled in ORCA. (b) Schematic of ORCA (c,d,e) Comparisons of ORCA output and historical observations for (c) daily Shasta Reservoir storage (d) Total monthly Delta exports (e) daily Delta outflow, with performance measured by Nash-Sutcliffe Efficiency (NSE).

1.3.2.1 Data sources

ORCA relies on several hydroclimatic time series as inputs. These include daily streamflows, precipitation, and air temperature, along with monthly snow water equivalent (SWE). For simulating historical operations, these data are obtained from the California Data Exchange Center (CDEC; 2018). CDEC provides streamflows for the Sacramento River, many of its tributaries, and several other rivers in the Sacramento-San Joaquin basin. Historical observations of approximate basin-averaged precipitation and SWE are drawn from several CDEC stations upstream of each reservoir. The CDEC database also provides historical reservoir releases, storage, and Delta pumping time-series used to test the accuracy of the model. While data availability varies slightly between locations, daily timestep data is generally available from 1997-present.

After confirming the ability of the model to reproduce historical operations, the observed hydroclimatic inputs are substituted with downscaled climate change projections. We use the downscaled CMIP5 Climate and Hydrology Projections from the United States Bureau of Reclamation (USBR) (Reclamation; 2013; Brekke et al.; 2014). These consist of 31 GCMs simulated for various emissions scenarios to generate 97 scenarios of precipitation and temperature on a daily timestep through 2100 (see Section 1.7.6 in the Appendix for GCM modeling center information). In the USBR study, outputs from these GCM simulations were routed through the Variable Infiltration Capacity (VIC) model (Liang et al.; 1994) calibrated for each basin, yielding additional streamflow and SWE projections to serve as model inputs. In cases where model inputs, such as SWE and precipitation, were averaged across multiple stations, the relevant information was extracted from the gridded VIC output to produce basin-wide spatial averages.

Visualizing these ensemble projections provides insight into what future hydrologic scenarios might entail. Figure 1.3 shows 50-year moving averages of (a) spatially averaged snowpack across the northern Sierra Nevada, (b) streamflow in the four tributaries of the Sacramento River, (c) the water year centroid, defined as the day of the water year at which

half of the total annual streamflow has occurred, and (d) streamflow during the flood season. All scenarios in the ensemble show a decline in snowpack, ranging from 20-90% of the historical average, which leads the water year centroid to shift earlier in the year. However, in several scenarios, the severity of snowpack decline and intra-annual shifts does not correspond directly to a decrease in overall flow (Figure 1.3b). Additionally, the seasonal shift leads to increased flood season streamflow in the majority of scenarios (Figure 1.3d), which even occurs in many scenarios which show some decrease in overall annual streamflow. This is indicated by the several scenarios with higher water availability that also have relatively low snowpack levels (Figure 1.3b). The end-of-century average annual flows range from -/+ 50% of historical values, indicating significant uncertainty in whether future scenarios will be wetter or drier. This uncertainty is shown in a very coarse statistic (the 50-year moving average), and an even higher degree of uncertainty would likely be seen in estimates of flood and drought frequency and severity that would cause system vulnerabilities.

1.3.2.2 Snowpack-to-streamflow forecasts

At each timestep of the simulation model, the first component to be evaluated is a seasonal snowpack-to-streamflow forecast. The expected cumulative inflow for the rest of the water year drives a number of key decisions in the system. ORCA uses a linear regression method to estimate this forecast value, aiming to reproduce the forecasting method developed by California state agencies (Rizzardo; 2016). The regression is computed for each of the k reservoirs based on a basin average of the snow water equivalent (SWE). For a given day of the water year dw_t, historical data is used to formulate a linear regression to predict the total volume of streamflow occurring through the rest of the water year, Qr_t, using the maximum to-date snow water equivalent, SWE_t^k, as the independent variable. An exceedance parameter $\bar{Z}_{wyt}^k$ is multiplied by the standard deviation of the regression residuals $\sigma_{dw_t}^k$ to perturb a forecast exceedance level, which varies based on the water year type (WYT) and reservoir. The exceedance level determines how conservative or aggressive the forecast

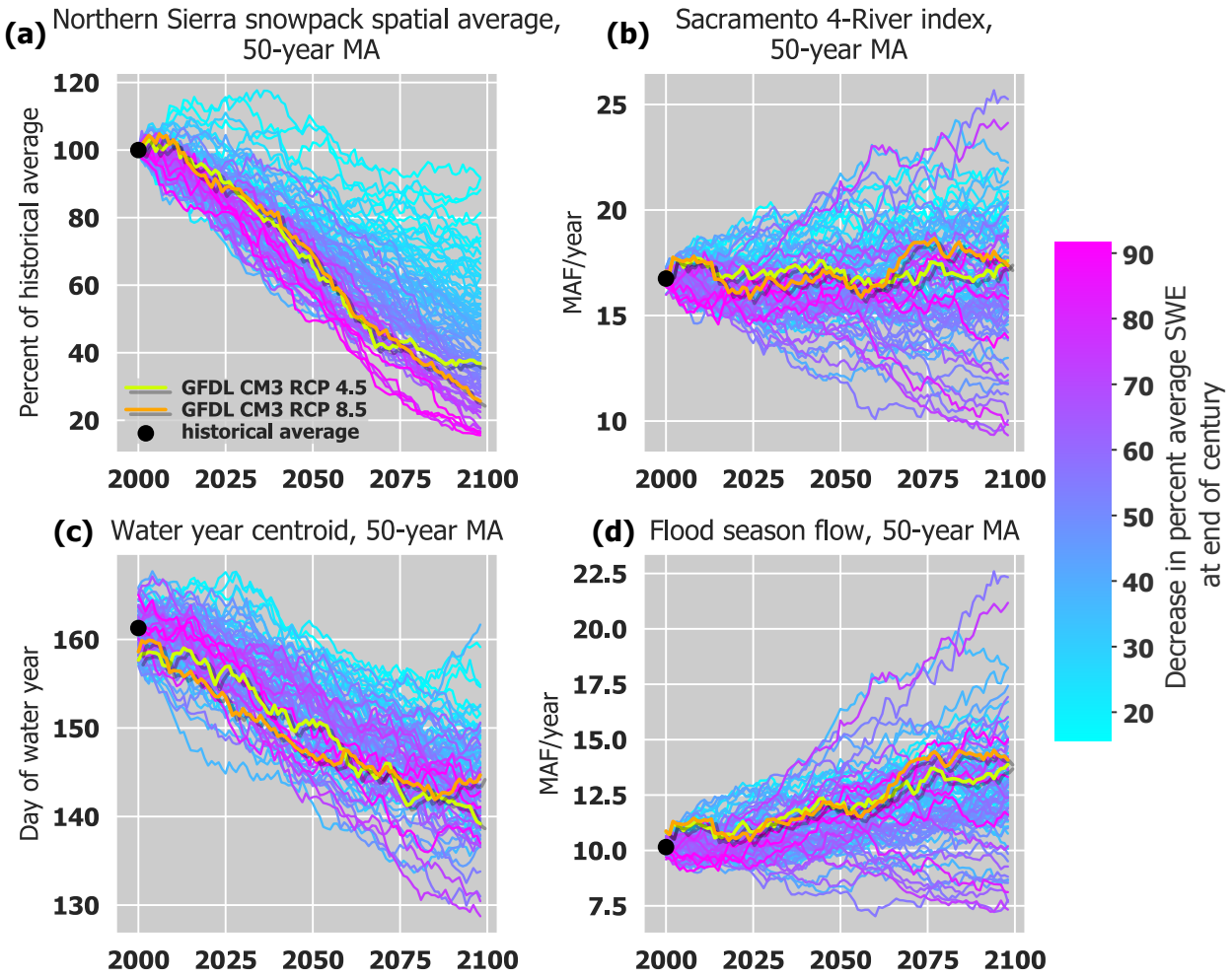

Figure 1.3: 50-year moving averages of CMIP5 scenarios for (a) snowpack in the northern Sierra Nevada, (b) streamflow in the four tributaries of the Sacramento River, (c) the water year centroid, defined as the day of the water year at which half of the total annual streamflow has been observed, and (d) the annual flood season flow (November - April) in the same four tributaries. The highlighted scenarios represent those presented in Figures 1.1 (RCP 8.5) and 1.8 (RCP 4.5).

will be. Historical operations use negative $\bar{Z}^k_{wyt}$ values corresponding to high exceedance levels, representing conservative operations to balance water supply and flood control. This regression is represented by:

$$Qr_t = \beta^k_{dw_t} SWE^k_t + \alpha^k_{dw_t} + \bar{Z}^k_{wyt} \sigma^k_{dw_t} \tag{1.1}$$

In simulating historical operations, values for the parameters $\beta^k_{dw_t}$, $\alpha^k_{dw_t}$, and $\sigma^k_{dw_t}$ are determined using the 22 years of historical SWE and streamflow data. When running the climate projections, these parameters are re-calibrated each water year using a 40-year trailing moving window. The choice of a 40-year window enables a sufficient sample size while also capturing the nonstationary relationship between the predictor and the predictand over time due to changes in precipitation phase patterns.

A similar forecasting technique is used to determine the Sacramento Valley water year type (WYT), classified as either wet (W), above normal (AN), below normal (BN), dry (D), or critical (C). The water year index (WYI) used to determine these classifications is based on observed flows to date in the Sacramento river and its tributaries:

$$\text{WYI}_y = 0.4 \times Q_{\text{Apr-Jul}} + 0.3 \times Q_{\text{Oct-Mar}} + 0.3 \times \text{WYI}_{y-1} \tag{1.2}$$

The WYI is first determined in December and updated through May. Thus, much of the flows used to determine the index must be forecasted, following a similar approach to the snowpack-to-streamflow forecasts. Section 1.7.1 in the Appendix gives further detail on this forecasting method, along with the classification rules to determine the water year type based on the water year index.

1.3.2.3 Reservoir and Delta simulation

The mass balance simulation occurs after forecasts are processed. The model has six main components: Shasta, Oroville, and Folsom reservoirs, the Delta, and Harvey and Tracy

pumping plants (Figure 1.2a,b). Feedbacks between these components drive reservoir and pumping operations. For each daily timestep t, storage S_t^k in reservoir k is updated based on inflows Q_k^t, evaporative losses L_t^k, and a release u_t^k:

$$S_t^k = S_{t-1}^k + Q_t^k - u_t^k - L_t^k \tag{1.3}$$

The release is dependent on both a release target RT_t^k and release curtailment c_t^k. The release target is the maximum of three required release requirements: flood control, environmental flow, and a water demand target:

$$RT_t^k = \max(u_{t,\text{environment}}^k, u_{t,\text{flood}}^k, u_{t,\text{demand}}^k) \tag{1.4}$$

The environmental flow requirement is based on a predetermined value for each month of the forecasted water year type:

$$u_{t,\text{environmental}}^k = E_{min}^k(m_t, wyt) \tag{1.5}$$

Flood control release requirements are based on seasonal flood pool curves (USACE; 1970, 1977, 1987). A flood control index FCI_t^k is computed each day based on the previous day's FCI, inflow, and basin averaged precipitation I_t. The flood control reservation target is then chosen using the dynamic top of conservation flood rule curve f_{tocs}^k and the flood control index. The flood control release is empirically determined as 20% of the difference between the current storage and the flood control reservation target f_{tocs}^k:

$$u_{t,\text{flood}}^k = 0.2\left(S_{t-1}^k + Q_t^k - f_{tocs}^k\left\{dw_t, h_k, FCI_t^k(Q_{t-1}^k, I_{t-1}^k, S_{t-1}^k)\right\}\right) \tag{1.6}$$

The demand requirement is based on the sum of north of Delta irrigation and municipal demands NOD_t^k, Delta outflow demands $D_{out,t}^k$, as well as demands for south of Delta pump-

ing exports SOD_t^k. The general locations for each of these three demands in the model are depicted in Figure 1.2b. These demands, which are shared by the three reservoirs, depend on several states of the system, including the month, water year type, storage in each of the three reservoirs, and gains G_t from other inflows to the Delta, which are estimated empirically. See Sections 1.7.2.1 and 1.7.3 in the Appendix for further explanation of this process. The total demand calculation for each reservoir is represented as:

$$u_{t,\text{demand}}^k = NOD^k(wyt, m_t) + SOD^k(m_t, G_t, S_{t-1}^{k=1}, S_{t-1}^{k=2}, S_{t-1}^{k=3}) + D_{out}^k(wyt, m_t) \tag{1.7}$$

Delta export volumes depend on a complex combination of environmental requirements and water demands, which we aim to replicate to the extent possible in the simulation model. The first goal is achieving a target Delta outflow $D_{out,t}^k$ to maintain adequate salinity levels in the Delta. The second objective is to pump exports south of the Delta: HRO_t for Banks pumping plant and TRP_t for Tracy pumping plant. Pumping targets and limits depend on storage in each projects' reservoirs, forecasted reservoir inflows, and river flows throughout the Delta (SWRCP; 2000a; NMFS; 2009). Overall, the Delta management interacts with reservoir releases in a feedback loop between pumping HRO_t and TRP_t, and demands $SOD^k(m_t, G_t, S_{t-1}^{k=1}, S_{t-1}^{k=2}, S_{t-1}^{k=3})$ to balance all of these requirements. The pumps will export as much of the target demand SOD^k function output as is allowed, given Delta outflow constraints.

A carryover storage target C_{wyt}^k at the end of the water year is set both for water supply and cold-pool storage to allow adequate release temperatures and storage through water years (DWR; 2017). If forecasted inflows for the rest of the year will not meet the carryover target C_{wyt}^k given projected rest-of-year releases $\sum_{d=t}^{t+365-dw_t} RT_d$, then releases are curtailed by a fraction c_t^k in order to meet this carryover target based on the available forecast. This process

serves as another integration of reservoir operations and snowpack-to-streamflow forecasts:

$$c^k_{t,5\leq m<10} = \min\left(1, \max\left\{\frac{\left(\beta^k_{dw_t} SWE^k_t + \alpha^k_{dw_t} + \bar{Z}^k_{wyt}\sigma^k_{dw_t}\right) + S^k_{t-1} - \sum_{d=t}^{t+365-dw_t} RT_d}{C^k_{wyt}}, c^k_{\max,wyt}\right\}\right) \quad (1.8)$$

Where the term $\left(\beta^k_{dw_t} SWE^k_t + \alpha^k_{dw_t} + \bar{Z}^k_{wyt}\sigma^k_{dw_t}\right)$ represents the rest-of-year inflow forecast. In the case that a perfect forecast is used, this term is replaced by the actual remaining rest-of-year inflow, enabling maximum available deliveries for water supply and Delta outflow given carryover target constraints.

The release for each reservoir is equal to the target release RT^k_t times the curtailment factor c^k_t:

$$u^k_t = RT^k_t \times c^k_t \quad (1.9)$$

See Sections **??** and 1.7.3 in the Appendix for a more detailed description of Reservoir and Delta management policies used in ORCA.

1.4 Computational experiment

1.4.1 Vulnerability assessment

ORCA is evaluated in parallel with each of the 97 downscaled climate scenarios in the USBR CMIP5 ensemble over the simulated period 1950-2099 as inputs. These initial runs include the same policies and parameters in the model uses to replicate historical operations. The outputs from these model runs serve as a baseline to highlight vulnerabilities that occur to many objectives in the system. These experiments are summarized in Table 1.3.

The analysis of these top down model runs isolates potential vulnerabilities that occur as snowpack decline levels become more severe. These vulnerabilities are thus separated from those that are more correlated with changes in streamflow levels. We compare rolling averages of carryover storage, minimum annual storage, water supply shortages, and Delta outflow with rolling averages in total annual streamflow and maximum annual SWE to find

Vulnerability	Symbol	Aggregation
Carryover storage	$S^k_{dw_t=365}$	End of water year
Minimum reservoir storage	$\min(S^k_{dw_t=1}, S^k_{dw_t=2}, \ldots S^k_{dw_t=365})$	Annual minimum
Water supply shortage	$\max(Bmax_t - SODcv_t, 0)$	Net annual
	$\max(Tmax_t - SODsw_t, 0)$	Net annual
Delta outflow	$Dout_t$	Net annual
Flood risk/maximum outflow	$\max(u^k_{dw_t=1}, u^k_{dw_t=2}, \ldots u^k_{dw_t=365})$	Maximum annual
Water year type forecast error	$WYT(\bar{Z}_{WYI})$	Cumulative classifications, perfect versus actual

Table 1.3: Summary of experiments performed for the vulnerability assessment. See the Appendix for more detailed explanation of variables.

Adaptation	Symbol	Impacts on
Floodpool shift	h^k	Carryover storage, shortage reductions
Inflow forecast exceedance	$\bar{Z}^k_{wyt}$	Reliability

Table 1.4: Summary of experiments performed for the adaptation study.

correlations between vulnerabilities to system objectives and hydrologic changes. Through this, we are able to separate which vulnerabilities are associated with snowpack decline and which are correlated with changes in streamflow. This influences the adaptation study by assigning the objectives to target in design of adaptations and objectives to analyze as the adaptations are implemented.

Along with system objective outputs, vulnerabilities in snowpack-to-streamflow forecasting are explored. Given a novel proposed adaptation directly related to changes in these forecasts, this becomes a necessary analysis. We compare perfect water year type forecasts with actual forecasts using snowpack and streamflow trends in climate scenarios to examine trends in water year type forecasting error through the century. By analyzing the changes in patterns of mis-forecasts between water year types through time, the vulnerabilities to accurate forecasts are pinpointed. This allows insight into potential alterations to forecasts as part of the adaptation study. These adaptations and their targeted impacts are summarized in Table 1.4.

1.4.2 Targeted adaptation

Two primary operational adaptations are considered in this study based on results of the vulnerability assessment: modifying the seasonal flood pool curve and the snowpack-to-streamflow forecasting method. These adaptations are hypothesized to directly address the challenges of seasonal reservoir management resulting from snowpack decline. We enumerate over parameters of these adaptations to explore their ability to mitigate vulnerabilities.

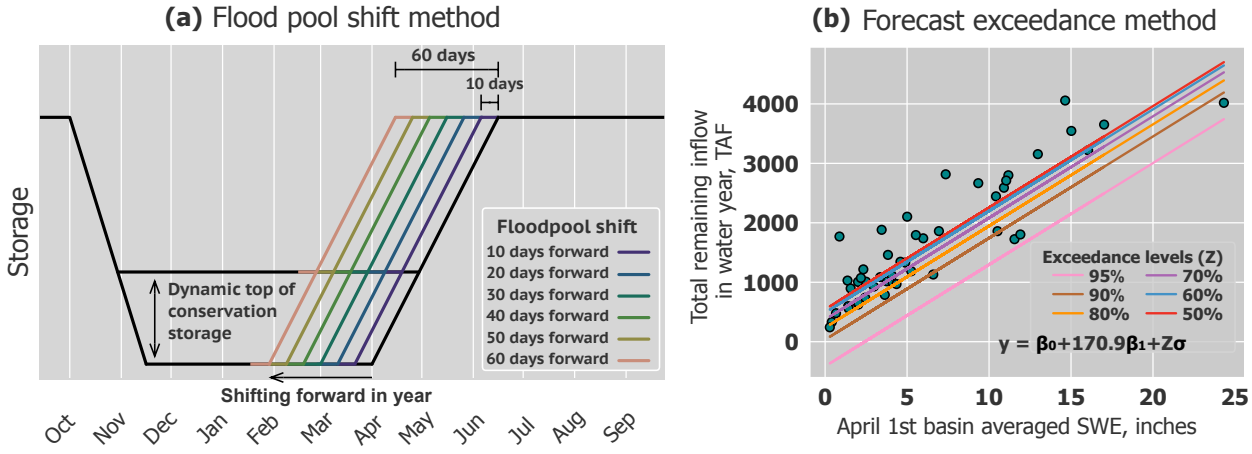

Figure 1.4: (a) Illustrative description of the flood pool shift adaptation method. The refill period is shifted earlier in the season by 10 day increments to enumeratively test the adaptation's effects on carryover storage and water supply. (b) Exceedance adaptation method for seasonal snowpack-to-streamflow forecasts.

1.4.2.1 Flood pool adaptation

As snowpack declines later in the century, current seasonal flood pool regulations may prevent refilling reservoir storage at the start of the irrigation season (Figure 1.1a). This adaptation proposes shifting the allowed refill period earlier in the year to increase reservoir storage in the absence of snowpack. Specifically, we shift the refill period earlier in 10 day increments, ranging from 10 to 60 days (Figure 1.4a). This is denoted by h_k, now an input to the flood control function in Equation 1.6. For each increment, we run each scenario of the ensemble through the model in parallel. The drawdown period and dynamic depth of the flood pool remain unchanged in this experiment. In the enumeration step, this results

in six separate parallel runs of ORCA with the input ensemble. We then compare results from the adaptation runs to the benchmark case to investigate whether the altered policies improve carryover storage and agricultural water supply without increasing flood risk.

1.4.2.2 Forecast adaptation

We consider adaptations for the statistically-based seasonal forecasts, whose trends in errors to possibly impact system performance were examined in terms of water year type forecasting. The improvements created by these adaptations are compared to perfect forecasts, where water year indices and rest-of-year inflows are assumed to be known with certainty. Simulations are run using the perfect forecasts of water year types along with perfect forecast of the rest-of-year inflow. To do this, we first consider benefits of using a perfect forecast in simulations. We then compare system outputs with a 99% exceedance level to those with a perfect forecast across all scenarios. These results are then used to observe the benefits of a perfect forecast for mitigating water supply shortages, and how these benefits differ between the first and second half of the century.

After determining the benefits of a perfect water year type and perfect inflow forecast across the ensemble, adaptations to the forecasting methods are explored. This is done to examine whether changing the forecast exceedance levels, $\bar{Z}^k_{wyt}$, can approximate the benefits from the perfect forecast. We first enumerate over several values of this parameter ranging from 50-95% (Figure 1.4b), holding them equal for each reservoir and each water year type. This is done separately for the first and second half of the century, to explore how patterns in agricultural water supply reliability by adaptations to seasonal forecasting methods will change as hydrology changes in the future.

1.5 Results

1.5.1 Vulnerability assessment

For the vulnerability assessment, results of reservoir carryover storage, minimum annual storage, water supply shortage, Delta outflow, flood risk, and forecast error are analyzed, assuming that operational rules remain unchanged in the future.

1.5.1.1 System objectives

Time series of the 50-year moving average of four system objectives for each scenario are displayed in Figure 1.5. The snowpack decline levels, in percent of historical average by end-of-century, are displayed for each scenario by color gradients. This indicates whether trends in objective values are significantly correlated with long-term snowpack decline in order to isolate this effect from other more uncertain hydrologic impacts of climate change. Table 1.5 presents these correlations and P-values to analyze if each specific system response is more correlated with streamflow or snowpack changes.

In Figure 1.5a, scenarios with greater snowpack decline will tend to have lower carryover storage progressing through the century. This is a key indicator that reduced snowmelt-driven spring and summer inflows will inhibit the abilities of reservoirs to refill after the flood season under current operating constraints. As reservoir releases occur during the summer season for irrigation and environmental purposes, the lack of snowmelt fed-inflows will cause relatively low storage by the end of the water year. This is further confirmed by the significant correlation between snowpack decline and reservoir carryover storage loss (Table 1.5). Interestingly, there is no significant correlation between long-term streamflow changes and carryover storage, suggesting that snowpack decline is the primary driver of this vulnerability, which is independent from changes in total runoff. The same pattern is present for minimum reservoir storage, which typically occurs near the end of the water year and thus aligns with carryover storage.

We quantify flood risk using the maximum annual daily reservoir outflow. Flood risk has a high positive correlation with streamflow, as it would be expected that futures with higher water availability also have higher peak flows in the flood season. However, these is little correlation with snowpack, denoting that streamflow is the major driver for this vulnerability. Not shown in Table 3 is that flood risk has a high Pearson correlation with water supply shortages, 0.74 (P=0). This is due to the fact that earlier streamflows will induce larger reservoir releases and during the beginning of the flood season, causing less storage available for deliveries into the irrigation season. This relationship highlights the projected increasing conflict between water supply and flood control.

Water supply shortages are more strongly correlated with streamflow changes than snowpack. . As would be expected, futures with lower overall water availability will yield the greatest shortages in water supply. However, their correlation with snowpack decline is also statistically significant The time series in Figure 1.5c show that, in general, shortages increase throughout the ensemble as the time horizon moves forward. Snowpack decline may have some effect on this, as it has been shown to be one of the climate change outcomes that is more prevalent throughout the majority of scenarios. For net annual Delta outflow, there is no correlation with snowpack decline. As expected, it is highly correlated with streamflows, as the majority of water entering the system flows out of the Delta to meet environmental and salinity requirements, and often includes flood pulses that exceed reservoir storage capacity. Lastly, not shown in Table 1.5 is that carryover storage is significantly correlated with water supply shortages, with a correlation coefficient of -0.73 and P-value of ~ 0 . This denotes that as carryover shortages decrease in scenarios in the ensemble, water supply shortages will increase, an important operational tradeoff that may be partially mitigated with revised policies.

Objective	Snowpack Pearson r	Snowpack P-value	Streamflow Pearson r	Streamflow P-value
Carryover storage	0.72	0	0.066	0.53
Minimum storage	0.69	0	0.017	0.87
Shortage	-0.35	5e-4	-0.63	0
Delta outflow	0.32	1.3e-3	0.96	0
Flood risk/maximum outflow	0.21	0.43	0.81	0

Table 1.5: Pearson-r correlation coefficients and P-values for each of the four objectives in the vulnerability assessment compared with snowpack and streamflow trends.

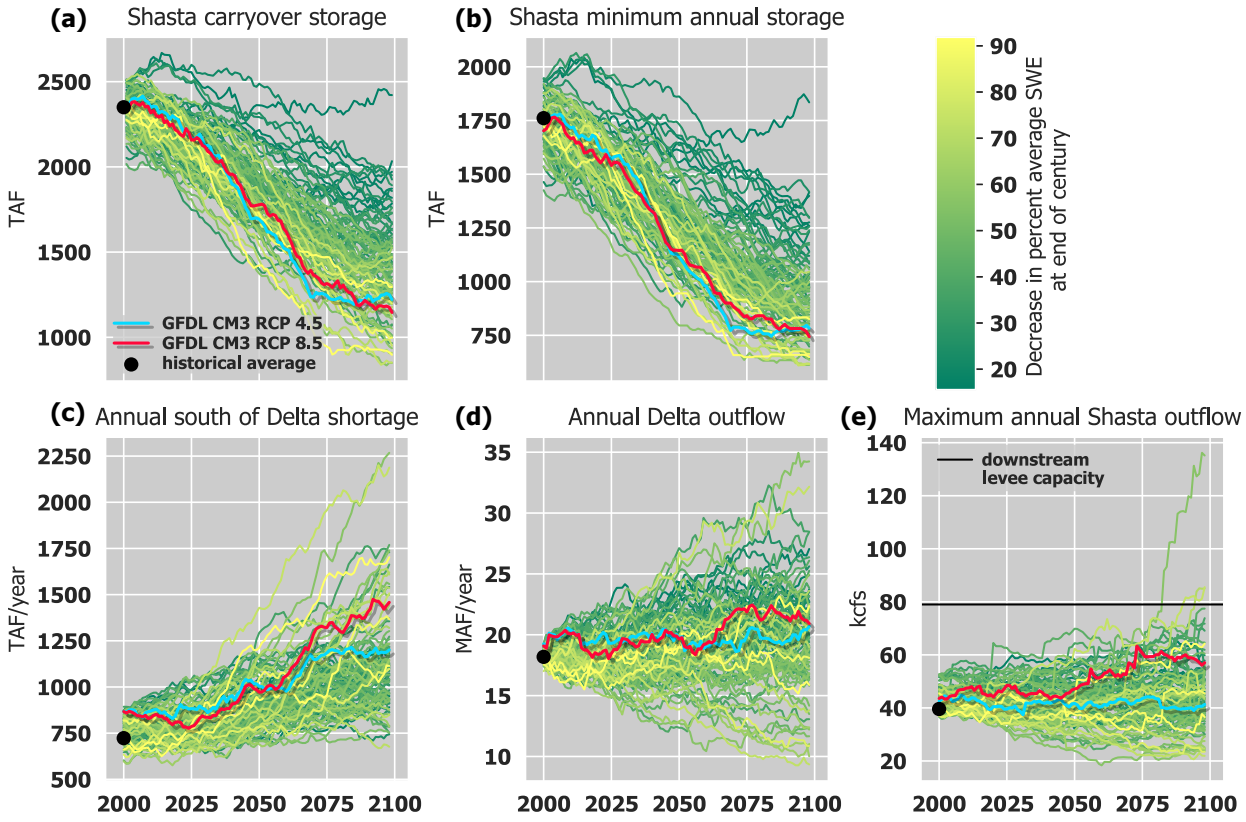

Figure 1.5: Time series of system objectives for each scenario in the ensemble used in the vulnerability assessment. These are displayed as 50 year moving averages of various transformations of these system outputs. This includes (a) Shasta storage on the last day of the water year, (b) Shasta minimum daily shortage in a water year, (c) total water supply shortages in a water year, (d) total Delta outflow for a water year, and (e) maximum annual outflow, a metric to quantify flood risk . The highlighted scenarios represent those presented in Figures 1.1 (RCP 8.5) and 1.8 (RCP 4.5)

1.5.1.2 Forecast errors

We also analyze water year type forecasting errors as part of the vulnerability assessment. The accuracy of these water year type forecasts–made on April 1st of each year–can be visualized with a confusion matrix (Figure 1.6a,b,c). These initial matrices represent all 150 years across each of the 97 scenarios in the ensemble, yielding a total of 14,550 water year type forecasts. The scenarios are divided into three 50-year periods, where each cell of the confusion matrix represents the percentage of years in which the combination of actual and perfect forecasts occurred. The distribution of correct water year types within each of the 50-year segments are presented in Table 1.6. In general, the extreme water year types (critical and wet) tend to be the most prevalent in the future climate projections. Over time the percentage of critical years increases, while the remaining water year types decrease slightly. These findings are consistent with those found in Null and Viers (2013). Based on their frequency as well as their impact to system operations, critical and wet years remain the most important to predict.

	Critical	Dry	Below Normal	Above Normal	Wet
1950-1999	19%	16%	17%	21%	27%
2000-2050	25%	15%	18%	16%	26%
2050-2099	29%	15%	16%	15%	25%

Table 1.6: Distribution of water year types over the full ensemble for the three 50 year time periods.

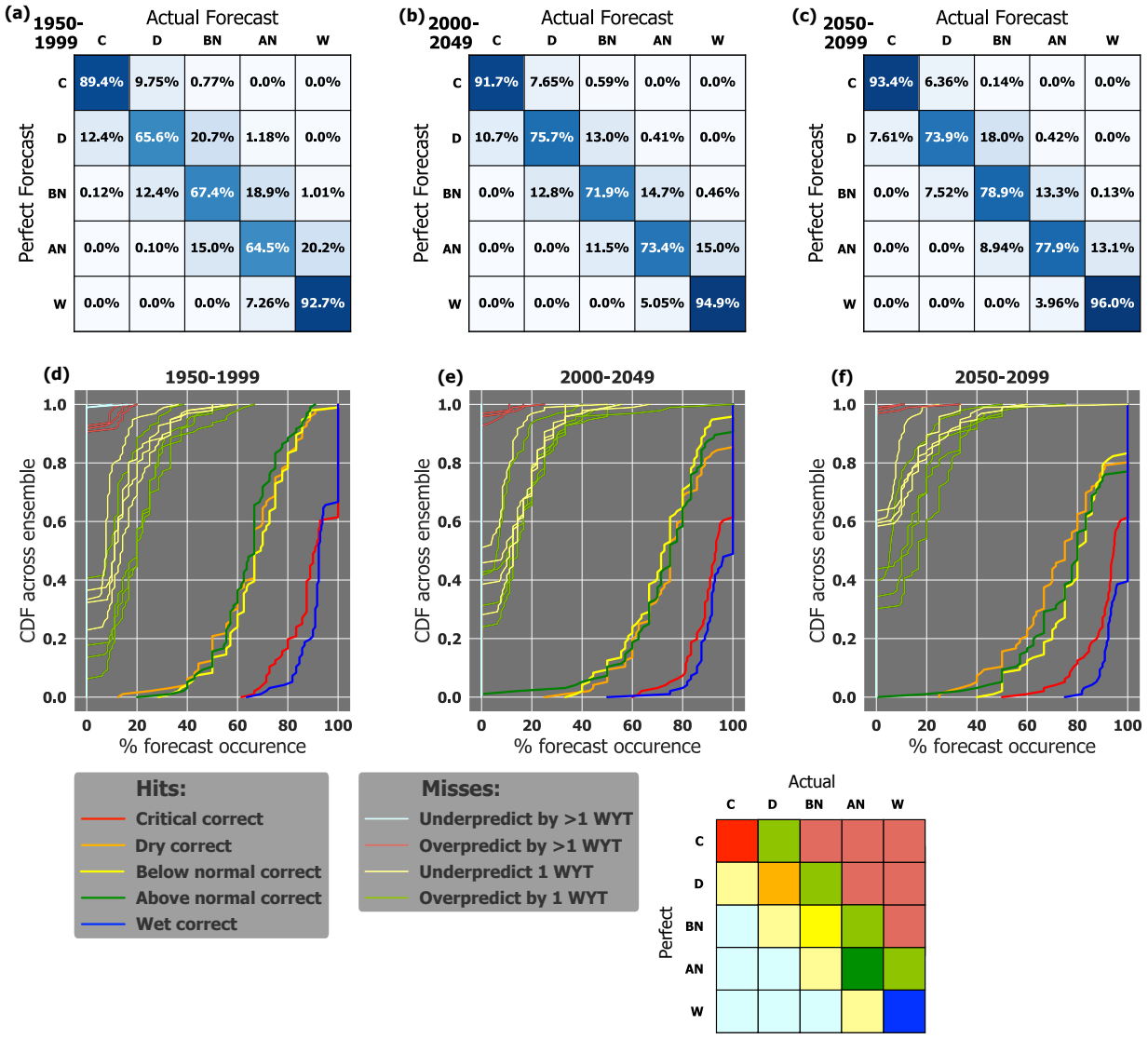

Figure 1.6: (a,b,c) Confusion matrices showing the distribution of average forecast performance for the full ensemble. Forecast accuracy trends across climate ensemble and through time as group of CDFs, each containing 97 values normalized on the y-axis. Water year type abbreviations correspond to Critical, Dry, Below Normal, Above Normal, and Wet, respectively.

For the confusion matrices, correct predictions on the diagonal tend to increase progressively over the three periods (Figure 1.6a,b,c). This can be attributed to two aspects of the forecasts. First, the 40-year moving window preserves the relationship between the predictor and predictand, which prevents forecasts accuracy from deteriorating. Second, the seasonal streamflow shift actually improves the accuracy of water year types. This occurs because a large proportion of the water year index calculation relies on the streamflow from October

through March. The fraction of annual flow during this period rises as streamflow shifts earlier in the year. Thus, the influence of streamflow already observed will increase, while the impact of inaccurate forecasts made in April is minimized. This leads to an overall increase in the April water year type forecast accuracy over time. However, the result in Figure 6 only applies to April forecasts, and it is expected that forecasts made earlier in the water year will degrade due to the loss of snowpack (Livneh and Badger; 2020).

While the confusion matrices show the trends in forecast error through time in terms of fractions of the ensemble, the distribution of each cell across the 97 scenarios should also be examined. Confusion matrices represent forecasts across the ensemble as group of CDFs each containing 97 values (Figure 1.6d,e,f). The x-axis represents the percentage of time that the specific correct forecasts or mis-forecast occurs. CDFs are assigned colors corresponding to correct predictions, slight overpredictions (light green), severe overpredictions (light red), slight underpredictions (light yellow), and severe underpredictions (light blue) (Figure 1.6g). This allows for visualization of the confusion matrices through both time and across the ensemble. In the ensemble from 1950-1999, correct forecasts occur for the , where only rarely does a scenario fall below 20% accuracy for any of the water year types. Of the mis-forecasted water year types, slight overpredictions are the most common during this time period, followed by slight underpredictions and severe overpredictions. Severe underpredictions are the least common of the mis-forecasted cases.

Progressing through the century (Figure 1.6e-f), the percentage of correct water year types still increases, but a small portion of scenarios may have more severe incorrect forecasts. This is seen by the elongation of the bottom tails of the correct forecast CDFs in the 2000-2049 and 2050-2099 periods. Despite overall increases in correct forecasts, there may be increasing uncertainty in forecast performance across scenarios caused by nonstationary snowpack and streamflow hydrology. Percentages of incorrect forecast CDFs also exhibit greater spread through time(Figure 1.6e-f), despite increasing averages (Figure 1.6a-c). Overall, this indicates that water year type forecasts by April 1 will generally improve

over time using the 40-year moving window, but there is still uncertainty as to whether this occurs for forecasts made earlier in the water year. Thus, there still exists a potential need for forecast adaptations.

1.5.2 Adaptation study

1.5.2.1 Flood pool adaptation

The results of the enumeration experiment for different levels of seasonal shifts in the flood pool refill period for various objectives are displayed in Figure 1.7. In the majority of scenarios, shifting the refill period earlier in the year benefits both carryover storage and agricultural water supply. Throughout the century, the ensemble mean and upper standard deviation level of the 50 year moving average of carryover storage increase in all three reservoirs (Figure 1.7a,b,c). This denotes that as streamflow shifts earlier in the year due to declining snowpack, shifting the flood pool increases the potential for carryover storage benefits. Additionally, larger shifts show more potential for carryover storage benefits in the mean and upper standard deviation, while only slightly decreasing the lower standard deviation level. The wide range of outcomes across the ensemble scenarios–with few showing reductions in carryover storage–suggests complex interactions between the intra-annual streamflow timing due to snowpack decline, and the broader trends in water availability in these downscaled scenarios, only some of which can be mitigated by shifting the refill period.

The same effect is also present for mitigating agricultural water supply shortages (Figure 1.7d). Again, the mean and upper standard deviation of the ensemble increase throughout the century as snowmelt loss becomes more severe, with greater benefits associated with larger flood pool shifts. Some compromise would have to be made between the benefits to water supply and potential flood risk brought about by a more extreme flood pool shift, which would result in more moderate shifts being implemented.

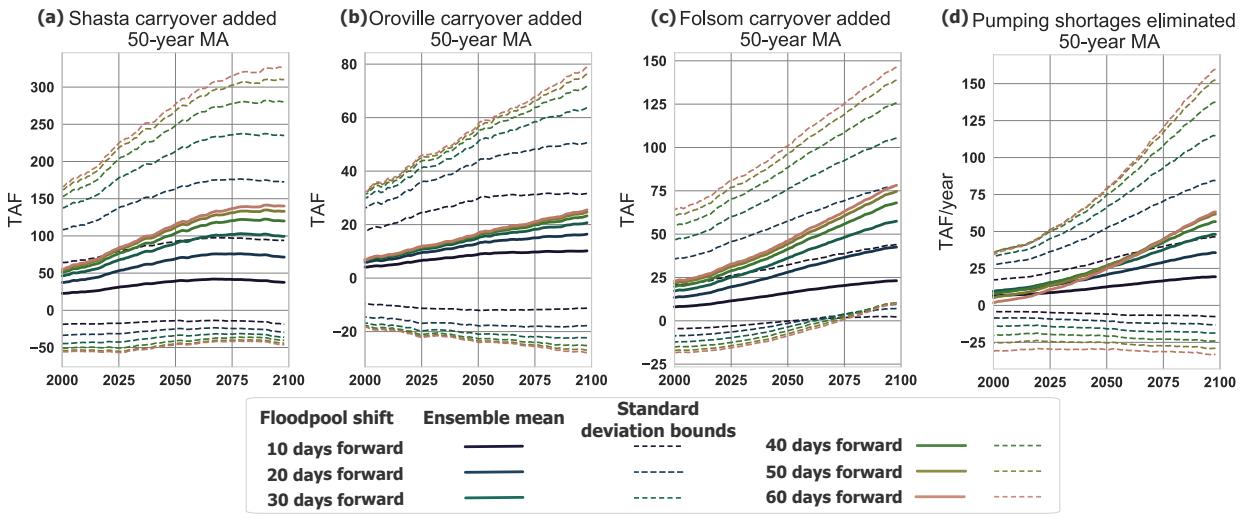

Figure 1.7: 50-year moving averages of additional carryover storage in ensemble resulting from flood pool shift for (a) Shasta, (b) Oroville, (c) Folsom. (d) Mitigated Delta export shortages from flood pool shift. Solid lines represent the ensemble mean, and dashed lines represent +/- one standard deviation.

It is also crucial to examine the effects of the flood pool shift adaptation on dynamics of the system. The behavior or these dynamics is viewed most easily through time series of reservoir storage and top of conservation targets over a single scenario (Figure 1.8). It is clear that shifting Shasta's flood pool refill period forward will increase reservoir storage throughout the water year, since dynamic top of conservation targets will become higher as the flood pool is shifted earlier in the year. This is especially evident during the refill period at the end of the flood season, and during the drawdown period through the irrigation season. The differences in this trend between years is also of importance, to see how the updated flood rule curve causes storage to respond to varying flow magnitudes and timing between water years.

The flood pool shift yields the greatest mitigation in shortages in years where runoff arrives earlier, a direct result of earlier melt and the shift in precipitation phase. In this particular scenario and time period, these include the majority of years in the decade shown. Even though these years have high flood peaks, the majority of precipitation is most likely in the form of rain rather than snowmelt, as evident by little or no flow present after April.

The flood pool shift allows the reservoir to store more inflows before April, making up for these lost spring and early summer flows. This benefits other objectives of the system, as seen by the decrease in water supply shortages from flood pool shifts (Figure 1.7d). More extreme shifts may send reservoirs to full storage during these years, with the tradeoff of increasing flood risks. However, more moderate shifts will send reservoirs back to desired levels at the end of the flood season, improving carryover storage and water deliveries while maintaining adequate flood control in the system.

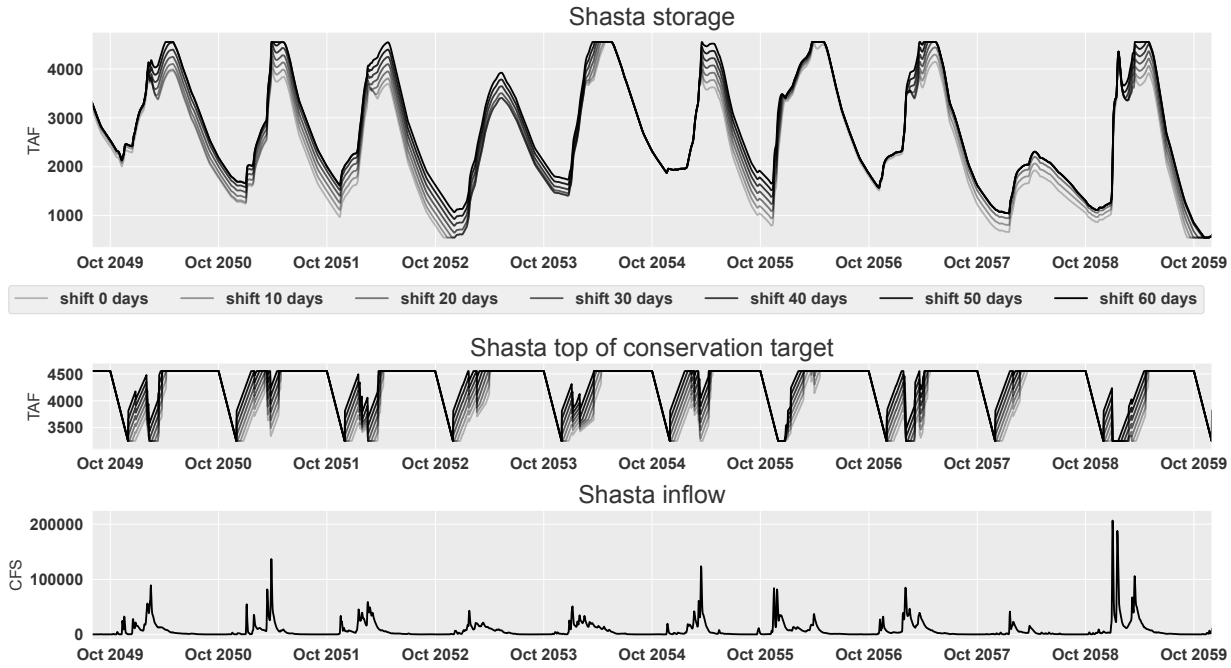

Figure 1.8: Time series of (a) Shasta storage, (b) Shasta dynamic top of conservation storage target, and (c) Shasta inflow, considering the flood pool shift adaptation for the former two. These time series come from ORCA outputs to the 2049-2059 time period in the NOAA GFDL-CM3, RCP 8.5 scenario.

In a few years in this time series, flow is more spread out in the winter and early spring seasons (specifically, these include the 2052, 2053, and 2055 water years). In these years, either too little precipitation occurs throughout the year to refill the reservoir, or more snowmelt-driven flow later in the flood season would result in higher storage regardless of flood pool shift. This shows that shifting the flood pool will not lead to large increases in

flood risk in years with relatively higher snowpack, but it will also not provide much benefit of increased storage during these years.

1.5.2.2 Forecast adaptation

We investigate the impacts of changing the exceedance level ($\bar{Z}^k_{wyt}$) of the snowpack-to-streamflow forecasts, rather than the water year index forecasts, where higher values represent a more conservative forecast. This conservative forecast is chosen as a baseline because the majority of mis-forecasts are shown to occur as overpredictions in water year type classifications. The perfect forecast enables the maximum available volume to be delivered for water supply through the irrigation season, given the constraint of meeting the carryover target. Benefits to water supply reliability from using a perfect forecast can range from 0.045 to 0.24 (Figure 1.9a). In the second half of the century, these reliability benefits decrease significantly for dry scenarios, and to a lesser degree for average and wetter scenarios (Figure 1.9c). The largest forecast benefit value occurs in average scenarios, where there is sufficient water available to meet demands, but only if it is managed well using the forecast. The perfect forecasts also give less improvement in the scenarios with high snowpack decline, regardless of the change in streamflow magnitudes (Figure 1.9a,c). This highlights the fact that reliability of the system is quite vulnerable to snowpack decline, and that perfect forecasts may have less potential to improve performance in the most extreme scenarios in terms of reduced snowpack.

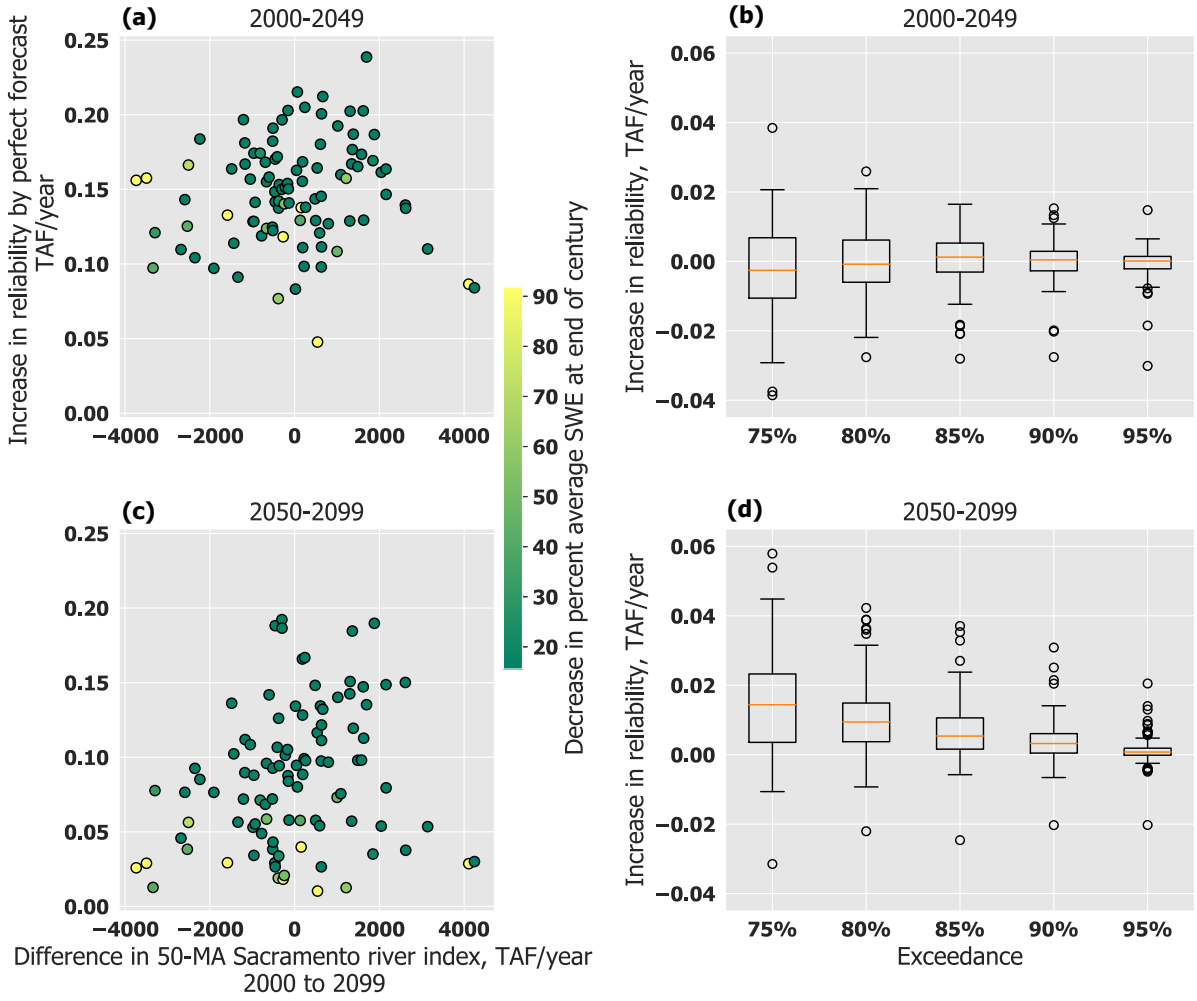

Figure 1.9: (a,c) Benefits to agricultural water supply reliability when using a perfect forecast compared to a 99% exceedance forecast. Each point represents the reliability increase for an individual scenario in the CMIP5 ensemble (y axis). The x-axis represents the difference over the century in the 50-year moving average of the Sacramento River and its three largest tributaries. This statistic is shown in Figure 1.3b as well. (b,d) Reliability increase from lowering exceedance levels, displayed as a boxplot for the whole ensemble.

Given that the perfect forecast provides advantages to water supply in all scenarios, we enumerate over the exceedance levels used in the forecasting method in an attempt to approach these benefits (Figure 1.9b,d). We then examine reliability changes as the exceedance level decreases (i.e., forecasts become less conservative). For the first half of the century, the effect of lowering forecast exceedance shows no clear pattern across the ensemble (Figure

1.9b), and the mean reliability change stays near zero. However, in the second half the century, lower exceedance levels tend to increase reliability in agricultural water supply (Figure 1.9d). This shows that with further hydrologic changes, there may be some possible benefits to water supply by making less conservative (low exceedance) reservoir inflow forecasts. This would cause higher projections of end-of-year carryover storage, eliminating unnecessary curtailments and providing flexibility in the system to allow excess releases to the Delta for water supply. However, the spread tends to increase as the exceedance is lowered, leading to more uncertainty in reliability as forecasts become less conservative. Fortunately, this increasing spread and increasing mean suggest a tendency toward larger deliveries in crucial months for water supply and irrigation.

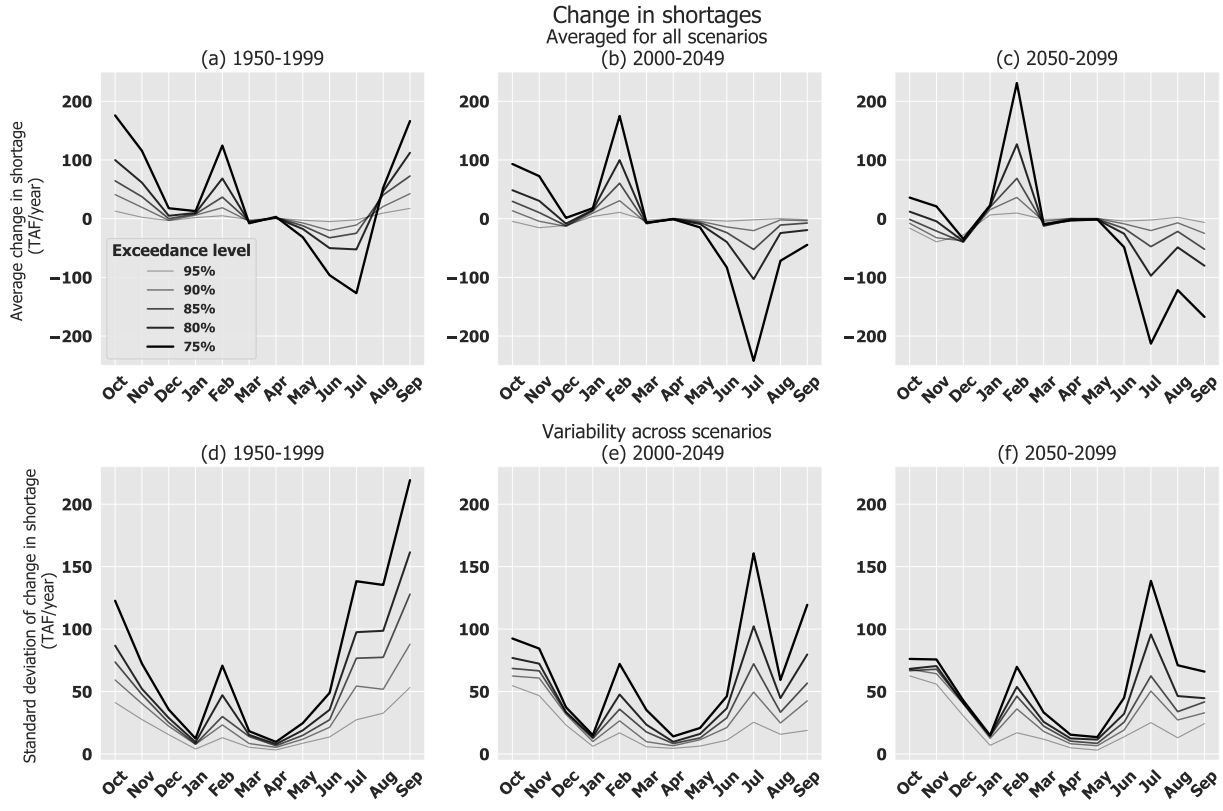

Figure 1.10: Top row: mean of scenario changes in shortage by month for exceedance level adaptation for the simulated time periods (a) 1950-1999, (b) 2000-2049, and (c) 2050-2099. Bottom row: Standard deviations of these changes across the ensemble scenarios for (d) 1950-1999, (e) 2000-2049, (f) 2050-2099.

While uncertainty does exist in the outcomes of forecast adaptations through the ensemble, analyzing the intra-annual dynamics of the altered forecasts provides insight into the timing of how forecast alterations impact system objectives, specifically water supply deliveries. Figure 1.10a-c show the average monthly shortage change across three time periods: 1950-1999, 2000-2049, and 2050-2099. Progressing through the time periods, the mean shortage change for each month becomes more extreme across all exceedance levels. In all three time periods, shortages decrease in the months of May through July. In these months, having a less conservative forecast will result in less curtailments, thus increasing water supply deliveries. Along with this, large storage in reservoirs increases significantly in February as more inflows must be captured, rather than released for flood control, to make up for the larger storage losses in the irrigation season. In the 1950-2000 time period, a more conservative forecast is potentially detrimental as it can increase shortages in August and September (Figure 1.10a). In this case, the overly-conservative forecast will cause an unnecessary increase in summer curtailments regardless of reservoir storage while snowmelt-fed inflows are still present. However, this adaptation is shown to be beneficial later in the century as the hydrology changes. Overall, each of the discussed patterns become more prevalent as forecasts become less conservative. In conclusion, this denotes that raising the forecast exceedance level (more conservative forecasts) will mitigate some of the intra-annual shifts in water supply shortage caused by snowpack loss further into the 21st century.

While these patterns are explained through the mean of the ensemble, there still exists significant variability in shortage changes across scenarios. Months with the largest change in mean shortage also show the greatest variability in changes (Figure 1.10e,f,g). This denotes uncertainty in the magnitude of these monthly changes across scenarios. The variability decreases in later time periods due to the lower snowpack, which will make patterns in forecast results more similar across scenarios. For the months with large changes (irrigation season and February), the standard deviations are less than the absolute value of the mean for their respective exceedance levels in all three time periods. Thus, for the majority of

scenarios, the direction of change in shortage will remain the same as that shown in the mean changes. In the last two time periods, irrigation season shortages will mostly decrease, while February shortages will generally increase across scenarios.

While much uncertainty exists in the net changes of annual shortages and reliability from the forecast adaptation (Figure 1.9), the effects of the forecast on intra-annual shortage changes is clear (Figure 1.10). Shortages in the irrigation season can have much more of a negative impact on system performance. Therefore, adapting forecast methods leads to reservoir operations that shift shortages from the irrigation season to the flood season. Even if the effect of this on overall annual shortages would be uncertain, significant decrease in summer shortages would have benefits to the system objectives, especially water supply.

1.6 Discussion and Conclusion

This paper contributes an approach to couple a top-down climate vulnerability assessment and adaptation study to isolate and adapt to specific physical impacts of climate change projected with confidence, namely snowpack decline, while also designing adaptations that respond to more uncertain impacts in total water availability. These methods contribute to the literature on top-down climate change adaptation in water resources systems while also providing adaptation policies generalizable to snowmelt-dominated systems in the Western U.S. and elsewhere. While this study focuses on the effects of snowpack decline, the approaches can be extended to isolate and adapt to impacts of many well-predicted aspects of climate change on water resources systems.

In a top-down approach, an analysis of the response of a reservoir model to an ensemble of downscaled climate scenarios cannot link vulnerabilities to specific hydrologic parameters unless these relationships are identified explicitly. While the relationships can be extrapolated from perturbed uncertainties in bottom-up studies, top-down approaches must discover them by leveraging physical processes and transient trends present in climate projections.

Through a statistical analysis, we accomplish this given the response of a model simulating the northern California reservoir system to an ensemble of climate scenarios. Considering snowpack and annual streamflows, results show that reservoir storage vulnerabilities are significantly correlated with snowpack decline, while environmental flows are significantly correlated with streamflow changes. Water supply shortages are correlated with both—more so with total streamflow changes, but also linked to snowpack decline due to the influence from reservoir operations. The transient trends in climate scenarios also indicate how these vulnerabilities change through time.

After analyzing vulnerabilities, the proposed adaptations are targeted to the specific impacts of snowpack decline, including seasonal streamflow shifts. This provides insight into which system outputs to monitor to identify the most significant changes caused by adapting system operations. The detailed long-term dynamics of the simulation outputs under climate change allow for further analysis of adaptations. For example, the upward trends in carryover storage from the flood pool adaptation through the end of the century in the majority of scenarios, as snowpack decline becomes more severe. The analysis of perfect forecasts benefits from isolation of snowpack decline levels and annual streamflow changes in each scenario. This utilizes the many hydrologic outputs from climate projections to show the adaptation's general performance related to both more widely predicted changes and to those that are more uncertain. The analysis gives insight for breaking down the uncertainty related to the adaptation's performance across an uncertain ensemble. Transient trends in climate projection output allow for the adaptations to be analyzed over multiple time horizons given varying extents in magnitudes of hydrologic changes. Lastly, we show that the intra-annual system dynamics of adaptations must be analyzed to gain better understanding of their effects on system objectives and vulnerabilities.

Going forward, this study can be extended in two ways. The first involves combining adaptations to forecasts and operations to improve robustness to uncertainty across the climate ensemble using formal policy search techniques to generate near-optimal adaptations

in multiple objectives. This is the subject of ongoing work. Second, the transient trends in these climate scenarios, particularly in streamflow, snowpack, and seasonal streamflow shifts, present an opportunity for dynamic adaptation. While this study analyzed the effects of these adaptations in different time periods throughout the century, there exists an opportunity to identify the conditions under which adaptations to operating rules should be implemented. This problem lends itself to a dynamic adaptation study considering multiple reversible operating policies, recognizing that certain hydrologic impacts are projected with higher certainty than others. Furthermore, there may be additional policies that can mitigate vulnerabilities to climate change beyond just snowpack decline. Examples of these adaptations include optimal combinations of the individual adaptations considered in this study, flood control operations utilizing short-term precipitation forecasts (e.g., Nayak et al.; 2018), and increased conjunctive use (e.g., Kourakos et al.; 2019). Overall, this study contributes an approach to targeting water system adaptations to specific physically-based impacts of climate change identified in a vulnerability assessment. The coupled approach to vulnerability assessment and adaptation is generalizable to other snowmelt-dominated water resources systems facing the loss of seasonal storage due to rising temperatures.

1.7 Appendix

1.7.1 Snowpack-to-streamflow forecasting

1.7.1.1 Rest-of-year inflow forecasting

The first component of the Operations of Reservoirs in California (ORCA) model is to forecast inflows before the simulation is performed. The snowpack-to-streamflow forecast is represented by the following regression:

$$\mathbf{y}_n = \beta_{m,n}\mathbf{X}_m + \alpha_{m,n} + \varepsilon \tag{1.10}$$

Where:

m = current month

n = month forecasting

y_n = foretasted flow remaining water year after month n

X_m = maximum snow water equivalent (SWE) in water year up to month m

$\beta_{m,n}$, $\alpha_{m,n}$ = regression coefficients

ε = forecast error

The same formulation is used for remaining flow forecasts after day d of the water year:

$$\mathbf{y}_d = \beta_d \mathbf{X}_d + \alpha_d + \varepsilon \tag{1.11}$$

The forecast formulation is rewritten as follows:

$$F_t = \sum_{365}^{dw_t} Q_t = \theta_{dw_t} SWE_t + \epsilon_{dw_t} + \bar{Z}_{wyt} \sigma_{dw_t} \tag{1.12}$$

Where θ_{dw_t} and ϵ_{dw_t} are the regression coefficients.

1.7.1.2 Water year type forecasting

A similar forecasting technique is used to determine the Sacramento Valley water year type, classified as either wet (W), above normal (AN), below normal (BN), dry (D), or critical (C). The water year index (WYI) used to determine these classifications is based on observed and forecasted flows in the sum of flows in the Sacramento, Feather, Yuba, and American rivers:

$$\text{WYI}_y = 0.4 \times Q_{\text{Apr-Jul}} + 0.3 \times Q_{\text{Oct-Mar}} + 0.3 \times \text{WYI}_{y-1} \tag{1.13}$$

This procedure is repeated each month in the water year from December through May. Thus, much of the flows in this calculation must be forecasted based on snowpack levels. This is done by using the same inflow forecast procedure from the previous section, combined with the water year index equation (Equation 1.13):

$$WYI_{y,m} = \begin{cases} 0.3WYI_{y-1,9} + 0.3\left[\sum_{i=10}^{m} Q_i + \sum_{i=m+1}^{12}\left(\mu_i + \bar{Z}\sigma_i\right) + \sum_{i=1}^{3}\left(\mu_i + \bar{Z}\sigma_i\right)\right] \\ \qquad +0.4\left[\sum_{i=4}^{7}\left(\mu_i + \bar{Z}\sigma_i\right)\right], & m \in \{10,11\} \\ 0.3WYI_{y-1,9} + 0.3\left[\sum_{i=10}^{11} Q_i + Q_m + \sum_{i=1}^{3}\left(\mu_i + \bar{Z}\sigma_i\right)\right] \\ \qquad +0.4\left[\sum_{i=4}^{7}\left(\theta_{m,i}SWE_m + \epsilon_{m,i} + \bar{Z}\sigma_i\right)\right], & m \in \{12\} \\ 0.3WYI_{y-1,9} + 0.3\left[\sum_{i=10}^{12} Q_i + \sum_{i=1}^{m} Q_i + \sum_{i=m+1}^{3}\left(\mu_i + \bar{Z}\sigma_i\right)\right] \\ \qquad +0.4\left[\sum_{i=4}^{7}\left(\theta_{m,i}SWE_i + \epsilon_{m,i} + \bar{Z}\sigma_i\right)\right], & m \in \{1,2\} \\ 0.3WYI_{y-1,9} + 0.3\left[\sum_{i=10}^{12} Q_i + \sum_{i=1}^{2} Q_i + Q_m\right] \\ \qquad +0.4\left[\sum_{i=4}^{7}\left(\theta_{m,i}SWE_i + \epsilon_{m,i} + \bar{Z}\sigma_i\right)\right], & m \in \{3\} \\ 0.3WYI_{y-1,9} + 0.3\left[\sum_{i=10}^{12} Q_i + \sum_{i=1}^{3} Q_i\right] \\ \qquad +0.4\left[Q_m + \sum_{i=m+1}^{7}\left(\theta_{m,i}SWE_i + \epsilon_{m,i} + \bar{Z}\sigma_i\right)\right], & m \in \{4\} \\ 0.3WYI_{y-1,9} + 0.3\left[\sum_{i=10}^{12} Q_i + \sum_{i=1}^{3} Q_i\right] \\ \qquad +0.4\left[Q_4 + Q_m + \sum_{i=m+1}^{7}\left(\theta_{m,i}SWE_i + \epsilon_{m,i} + \bar{Z}\sigma_i\right)\right], & m \in \{5\} \\ WYI_{y,5}, & m \in \{6,7,8,9\} \end{cases} \tag{1.14}$$

Where WYI_{y-1} is the previous year's water year index.

Q_i is month i's flow.

μ_i is the mean flow in month i.

σ_i is the standard deviation of month i's flow.

$\theta_{m,i}$ and $\epsilon_{m,i}$ are the regression coefficients.

The water year type is then defined based on the following water year index ranges:

$$wyt_t = \left\{ \begin{array}{ll} W & WYI_t \geq 9.2 \\ AN & 9.2 > WYI_t > 7.8 \\ BN & 7.8 \geq WYI_t > 6.5 \\ D & 6.5 \geq WYI_t > 5.4 \\ C & 5.4 \geq WYI_t \end{array} \right\} \tag{1.15}$$

1.7.2 Reservoir releases

1.7.2.1 Mass balance

The reservoir mass balance is the main engine of the simulation model. The equations in this section are applicable to each of the model's three reservoirs. The state variable, storage, is updated each daily time step via Equation 3.1.

$$S_t = S_{t-1} + Q_t - u_t - E_t \tag{1.16}$$

Where S_t is reservoir storage, Q_t is reservoir inflow, u_t is the reservoir release, and E_t evaporation. A quadratic multivariate regression based on reservoir storage elevation and temperature T_t generates evaporation values:

$$E_t = \theta_0 T_t + \theta_1 S_{t-1} + \theta_2 T_t S_{t-1} + \theta_3 T_t^2 + \theta_4 S_{t-1}^2 + \epsilon \tag{1.17}$$

Where $\theta_{0,1,...4}$ are the regression coefficients, and ϵ is the intercept. Additionally, reservoir spillage W_t is described by Equation 1.18. S_{max} is the maximum reservoir capacity.

$$W_t = \left\{ \begin{array}{ll} S_t - S_{\max}, & S_t > S_{\max} \\ 0, & S_t \leq S_{\max} \end{array} \right\} \tag{1.18}$$

Reservoir releases, the control variable, are defined as the maximum of a target release, a maximum release capacity, and dead pool constraint:

$$u_t = \max\{RT_t, S_{t-1} + Q_t - DP, Mout\} + W_t \tag{1.19}$$

Where RT_t is the release target, DP is the reservoir's dead pool and $Mout$ is the maximum possible reservoir release. The only dynamic variable of these three is the release target. The remainder of Section 1.7.2 will describe how the release target is determined.

1.7.2.2 Release targets

The release target is defined differently for each month, depending on whether releases must be curtailed for carryover storage. It is defined as the maximum of a required flood control release (see Section 1.7.2.3), total demands, and a minimum environmental flow. The release target is often curtailed by a carryover multiplier, in order to meet an end-of-water-year carryover target. The release target, for days in each month, is described by Equation 1.20.

$$RT_t = \left\{ \begin{array}{ll} \max\{FCR_t, NOD_t + SOD_t + Dout_t, Emin_t\} & m \in \{10,11,12,1,2,3,4\} \\ \max\{FCR_t, NOD_t + SOD_t + Dout_t, Emin_t\} & m \in \{5,6,7,8,9\} \wedge FCS > CT_{wyt} \\ \max\{FCR_t, NOD_t + SOD_t + Dout_t, Emin_t\} * CC_t & m \in \{5,6,7,8,9\} \wedge FCS < CT_{wyt} \end{array} \right\} \tag{1.20}$$

Where FCR_t is the required flood control release, NOD_t are north of Delta demands for the given reservoir, SOD_t are calculated south of Delta demands for the given reservoir, $Dout_t$ is the amount of Delta outflow allocated to that reservoir, $Emin_t$ is the minimum environmental flow requirement downstream of the reservoir, and CC_t is a carryover curtailment multiplier.

For months May through September, if the carryover target will not be met based on the current release from Equation 1.19 and forecasted inflows, the release is curtailed by a carryover curtailment ratio CC_t (see Section 1.7.2.4).

The minimum outflow for the Delta, $Dout_t$, has a portion allocated to each reservoir based on a required fraction of its project (either CVP or SWP):

$$Dout_t = Dmin_t * Dpct_r \tag{1.21}$$

Where $Dmin_t$ is the overall required minimum Delta outflow for either CVP or SWP, and $Dpct_r$ is the percent of that outflow to be supplied by each reservoir. Next, south-of-Delta demands SOD_t are calculated:

$$SOD_t = SODproj_t * SODprct_t * SODc_{wyt} \tag{1.22}$$

Where $SODproj_t$ is the South of Delta demands for the reservoir's project as calculated in the Delta requirements. $SODprct_t$ is percent of the south of Delta demands that the reservoir contributes to its projects. For Oroville this is 100%. For CVP reservoirs $SODc_{wyt}$ is a curtailment factor for the reservoirs south of Delta demands based on water year type (see Equations 1.56 and 1.57).

North-of-Delta Demands are dependent only on the day of the water year:

$$NOD_t = NOD_{dw_t} \tag{1.23}$$

Minimum environmental flows are dependent on the current month and water year type:

$$Emin_t = Emin_{wyt_m} \tag{1.24}$$

The release to the Delta is then formulated as:

$$RDelta_t = \left\{ \begin{array}{ll} R_t - NOD_t, & m \in \{10,11,12,1,2,3,4\} \\ R_t - NOD_t, & m \in \{5,6,7,8,9\} \wedge FCS_t > CT_{wyt} \\ R_t - NOD_t * CC_t, & m \in \{5,6,7,8,9\} \wedge FCS_t < CT_{wyt} \end{array} \right\} \tag{1.25}$$

The model assumes that north of Delta demands are met unless releases are curtailed for carryover targets. Thus the reservoirs release to the Delta, $RDelta_t$, is equal to its release minus north of Delta demands. The north of Delta deliveries are multiplied by the carryover curtailment factor if it is also used for reservoir releases.

1.7.2.3 Flood control

Flood control in the reservoirs is primarily based upon a flood control index FCI_t and a top of conservation function $f_{tocs}\{dw_t, FCI_t\}$ (USACE; 1970, 1977, 1987) . The top of conservation curve is based on the flood control index and and the day of the water year. The reservoir must the empty 20% of the difference between the current storage and top of conservation rule (Equation 1.26):

$$FCR_t = 0.2\left(S_{t-1} + Q_t - f_{tocs}\{dw_t, FCI_t\}\right) \tag{1.26}$$

Shasta's FCI is based the reservoir inflow and the previous day's FCI:

$$FCIsha_t = Qsha_t + 0.95FCIsha_{t-1} \tag{1.27}$$

FCIs for Oroville and Folsom differ in that they are based on precipitation:

$$FCIoro_t = I_t + 0.97FCIoro_{t-1} \tag{1.28}$$

Where I_t is the spatially averaged daily precipitation in the Feather River basin.

$$FCIfol_t = I_t + 0.95FCIfol_{t-1} \tag{1.29}$$

Where I_t is the spatially averaged daily precipitation in the American River basin.

1.7.2.4 Carryover storage

Using the release target and day of water year, potential carryover storage FCS_t is predicted in Equation 1.7.2.4:

$$FCS = F_t + S_{t-1} - RT_t\,(365 - dw_t) \tag{1.30}$$

If this FCS_t value is less than the carryover target CT_{wyt}, a curtailment percentage CCP_t is calculated based on that carryover target:

$$CCP_t = F_t + S_{t-1} - RT_t \left(\frac{365 - dw_t}{CT_{wyt}} \right) \tag{1.31}$$

However in drier conditions this curtailment can lead to severely low flows into the Delta. Thus a maximum curtailment factor CCP_{max} must be implemented:

$$CC_t = \left\{ \begin{array}{ll} CCP_t, & CCP_t \leq CCP_{\max_{wyt}} \\ CCP_{\max}, & CCP_t > CCP_{\max_{wyt}} \end{array} \right\} \tag{1.32}$$

1.7.2.5 Available storage calculations

The forecasted available rest-of-year storage in each reservoir is important for Delta mass balance and assignment of south of Delta demands. Available storage is the first object simulated at the beginning of each time step, as it is necessary to calculate south of Delta demands (see Section 1.7.3.2).

A cumulative minimum release CMT_t the remainder of the water year is predicted. The release target is defined by the sum of the maximum of the minimum environmental flow,

north of Delta demand, and temperature release standards for flows further downstream (TR_t), for each remaining day in the water year:

$$CMR_t = \sum_{365}^{dw_t} \max\{Emin_t, NOD_t, TR_t\} \tag{1.33}$$

$$TR_t = TR_{wyt_m} \tag{1.34}$$

Using the projected cumulative minimum release, the available storage is calculated using the carryover storage, cumulative minimum release, and inflow forecasts:

$$AS_t = \left\{ \begin{array}{ll} S_{t-1} - CT_{wyt}\bar{Z}_{wyt} + F_t - CMR_t, & S_{t-1} + F_t > CT_{wyt}\bar{Z}_{wyt} + CMR_t \\ 0, & S_{t-1} + F_t \leq CT_{wyt}\bar{Z}_{wyt} + CMR_t \end{array} \right\} \tag{1.35}$$

1.7.3 Delta rules

1.7.3.1 Mass balance

Pumping is simulated once reservoir releases have been determined. The release to the Delta from CVP reservoirs is denoted by $Rcvp_t$. It is equal to the sum of the releases to the Delta from Folsom and Shasta (see Equation 1.25).

$$Rcvp_t = Rfol_t + Rsha_t \tag{1.36}$$

Likewise, the release to the Delta for the SWP is just that from Oroville.

$$Rswp_t = Roro_t \tag{1.37}$$

Thus the total inflow to the Delta Qin_t is equal to releases from the two projects plus statistically modeled gains G_t:

$$Qin_t = G_t + Rcvp_t + Rswp_t \tag{1.38}$$

The minimum Delta outflow requirement $Qmin_t$ is based on the month and water year type:

$$Qmin_t = Qmin_{wyt_m} \tag{1.39}$$

$Qreq_t$ is a required outflow from the Delta:

$$Qreq_t = \max\left\{Qmin_t, Qin_t\left(1 - \nu_t\right)\right\} \tag{1.40}$$

Where ν_t represents an export ratio based on month:

$$\nu_t = \nu_{wyt_m} \tag{1.41}$$

The projects can pump no more than this ratio multiplied by the Delta inflow. Thus, if the remainder of the inflow that cannot be pumped is greater than $Qmin_t$, it becomes the required outflow $Qreq_t$.

If the gains to the Delta are greater than the required outflow, then there there is a surplus (s_t) available for pumping.

$$s_t = G_T - Qreq_t \tag{1.42}$$

However if the gains are less than the required outflow this value becomes negative and there is a deficit (see Equation 1.45).

If there is surplus available for pumping such that the gains cover both $Qmin_t$ and the export ratio requirement, CVP can pump 55 % of this surplus water and SWP can pump the other 45%. This is denoted in Equations 1.43 and 1.44 (SWRCP; 2000b):

$$TRP_t = \left\{ \begin{array}{ll} \max\left\{Rcvp_t + 0.55s_t, 0\right\} & Rcvp_t + 0.55s_t < Tmax_t \\ \max\left\{Tmax_t, 0\right\} & Rcvp_t + 0.55s_t \geq Tmax_t \end{array} \right\} \tag{1.43}$$

$$HRO_t = \left\{ \begin{array}{ll} \max\{Rswp_t + 0.45s_t, 0\} & Rswp_t + 0.45s_t < Bmax_t \\ \max\{Bmax_t, 0\} & Rswp_t + 0.45s_t \geq Bmax_t \end{array} \right\} \tag{1.44}$$

Where TRP_t is pumping from the Tracy plant (CVP) and HRO_t is pumping for the Harvey O. Banks plant (SWP). $Tmax_t$ and $Bmax_t$ are the maximum allowed pumping for the Tracy and Banks plant, respectively. These are calculated in Section 1.7.3.4.

If the gains to the Delta are less than the requirement, a deficit d_t is defined:

$$d_t = -s_t \tag{1.45}$$

The deficit must be made up by the reservoir releases. In this case 75% of these releases come from CVP reservoirs and 25% come from SWP reservoirs (SWRCP; 2000b). This results in the constraints on pumping $CVPp_t$ and $SWPp_t$:

$$CVPp_t = \left\{ \begin{array}{ll} Rcvp_t - 0.75d_t, & Rcvp_t - 0.75d_t > 0 \\ 0, & Rcvp_t - 0.75d_t \leq 0 \end{array} \right\} \tag{1.46}$$

$$SWPp_t = \left\{ \begin{array}{ll} \max\{Rswp_t + d_t - Rcvp_t, 0\} & CVPp_t = 0 \\ \max\{Rswp_t - 0.25d_t, 0\} & CVPp_t > 0 \end{array} \right\} \tag{1.47}$$

The environmental constraints for pumping from Section 1.7.3.4 are then implemented:

$$TRP_t = \left\{ \begin{array}{ll} CVPp_t & CVPp_t \geq Tmax_t \wedge CVPp_t \geq 0 \\ 0 & CVPp_t \geq Tmax_t \wedge CVPp_t < 0 \\ Tmax_t & CVPp_t < Tmax_t \wedge Tmax_t \geq 0 \\ 0 & CVPp_t < Tmax_t \wedge Tmax_t < 0 \end{array} \right\} \tag{1.48}$$

$$HRO_t = \left\{ \begin{array}{ll} SWPp_t & SWPp_t \geq Bmax_t \wedge SWPp_t \geq 0 \\ 0 & SWPp_t \geq Bmax_t \wedge SWPp_t < 0 \\ Tmax_t & SWPp_t < Bmax_t \wedge Bmax_t \geq 0 \\ 0 & SWPp_t < Bmax_t \wedge Bmax_t < 0 \end{array} \right\} \tag{1.49}$$

The outflow from the Delta $Qout_t$ is equal to the Delta inflow minus the exports from pumping.

$$Qout_t = Qin_t - TRP_t - HRO_t \tag{1.50}$$

Additionally, we define $X2_t$, the point identified by its distance from the Golden Gate Bridge where salinity at the river's bottom is about 2 parts per thousand (ppt):

$$X2_{t+1} = \left\{ \begin{array}{ll} 10.16 + 0.945X2_t - 1.487\log\{Qout_t\} & Qout_t > 0 \\ 10.16 + 0.945X2_t - 1.487\log\{50\} & Qout_t > 0 \end{array} \right\} \tag{1.51}$$

1.7.3.2 South-of-Delta demands

The south-of-Delta demands are determined before reservoir releases. A target pumping for each day of the water year is defined for Tracy and Banks as TT_t and BT_t, respectively

$$\begin{aligned} TT_t &= TT_{dw_t} \\ BT_t &= BT_{dw_t} \end{aligned} \tag{1.52}$$

The additional flow needed if a deficit exists is Qa_t, defined by the minimum flow rule and export ratio:

$$Qa_t = Qmin_t \frac{\nu_t}{1 - \nu_t} \tag{1.53}$$

If a surplus exists, then the projects need not include their excess water available for pumping in their south of Delta release targets $SODcv_t$ and $SODsw_t$ (see the first piece-wise segment in Equations 1.54 and 1.55). If a deficit exists, a portion of this deficit is added to the

south-of-Delta release target for each project if the total maximum possible pumping will overcome the additional flow requirement (see second two piece-wise segments in Equations 1.54 and 1.55).

$$SODcv_t = \begin{Bmatrix} \max\left(TT_t - 0.55\frac{G_t - Qmin_t}{\nu_t}, 0\right) & G_t > Qmin_t \\ TT_t & G_t < Qmin_t \wedge TT_t + BT_t < Qa_t \\ 0.75Qa_t + \frac{TT_t - 0.75Qa_t}{\nu_t} & G_t < Qmin_t \wedge TT_t + BT_t > Qa_t \end{Bmatrix} \quad (1.54)$$

$$SODsw_t = \begin{Bmatrix} \max\left(BT_t - 0.45\frac{G_t - Qmin_t}{\nu_t}, 0\right) & G_t > Qmin_t \\ BT_t & G_t < Qmin_t \wedge TT_t + BT_t < Qa_t \\ 0.25Qa_t + \frac{TT_t - 0.25Qa_t}{\nu_t} & G_t < Qmin_t \wedge TT_t + BT_t > Qa_t \end{Bmatrix} \quad (1.55)$$

In Equations 1.56 and 1.57, percentages of the CVP demand to Folsom and Shasta ($SODf_t$ and $SODs_t$, respectively) are allocated. This is based on the ratio of their available storage values to overall CVP available storage:

$$SODf_t = \begin{Bmatrix} \frac{ASf}{ASf + ASs} & ASf > 0 \wedge ASs > 0 \\ 0 & \wedge ASf < 0 \wedge ASs > 0 \\ 1 & \wedge ASf > 0 \wedge ASf < 0 \end{Bmatrix} \quad (1.56)$$

$$SODs_t = 1 - SODf_t \quad (1.57)$$

1.7.3.3 Target pumping based on Biological Opinions Requirements

Target pumping is calculated based on the following equations for using San Joaquin allocations:

$$\begin{aligned} SJswp &= \frac{SWPmax_{m=5}}{\nu_{wyt_{m=8}}} \\ SJcvp &= \frac{CVPmax_{m=5}}{\nu_{wyt_{m=8}}} \end{aligned} \quad (1.58)$$

$$N_t = \left\{ \begin{array}{ll} 92, & dw_t \geq 274 \\ 1, & dw_t < 274 \end{array} \right\} \tag{1.59}$$

Where N_t is the number of days to save, based on days of July, August, and September.

$$SJie_t = Isj_t \chi_t \tag{1.60}$$

Where Isj_t is the San Joaquin river inflow to the Delta, and χ_t San Joaquin is the import-export ratio (NMFS; 2009). Maximum allowed pumping exports are then calculated:

$$SWPdmax = \left\{ \begin{array}{ll} Bin_{dw_t} + SJa_t, & Bin_{dw_t} + SJa_t > 0.45SJie \\ 0.45SJie, & Bin_{dw_t} + SJa_t < 0.45SJie \end{array} \right\} \tag{1.61}$$

$$CVPdmax = \left\{ \begin{array}{ll} Tin_{dw_t}, & Tin_{dw_t} + SJa_t > 0.55SJie \\ 0.55SJie, & Tin_{dw_t} < 0.55SJie \end{array} \right\} \tag{1.62}$$

$$BP_t = \left\{ \begin{array}{ll} \min\{SWPdmax_t, SWPmax_t\} & ASo_t > SJswN_t \\ 0, & ASo_t \leq SJswN_t \end{array} \right\} \tag{1.63}$$

$$TP_t = \left\{ \begin{array}{ll} \min\{CVPdmax_t, CVPmax_t\} & ASs_t + ASf_t > SJvoN_t \\ 0, & ASs_t + ASf_t \leq SJvoN_t \end{array} \right\} \tag{1.64}$$

1.7.3.4 Old and Middle River requirements

The maximum allowed pumping is lastly constrained based on requirements for the Old and Middle Rivers, which are most effected by the pumping stations.

Equation 1.65 determines maximum total pumping allowed based on alteration of the natural flow in the Old and Middle rivers $OMRn_t$. For each day of the water year, a threshold $OMRr_{dw_t}$ exists. Pumping cannot alter the natural flow below this threshold.

Thus, the overall total pumping $Pmax_t$ cannot be greater than this difference:

$$Pmax = \left\{ \begin{array}{ll} OMRn_t - OMRr_{dw_t} & OMRn > OMRr_{dw_t} \\ 0 & OMRn < OMRr_{dw_t} \end{array} \right\} \tag{1.65}$$

Next the amount of $Pmax_t$ allocated to each project must be determined. For this we use the constrained pumping values BP_t and TP_t from Section 1.7.3.3. This new OMR constraint is defined in Equations 1.66 and 1.67:

$$Bmax_t = \left\{ \begin{array}{ll} Pmax_t - TP_t & TP_t < 0.55Pmax_t \\ 0.45Pmax_t & TP_t \geq 0.55Pmax_t \wedge BP_t \geq 0.45Pmax_t \\ BP_t & TP_t < 0.55Pmax_t \wedge BP_t < 0.45Pmax_t \end{array} \right\} \tag{1.66}$$

$$Tmax_t = \left\{ \begin{array}{ll} Pmax_t - BP_t & BP_t < 0.45Pmax_t \\ 0.55Pmax_t & TP_t \geq 0.55Pmax_t \wedge BP_t \geq 0.45Pmax_t \\ TP_t & TP_t < 0.55Pmax_t \wedge BP_t < 0.45Pmax_t \end{array} \right\} \tag{1.67}$$

Where $Bmax_t$ is the maximum overall pumping allowed from the Banks plant and $Tmax_t$ is the maximum overall pumping allowed from the Tracy plant. These are the final values pumped at each timestep.

1.7.4 Model fits

Using historical data from CDEC as inputs, ORCA can be validated to historical operations. Model fits for various components are shown in this section.

1.7.4.1 Reservoir storage

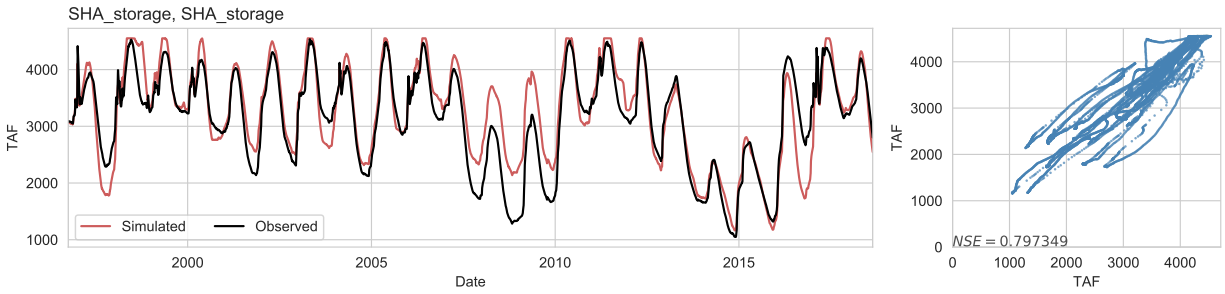

Figure 1.11: Daily Shasta storage

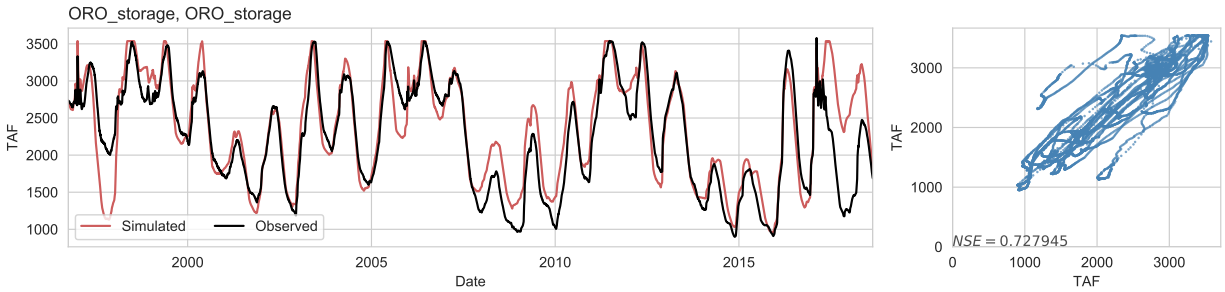

Figure 1.12: Daily Oroville storage

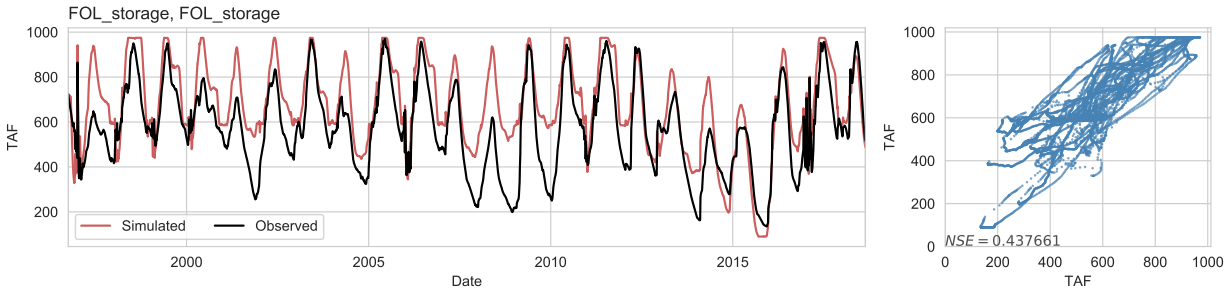

Figure 1.13: Daily Folsom storage

1.7.4.2 Reservoir outflow

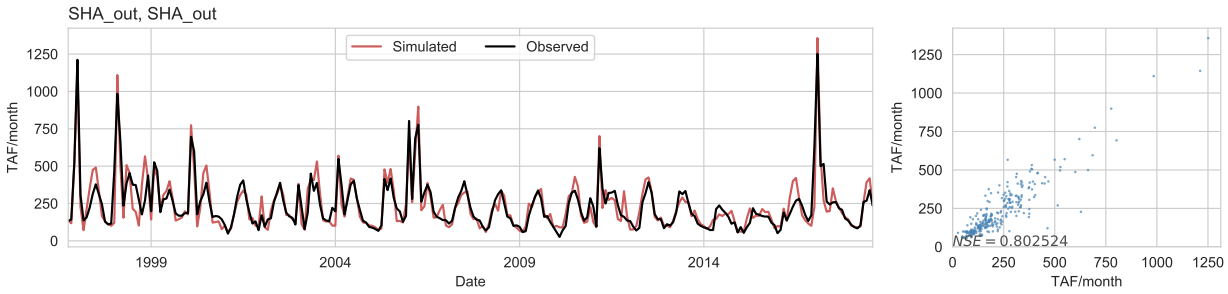

Figure 1.14: Monthly Shasta outflow

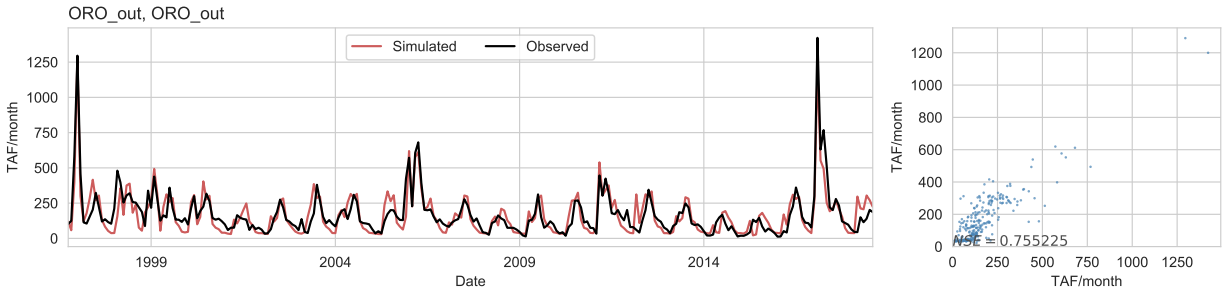

Figure 1.15: Monthly Oroville outflow

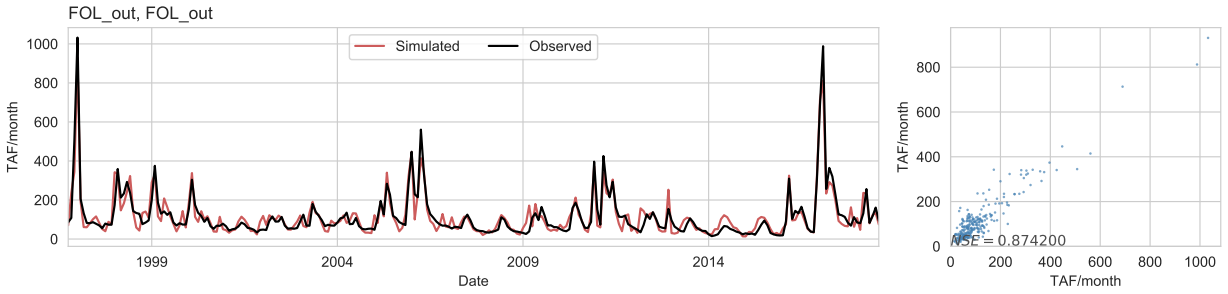

Figure 1.16: Monthly Folsom outflow

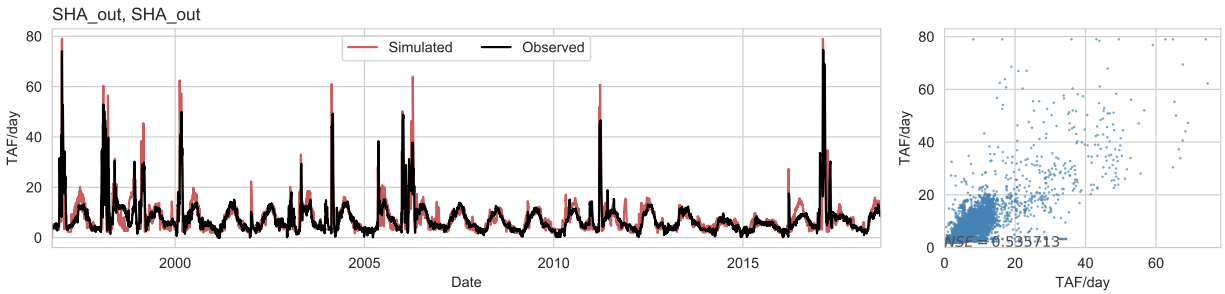

Figure 1.17: Daily Shasta outflow

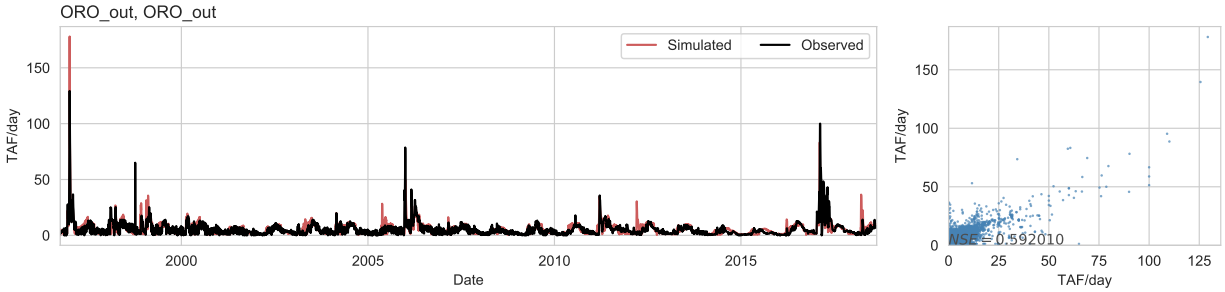

Figure 1.18: Daily Oroville outflow

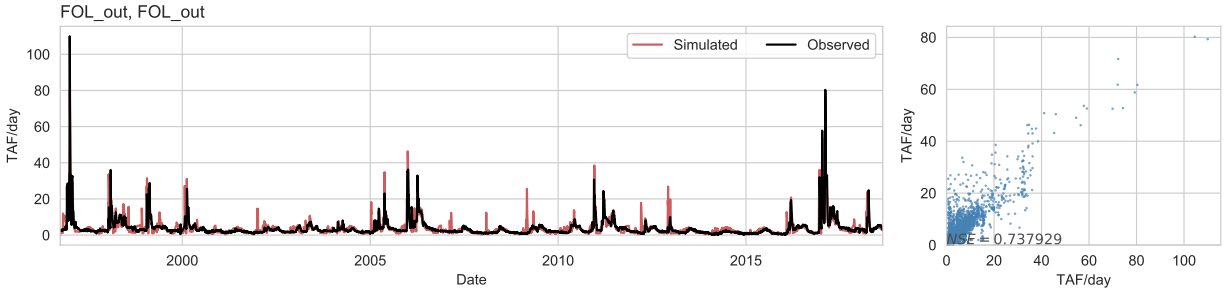

Figure 1.19: Daily Folsom outflow

1.7.4.3 Delta flows

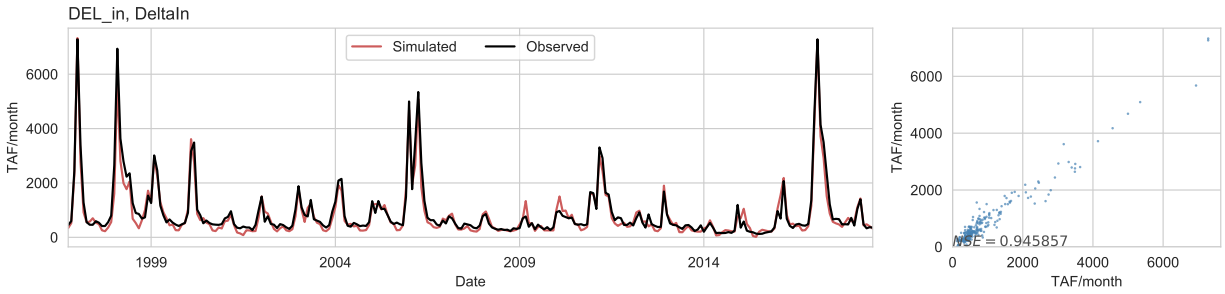

Figure 1.20: Monthly Delta inflow

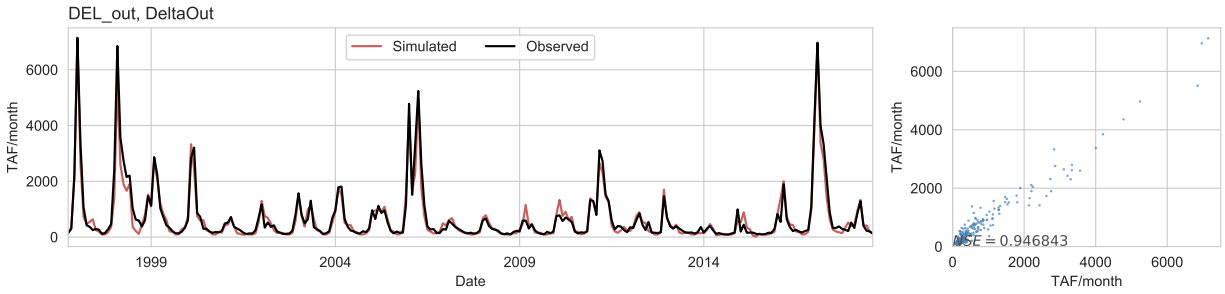

Figure 1.21: Monthly Delta outflow

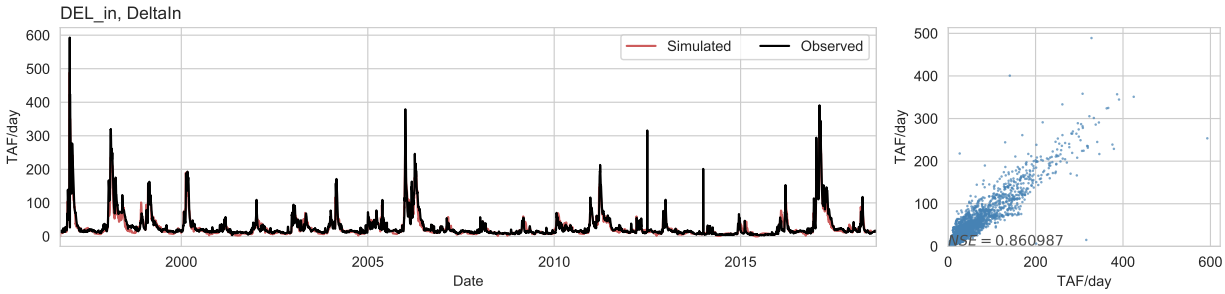

Figure 1.22: Daily Delta inflow

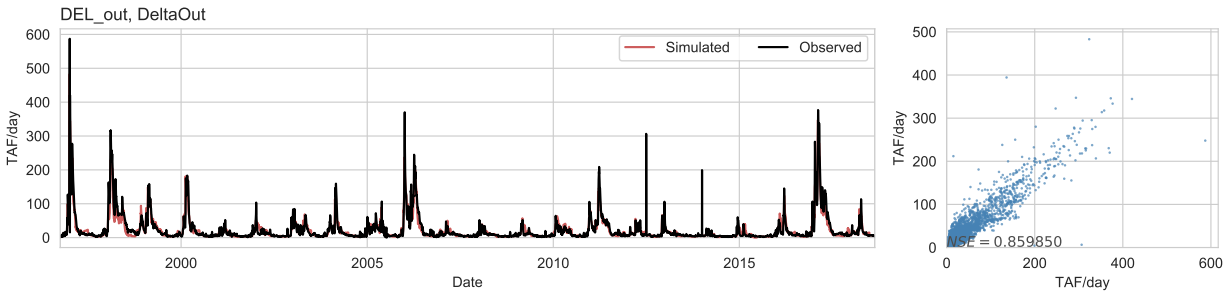

Figure 1.23: Daily Delta outflow

1.7.5 Environmental requirements

Environmental requirement and flow values are obtained from SWRCP (1990), SWRCP (2000b), NMFS (2009), SWF (2015), and FERC (2016).

1.7.5.1 In-stream environmental flows

Month	Wet	Above Normal	Below Normal	Dry	Critical
October - November	3250	3250	3250	2800	2800
December - February	3250	3250	3250	2000	2000
March - August	2300	2300	2300	2300	2300
September	3250	3250	3250	2800	2800

Table 1.7: Minimum release requirements below Shasta Dam, by month and water year type (in CFS)

Month	Wet	Above Normal	Below Normal	Dry	Critical
October - February	1700	1700	1700	1200	1200
March	1700	1700	1700	1000	1000
April - September	1000	1000	1000	1000	1000

Table 1.8: Minimum release requirements below Oroville Dam, by month and water year type (CFS)

Month	Wet	Above Normal	Below Normal	Dry	Critical
October	1500	1500	1250	800	500
November	1750	1500	1250	800	500
December	2000	1750	1250	800	550
January – February	1750	1487	1250	800	250
March – September	1750	155	1050	800	250

Table 1.9: Minimum release requirements below Folsom Dam, by month and water year type (CFS)

1.7.5.2 Delta outflow and salinity

Month	Wet	Above Normal	Below Normal	Dry	Critical
October	4000	4000	4000	4000	4000
November - December	4500	4500	4500	4500	3500
January – June	4500	4500	4500	4500	4500
July	8000	8000	6500	5000	4000
August	4000	4000	4000	3500	3000
September	3000	3000	3000	3000	3000

Table 1.10: Minimum Delta outflow requirements, by month and water year type (CFS)

Month	Wet	Above Normal	Below Normal	Dry	Critical
October	74	82	87.5	87.5	90
November - December 15	87.5	87.5	87.5	87.5	90
December 16 - March	85	85	85	85	85
April – June 15	77	77	77	77	90
June 16 - June 30	77	77	80	89	90
July - August 15	77	80	80	89	90
August 16 - August 30	90	90	90	90	90
September	77	82	77	90	90

Table 1.11: Minimum Delta X2, by month and water year type

1.7.5.3 Carryover and cold-pool storage targets

Reservoir	Wet	Above Normal	Below Normal	Dry	Critical
Shasta (TAF)	2200	2200	2200	1900	1600
Oroville (TAF)	1000	1000	1000	1000	1000
Folsom (TAF)	690	660	440	395	260

Table 1.12: Minimum carryover storage targets based on water year type (TAF)

1.7.6 CMIP5 modeling centers

Modeling Center (or Group)	Institute ID	Model Name
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	CSIRO-BOM	ACCESS1.0
Beijing Climate Center, China Meteorological Administration	BCC	BCC-CSM1.1 BCC-CSM1.1(m)
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
National Center for Atmospheric Research	NCAR	CCSM4
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1(BGC) CESM1(CAM5)
Euro-Mediterranean Center on Climate Change	CMCC	CMCC-CM
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-QCCCE	CSIRO-Mk3.6.0
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS,Tsinghua University	LASG-CESS	FGOALS-g2
The First Institute of Oceanography, SOA, China	FIO	FIO-ESM
NASA Global Modeling and Assimilation Office	NASA GMAO	GEOS-5
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-CM3 GFDL-ESM2G GFDL-ESM2M
NASA Goddard Institute for Space Studies	NASA GISS	GISS-E2-H-CC GISS-E2-R GISS-E2-R-CC
National Institute of Meteorological Research/Korea Meteorological Administration	NIMR/KMA	HadGEM2-AO
Met Office Hadley Centre	MOHC	HadGEM2-CC HadGEM2-ES
Institute for Numerical Mathematics	INM	INM-CM4
Institute Pierre-Simon Laplace	IPSL	IPSL-CM5A-MR IPSL-CM5B-LR
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC	MIROC-ESM MIROC-ESM-CHEM
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC	MIROC5
Max Planck Institute for Meteorology	MPI-M	MPI-ESM-MR MPI-ESM-LR
Meteorological Research Institute	MRI	MRI-CGCM3
Norwegian Climate Centre	NCC	NorESM1-M

Figure 1.24: CMIP5 modeling information

1.8 Data availability

Some or all data, models, or code generated or used during the study are available in a repository online:

ORCA:

https://github.com/jscohen4/orca/tree/cohen-2020-adaptation-reservoir-ops-snowpack-decli

ORCA CMIP5 inputs:

URL: https://github.com/jscohen4/orca_cmip5_inputs

1.9 Acknowledgements

This work was partially supported by the U.S. National Science Foundation grant and CBET-1803589 INFEWS grant CNS-1639268. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views or policies of the NSF. We further acknowledge the World Climate Research Program's Working Group on Coupled Modeling and the climate modeling groups listed in the Appendix for producing and making available their model output.

Bibliography

Anderson, J., Chung, F., Anderson, M., Brekke, L., Easton, D., Ejeta, M., Peterson, R. and Snyder, R. (2008). Progress on incorporating climate change into management of California's water resources, *Climatic Change* **87**(1): 91–108.

Barnett, T. P., Adam, J. C. and Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature* **438**(7066): 303.

Barnett, T. P., Pierce, D. W., Hidalgo, H. G., Bonfils, C., Santer, B. D., Das, T., Bala, G., Wood, A. W., Nozawa, T., Mirin, A. A. et al. (2008). Human-induced changes in the hydrology of the western United States, *science* **319**(5866): 1080–1083.

Belmecheri, S., Babst, F., Wahl, E. R., Stahle, D. W. and Trouet, V. (2016). Multi-century evaluation of Sierra Nevada snowpack, *Nature Climate Change* **6**(1): 2.

Brekke, L. D., Maurer, E. P., , J. D., Dettinger, M. D., Townsley, E. S., Harrison, A. and Pruitt, T. (2009). Assessing reservoir operations risk under climate change, *Water Resources Research* **45**(4).

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

Cayan, D. R. (1996). Interannual climate variability and snowpack in the western United States, *Journal of Climate* **9**(5): 928–948.

Cayan, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M. and Peterson, D. H. (2001). Changes in the onset of spring in the western United States, *Bulletin of the American Meteorological Society* **82**(3): 399–416.

CDEC (2018). *California Data Exchange Center*, California Department of Water Resources.

Christensen, N. and Lettenmaier, D. P. (2006). A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, *Hydrology and Earth System Sciences Discussions* **3**(6): 3727–3770.

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P. and Palmer, R. N. (2004). The effects of climate change on the hydrology and water resources of the Colorado River Basin, *Climatic change* **62**(1-3): 337–363.

Culley, S., Noble, S., Yates, A., Timbs, M., Westra, S., Maier, H., Giuliani, M. and Castelletti, A. (2016). A bottom-up approach to identifying the maximum operational adaptive capacity of water resource systems to a changing climate, *Water Resources Research* **52**(9): 6751–6768.

Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., Dunn, R., Willett, K., Aguilar, E., Brunet, M., Caesar, J. et al. (2013). Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset, *Journal of Geophysical Research: Atmospheres* **118**(5): 2098–2118.

Draper, A. J. and Lund, J. R. (2004). Optimal hedging and carryover storage value, *Journal of Water Resources Planning and Management* **130**(1): 83–87.

DWR (2017). *Management of the California state water project*, Vol. 132.

FERC (2016). *Endangered Species Act Section 7(a)(2) Biological Opinion, and Magnuson-Stevens Fishery Conservation and Management Act Essential Fish Habitat Response and Fish and Wildlife Coordination Act Recommendations for Relicensing the Oroville Facilities Hydroelectric Project, Butte County California (FERC Project No. 2100-134).*, United States Federal Energy Regulatory Commission.
URL: *https://www.westcoast.fisheries.noaa.gov/publications/CentralValley/Sacramento%20River/201*

Georgakakos, A., Yao, H., Kistenmacher, M., Georgakakos, K., Graham, N., Cheng, F.-Y., Spencer, C. and Shamir, E. (2012). Value of adaptive water resources management in northern California under climatic variability and change: Reservoir management, *Journal of Hydrology* **412**: 34–46.

Gleick, P. H. (2002). Water management: Soft water paths, *Nature* **418**(6896): 373.

Goharian, E., Burian, S. J., Bardsley, T. and Strong, C. (2016). Incorporating potential severity into vulnerability assessment of water supply systems under climate change conditions, *Journal of Water Resources Planning and Management* **142**(2): 04015051.

Hamlet, A. F. and Lettenmaier, D. P. (1999). Effects of climate change on hydrology and water resources in the Columbia River Basin 1, *JAWRA Journal of the American Water Resources Association* **35**(6): 1597–1623.

Hayhoe, K., Cayan, D., Field, C. B., Frumhoff, P. C., Maurer, E. P., Miller, N. L., Moser, S. C., Schneider, S. H., Cahill, K. N., Cleland, E. E. et al. (2004). Emissions pathways, climate change, and impacts on California, *Proceedings of the national academy of sciences* **101**(34): 12422–12427.

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M. and Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty, *Water Resources Research* p. e24389.

Herman, J. D., Reed, P. M., Zeff, H. B. and Characklis, G. W. (2015). How should robustness be defined for water systems planning under change?, *Journal of Water Resources Planning and Management* **141**(10): 04015012.

Huang, X., Hall, A. D. and Berg, N. (2018). Anthropogenic warming impacts on today's Sierra Nevada snowpack and flood risk, *Geophysical Research Letters* **45**(12): 6215–6222.

Kapnick, S. and Hall, A. (2010). Observed climate–snowpack relationships in California and their implications for the future, *Journal of Climate* **23**(13): 3446–3456.

Karamouz, M., Goharian, E. and Nazif, S. (2013). Reliability assessment of the water supply systems under uncertain future extreme climate conditions, *Earth Interactions* **17**(20): 1–27.

Klos, P. Z., Link, T. E. and Abatzoglou, J. T. (2014). Extent of the rain-snow transition zone in the western US under historic and projected climate, *Geophysical Research Letters* **41**(13): 4560–4568.

Knowles, N., Cronkite-Ratcliff, C., Pierce, D. and Cayan, D. (2018). Responses of unimpaired flows, storage, and managed flows to scenarios of climate change in the San Francisco Bay-Delta Watershed, *Water Resources Research* **54**(10): 7631–7650.

Knowles, N., Dettinger, M. D. and Cayan, D. R. (2006). Trends in snowfall versus rainfall in the western United States, *Journal of Climate* **19**(18): 4545–4559.

Koster, R. D., Mahanama, S. P., Livneh, B., Lettenmaier, D. P. and Reichle, R. H. (2010). Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow, *Nature Geoscience* **3**(9): 613.

Kourakos, G., Dahlke, H. E. and Harter, T. (2019). Increasing groundwater availability and seasonal base flow through agricultural managed aquifer recharge in an irrigated basin, *Water Resources Research* **55**(9): 7464–7492.

Lee, S.-Y., Hamlet, A. F., Fitzgerald, C. J. and Burges, S. J. (2009). Optimized flood control in the Columbia River Basin for a global warming scenario, *Journal of water resources planning and management* **135**(6): 440–450.

Lettenmaier, D. P. and Sheer, D. P. (1991). Climatic sensitivity of California water resources, *Journal of Water Resources Planning and Management* **117**(1): 108–125.

Leung, L. R., Qian, Y., Bian, X., Washington, W. M., Han, J. and Roads, J. O. (2004). Mid-century ensemble regional climate change scenarios for the western United States, *Climatic Change* **62**(1-3): 75–113.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres* **99**(D7): 14415–14428.

Livneh, B. and Badger, A. M. (2020). Drought less predictable under declining future snowpack, *Nature Climate Change* pp. 1–7.

Mahanama, S., Livneh, B., Koster, R., Lettenmaier, D. and Reichle, R. (2012). Soil moisture, snow, and seasonal streamflow forecasts in the United States, *Journal of Hydrometeorology* **13**(1): 189–203.

Mateus, M. C. and Tullos, D. (2017). Reliability, sensitivity, and vulnerability of reservoir operations under climate change, *Journal of Water Resources Planning and Management* **143**(4): 04016085.

McCabe, G. J., Clark, M. P. and Hay, L. E. (2007). Rain-on-snow events in the western United States, *Bulletin of the American Meteorological Society* **88**(3): 319–328.

Medellín-Azuara, J., Harou, J. J., Olivares, M. A., Madani, K., Lund, J. R., Howitt, R. E., Tanaka, S. K., Jenkins, M. W. and Zhu, T. (2008). Adaptability and adaptations of California's water supply system to dry climate warming, *Climatic Change* **87**(1): 75–90.

Minville, M., Brissette, F. and Leconte, R. (2010). Impacts and uncertainty of climate change on water resource management of the Peribonka River system (Canada), *Journal of Water Resources Planning and Management* **136**(3): 376–385.

Mote, P. W., Hamlet, A. F., Clark, M. P. and Lettenmaier, D. P. (2005). Declining mountain snowpack in western North America, *Bulletin of the American meteorological Society* **86**(1): 39–50.

Nayak, M. A., Herman, J. D. and Steinschneider, S. (2018). Balancing flood risk and water supply in California: Policy search integrating short-term forecast ensembles with conjunctive use, *Water Resources Research* **54**(10): 7557–7576.

NMFS (2009). *Biological Opinion and Conference Opinion on the Long-Term Operations of the Central Valley Project and the State Water Project.*, National Marine and Fisheries Service.

Null, S. E. and Viers, J. H. (2013). In bad waters: Water year classification in nonstationary climates, *Water Resources Research* **49**(2): 1137–1148.

Pederson, G. T., Gray, S. T., Woodhouse, C. A., Betancourt, J. L., Fagre, D. B., Littell, J. S., Watson, E., Luckman, B. H. and Graumlich, L. J. (2011). The unusual nature of recent snowpack declines in the north american cordillera, *science* **333**(6040): 332–335.

Qin, Y., Abatzoglou, J. T., Siebert, S., Huning, L. S., AghaKouchak, A., Mankin, J. S., Hong, C., Tong, D., Davis, S. J. and Mueller, N. D. (2020). Agricultural risks from changing snowmelt, *Nature Climate Change* pp. 1–7.

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M. and Brown, C. (2020). Vulnerability and risk: climate change and water supply from California's Central Valley water system, *Climatic Change* pp. 1–23.

Reclamation (2013). Downscaled CMIP3 and CMIP5 climate and hydrology projections: Release of downscaled CMIP5 climate projections, comparison with preceding information, and summary of user needs.

Rhoades, A. M., Jones, A. D. and Ullrich, P. A. (2018). Assessing mountains as natural reservoirs with a multimetric framework, *Earth's Future* **6**(9): 1221–1241.

Rizzardo, D. (2016). California's water supply forecasting.
URL: *https://www.watereducation.org/sites/main/files/file-attachments/rizzardo_water_forecasting.pdf*

Simpson, J. J., Dettinger, M. D., Gehrke, F., McIntire, T. J. and Hufford, G. L. (2004). Hydrologic scales, cloud variability, remote sensing, and models: Implications for forecasting snowmelt and streamflow, *Weather and forecasting* **19**(2): 251–276.

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

Sterle, K., Jose, L., Coors, S., Singletary, L., Pohll, G. and Rajagopal, S. (2020). Collaboratively modeling reservoir reoperation to adapt to earlier snowmelt runoff, *Journal of Water Resources Planning and Management* **146**(1): 05019021.

Stewart, I. T., Cayan, D. R. and Dettinger, M. D. (2005). Changes toward earlier streamflow timing across western North America, *Journal of climate* **18**(8): 1136–1155.

Sturm, M., Goldstein, M. A. and Parr, C. (2017). Water and life from snow: A trillion dollar science question, *Water Resources Research* **53**(5): 3534–3544.

Surfleet, C. G. and Tullos, D. (2013). Variability in effect of climate change on rain-on-snow peak flow events in a temperate climate, *Journal of Hydrology* **479**: 24–34.

SWF (2015). *The Lower American River Modified Flow Management Standard: A Drought Buffer for the Environment and Local Water Supplies*, Sacramento Water Forum.
URL: *http://www.waterforum.org/wp-content/uploads/2017/04/WF-Modified-FMS-$10_8 f inal_S ingle.pdf$*

SWRCP (1990). *Water Rights Order 90-05: Order Setting Terms and Conditions for Fishery Protection and Setting a Schedule for Completion of Tasks. United States Bureau of Reclamation*, United States Bureau of Reclamation.
URL: *https://www.waterboards.ca.gov/waterrights/board_decisions/adopted_orders/orders/1990/wro90-05.pdf*

SWRCP (2000a). *Revised Water Right Decision 1641: Implementation of Water Quality Objectives for the San Francisco Bay/Sacramento-San Joaquin Delta Estuary*, California EPA.

SWRCP (2000b). *Revised Water Right Decision 1641: Implementation of Water Quality Objectives for the San Francisco Bay/Sacramento-San Joaquin Delta Estuary. California EPA.*, California EPA.
URL: *http://www.waterboards.ca.gov/waterrights/board_decisions/adopted_orders/decisions/d1600_d1649/wrd1641_1999dec29.pdf*

Tanaka, S. K., Zhu, T., Lund, J. R., Howitt, R. E., Jenkins, M. W., Pulido, M. A., Tauber, M., Ritzema, R. S. and Ferreira, I. C. (2006). Climate warming and water management adaptation for California, *Climatic Change* **76**(3-4): 361–387.

USACE (1970). *Oroville Dam and Reservoir. Feather River, California : Report on Reservoir Regulation for Flood Control*, Department of the Army, Sacramento District, Corps of Engineers, Sacramento, California.

USACE (1977). *Shasta Dam and Lake. Sacramento River, California: Report on Reservoir Regulation for Flood Control*, Department of the Army, Sacramento District, Corps of Engineers, Sacramento, California.

USACE (1987). *Folsom Dam and Lake: Water Control Manual*, Department of the Army, Sacramento District, Corps of Engineers, Sacramento, California.

VanRheenen, N. T., Wood, A. W., Palmer, R. N. and Lettenmaier, D. P. (2004). Potential implications of PCM climate change scenarios for Sacramento–San Joaquin River basin hydrology and water resources, *Climatic change* **62**(1-3): 257–281.

Vicuna, S. and Dracup, J. (2007). The evolution of climate change impact studies on hydrology and water resources in California, *Climatic Change* **82**(3-4): 327–350.

Vicuna, S., Maurer, E. P., Joyce, B., Dracup, J. A. and Purkey, D. (2007). The sensitivity of California water resources to climate change scenarios, *JAWRA Journal of the American Water Resources Association* **43**(2): 482–498.

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

Wilby, R. L. and Dessai, S. (2010). Robust adaptation to climate change, *Weather* **65**(7): 180–185.

Wilby, R. L. and Keenan, R. (2012). Adapting to flood risk under climate change, *Progress in Physical Geography* **36**(3): 348–378.

Willis, A. D., Lund, J. R., Townsley, E. S. and Faber, B. A. (2011). Climate change and flood operations in the Sacramento Basin, California, *San Francisco Estuary and Watershed Science* **9**(2).

Yao, H. and Georgakakos, A. (2001). Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. reservoir management, *Journal of Hydrology* **249**(1-4): 176–196.

Chapter 2

How do the properties of training scenarios influence the robustness of reservoir operating policies to climate uncertainty?[1]

2.1 Abstract

Reservoir control policies provide a flexible option to adapt to the uncertain hydrologic impacts of climate change. This challenge requires robust policies capable of navigating scenarios that are wetter, drier, or more variable than anticipated. While a number of prior studies have trained robust policies using large scenario ensembles, there remains a need to understand how the properties of training scenarios impact policy robustness. Specifically, this study investigates scenario properties including annual runoff, snowpack, and baseline regret—the difference between baseline policy and perfect foresight performance in an individual scenario. Results indicate that policies trained to scenario subsets with high baseline regret outperform those generated with other training sets in both wetter and drier futures, largely by adopting an intra-annual hedging strategy. The approach highlights the potential to improve the efficiency and robustness of policy training by considering both the hydrologic

[1]This chapter has been published: Cohen, J. S., Zeff, H. B. and Herman, J. D. (2021). "How do the properties of training scenarios influence the robustness of reservoir operating policies to climate uncertainty?". *Environmental Modelling & Software* p. 105047.

properties and baseline regret of the training ensemble.

2.2 Introduction

Adaptation to the multi-scale impacts of climate change in water resources systems is challenged by substantial uncertainty in future hydrologic projections, particularly in flood and drought risks (Wilby and Dessai; 2010; Asadieh and Krakauer; 2017; Dottori et al.; 2018). This hinders the ability to use traditional prediction-based planning methods and has resulted in the recent consensus toward robust planning (Dessai and Hulme; 2004; Wilby and Dessai; 2010). Robust and adaptive planning have been widely considered for both expansion of water resources infrastructure (e.g. Haasnoot et al.; 2013; Beh et al.; 2015; Zeff et al.; 2016; Kwakkel et al.; 2016; Maier et al.; 2016; Trindade et al.; 2017) as well as changes to reservoir control policies (e.g. Giuliani et al.; 2014; Quinn et al.; 2018; Herman and Giuliani; 2018). Of the two alternatives, control policies provide a more flexible "soft path" approach, as they can be reversed if the future unfolds differently than predicted (Gleick; 2002; Fletcher et al.; 2017). In this case, the physical constraints of the existing system establish the range of uncertain scenarios that can be adapted to before new infrastructure is needed (e.g. Culley et al.; 2016).

Robust planning of reservoir control policies generally consists of two phases that have been studied using a variety of different approaches: policy design and robustness analysis. A number of studies have focused on the robustness of current system operations to a range of future climate changes represented either by downscaled Global Circulation Model (GCM) scenarios (e.g., Brekke et al.; 2009; Karamouz et al.; 2013; Knowles et al.; 2018) or synthetically generated scenarios based on perturbed statistics of hydrologic timeseries (e.g., Prudhomme et al.; 2010; Brown et al.; 2012; Weaver et al.; 2013; Turner et al.; 2014). Both approaches often serve as precursors to adaptation studies in which a discrete set of proposed management alternatives are tested to mitigate vulnerabilities in future scenar-

ios (e.g., Groves et al.; 2013; Steinschneider, McCrary, Wi, Mulligan, Mearns and Brown; 2015; Mateus and Tullos; 2017). In this case, the policies are not trained or optimized to a particular set of scenarios, but instead arise from stakeholder expertise and negotiations.

An alternative approach is to generate candidate alternatives via optimization approaches (Kasprzyk et al.; 2013). In this case, policy design and robustness analysis are analogous to the train-test terminology often used in machine learning (e.g., Russell and Norvig; 2002), and recently in the water resources field (e.g., Brodeur et al.; 2020). Policy design (training) involves optimizing learned policy parameters to a specific set of input data (the training set). In robustness analysis (testing), a test set consisting of input data separate from the training set are used to assess the performance of an optimized policy. In the context of reservoir control under climate change, the most relevant optimization approach is policy search, in which parameterized operating rules are optimized for system performance objectives under a set of training scenarios (Koutsoyiannis and Economou; 2003; Giuliani, Castelletti, Pianosi, Mason and Reed; 2015; Giuliani et al.; 2017). This heuristic approach functions as both a simulation-optimization problem (Salazar et al.; 2016) and an information selection problem for the policy inputs (Giuliani, Pianosi and Castelletti; 2015; Nayak et al.; 2018). While the training performance of a policy on an individual scenario represents a best-case outcome with perfect foresight, the key challenge is whether the policy can generalize to a different test set, which is also the case for any optimization method applied in the context of climate adaptation. Test sets often include additional stochastic realizations of the same uncertainties used in training, i.e., to obtain a thorough representation of sampling uncertainty (e.g., Quinn et al.; 2017; Trindade et al.; 2017). They may also include scenarios representing a different characterization of uncertainty altogether (Watson and Kasprzyk; 2017; Eker and Kwakkel; 2018).

The robustness of alternatives generated by policy search therefore strongly depends on the choice of scenarios used for training and testing, both in terms of coarse-scale statistics (wet vs. dry) and the realizations of natural variability that lead to extreme events (Herman

et al.; 2020). This includes the case where the training data represent a baseline or historical scenario (e.g., Kasprzyk et al.; 2013; Giuliani and Castelletti; 2016; Quinn et al.; 2018). These studies evaluate resulting alternatives over test scenarios spanning a wide range of potential future hydrology, but without analyzing the influence of the training scenarios on robustness. Robust optimization studies overcome this by optimizing robustness metrics over many samples of uncertain scenarios (e.g., Hamarat et al.; 2014; Kwakkel et al.; 2015). While they optimize over a large range of training scenarios, studies using robust optimization generally have not considered how well solutions can generalize out-of-sample. Robustness measures will typically include either regret, which quantifies the cost of choosing an incorrect solution, or satisficing, which calculates the fraction of scenarios in which a policy meets a set of performance criteria (Lempert and Collins; 2007; Herman et al.; 2015). While both of these methods are effective in evaluating the performance of individual solutions, there is also the consideration of the robustness of the Pareto set as a whole to quantify deviations in multi-objective performance.

While the properties of the test scenarios have been the focus of many prior studies using scenario discovery and related methods, the properties of the training scenarios and their influence on policy robustness have received relatively less attention. However, several recent studies have begun to analyze the effect of the choice of training scenarios in optimization problems. For example, Watson and Kasprzyk (2017) extend many-objective robust decision making by optimizing to several different sets of scenarios with varying properties. They then re-evaluate solutions in out-of-sample scenarios, quantifying robustness for individual solutions using the satisficing metric. Eker and Kwakkel (2018) optimize to scenarios with maximum diversity and policy relevance and re-evaluate solutions under the same uncertainty characterization used in training. Giudici et al. (2020) propose an algorithm to select the smallest subset of training scenarios which can be used to generate robust solutions when re-evaluated against the full set to minimize computational cost. These studies all effectively aim to find training scenarios (scenario selection) that lead to robust out of sample perfor-

mance. However, studies to date have not attributed multi-objective policy robustness to the hydrologic properties of the training scenarios, which holds significant implications for the design of robust policies under climate uncertainty.

This study proposes an experimental design to determine how the properties of forcing scenarios influence the robustness of multi-objective policy alternatives across several combinations of test scenarios. This framework is generalizable to any environmental planning problem that includes a no action case, an optimization component, and an ensemble of forcing scenarios exhibiting uncertainty. For the initial application, we focus specifically on how the hydrologic properties of climate scenarios influence reservoir policy alternatives. One additional scenario property is the baseline regret, which quantifies the extent to which policy search can improve upon the status quo based on a perfect foresight optimization for an individual scenario. Scenarios are clustered into groups with similar hydrology via unsupervised learning, and split into different combinations of training and test sets. The robustness of the resulting policies is quantified relative to the perfect foresight and baseline solutions to ensure that, at a minimum, all solutions outperform the baseline no-action policy. This is done with a normalized hypervolume metric to represent the robustness of the Pareto-set as a whole, which simultaneously quantifies the changes in the performance of the solutions as well as their diversity when re-evaluated on a given test set. Finally, we perform hypothesis tests on several iterations of the train-test split to identify the properties of training scenarios that lead to the most robust results for each test set, with a particular focus on training policies to scenarios which have high baseline regret. We demonstrate this approach using a simulation model of the northern California reservoir system coupled with an ensemble of transient downscaled climate scenarios.

2.3 Case study

2.3.1 Northern California reservoir system

To support urban and agricultural growth amid intense intra- and inter-annual variability in hydrology, California has built a complex system of water resources infrastructure. Reservoirs in the foothills of the Sierra Nevada range capture winter and spring flood season flows to be delivered for agriculture and municipal supply, particularly during summer months. The State Water Project (SWP) and federal Central Valley Project (CVP) consist of a number of reservoirs and aqueducts throughout the Sacramento-San Joaquin river basin. The terminal delta of this system is the site of pumped water exports from north to south, which support agriculture and municipal supply in the southern portion of the state. Critical environmental requirements related to the salinity of outflows from the Delta are a major constraint on these exports. Delta exports are a key metric for water supply reliability in the state and have been found vulnerable to climate change, due to both changes in precipitation levels and seasonal runoff timing (Anderson et al.; 2008; Ray et al.; 2020).

In the Sacramento River Basin, three of the largest Sierra foothill reservoirs by volume (Shasta, Oroville, and Folsom) combine to a total of 9 million acre-feet (11.1 km^3) of storage in parallel. These reservoirs play a major role in balancing the state's human and environmental water needs. Carryover storage in these reservoirs, measured at the end of the water year (September), is a strong indicator of overall system performance and potential economic vulnerabilities (Draper and Lund; 2004). Uncertainty in changing inter-annual precipitation patterns and reduced snowpack levels has the potential to be detrimental to carryover storage levels and their economic benefits (Medellín-Azuara et al.; 2008). Under a variety of projected changes to the hydrologic regime, operational adaptations are needed to maintain carryover storage levels to support multiple environmental and water supply related objectives while continuing to provide adequate flood control functions (Cohen et al.; 2020).

2.3.2 Simulation model (ORCA)

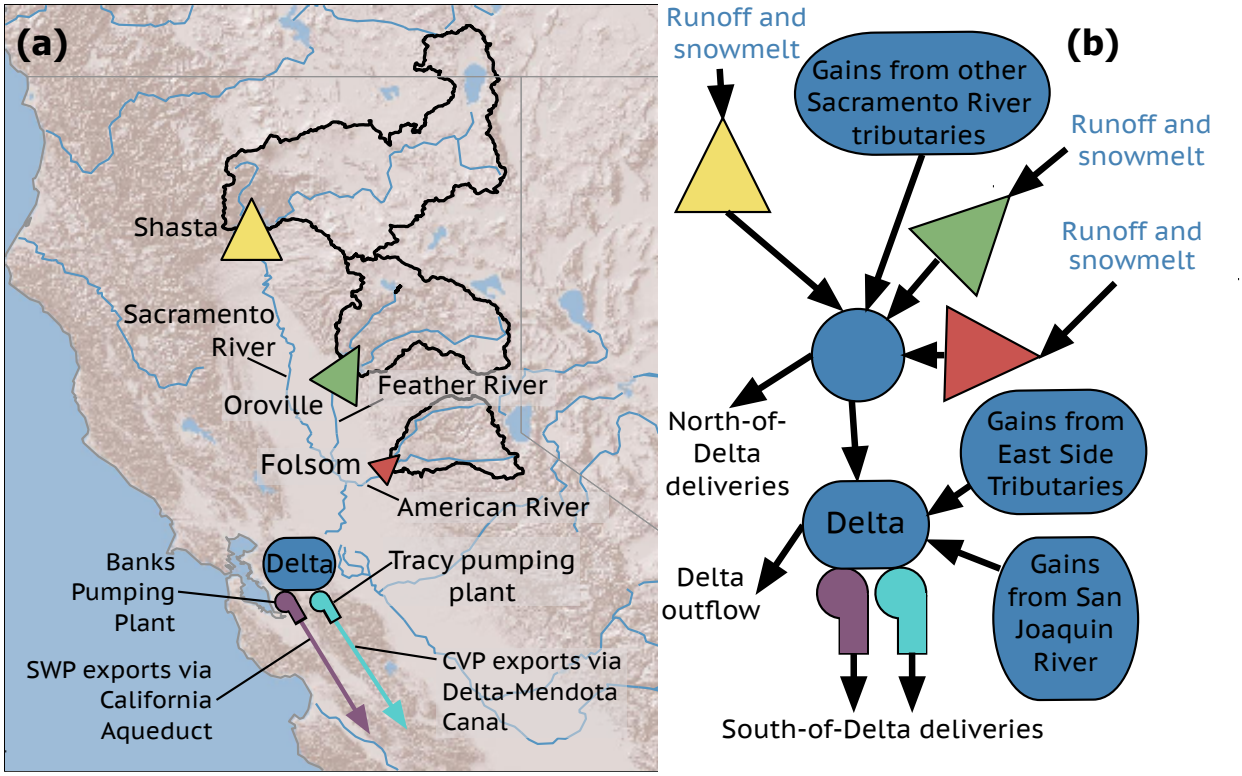

Figure 2.1: (a) Map of northern California reservoir system modeled in ORCA. (b) Model schematic showing primary storage and pumping infrastructure.

We use the open source model Operation of Reservoirs in California (ORCA) to simulate the northern California reservoir system (https://github.com/jscohen4/orca/tree/cohen-2021-properties-training-scenarios). ORCA is a simulation model that runs on a daily timestep and accurately reproduces historical operations (Cohen et al.; 2020). The operating rules that drive ORCA are used as the baseline policy in the current study. The model simulates the major components of the California system north of the Delta, including the Shasta, Oroville, and Folsom Reservoirs, and Delta water supply exports via the Harvey O. Banks (SWP) and Tracy (CVP) pumping plants (Figure 2.1a,b). While not as spatially comprehensive as several other statewide models, ORCA is a pure simulation model, which allows for flexible adjustments to operating rules and straightforward evalu-

ation of alternative hydrologic scenarios, as required by the proposed set of policy search experiments. ORCA is driven by a basic mass balance update for each reservoir. Based on timestep t, storage S_t^r in reservoir r is updated based on inflows Q_r^t, evaporative losses L_t^r, and a release u_t^r:

$$S_t^r = S_{t-1}^r + Q_t^r - u_t^r - L_t^r \tag{2.1}$$

A target release RT_t^r is determined by the greatest of three minimum operating requirements that must be satisfied for each reservoir:

$$RT_t^r = \max(u_{t,\text{environment}}^r, u_{t,\text{flood}}^r, u_{t,\text{demand}}^r) \times c_t^r \tag{2.2}$$

The first is a minimum environmental flow requirement $u_{t,\text{environment}}^r$ that varies based on the time of year and water year type. The second is a flood control release target $u_{t,\text{flood}}^r$. The flood control release depends on a dynamic flood control rule curve, which is determined by a flood control index based on the previous day's precipitation, and current reservoir storage. Finally, the minimum demand release $u_{t,\text{demand}}^r$ consists of water supply demands north of the Delta, south of Delta demands to be delivered by Banks and Tracy pumping plants, and a Delta outflow demand for environmental benefits and salinity control. These water demands are also partially controlled by current and projected reservoir storage, creating a feedback between reservoir operations and downstream Delta management.

A snowpack-to-streamflow forecast enables projections of reservoir inflows throughout the irrigation season. This forecast also determines the water year type classification, which influences both environmental flow requirements and water supply demands along with other operational parameters. The snowpack-to-streamflow forecast is altered by an exceedance level $\bar{Z}_{WYI}$ for the water year type prediction and $\bar{Z}_{wyt}^r$ for individual reservoir inflows. The exceedance represents the confidence in the forecast. A lower exceedance level indicates a more conservative forecast, resulting in lower inflow forecasts and drier water year type classifications, and likely hedged reservoir releases. The forecast is updated each day in the

simulation via Equation 2.3:

$$Qf_t^r = \beta_{dw_t}^r SWE_t^r + \alpha_{dw_t}^r + \bar{Z}_{wyt}^r \sigma_{dw_t}^r \tag{2.3}$$

Where Qf_t is the forecasted inflow for the remainder of the water year at reservoir r, SWE_t is the snow water equivalent at day t, $\beta_{dw_t}^k$ and $\alpha_{dw_t}^k$ are regression coefficients for day of water year dw_t based on historical streamflow records, and $\sigma_{dw_t}^k$ is the standard deviation of remaining streamflow on dw_t, also based on the historical record.

A curtailment multiplier c_t^r can hedge releases in cases where the system is not projected to meet a carryover storage target C_{wyt}^r at the end of the water year. The forecasted flow and current reservoir storage are used to determine what curtailment multiplier would be necessary to meet the carryover target (Equation 2.4). The curtailment multiplier is also constrained by a maximum curtailment allowance $c_{\max,wyt}^r$. A higher maximum curtailment will allow for lower releases to occur in the irrigation season to maintain the cold-pool carryover storage. The daily curtailment multiplier during between May and September ($5 \leq M \leq 9$) is determined at each timestep via Equation 2.4:

$$c_{t,5\leq M\leq 9}^r = \min\left(1, \max\left\{\frac{Qf_t^r + S_{t-1}^r - \sum_{dw_t}^{365} RT_t)}{C_{wyt}^r}, c_{\max,wyt}^r\right\}\right) \tag{2.4}$$

The release for each reservoir is then equal to the target release RT_t^k times the curtailment factor c_t^k:

$$u_t^r = RT_t^r \times c_t^r \tag{2.5}$$

Further details concerning operations modeled in ORCA are described in Cohen et al. (2020).

2.3.3 Data sources

Several hydroclimatic time series are used as inputs for the simulation model. These include daily streamflows, spatially gridded and site-specific precipitation, and air tempera-

ture, along with spatially averaged and site specific monthly spatial snow water equivalent (SWE). For simulating historical operations, these data are obtained from the California Data Exchange Center (CDEC; 2018). Downscaled CMIP5 Climate and Hydrology Projections are obtained from the United States Bureau of Reclamation (USBR) (Reclamation; 2013; Brekke et al.; 2014). These consist of 31 GCMs run for various emissions scenarios to generate 97 scenarios of precipitation and temperature on a daily timestep through 2100 (see Section 2.7.1 in the Appendix for a list of institutions providing GCM projections). In the USBR study, outputs from these GCM simulations were routed through the Variable Infiltration Capacity (VIC) model (Liang et al.; 1994) calibrated for each basin, yielding additional streamflow and SWE projections to serve as model inputs. The choice to use a GCM ensemble in this study reflects several considerations. First, it provides the best available representation of physically-based transient changes to hydrology, including extreme events, despite the known limitations of GCM projections (Herman et al.; 2020). Second, it provides an accurate link between hydrologic variables across space and time, linking precipitation, temperature, streamflow, and snowpack in multiple basins. This is difficult to achieve with synthetic generators, though these are rapidly improving for this purpose (e.g., Steinschneider et al.; 2019).

This ensemble exhibits the high degree of uncertainty associated with future precipitation and, to a lesser extent, temperature. In Figure 2.2a, the trajectories of annual streamflow show an end-of-century average annual flow ranging from -/+ 50% of historical values. This creates a challenge for how to best adapt operations to balance the tradeoff between flood control and water supply (Herman and Giuliani; 2018; Nayak et al.; 2018). All scenarios in the ensemble show a decline in snowpack, ranging from 20-90% of the historical average. This is one of the best predicted aspects of climate change, although uncertainties exist in the extent and severity of this thermodynamic-hydrologic change (Cayan et al.; 2001; Klos et al.; 2014). This can be particularly impactful in mountainous regions where snowpack has historically functioned as a natural reservoir (Rhoades et al.; 2018). As a result, the primary impact

of rising temperatures in the region is earlier spring snowmelt timing (Knowles et al.; 2006; Kapnick and Hall; 2010). These intra-annual streamflow shifts are predicted throughout the CMIP5 ensemble based on the water year centroid, a representation of the center of mass of the annual hydrograph (Figure 2.2c). The ensemble also shows uncertainty in the extent of flood risk changes (Figure 2.2d) based on both uncertain dynamic climate changes as well as the potential increases given a shift from snow to rain and more rain-on-snow events (McCabe et al.; 2007; Surfleet and Tullos; 2013; Huang et al.; 2018). Lastly, the ensemble shows severity in uncertainty related to changes in drought patterns (Figure 2.1e). Neither changes in drought nor flood statistics show a reliable relationship to the overall changes in total annual streamflow. Overall, the downscaled projection ensemble exhibits the significant uncertainty typical of climate adaptation studies, requiring careful attention to the choice of scenarios under which reservoir control policies are trained and tested.

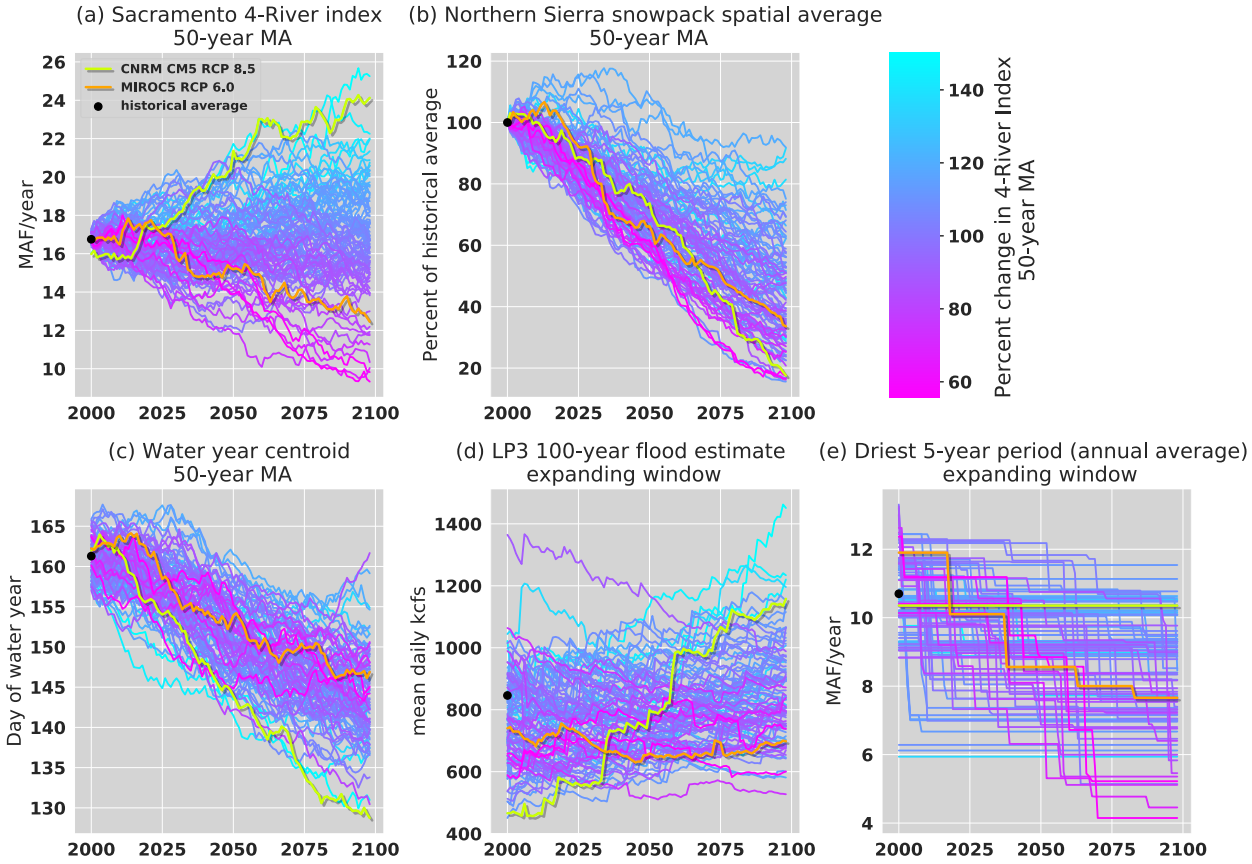

Figure 2.2: 50-year moving averages of CMIP5 projections showing a wide range of uncertainty in future flood and drought risk. (a) Streamflow in the Sacramento River downstream of its three largest tributaries (the Feather, Yuba, and American Rivers), denoting the 4-river index. (b) snowpack in the northern Sierra Nevada. (c) water year centroid, defined as the day of the water year at which half of the total annual streamflow has been observed. (d) Log-Pearson Type III distribution (LP3) 100-year flood estimate for the Sacramento River flow below the American River. (e) The driest 5-year period in the 4-river index. The scenarios shown in yellow (RCP 8.5) and orange (RCP 6.0) are examined later in the analysis (Section 2.4.3). See Section 2.7.1 in the Appendix for institutions providing climate models and model abbreviations.

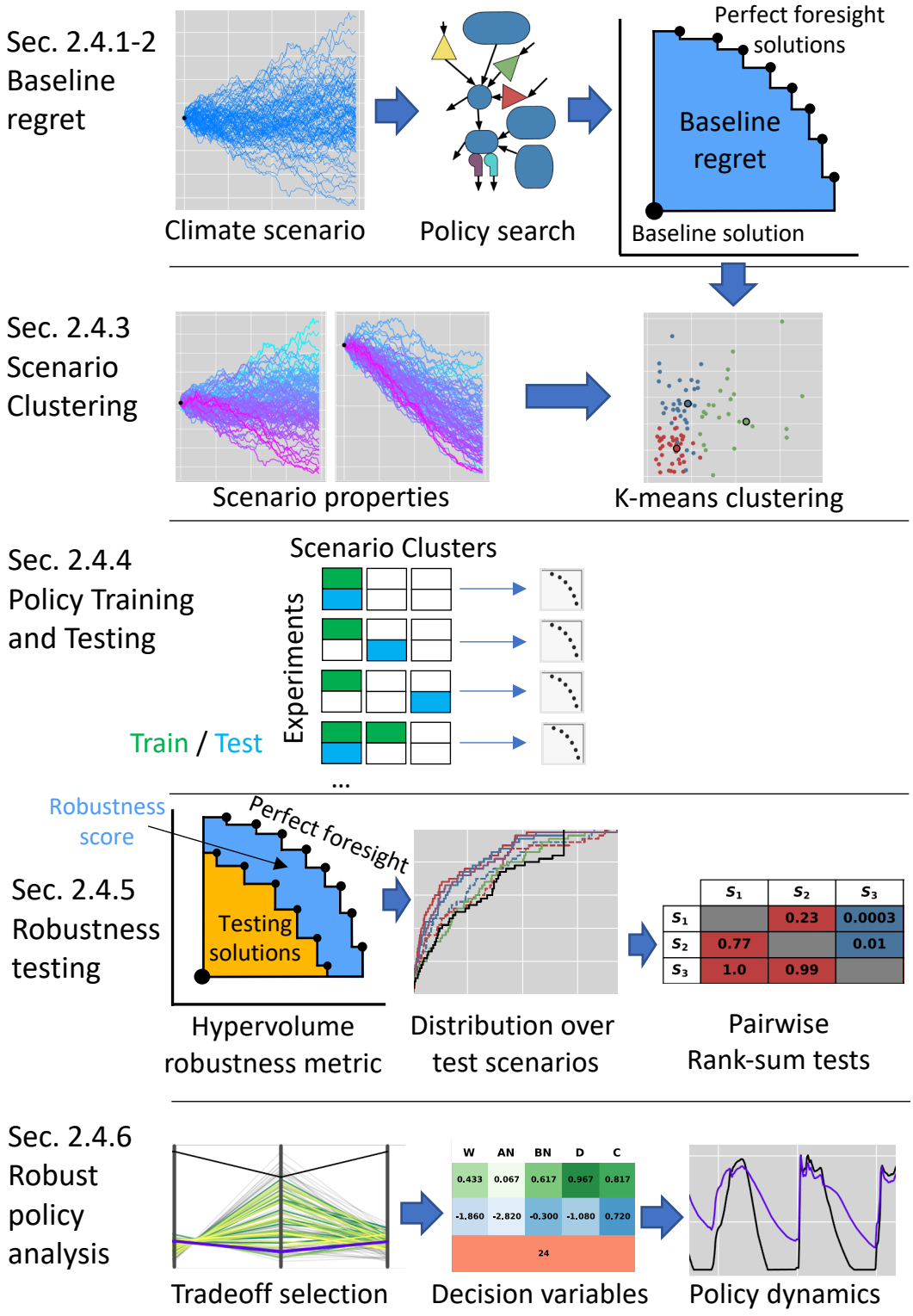

Figure 2.3: Overview of methods

2.4 Methods

The proposed experiments aim to analyze reservoir policy performance on held-out climate projections by selecting different subsets of training scenarios based on their hydrologic properties and a baseline regret property. The experiments require several components (Figure 2.3): (1) policy search, which is used at several steps throughout the experiment; (2) multi-objective baseline regret, which uses perfect foresight optimization to determine the upper bound of system performance in each scenario; (3) unsupervised clustering of the scenario ensemble, taking into account baseline regret as well as hydrologic properties to determine train-test splits; and (4) the robustness of policies trained to one set of scenarios when evaluated on another set. Finally, we analyze the decision variables and dynamics of several robust policies identified using different training sets in the policy search.

2.4.1 Policy search

We employ multi-objective policy search throughout this study to identify operational adaptations by parameterizing the structure of existing rules. In this study, policy search aims to solve the optimization problem:

$$\boldsymbol{\theta}^* = \text{argmax}_\theta \left(\frac{1}{n} \sum_{\boldsymbol{S}} \boldsymbol{J}(\theta, s) \right) \tag{2.6}$$

Where $s \in \boldsymbol{S}$ are the training scenarios over which the particular optimization occurs, and n is the number of scenarios in set S. θ is the vector of decision variables representing the parameters of the operating policy, and $\boldsymbol{J}$ is the vector of objective functions. $\boldsymbol{\theta}^*$ is the set of policies which correspond to the Pareto-optimal solutions.Thus, the policy search attempts to maximize the expected value of objectives $\boldsymbol{J}$ in scenarios s across set $\boldsymbol{S}$. In the specific case of a perfect foresight optimization, $\boldsymbol{S}$ includes only one training scenario.

The alternatives represented by the decision variables include: a revised snowpack-to-streamflow forecasting method, updated release curtailment rules, and changes in the timing

of a dynamic flood control curve. Specifically:

$$\theta = \begin{bmatrix} \bar{Z}_{WYI} & \\ \bar{Z}^r_{wyt} & \forall r, wyt \\ c^r_{\max,wyt} & \forall r, wyt \\ FS^r & \forall r \end{bmatrix} \tag{2.7}$$

Where $\bar{Z}_{WYI}$ is the forecast exceedance level to determine water year indices and $\bar{Z}^r_{wyt}$ is the rest-of-year inflow forecast exceedance level for each reservoir r in each water year type wyt. $c^r_{\max,wyt}$ is the maximum curtailment ratio, and FS^r represents a shift of the reservoir flood-control refill period earlier in the water year. Given the $n_r = 3$ reservoirs and $n_{wyt} = 5$ water year types, this leads to a total of $1+n_r(1+2n_{wyt}) = 34$ decision variables for each optimization. In prior work these individual actions have been shown to improve system performance by enumeration as snowpack decline continues later in the century (Cohen et al.; 2020), but their effect in tandem has not yet been analyzed as a policy search problem. The decision variables are optimized using a normalized set, with 60 discrete values, in order for consistency with alternatives defined in Cohen et al. (2020). Values from the normalized set are transformed in the model to reflect the actual bounds for each variable from the previous study:$\bar{Z}_{WYI}, \bar{Z}^r_{wyt} \in [-3.6, 3.6]$, $c^r_{\max,wyt} \in [0, 1]$, and $FS^r \in [0, 60]$. The choice of decision variables reflecting system parameters rather than a universal approximator function, such as a neural network (e.g., Salazar et al.; 2016; Giuliani et al.; 2017), is intended to support the interpretability of the resulting policies, as well as their compatibility with already in-place system operations. However, there are currently efforts to formulate methods that improve the interpretability of neural network-based policies, for example via sensitivity analysis (Quinn et al.; 2019).

The objectives $\boldsymbol{J}$ contain five performance metrics, including flood control, reservoir carryover storage at the end of the water year, Delta outflow representing salinity control and environmental benefits, Delta exports for water supply, and hydropower generation. The

expectation of each objective across a scenario set is to be maximized (Equation 2.6). To objective values over each scenario are calculated according to:

$$J_{\text{Flood}}(\theta) = -\sum_{t=1}^{T}\sum_{r=1}^{3} \max\left(u_t^r - DQ^r, 0\right)^2 \tag{2.8}$$

$$J_{\text{Carryover}}(\theta) = \sum_{y=1}^{N}\sum_{r=1}^{3} Cr_y^r \tag{2.9}$$

$$J_{\text{Outflow}}(\theta) = \sum_{t=1}^{T} \left[Q_{in,t} - (TRP_t + HRO_t)\right] \tag{2.10}$$

$$J_{\text{Exports}}(\theta) = \sum_{t=1}^{T} \left[TRP_t + HRO_t\right] \tag{2.11}$$

$$J_{\text{Hydro}}(\theta) = \sum_{t=1}^{T}\sum_{r=1}^{3} HP_t^r \tag{2.12}$$

Where T is the number of days t in the simulation period, while N is the number of water years y. In the flooding objective, DQ^r is the downstream levee capacity of reservoir r. Note in Equation 2.8 we maximize negative flooding for consistency with maximization of the other objectives. Cr_y^r is the carryover storage in reservoir r at the end of water year y. $Q_{in,t}$ is the Delta inflow on day t, while TRP_t (Tracy pumping plant) and HRP_t (Harvey O. Banks pumping plant) are exports from the Delta to the Central Valley Project and State Water Project, respectively. Lastly, HP_t^r represents the hydropower production from reservoir r on day t. These objective functions are intended to capture the necessary balance between key aspects of system performance.

The optimization is performed using the Non-Dominated Genetic Sorting Algorithm (NSGAIII) (Deb and Jain; 2013) via the open source Platypus library (Hadka; 2015). To support this choice, algorithm performance was tested over 70 scenarios with three random trials of 50,000 maximum number of function evaluations (NFE) each, obtaining similar results for each trial. The results were compared with alternative MOEAs, including ϵ-MOEA, NSGAII, and SPEA2, which showed no significant improvement over NSGAIII for this problem. In

further instances where this problem is solved, 10,000 NFE are used when optimizing to a single scenario (perfect foresight), while all other optimizations with various train-test splits use 50,000 NFE. As the number of scenarios increases, the optimization is slower to converge visually based on hypervolume, influencing the choice of 10,000 vs 50,000 NFE. All optimization runs were performed on the HPC1 cluster at UC Davis, which includes 60 nodes with 16 cores each running at 2.4 GHz.

2.4.2 Baseline regret

Multi-objective baseline regret quantifies the maximum level to which system performance can be improved for a particular scenario, within the constraints imposed by the policy function and existing infrastructure. This property is scenario-specific: each time it is calculated the baseline policy is held constant, while the scenario differs. This concept draws from the Expected Value of Perfect Information (EVPI) metric proposed by Giuliani, Pianosi and Castelletti (2015) for multi-objective problems. By incorporating the results of a perfect foresight optimization, it couples the EVPI approach with a regret metric (Savage; 1951), which describes the performance of a policy based on its distance from the best possible alternative. Traditional decision making under uncertainty problems often uses the minimax regret approach, in which the goal is to choose the alternative that minimizes the maximum regret across all scenarios (e.g., Giuliani and Castelletti; 2016). The baseline regret metric differs from minimax regret because it applies only to the no action case of an individual scenario, rather than policy alternatives, reflecting the EVPI approach. It is based on the performance of both a baseline solution and perfect foresight solution set for each scenario, as the performance of any other effective policy solution is expected to be bounded by these two. As a result, the baseline regret partially depends on the suitability of the baseline policy for each climate scenario. However, this still reflects the ability of the system to adapt to future change, even if it is starting from a poor baseline.

2.4.2.1 Baseline policy performance and perfect foresight optimization

The baseline policy simulation uses parameters θ_B to best represent the dynamics of the system shown in historical observations. The resulting baseline policy solution performance is denoted as $J_B(s) = J(\theta_B, s)$, a one-dimensional vector containing a single value for each objective, rather than a full Pareto set. Under the baseline policy, this performance is not optimized; it should be viewed as a simplified representation of the several performance considerations of a real-world system operator.

The upper bound performance is established by a perfect foresight optimization. In this case, the policy search is performed over each scenario individually to determine the objective values if the future were known exactly. We define the perfect foresight performance metric as $\boldsymbol{J_P}(s) = \boldsymbol{J_P}(\boldsymbol{\theta^*}, s)$, where the optimized parameters $\boldsymbol{\theta^*}$ are specific to the training set consisting of the single scenario s.

2.4.2.2 Hypervolume metric

We use a hypervolume metric to quantify the baseline regret of each scenario s in the ensemble. The hypervolume is defined as the volume in the objective space between the perfect foresight Pareto set $\boldsymbol{J_P}(s)$ and the baseline policy performance J_B, which is used as a reference point. While baseline regret is calculated in a five-dimensional objective space for this application, the hypervolume concept is illustrated in two dimensions in the top row of Figure 2.3. Solutions in the perfect foresight Pareto front $\boldsymbol{J_P}(s)$ are anticipated to dominate the baseline policy performance $J_B(s)$. The rare solutions for which this does not occur are not considered in the remainder of the calculations. In general, a larger hypervolume value indicates improvement over the baseline policy as well as a higher variety in alternatives among the Pareto set.

Initially disregarding the baseline solution, we normalize all objective values in $\boldsymbol{J_P}(s) \in [0, 1]$ to reduce scaling issues between the objectives. The baseline policy performance is normalized accordingly to $J_B \in (-\infty, 0]$ to allow for consistent comparison of baseline regret

across scenarios. We then calculate the baseline regret $R(s)$ based on the hypervolume $h(\dots)$ between the perfect foresight Pareto set and the baseline policy reference point, such that:

$$R(s) = h\left(\boldsymbol{J_P}(s), J_B(s)\right) \tag{2.13}$$

The baseline regret describes the performance of a perfect foresight optimization relative to the baseline for each climate scenario. Because the objective values are normalized, it provides an upper bound performance metric that can be directly compared across scenarios for a given policy.

2.4.3 Scenario clustering

Unsupervised clustering provides the basis for separating climate projections into training and test sets for the policy search. The clustering is based on three features: averaged annual streamflow, averaged peak annual snow-water equivalent, and the baseline regret metric described above. These features are calculated on the time horizon 2070-2100, which is chosen as the period of analysis due to its large deviation from historical hydrologic conditions and thus high-regret (Figure 2.10 in the Appendix). This time period also contains much more variability in hydrologic properties than do earlier periods in the projected time horizon (Figure 2.2).

The three features are calculated for each of the 97 scenarios in the ensemble and clustered using the K-means algorithm with $K = 3$, equal to the number of features. This allows for a minimally complex characterization of scenario properties based on cluster centroids. The resulting clusters are denoted as C_1, C_2, and C_3.

2.4.4 Training and test sets

We first split each cluster C_i randomly into roughly equal training and test subsets, S and S_t, respectively. Various combinations of these training and test subsets make up the

overall training and test sets, S_i and S_{tj}, respectively (Table 2.1). While the goal of this division is to ensure that test information is never used in training, we acknowledge the possibility for interdependence among the ensemble of climate scenarios, for example using the same model or emissions scenario, or different models relying on the same components (Steinschneider, McCrary, Mearns and Brown; 2015). Additionally,

2.4.4.1 Training and testing

Policy search runs separately for each training set to identify the Pareto set of policies $\boldsymbol{\theta}^*_{S_i}$ corresponding to each training set of scenarios S_i:

$$\boldsymbol{\theta}^*_{S_i} = \text{argmax}_\theta J(\theta, S_i) \tag{2.14}$$

In order to increase the extent and continuity of the Pareto-optimal solutions, three trials of each optimization are run using varying random seeds. The use of baseline regret as a scenario property links the perfect foresight optimization to various combinations of training scenarios without explicitly using perfect foresight to inform the choice of all training scenarios.

We next re-evaluate the policies trained to set S_i over each scenario in the test set S_{tj}, resulting in a set of objectives $\boldsymbol{J}_{S_i}(s)$ for each test scenario:

$$\boldsymbol{J}_{S_i}(s) = J(\boldsymbol{\theta}^*_{S_i}, s) \quad \forall s \in S_{tj} \tag{2.15}$$

We consider only solutions that also dominate the baseline policy for all scenarios in the test set. This is achieved via a filtering step which identifies the solutions that will at a minimum outperform the status quo in all re-evaluations. We identify these particular policies and solutions as policy set $\boldsymbol{\theta}_{i,j}$ and solution set $\boldsymbol{J}_{i,j}$. This process results in a total of 28 pairwise combinations of training and test scenarios.

2.4.4.2 Set diversity

While the training and test sets are delineated via K-means clustering, the diversity of each set should also be considered for the analysis. This can help determine if policy performance across train-test set combinations is influenced by the scenario diversity in each set. Diversity is determined via Equations 2.16 and 2.17, adapted from Carlsen et al. (2016) and Eker and Kwakkel (2018). In Equation 2.16, D_S represents the diversity of set S. w, the weight assigned to the extent the mean distance, is 0.5 in this case. $d_{l,k}$ is the Euclidean distance based on the $m = 3$ features (FNF, SWE, baseline regret) for scenarios l and k. In Equation 2.17 $\bar{f}_{m,l}$ and $\bar{f}_{m,k}$ are the values for these features. To ensure equal weighting of all features, $\bar{f}_{m,l}$ and $\bar{f}_{m,k}$ are normalized from 0 to 1 across all scenarios.

$$D_S = (1 - w) \min_{\forall l,k \in S} \{d_{l,k}\} + w \underset{\forall l,k \in S}{\text{mean}} \{d_{l,k}\} \tag{2.16}$$

$$d_{l,k} = \sqrt{\sum_m \left(\bar{f}_{m,l} - \bar{f}_{m,k}\right)^2} \tag{2.17}$$

2.4.5 Policy robustness

We would like to evaluate the robustness of policies trained to set S_i when re-evaluated over each scenario s in set S_{tj}, for all combinations of (i, j). Because the performance is multi-objective across a range of hydrologic scenarios, we use a hypervolume metric normalized by the baseline regret (Section 2.4.2) to represent the robustness of the policy set as a whole.

2.4.5.1 Hypervolume robustness metric

Robustness is represented by a normalized hypervolume metric for each solution set. The hypervolume for a solution set of train-test set combination (i, j) applied to scenario s is defined as that between the baseline reference point $J_B(s)$ and solution set $\boldsymbol{J}_{i,j}$: $h\left(\boldsymbol{J}_{i,j}(s), J_B(s)\right)$. This is normalized by the baseline regret $R(s)$, giving the hypervolume robustnesss metric

$HR_{i,j}(s)$:

$$HR_{i,j}(s) = \frac{h(\boldsymbol{J}_{i,j}(s), J_B(s))}{R(s)} \tag{2.18}$$

Thus the hypervolume robustness metric will always be a fraction of the baseline regret $R(s)$, ensuring that it can be appropriately compared across train-test combinations (see fourth row in Figure 2.3). A higher normalized hypervolume metric denotes a more robust policy set, with a value of 1 equaling the performance of perfect foresight policies. This ensures that the robustness of a policy set is not measured only by its ability to improve system performance relative to the baseline, but also the extent to which the policies are able to reach the maximum attainable level of system performance.

For each train test-combination, we can obtain a set of hypervolume robustness metric values $HRT_{i,j}$, where:

$$HRT_{i,j} = \{HR_{i,j}(s_1), HR_{i,j}(s_2), \ldots, HR_{i,j}(s_n)\} \quad \forall s \in S_{tj} \tag{2.19}$$

In general, larger hypervolume robustness metrics will indicate two properties of the objective outputs. The first is that as hypervolumes increase, the distance between the baseline policy $J_B(s)$ and policy performance set $\boldsymbol{J}_{i,j}(s)$ will increase, indicating higher performance improvements compared with the baseline policy. Additionally, a larger hypervolume indicates a higher diversity of solutions across the Pareto front.

2.4.5.2 Rank-sum tests

For each pair of train-test combinations with identical test sets, we perform a one-sided Mann-Whitney U test (Mann and Whitney; 1947) to determine if the hypervolume of a given training set exceeds that of a second training set when evaluated on the same test set. This test aims to determine if policies trained to a test set with particular properties are significantly more robust. With $p \leq 0.05$, we reject the null hypothesis and conclude that

the distribution of hypervolume across test scenarios S_{tj} in sample $HRT_{1,j}$ is greater than that in sample $HRT_{2,j}$ with statistical significance.

2.4.6 Policy analysis

Finally, we analyze individual policies chosen from the most robust training sets by considering tradeoffs between the objective values. The decision variables of these policies are compared to the baseline policy to understand what combinations of adaptations system operations could be promising under a range of future climates. We then compare the dynamics in terms of reservoir storage and water supply exports to those obtained by simulating the baseline policy on the same hydrologic inputs, and then relate key differences to the decision variables interpreted in the context of the system.

2.5 Results and Discussion

2.5.1 Scenario clusters

Scenarios are divided into three clusters based on their streamflow, snowpack, and baseline regret, as shown in Figure 2.4. Based on the cluster centroids, we define them as high-regret, low-regret wet, and low-regret dry. The high-regret scenarios contain a mix of streamflow and snowpack values distributed throughout their respective ranges, indicating that baseline regret does not solely depend on annual hydrologic properties. The clear separation between the high-regret and low-regret clusters suggests the possible utility of this metric in determining combinations of training scenarios in policy search experiments.

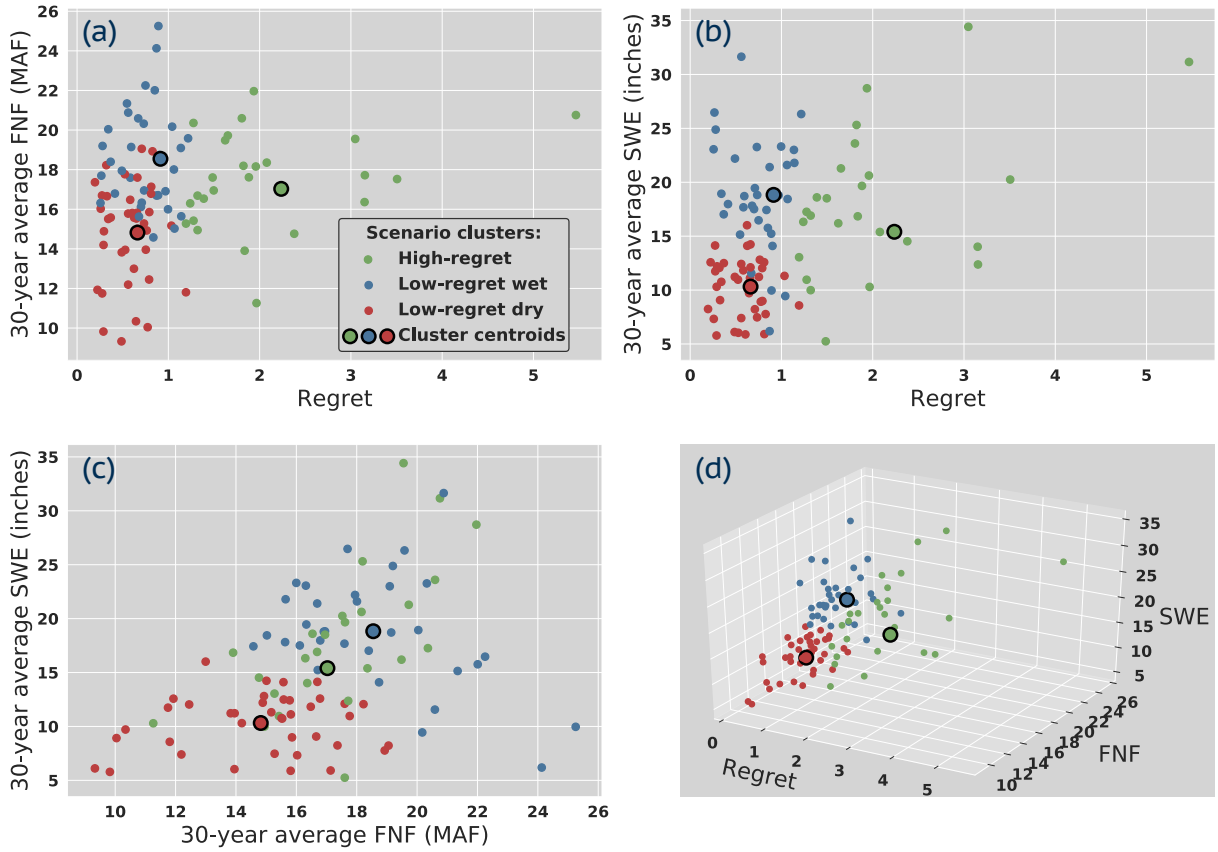

Figure 2.4: Scenario clusters in (a,b,c) two-dimensional projections and (d) all three properties: full natural flow (streamflow, FNF), snow-water equivalent (snowpack, SWE), and baseline regret.

The low-regret scenarios occur in both wet and dry clusters. However, the ranges of streamflow and snowpack values overlap across these two clusters. (Figure 2.4a,b). Specifically, this occurs in two cases: first, some wetter scenarios may also show high levels of snowpack decline due to severely warmer temperatures; second, there exist dry scenarios with relatively higher snowpack values than other low-flow scenarios due to less warming. This overlap, along with the much clearer separation between high-regret and low-regret clusters, supports the choice of $K = 3$ clusters to minimize complexity.

2.5.2 Training set robustness comparison

The three clusters, each split randomly into training and test subsets, are combined to create different train-test splits. The overall proportion of training to test sets out of all available scenarios is 50:47 (Table 2.1). From all possible combinations combined to create different train-test splits created in this process, a total of seven training sets and four test sets are chosen to demonstrate the training-testing process. These are described in Table 2.1.

Training Sets S_i	# of scenarios	Diversity	Test Sets S_{tj}	# of scenarios	Diversity
S_1 High-regret	13	0.241	S_{t1} High-regret	12	0.206
S_2 Low-regret wet	17	0.199	S_{t2} Low-regret wet	16	0.162
S_3 Low-regret dry	20	0.130	S_{t3} Low-regret dry	19	0.131
S_4 High-regret/low-regret wet $(S_1 \cup S_2)$	30	0.233	S_{t4} All test scenarios $(S_{t1} \cup S_{t2} \cup S_{t3})$	47	0.202
S_5 High-regret/low-regret dry $(S_1 \cup S_3)$	33	0.230			
S_6 Low-regret wet/dry $(S_2 \cup S_3)$	37	0.197			
S_7 All training scenarios $(S_1 \cup S_2 \cup S_3)$	50	0.233			

Table 2.1: Outline and properties of training and test sets.

These sets are used to determine the performance of policies optimized to each training set when re-evaluated in each test set, measured according to the hypervolume robustness metric. Figure 2.5 shows the distributions of the resulting hypervolume metric for each train-test split, plotted as cumulative distributions.

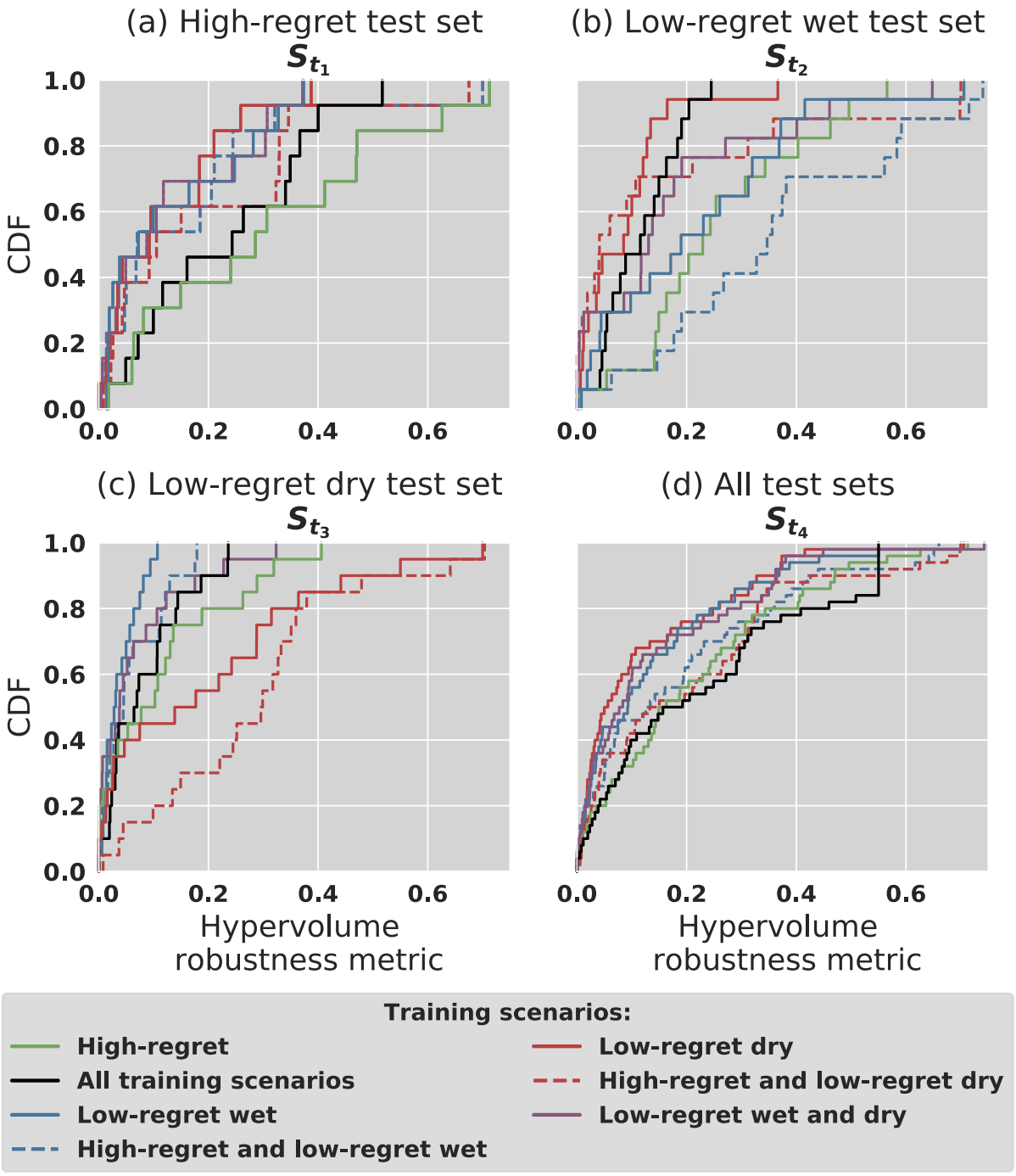

Figure 2.5: Cumulative distributions of the hypervolume metric evaluated on each test set (A-D). Each CDF represents the distribution of performance over all scenarios in the test set for the Pareto front of policies trained to the scenarios identified by the line style.

Distributions shifted further right indicate higher robustness of the policy sets over the test set. While these distributions support the interpretation of the performance differences

between policy sets trained on different scenarios, the rankings of policy sets must be shown to be statistically significant. These conclusions are made using the Mann-Whitney U test between each pair of train sets over each test set, with results shown in Figure 2.6.

(a) High-regret test set (S_{t_1})

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1		0.01	0.01	0.02	0.02	0.01	0.15
S_2	0.99		0.5	0.38	0.64	0.48	0.96
S_3	0.99	0.5		0.57	0.71	0.56	0.97
S_4	0.98	0.62	0.43		0.75	0.62	0.96
S_5	0.98	0.36	0.29	0.25		0.38	0.96
S_6	0.99	0.52	0.44	0.38	0.62		0.96
S_7	0.85	0.04	0.03	0.04	0.04	0.04	

(b) Low-regret wet test set (S_{t_2})

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1		0.23	0.0003	0.96	0.01	0.05	0.13
S_2	0.77		0.01	0.98	0.01	0.33	0.43
S_3	1.0	0.99		1.0	0.14	0.99	1.0
S_4	0.04	0.02	8e-05		0.002	0.01	0.01
S_5	0.99	0.99	0.86	1.0		0.97	0.98
S_6	0.95	0.67	0.01	0.99	0.03		0.98
S_7	0.87	0.57	0.0003	0.99	0.02	0.02	

(c) Low-regret dry test set (S_{t_3})

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1		0.04	0.86	0.11	0.97	0.14	0.35
S_2	0.96		0.99	0.71	1.0	0.43	0.98
S_3	0.14	0.01		0.02	0.96	0.02	0.10
S_4	0.89	0.29	0.98		1.0	0.47	0.85
S_5	0.03	1e-05	0.04	0.0005		0.0002	0.0005
S_6	0.86	0.57	0.98	0.53	1.0		0.87
S_7	0.65	0.02	0.9	0.15	1.0	0.13	

(d) All test scenarios (S_{t_4})

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
S_1		0.005	0.01	0.11	0.19	0.02	0.64
S_2	1.0		0.67	0.86	0.86	0.68	1.0
S_3	0.99	0.33		0.77	0.81	0.47	0.99
S_4	0.89	0.14	0.24		0.57	0.04	0.92
S_5	0.81	0.14	0.19	0.43		0.03	0.93
S_6	0.98	0.32	0.53	0.96	0.97		0.99
S_7	0.36	0.001	0.01	0.08	0.07	0.01	

Reject null (row > column) Fail to reject (row ≯ column)

Figure 2.6: Results of pairwise Mann-Whitney U rank-sum tests. Each test has the null hypothesis that the hypervolume metric associated with the training set in each row is less than or equal to that associated with the column. The subplots correspond to the four test sets. A row where the null hypothesis is rejected (blue) for each cell denotes a robust policy set that ranks highest for the particular test set.

For the high-regret test set (Figure 2.6a), the most robust policies are those optimized to the high-regret and all-scenario training sets, where the latter contains the former. Neither of these significantly outperforms the other. This finding is not surprising, as the policies trained to scenarios with similar properties demonstrate the best out-of-sample performance. However, this result does not always hold for the other test sets. For example, in the test set consisting of low-regret wet scenarios (Figure 2.6b), the best-performing set of policies are those trained to a mix of high-regret and low-regret wet scenarios (S_4), which ranks higher than every other training set. The training sets containing dry scenarios and lacking wet scenarios (S_3 and S_5) perform worst for the high-regret test set. Set S_1, consisting of only high-regret scenarios, outperforms set S_6, which consists of wet and dry low-regret scenarios.

This indicates that training to only high-regret scenarios may be more effective than training to low-regret scenarios regardless of the variability in scenarios' hydrologic properties. This result shows that adding high-regret scenarios to the training set—whether they are wet or dry—improves the robustness of the optimized policies when tested in out-of-sample wet scenarios. Additionally, including low-regret dry scenarios in training sets for policies tested on low-regret wet scenarios degrades policy performance. Similar results are shown for the low-regret dry test set (Figure 2.6c), where the highest ranking training set is again not only the dry scenarios (S_3), but also a mix of high-regret and low-regret dry (S_5) scenarios. In addition, training sets including low-regret wet scenarios (S_2 and S_4) have the lowest ranking when their corresponding policies are simulated over the low-regret dry test set.Lastly, the high-regret training set continues to outperform the low-regret wet/dry training set for S_{t4}, further highlighting the good training value of high-regret scenarios.

For the final test set S_{t4}, which includes all testing scenarios (Figure 2.6d), the majority of Mann-Whitney U tests fail to reject the null hypothesis. However, results indicate that the high-regret training set (S_1) outperforms the low-regret training sets (S_2, S_3, and S_6), as does the all-scenarios training set (S_7). Especially notable is that the high-regret training set S_1 outperforms the combined wet-dry low regret training set S_6 when testing to all scenarios. Since both of these training sets have wide ranges for the hydrologic properties, this further highlights the benefit of high-regret training scenarios over low-regret scenarios.

The diversity of the training sets can be analyzed in tandem with these results. The high-regret (S_1) and all training scenario (S_7) sets are the most diverse (Table 2.1). Based on our specific quantification of set diversity, this is an artifact specifically of the high-regret values, which contain more outliers and a more skewed distribution across all scenarios (Figure 2.4a,b). The high mean distances that occur from this cause the diversity values to be larger whenever the high-regret scenarios are included in a set. This leads to the fact that the three sets which do not contain the high-regret scenarios (S_2, S_3, and S_6) are the least diverse of the sets. It could then be concluded that the larger diversity of the high-regret set influences

its good performance. However, since this value is skewed by just a few outliers, it should not be considered the only reason for the effective training value of the high-regret sets.

Training sets S_1 (high-regret only) and S_6 (low-regret wet/dry) have similar ranges across both hydrologic properties. The lower diversity of set S_6 is influenced by its small range in baseline regret values, as well as the fact that is has several scenarios in close proximity in terms of hydrologic properties (Figure 2.4c), leading to a skewed minimum distance value in the diversity calculation. Set S_6 has almost three times the nubmer of scenarios as set S_1, which contributes to its low diversity calculation. Several close-proximity scenarios could be omitted to make set S_6 more diverse. This would not improve the performance as the set would lose valuable training data and potential for overfitting would increase. Therefore, the high diversity of set S_1 is not the only factor controlling the set's good performance. Its high baseline regret values will enable the policy search to find solutions more robust to vulnerable conditions. Additionally, there may be many other scenario properties that are not examined in this study which contribute to set performance and scenario training value. These include hydroclimatic properties such as temperature rise, flood frequencies, flow timing, precipitation, drought patterns, soil moisture, and evapotranspiration.

Because the high-regret training set performs no worse than training to all scenarios, the strategy of designing a training set around scenarios with high baseline regret may serve to reduce the computational cost of policy search for large-ensemble cases, and/or to reserve more scenarios for testing. To support this point, Table 2.2 compares the computational cost for different aspects of policy training in this study. Training to scenarios with high baseline regret (which includes the perfect foresight optimizations) required 9,733 computing hours, roughly three times less than training to all scenarios. Training to scenarios with high baseline regret improves the efficiency of policy search without sacrificing robustness relative to the case of training to all scenarios. This denotes the benefit of analyzing the hydrology and baseline regret of scenarios before a train/test split is determined.

Thus, it is also possible to determine the conditions under which a high baseline regret

set will give computational benefits by generalizing the requirements outlined in Table 2.2. This condition is described as:

$$f_p\rho_p + \eta f_r\rho < f_A\rho_A \tag{2.1}$$

Where η represents fraction of overall scenarios which are in the high-regret set, f_P, f_R, and f_A denote the number of function evaluations, and ρ_P, ρ_R, and ρ_A denote the number of random seeds for each of the perfect foresight, high-regret only, and all training scenario sets, respectively. This generalization can potentially be applied to other planning problems in which the baseline regret is determined a priori, and where there is a choice about how many high-regret solutions to include in the training set.

This analysis has important implications for the generalizability of this approach. Several variables may be degrees of freedom, for instance numbers of random seeds ρ and function evaluations f necessary for convergence to diverse and near-optimal Pareto-solutions sets will vary across models. The fraction of high regret scenarios η may differ based on the number of clusters chosen. In some instances, if the level of baseline regret is not a significant source of variation among scenarios, it may not provide a way of separating different training sets using a clustering approach. Furthermore, differences in performance among training sets may be due to confounding factors not reflected in the abstracted scenario properties, especially for hydrologic timeseries which can be summarized in a number of different ways. However, the proposed clustering and train/test methodology is still generalizable across environmental planning applications to pinpoint the most important scenario properties for policy training and out-of-sample performance, therefore discovering conditions for computational benefits.

Furthermore, results presented in Figures 2.5 and 2.6 must be interpreted in light of the fact that the future climate trajectory is uncertain. It is likely that more information about future hydrology will be collected over time, and this process could complement policy search methods in the context of dynamic planning (e.g., Hui et al.; 2018; Fletcher et al.; 2019). Therefore, in this study the methodology aims to identify a training strategy that leads to robust outcomes to both uncertain and clustered future climate, measured according

	Max NFE per optimization (f)	Scenarios per optimization (n)	Time per function evaluation	# of random seeds (ρ)	Total computing hours
Perfect foresight (P)	10,000	1	12s	97 (individual scenario trials)	3,233h
High-regret only (R)	50,000	13	156s	3 (random seed trials)	6,500h
All training scenarios (A)	50,000	49	588s	3 (random seed trials)	24,500h

Table 2.2: Description of computing requirements for several optimizations included in this study. Note that less NFEs are required for a perfect foresight, as these optimizations are quicker to converge. Times per function evaluation and total computing hours are specific to the UC Davis HPC1 computing cluster.

to multi-objective performance bounded by the baseline policy and perfect foresight cases. We find that training to scenarios with high baseline regret is competitive with training to all scenarios across a range of future climates, and often leads to the best out-of-sample performance. This is likely due to higher inter-annual variability in these scenarios. Based on a higher diversity of extreme events across individual scenarios and potential poor baseline performance in the high-regret cluster, solutions will give both a wider variety of tradeoffs in objectives and improvements relative to baseline policy performance. These findings extend to both wet and dry futures, where the inclusion of high-regret scenarios in the training set outperforms using exclusively either wet or dry training scenarios. This result links to the importance of evaluating perfect foresight policies in individual scenarios when designing the training set to establish an upper bound for system performance.

2.5.3 Policy analysis

The final step of the analysis is to determine what specific adaptations are implemented by the robust policies. This analysis focuses on six specific train-test splits, chosen based on their high-ranking performance: (1) policies trained on set S_4 and tested on set S_{t2}, (2) policies trained on the set S_5 tested on set S_{t3}, (3,4) policies trained on set S_1 tested on sets S_{t2} and S_{t3}, and (5,6) policies trained on set S_6 tested on sets S_{t2} and S_{t3}. The average performance measures across all scenarios for these sets are shown by the highlighted solutions on the parallel-axis plots in Figures 2.7. While the expected value of all highlighted solutions dominates the baseline policy, there are still several significant tradeoffs between

the objectives, indicated by their nonlinear correlations (see Section 2.7.3 in the Appendix). For the S_4/S_{t2} train-test combination, these include statistically significant tradeoffs between hydropower and flooding ($\rho = 0.53$) and water supply and Delta outflow ($\rho = 0.97$). The same tradeoffs exist in the dry test scenarios, which also exhibit tradeoffs between carryover storage and flooding ($\rho = 0.49$). In general, these relationships reflect the fact that higher storage levels benefit several of the proposed objectives, although they can be detrimental to the flooding objective, which is to be minimized. While this high water elevation benefits the hydropower and carryover storage objectives, it can induce larger releases if large storms occur later in the spring.

We have shown that training set S_4 (high-regret/low-regret wet) will yield the best performing policies for the low-regret wet test set S_{t2} (Figure 2.6b). Likewise, training set S_5 (high-regret/low-regret dry) will yield the best performing policies for the low-regret wet test set S_{t3} (Figure 2.6b). This is reflected in Figure 2.7(a,b,c),(e,f,g) where the highlighted Pareto solutions for S_4 and S_5 are shifted higher than S_1 and S_6 over their particular test sets, as shown by the higher maximum percent of baseline values (for the flood objective lower minimum) in Figure 2.7a, e. These ranges in Figure 2.7(a,b),(e,f) also reflect the better performance of the high-regret training set S_1 over the low-regret wet/dry training set S_6 for both test sets.

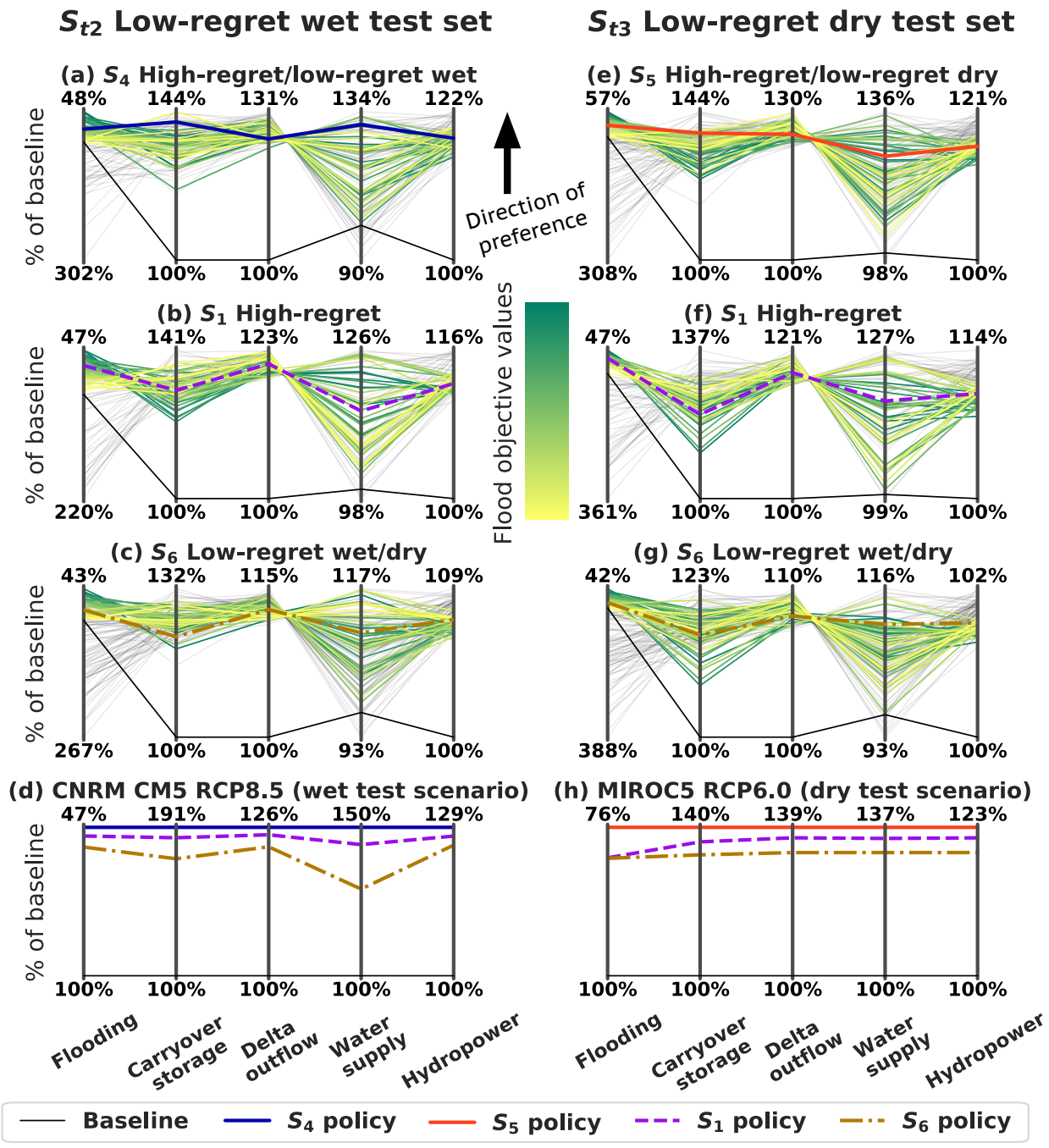

Figure 2.7: Parallel axis plots displaying various results of train-test combinations. (a,b,c) policies trained on sets S_4, S_1, and S_6, respectively, tested on set S_{t2}. (e,f,g) policies trained on sets S_5, S_1, and S_6, respectively, tested on set S_{t3}. Solutions highlighted by the yellow-green gradient represent solutions for which the expected value of solutions across the testing scenarios dominate the baseline policy solution. This gradient represents the solutions' ranks for the flood objective in their particular Pareto set. Grey solutions dominate the baseline in terms of expected value, but do not dominate the baseline for each individual scenario in the testing set. The individual highlighted solutions denote four different compromise policies (S_4, S_5, S_1, and S_6 policies) that are analyzed in detail. Subplots (d) and (h) show robust performance of the four compromise policies over an individual scenario in the corresponding test set. These particular scenarios are also highlighted in Figure 2.2.

We next examine the four compromise policies that balance the tradeoffs in performance measures, denoted as the S_4, S_5, S_1, and S_6 policies in Figure 2.7. The S_4 and S_5 policies, coming from the most robust training sets for the respective test sets, also give the best performance on the individual scenarios (Figure 2.7d,h). The alternatives that these policies employ are shown in Figure 2.8 along with a comparison to the decision variables of the baseline policy. Each column in the tables represents the decision variable which occurs for that specific water year type.

In the S_4 policy, Shasta and Folsom reservoirs have higher maximum allowable curtailments $c^r_{max,wyt}$ than in the baseline policy. These higher maximum curtailment levels will allow for increased hedging of releases. The curtailments for Oroville reservoir are higher in wet, above and below normal years, but lower in dry and critical years. All three reservoirs also have a flood pool shift of at least 10 days forward in the water year for the S_4 policy. In wet and above normal years, Shasta and Oroville use low $\bar{Z}^r_{wyt}$ values, indicating a very conservative forecast with a high exceedance level. In drier water year types, the $\bar{Z}^r_{wyt}$ values are generally close to or greater than the baseline exceedance levels. For Folsom reservoir, these values vary much more across water year types. The differences between operational adaptations at each reservoir highlight the complexity of managing the multi-reservoir system, and the potential to design adaptations for system-wide benefit.

Figure 2.9 shows the system dynamics of the baseline policy compared to the compromise policies in a time series over one scenario from each corresponding test set: an RCP 8.5 scenario (CNRM-CM5) for low-regret wet, and an RCP 6.0 (MIROC5) scenario for low-regret dry. Under the baseline policy, reservoir storage levels are vulnerable to snowmelt loss regardless of water year type, evidenced by low storage levels in the irrigation season even in wetter years. The S_4 policy mitigates this vulnerability via an intra-annual hedging, resulting in higher reservoir storage during the early irrigation season (May-June); The S_5 policy functions similarly. For both policies, this intra-annual hedging dynamic is supported by the adapted snowpack-to-streamflow forecasts, where underpredictions will cause some release

curtailments to conserve for potential low inflows later in the season. However, curtailments can be partially avoided with higher carryover storage due to the flood pool shift. This seasonal shift is also reflected in the Delta exports (Figure 2.9d,h), which maximize total volume by shifting throughout the year. The remaining S_6 and S_1 policies also exhibit the intra-annual hedging strategy (see Section 2.7.4 in the Appendix for these policies' decision variables). However, given that reservoir storage becomes higher in the flood season and carryover storage drops lower when these policies are deployed, they are slightly less effective (Figure 2.9b,c,f,g). Additionally, they often will have periods of low Delta exports (Figure 2.9d,h). This highlights that a policy from the best performing training set for a particular test set may be more likely to give better performance for scenarios in that test set.

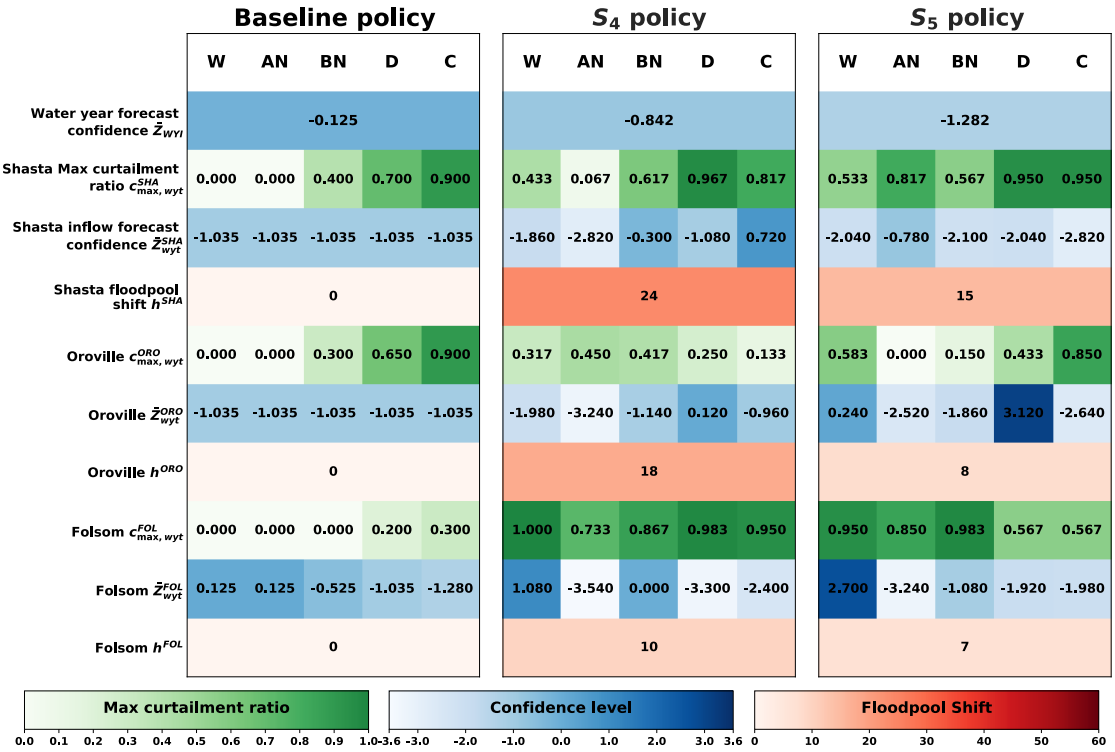

Figure 2.8: Policy tables showing decision variables for: the baseline policy, the S_4 policy (high-regret/low-regret wet set), and the S_5 policy (high-regret/low-regret dry set). The columns denote water year type classifications associated with each decision variable, corresponding to wet, above normal, below normal, dry, and critical. Policy tables for the S_1 and S_6 policies can be found in Section 2.7.4 of the Appendix.

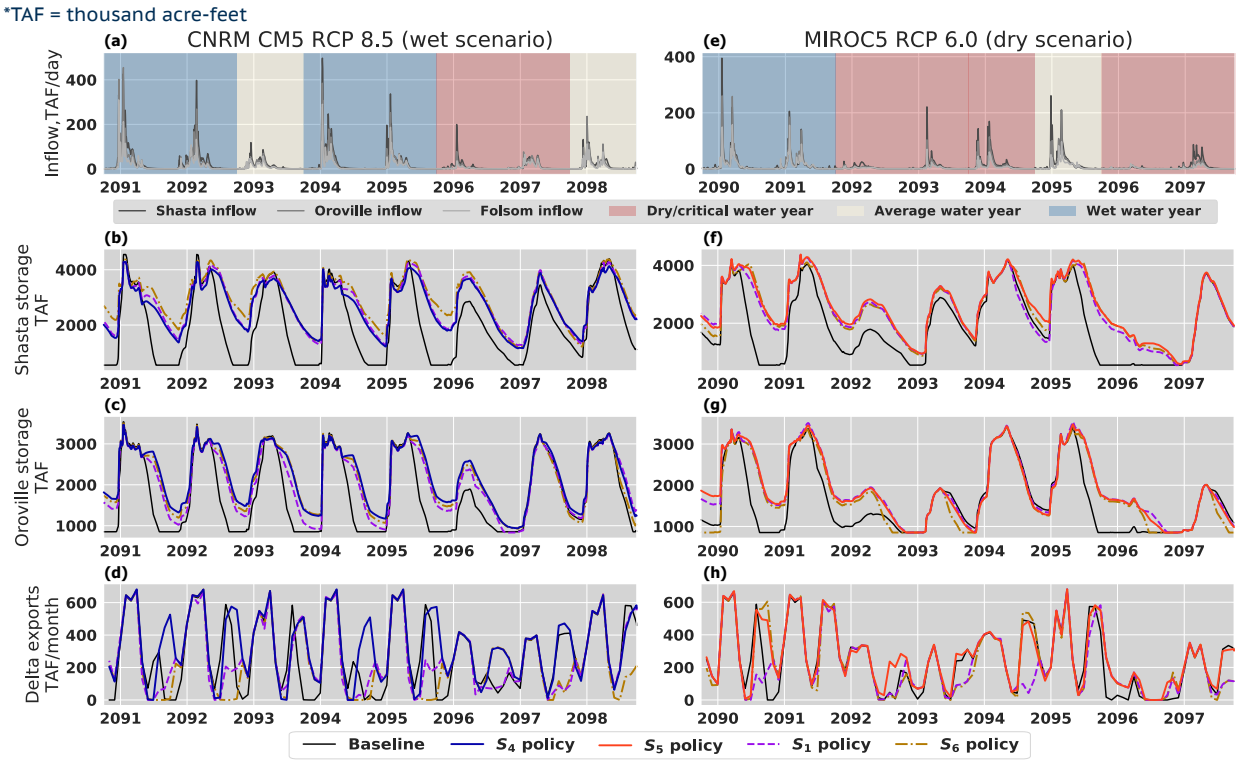

Figure 2.9: Time series of system states and flows for baseline and adaptation policies in individual test scenarios. The left and right columns show results from the individual wet and dry scenarios, respectively. (a,e) Daily inflows to each reservoir with water year types highlighted; (b,c,f,g) Reservoir storage; (d,h) monthly Delta exports, primarily for agricultural and municipal water supply.

There are two major differences between the the S_4 and S_5 policies stemming from the hydrologic properties of their respective training scenarios. The first is that the S_5 policy tends to curtail releases more during dry and critical years, reflected in its conservative forecasts and high maximum curtailment allowances. The S_4 policy hedges less during dry and critical years, and instead relies on larger storage brought about by intra-annual hedging. This is further driven by the low maximum curtailment allowances for Oroville during these water year types. The second difference is that the S_4 policy tends to hold less storage during the flood season than the baseline policy, while the S_5 policy does not. For the S_4 policy, this makes curtailment less necessary later in the mid-to late summer, and reduces flood vulnerabilities. The fact that policies exist that can improve upon both of these objectives via the same policy parameters is the main reason why flood control

and carryover storage do not have a significant tradeoff in the wet test set. In summary, analysis of these two compromise policies shows how training to scenarios with high baseline regret can yield policies with improved performance on out-of-sample hydrology to balance conflicting objectives

2.6 Conclusions

This study advances the design and testing of robust control policies as an adaptation to uncertainty in environmental planning problems, contributing an experimental design to better understand the influence of the forcing scenario properties and baseline regret of training scenarios on the robustness of resulting policies. We demonstrate this approach for the northern California reservoir system to determine how transient downscaled climate scenarios impact tradeoffs between water supply, flood control, environmental flows, and hydropower generation. Results indicate that policies trained to scenario sets with high baseline regret tend to outperform those generated with other training sets in both wetter and drier futures. Additionally, the policies adapted under these conditions develop an intra-annual hedging strategy to mitigate the effects of snowpack decline under rising temperatures. The approach highlights the general importance of considering the specific properties of training scenarios in the design of robust control policies.

Beyond the pairwise comparison of train-test splits, this analysis also highlights the general difficulty of maintaining out-of-sample performance for reservoir control policies. This is driven primarily by extreme events that occur infrequently by definition and which may be the result of natural variability rather than anthropogenic change, creating a risk of overfitting to the training set. The baseline regret, based on perfect foresight optimization, provides a measure of regret to place this performance degradation in context. Unlike the traditional minimax regret strategy, where the alternative that minimizes the maximum regret across all scenarios is chosen, our approach uses a regret metric to choose training

scenarios rather than optimal alternatives. We show that optimal policies benefit from training to sets of scenarios with a high regret for the baseline solution. Our methodology also provides a way to group ensembles of scenarios using an unsupervised learning approach, along with other hydrologic properties including streamflow and snowpack, to create an experiment which maps the relationship between training and test scenarios to the outcome of policy robustness considering both the performance and diversity of solutions. The latter is particularly important given the concern with reversible adaptations to operations which can be changed over time (Herman et al.; 2020).

There is a view that the effectiveness of management decisions may be obscured in studies which focus too narrowly on annual or even monthly statistics (e.g. Persad et al.; 2020; Swain et al.; 2018). In this study policies are trained to many different hydrologic variables in projections, thus we do not necessarily consider only these annual and monthly statistics. However, there is potential to include more agreed upon shifts in hydroclimate as features in this type of analysis. In California, these may include increased extreme and heavy precipitation (e.g. Surfleet and Tullos; 2013; Berg and Hall; 2015; Huang et al.; 2018; Swain et al.; 2018) more variability in and increased wet and dry years (e.g. Persad et al.; 2020; Gershunov et al.; 2019; Polade et al.; 2017), earlier streamflow timing (e.g. Knowles et al.; 2006; Kapnick and Hall; 2010), and subsequent increased flood flows (e.g. McCabe et al.; 2007; Das et al.; 2011, 2013). Along these lines, Cohen et al. (2020) presents an example of adaptation particularly to declining snowpack and seasonal shifts in streamflow timing, and Herman and Giuliani (2018) present an example of adaptation focused on increased frequencies and variability of extreme wet and dry years.

While this study considers uncertainty in hydrology due to climate change across downscaled model projections, it could further test the robustness of the resulting policies against more realizations of sampling variability from a synthetic generator, or supplement the training set with the same. Increasing the number of scenario realizations would allow for additional hydrologic variables to be included in clustering, such as changes in flood and drought

frequencies and intra-annual streamflow shifts. Additionally, policy training might be improved with a more flexible policy structure beyond parameterizing the existing system, such as a neural network—though this may also increase the potential for overfitting due to increased degrees of freedom. Policy training can also be coupled with infrastructure design (e.g. Bertoni et al.; 2020), which in many regions will be required to cope with the more extreme projections of hydrologic change. Lastly, while our approach is demonstrated with an exampled from the water resources management management field, it can generalize to any environmental, natural resources, or infrastructure planning problem which includes a no action case, an optimization component, and a forcing scenario ensemble. Future work should explore the impacts of these additional experimental components in combination with the analysis of the training scenarios properties presented here to further improve robust policy search under uncertainty.

2.7 Appendix

2.7.1 CMIP5 modeling centers

Modeling Center (or Group)	Institute ID	Model Name
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	CSIRO-BOM	ACCESS1.0
Beijing Climate Center, China Meteorological Administration	BCC	BCC-CSM1.1 BCC-CSM1.1(m)
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
National Center for Atmospheric Research	NCAR	CCSM4
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1(BGC) CESM1(CAM5)
Euro-Mediterranean Center on Climate Change	CMCC	CMCC-CM
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-QCCCE	CSIRO-Mk3.6.0
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS,Tsinghua University	LASG-CESS	FGOALS-g2
The First Institute of Oceanography, SOA, China	FIO	FIO-ESM
NASA Global Modeling and Assimilation Office	NASA GMAO	GEOS-5
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-CM3 GFDL-ESM2G GFDL-ESM2M
NASA Goddard Institute for Space Studies	NASA GISS	GISS-E2-H-CC GISS-E2-R GISS-E2-R-CC
National Institute of Meteorological Research/Korea Meteorological Administration	NIMR/KMA	HadGEM2-AO
Met Office Hadley Centre	MOHC	HadGEM2-CC HadGEM2-ES
Institute for Numerical Mathematics	INM	INM-CM4
Institute Pierre-Simon Laplace	IPSL	IPSL-CM5A-MR IPSL-CM5B-LR
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC	MIROC-ESM MIROC-ESM-CHEM
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC	MIROC5
Max Planck Institute for Meteorology	MPI-M	MPI-ESM-MR MPI-ESM-LR
Meteorological Research Institute	MRI	MRI-CGCM3
Norwegian Climate Centre	NCC	NorESM1-M

Table 2.3: CMIP5 modeling information

2.7.2 Adaptation potential

The following figure (Figure 2.10) shows the distributions of adaptation potential and hypervolume robustness metrics for each projection in the ensemble, evaluated in 10-year increments from 2020-2099. Results in subplot (a) correspond to the distribution of adaptation potential evaluated by a perfect foresight optimization for each scenario in each 10-year increment. To develop subplot (b), perfect foresight optimizations were performed over each scenario for the full 2020-2099 time period. The resulting policies were then re-evaluated over each 10-year increment for each corresponding scenario to calculate the distribution of hypervolume robustness metrics through time. In both instances, the three greatest distributions occur in the last three decades.

Figure 2.10: (a) distribution of adaptation potential for each projection in the ensemble evaluated in 10-year increments. (b) distribution of hypervolume robustness metrics for perfect foresight policies trained from 2020-2099 and re-evaluated over the same scenario in 10-year increments.

2.7.3 Multi-objective tradeoffs

The following figures show correlations for system performance metrics in re-evaluation of policy sets. These include (1) policies trained with high-potential/low-potential wet training set S_4 and tested on low-potential wet test set S_{t2} (Figure 2.11) and (2) policies trained with high-potential/low-potential dry training set S_5 tested on low-potential dry test set S_{t3} (Figure 2.12).

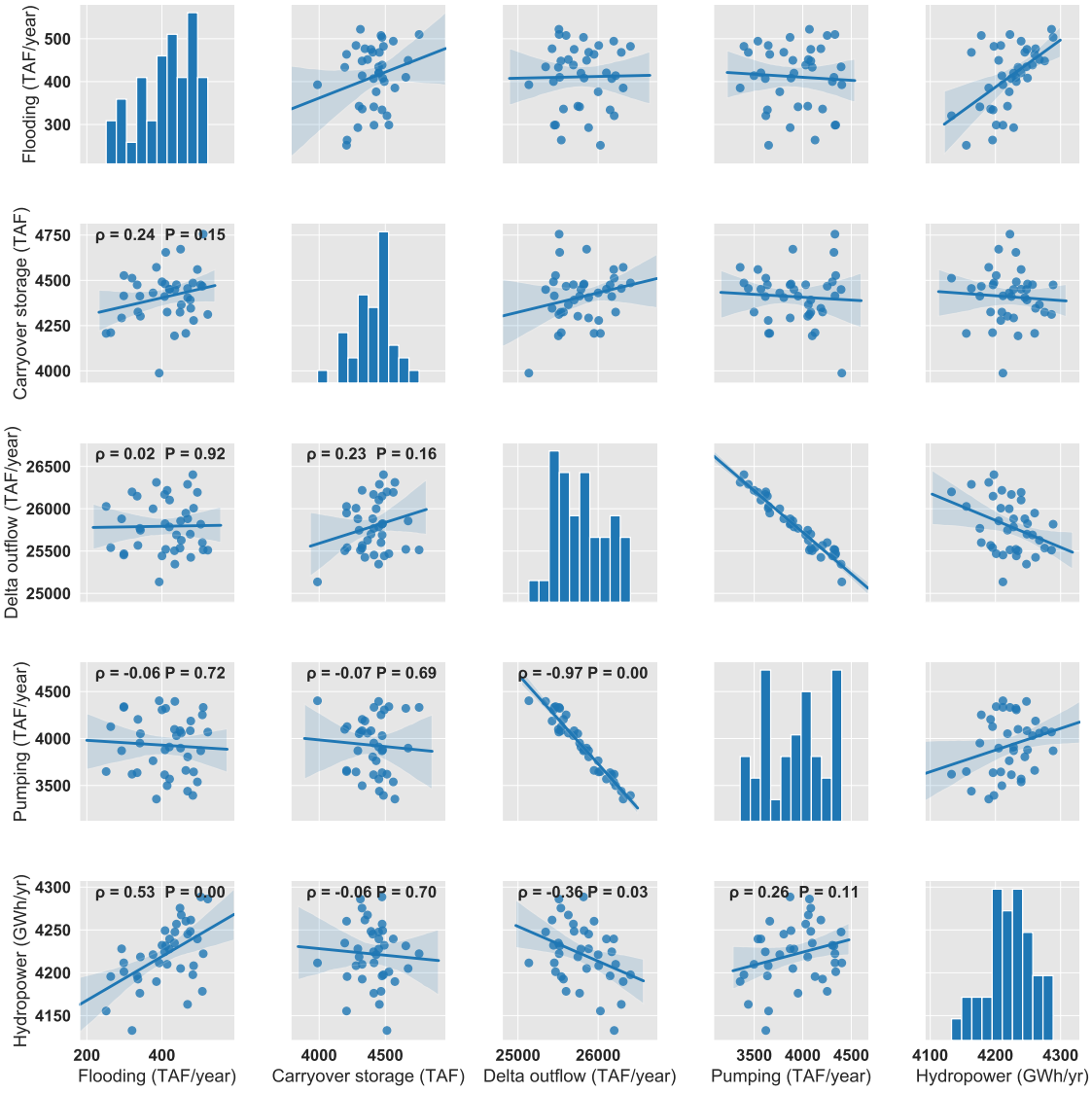

Figure 2.11: Pearson correlations for performance metrics of policies trained with high-potential/low-potential wet training set S_4 and tested on low-potential wet test set S_{t2}. These correspond to the performance metric solutions displayed in Figure 7a in the manuscript.

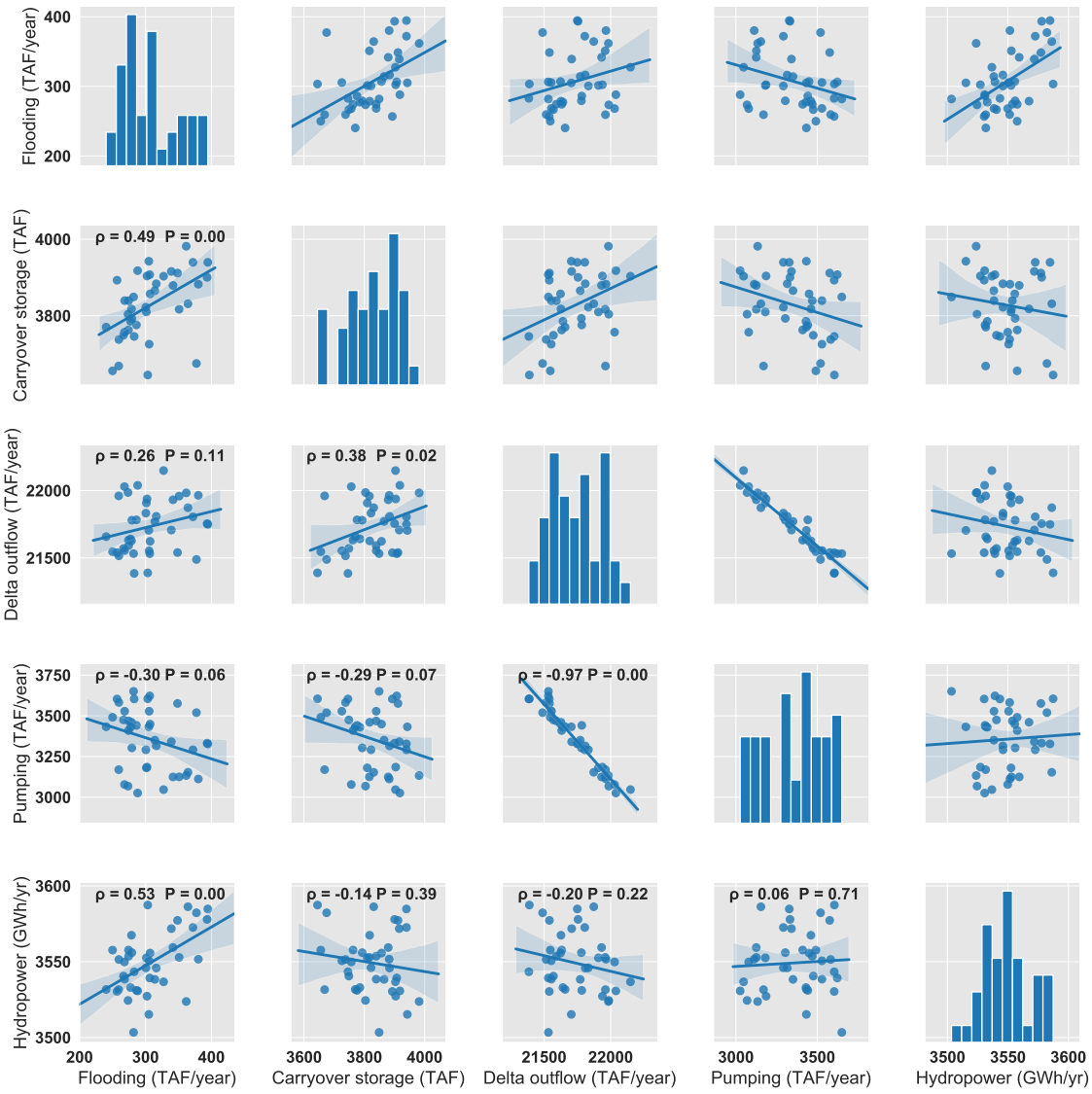

Figure 2.12: Pearson correlations for performance metrics of policies trained with high-potential/ low-potential dry training set S_5 and tested on low-potential dry test set S_{t3}. These correspond to the performance metric solutions displayed in Figure 7b in the manuscript.

2.7.4 Additional policy tables

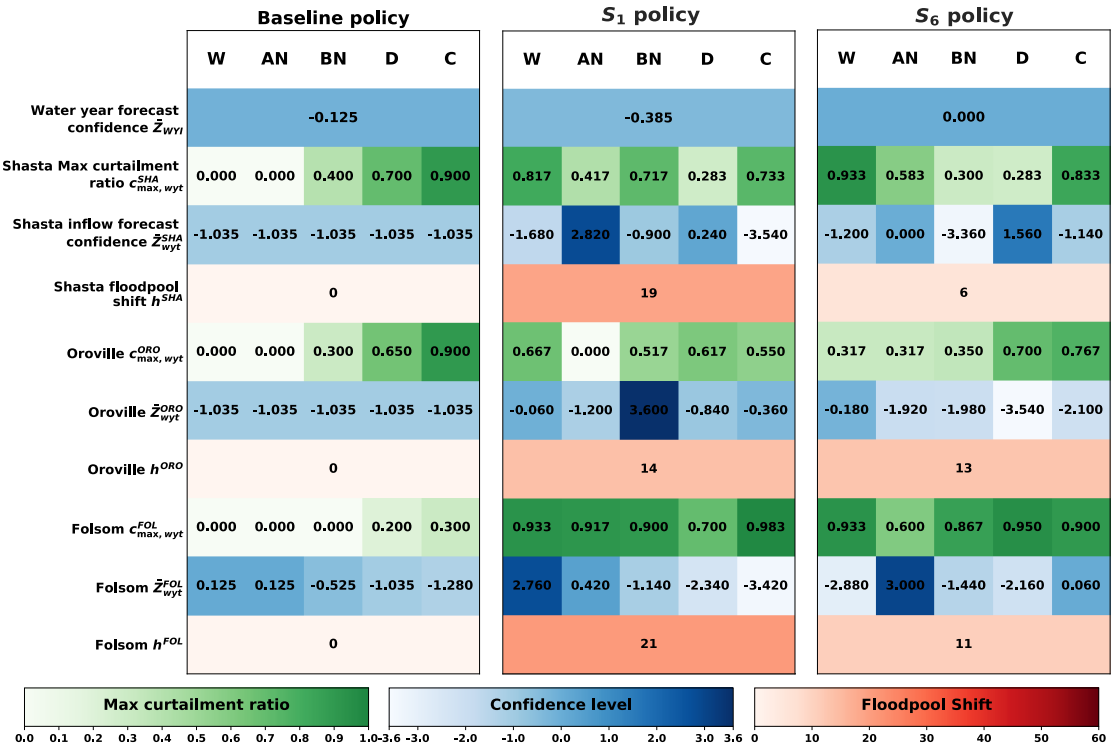

Figure 2.13: Policy tables showing decision variables for: the baseline policy, the S_1 policy (high-regret only set), and the S_6 Policy (low-regret wet/dry set). The columns denote water year type classifications associated with each decision variable, corresponding to wet, above normal, below normal, dry, and critical.

2.8 Data availability

Some or all data, models, or code generated or used during the study are available in a repository online:

ORCA:

https://github.com/jscohen4/orca/tree/cohen-2021-properties-training-scenarios

ORCA CMIP5 inputs:

URL: https://github.com/jscohen4/orca_cmip5_inputs

2.9 Acknowledgements

This work was partially supported by the U.S. National Science Foundation grant CBET-1803589 and INFEWS grant CNS-1639268. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views or policies of the NSF. We further acknowledge the World Climate Research Program's Working Group on Coupled Modeling and the climate modeling groups listed in the Appendix for producing and making available their model output.

Bibliography

Anderson, J., Chung, F., Anderson, M., Brekke, L., Easton, D., Ejeta, M., Peterson, R. and Snyder, R. (2008). Progress on incorporating climate change into management of California's water resources, *Climatic Change* **87**(1): 91–108.

Asadieh, B. and Krakauer, N. Y. (2017). Global change in streamflow extremes under climate change over the 21st century, *Hydrology and Earth System Sciences* **21**(11): 5863.

Beh, E. H., Maier, H. R. and Dandy, G. C. (2015). Adaptive, multiobjective optimal sequencing approach for urban water supply augmentation under deep uncertainty, *Water Resources Research* **51**(3): 1529–1551.

Berg, N. and Hall, A. (2015). Increased interannual precipitation extremes over california under climate change, *Journal of Climate* **28**(16): 6324–6334.

Bertoni, F., Giuliani, M. and Castelletti, A. (2020). Integrated design of dam size and operations via reinforcement learning, *Journal of Water Resources Planning and Management* **146**(4): 04020010.

Brekke, L. D., Maurer, E. P., , J. D., Dettinger, M. D., Townsley, E. S., Harrison, A. and Pruitt, T. (2009). Assessing reservoir operations risk under climate change, *Water Resources Research* **45**(4).

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

Brodeur, Z. P., Herman, J. D. and Steinschneider, S. (2020). Bootstrap aggregation and cross-validation methods to reduce overfitting in reservoir control policy search, *Water Resources Research* **56**(8): e2020WR027184.

Brown, C., Ghile, Y., Laverty, M. and Li, K. (2012). Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector, *Water Resources Research* **48**(9).

Carlsen, H., Lempert, R., Wikman-Svahn, P. and Schweizer, V. (2016). Choosing small sets of policy-relevant scenarios by combining vulnerability and diversity approaches, *Environmental Modelling & Software* **84**: 155–164.

Cayan, D. R., Kammerdiener, S. A., Dettinger, M. D., Caprio, J. M. and Peterson, D. H. (2001). Changes in the onset of spring in the western United States, *Bulletin of the American Meteorological Society* **82**(3): 399–416.

CDEC (2018). *California Data Exchange Center*, California Department of Water Resources.

Cohen, J. S., Zeff, H. B. and Herman, J. D. (2020). Adaptation of multiobjective reservoir operations to snowpack decline in the western United States, *Journal of Water Resources Planning and Management* **146**(12): 04020091.

Culley, S., Noble, S., Yates, A., Timbs, M., Westra, S., Maier, H., Giuliani, M. and Castelletti, A. (2016). A bottom-up approach to identifying the maximum operational adaptive capacity of water resource systems to a changing climate, *Water Resources Research* **52**(9): 6751–6768.

Das, T., Dettinger, M. D., Cayan, D. R. and Hidalgo, H. G. (2011). Potential increase in floods in california's sierra nevada under future climate projections, *Climatic Change* **109**(1): 71–94.

Das, T., Maurer, E. P., Pierce, D. W., Dettinger, M. D. and Cayan, D. R. (2013). Increases in flood magnitudes in california under warming climates, *Journal of Hydrology* **501**: 101–110.

Deb, K. and Jain, H. (2013). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: solving problems with box constraints, *IEEE transactions on evolutionary computation* **18**(4): 577–601.

Dessai, S. and Hulme, M. (2004). Does climate adaptation policy need probabilities?, *Climate policy* **4**(2): 107–128.

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R. A. et al. (2018). Increased human and economic losses from river flooding with anthropogenic warming, *Nature Climate Change* **8**(9): 781–786.

Draper, A. J. and Lund, J. R. (2004). Optimal hedging and carryover storage value, *Journal of Water Resources Planning and Management* **130**(1): 83–87.

Eker, S. and Kwakkel, J. H. (2018). Including robustness considerations in the search phase of many-objective robust decision making, *Environmental Modelling & Software* **105**: 201–216.

Fletcher, S., Lickley, M. and Strzepek, K. (2019). Learning about climate change uncertainty enables flexible water infrastructure planning, *Nature communications* **10**(1): 1782.

Fletcher, S. M., Miotti, M., Swaminathan, J., Klemun, M. M., Strzepek, K. and Siddiqi, A. (2017). Water supply infrastructure planning: decision-making framework to classify multiple uncertainties and evaluate flexible design, *Journal of Water Resources Planning and Management* **143**(10): 04017061.

Gershunov, A., Shulgina, T., Clemesha, R. E., Guirguis, K., Pierce, D. W., Dettinger, M. D., Lavers, D. A., Cayan, D. R., Polade, S. D., Kalansky, J. et al. (2019). Precipitation regime change in western north america: The role of atmospheric rivers, *Scientific reports* **9**(1): 1–11.

Giudici, F., Castelletti, A., Giuliani, M. and Maier, H. R. (2020). An active learning approach for identifying the smallest subset of informative scenarios for robust planning under deep uncertainty, *Environmental Modelling & Software* p. 104681.

Giuliani, M. and Castelletti, A. (2016). Is robustness really robust? how different definitions of robustness impact decision-making under climate change, *Climatic Change* **135**(3-4): 409–424.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E. and Reed, P. M. (2015). Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations, *Journal of Water Resources Planning and Management* **142**(2): 04015050.

Giuliani, M., Herman, J., Castelletti, A. and Reed, P. (2014). Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management, *Water resources research* **50**(4): 3355–3377.

Giuliani, M., Pianosi, F. and Castelletti, A. (2015). Making the most of data: an information selection and assessment framework to improve water systems operations, *Water Resources Research* **51**(11): 9073–9093.

Giuliani, M., Quinn, J. D., Herman, J. D., Castelletti, A. and Reed, P. M. (2017). Scalable multiobjective control for large-scale water resources systems under uncertainty, *IEEE Transactions on Control Systems Technology* **26**(4): 1492–1499.

Gleick, P. H. (2002). Water management: Soft water paths, *Nature* **418**(6896): 373.

Groves, D. G., Fischbach, J. R., Bloom, E., Knopman, D. and Keefe, R. (2013). *Adapting to a changing Colorado River: Making future water deliveries more reliable through robust management strategies*, RAND corporation.

Haasnoot, M., Kwakkel, J. H., Walker, W. E. and ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world, *Global environmental change* **23**(2): 485–498.

Hadka, D. (2015). Platypus-multiobjective optimization in python.

Hamarat, C., Kwakkel, J. H., Pruyt, E. and Loonen, E. T. (2014). An exploratory approach for adaptive policymaking by using multi-objective robust optimization, *Simulation Modelling Practice and Theory* **46**: 25–39.

Herman, J. D. and Giuliani, M. (2018). Policy tree optimization for threshold-based water resources management over multiple timescales, *Environmental modelling & software* **99**: 39–51.

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M. and Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty, *Water Resources Research* p. e24389.

Herman, J. D., Reed, P. M., Zeff, H. B. and Characklis, G. W. (2015). How should robustness be defined for water systems planning under change?, *Journal of Water Resources Planning and Management* **141**(10): 04015012.

Huang, X., Hall, A. D. and Berg, N. (2018). Anthropogenic warming impacts on today's Sierra Nevada snowpack and flood risk, *Geophysical Research Letters* **45**(12): 6215–6222.

Hui, R., Herman, J., Lund, J. and Madani, K. (2018). Adaptive water infrastructure planning for nonstationary hydrology, *Advances in water resources* **118**: 83–94.

Kapnick, S. and Hall, A. (2010). Observed climate–snowpack relationships in California and their implications for the future, *Journal of Climate* **23**(13): 3446–3456.

Karamouz, M., Goharian, E. and Nazif, S. (2013). Reliability assessment of the water supply systems under uncertain future extreme climate conditions, *Earth Interactions* **17**(20): 1–27.

Kasprzyk, J. R., Nataraj, S., Reed, P. M. and Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change, *Environmental Modelling & Software* **42**: 55–71.

Klos, P. Z., Link, T. E. and Abatzoglou, J. T. (2014). Extent of the rain-snow transition zone in the western US under historic and projected climate, *Geophysical Research Letters* **41**(13): 4560–4568.

Knowles, N., Cronkite-Ratcliff, C., Pierce, D. and Cayan, D. (2018). Responses of unimpaired flows, storage, and managed flows to scenarios of climate change in the San Francisco Bay-Delta Watershed, *Water Resources Research* **54**(10): 7631–7650.

Knowles, N., Dettinger, M. D. and Cayan, D. R. (2006). Trends in snowfall versus rainfall in the western United States, *Journal of Climate* **19**(18): 4545–4559.

Koutsoyiannis, D. and Economou, A. (2003). Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems, *Water Resources Research* **39**(6).

Kwakkel, J. H., Haasnoot, M. and Walker, W. E. (2015). Developing dynamic adaptive policy pathways: a computer-assisted approach for developing adaptive strategies for a deeply uncertain world, *Climatic Change* **132**(3): 373–386.

Kwakkel, J. H., Haasnoot, M. and Walker, W. E. (2016). Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty, *Environmental Modelling & Software* **86**: 168–183.

Lempert, R. J. and Collins, M. T. (2007). Managing the risk of uncertain threshold responses: comparison of robust, optimum, and precautionary approaches, *Risk Analysis: An International Journal* **27**(4): 1009–1026.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres* **99**(D7): 14415–14428.

Maier, H. R., Guillaume, J. H., van Delden, H., Riddell, G. A., Haasnoot, M. and Kwakkel, J. H. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together?, *Environmental Modelling & Software* **81**: 154–164.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other, *The annals of mathematical statistics* pp. 50–60.

Mateus, M. C. and Tullos, D. (2017). Reliability, sensitivity, and vulnerability of reservoir operations under climate change, *Journal of Water Resources Planning and Management* **143**(4): 04016085.

McCabe, G. J., Clark, M. P. and Hay, L. E. (2007). Rain-on-snow events in the western United States, *Bulletin of the American Meteorological Society* **88**(3): 319–328.

Medellín-Azuara, J., Harou, J. J., Olivares, M. A., Madani, K., Lund, J. R., Howitt, R. E., Tanaka, S. K., Jenkins, M. W. and Zhu, T. (2008). Adaptability and adaptations of California's water supply system to dry climate warming, *Climatic Change* **87**(1): 75–90.

Nayak, M. A., Herman, J. D. and Steinschneider, S. (2018). Balancing flood risk and water supply in California: Policy search integrating short-term forecast ensembles with conjunctive use, *Water Resources Research* **54**(10): 7557–7576.

Persad, G. G., Swain, D. L., Kouba, C. and Ortiz-Partida, J. P. (2020). Inter-model agreement on projected shifts in california hydroclimate characteristics critical to water management, *Climatic Change* **162**(3): 1493–1513.

Polade, S. D., Gershunov, A., Cayan, D. R., Dettinger, M. D. and Pierce, D. W. (2017). Precipitation in a warming world: Assessing projected hydro-climate changes in california and other mediterranean climate regions, *Scientific reports* **7**(1): 1–10.

Prudhomme, C., Wilby, R. L., Crooks, S., Kay, A. L. and Reynard, N. S. (2010). Scenario-neutral approach to climate change impact studies: application to flood risk, *Journal of Hydrology* **390**(3-4): 198–209.

Quinn, J. D., Reed, P. M., Giuliani, M., Castelletti, A., Oyler, J. W. and Nicholas, R. E. (2018). Exploring how changing monsoonal dynamics and human pressures challenge multireservoir management for flood protection, hydropower production, and agricultural water supply, *Water Resources Research* **54**(7): 4638–4662.

Quinn, J. D., Reed, P. M. and Keller, K. (2017). Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points, *Environmental modelling & software* **92**: 125–141.

Quinn, J., Reed, P., Giuliani, M. and Castelletti, A. (2019). What is controlling our control rules? opening the black box of multi-reservoir operating policies using time-varying sensitivity analysis, *Water Resources Research* .

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M. and Brown, C. (2020). Vulnerability and risk: climate change and water supply from California's Central Valley water system, *Climatic Change* pp. 1–23.

Reclamation (2013). Downscaled CMIP3 and CMIP5 climate and hydrology projections: Release of downscaled CMIP5 climate projections, comparison with preceding information, and summary of user needs.

Rhoades, A. M., Jones, A. D. and Ullrich, P. A. (2018). Assessing mountains as natural reservoirs with a multimetric framework, *Earth's Future* **6**(9): 1221–1241.

Russell, S. and Norvig, P. (2002). Artificial intelligence: a modern approach.

Salazar, J. Z., Reed, P. M., Herman, J. D., Giuliani, M. and Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control, *Advances in water resources* **92**: 172–185.

Savage, L. J. (1951). The theory of statistical decision, *Journal of the American Statistical association* **46**(253): 55–67.

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

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

Steinschneider, S., Ray, P., Rahat, S. H. and Kucharski, J. (2019). A weather-regime-based stochastic weather generator for climate vulnerability assessments of water systems in the Western United States, *Water Resources Research* **55**(8): 6923–6945.

Surfleet, C. G. and Tullos, D. (2013). Variability in effect of climate change on rain-on-snow peak flow events in a temperate climate, *Journal of Hydrology* **479**: 24–34.

Swain, D. L., Langenbrunner, B., Neelin, J. D. and Hall, A. (2018). Increasing precipitation volatility in twenty-first-century california, *Nature Climate Change* **8**(5): 427–433.

Trindade, B., Reed, P., Herman, J., Zeff, H. and Characklis, G. (2017). Reducing regional drought vulnerabilities and multi-city robustness conflicts using many-objective optimization under deep uncertainty, *Advances in water resources* **104**: 195–209.

Turner, S. W., Marlow, D., Ekström, M., Rhodes, B. G., Kularathna, U. and Jeffrey, P. J. (2014). Linking climate projections to performance: A yield-based decision scaling assessment of a large urban water resources system, *Water Resources Research* **50**(4): 3553–3567.

Watson, A. A. and Kasprzyk, J. R. (2017). Incorporating deeply uncertain factors into the many objective search process, *Environmental Modelling & Software* **89**: 159–171.

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

Wilby, R. L. and Dessai, S. (2010). Robust adaptation to climate change, *Weather* **65**(7): 180–185.

Zeff, H. B., Herman, J. D., Reed, P. M. and Characklis, G. W. (2016). Cooperative drought adaptation: Integrating infrastructure development, conservation, and water transfers into adaptive policy pathways, *Water Resources Research* **52**(9): 7327–7346.

Chapter 3

Dynamic adaptation of water resources systems under uncertainty using policy tree optimization[1]

3.1 Abstract

The challenge of adapting water resources systems to uncertain hydroclimatic and socioeconomic conditions warrants a dynamic planning approach. Recent studies have designed policies with structures linking infrastructure and management actions to threshold values of indicator variables observed over time. Typically, one or more of these components are held fixed while the others are optimized, constraining the flexibility of policy generation. Here we develop a framework to address this challenge based on multi-objective policy tree optimization, a heuristic search method that combines relevant indicators, actions, and thresholds in a flexible policy structure. The approach is demonstrated for a case study of northern California, where a mix of infrastructure, management, and operational adaptations are considered over time in response to an ensemble of nonstationary hydrology, water demand, and economic conditions. We first identify a subset of non-dominated policies that are robust to held-out scenarios, and then analyze their most common actions and indicators compared to the non-robust policies. Results show that the robust policies are not differentiated by

[1]This chapter will be submitted to *Water Resources Research*: Cohen, J. S. and Herman, J. D. "A policy tree optimization approach to dynamic planning under future uncertainty".

the actions they select, but show substantial differences in their indicator variables, which can be interpreted in the context of physical hydrologic trends. In particular, the statistical transformations of the selected indicator variables highlight the balance between adapting quickly versus correctly, which the non-robust policies fail to achieve. Additionally, we determine the indicators most frequently associated with each action, as well as the distribution of action timing across the scenario ensemble through the end of the century. This study presents a new and transferable problem framing for adaptation under uncertainty in which indicator variables, actions, and policy structure are identified simultaneously during the optimization.

3.2 Introduction

Adaptation to the multi-scale impacts of climate change in water resources systems is challenged by substantial uncertainty in future hydrologic projections (Wilby and Dessai; 2010; Asadieh and Krakauer; 2017; Dottori et al.; 2018), as well as land use and water demand (Clarke et al.; 2018). Under these conditions, dynamic planning provides a basis for responding to new observations as they occur, aiming to prevent both over- and under-investment (De Neufville and Scholtes; 2011; Walker et al.; 2013). The implementation of these decisions depends on a policy mapping observed indicator variables to actions, which can be optimized to determine the sequence, timing, and/or threshold values on which actions are conditioned (Herman et al.; 2020). This approach can be supported by optimal control methods such as stochastic dynamic programming (Hui et al.; 2018; Fletcher et al.; 2019) or policy search (e.g. Kwakkel et al.; 2015; Zeff et al.; 2016), potentially integrated within decision support frameworks such as Dynamic Adaptative Policy Pathways (Haasnoot et al.; 2013).

The use of optimal control methods for long-term adaptation has typically required some components of a policy to be held fixed while others are optimized. For example, several

studies have optimized a sequence of discrete actions to be implemented on defined intervals (Beh et al.; 2015; Kwakkel et al.; 2015), while others have optimized the timing and magnitude of actions directly (Beh et al.; 2014; Borgomeo et al.; 2016; Beh et al.; 2017; Borgomeo et al.; 2018). While these approaches represent uncertainty implicitly through the forcing scenarios used for optimization, policies can instead be conditioned directly on observations of system states and fluxes (indicators) to navigate uncertain futures. Several studies have used hydroclimatic variables as policy indicators (Haasnoot et al.; 2015; Kwakkel et al.; 2016; Fletcher et al.; 2019), while others have derived indicators from system states and estimates of future risk (Mortazavi-Naeini et al.; 2015; Zeff et al.; 2016; Trindade et al.; 2017; Erfani et al.; 2018; Gold et al.; 2019; Trindade et al.; 2020). In these cases, the optimization problem is to identify the triggers for adaptation, while the indicators and policy structure remain fixed. However, there is an opportunity to improve the flexibility of optimal control methods for long-term adaptation by simultaneously optimizing the set of indicators, actions, threshold values, timing, and policy structure.

A major motivation for more flexible policy optimization is the ability to identify informative indicators drawn from a combination of multi-scale observations of climate, hydrology, and demand (Kenney et al.; 2018). This could be especially important for multi-objective planning problems, where the actions needed to improve different objectives will likely be informed by different types of observations (Quinn et al.; 2019; Culley et al.; 2021). A useful indicator could represent a tipping point into a vulnerable system state (Haasnoot et al.; 2015, 2018), or a prediction of a future vulnerable state (Robinson and Herman; 2019). Indicators will ideally separate signal from noise (Hegerl and Zwiers; 2011), either by aggregating observations over a longer time window, or by focusing on changes that are projected with more certainty, such as temperature-driven impacts on sea level rise (Ceres et al.; 2017) and snowpack decline (Cohen et al.; 2020). Still, most water infrastructure investments depend on projections of extreme floods and droughts, which are subject to natural variability more than emissions scenario or model structure (Pielke Sr et al.; 2012; Lopez-Cantu et al.; 2020).

The need to identify and adapt to trends in these interrelated variables points to the value of ensembles of transient, physically-based model projections (e.g., Shortridge and Zaitchik; 2018; Taner et al.; 2019), which introduce their own uncertainties but perhaps no more so than stochastically generated alternatives (Quinn et al.; 2020; McPhail et al.; 2020).

Given the role of natural variability in transient scenarios of nonstationary climate and demand, any policy search approach for long-term adaptation must be tested for overfitting against held-out data. This is a specific instance of the broader challenge recognized by Many-Objective Robust Decision Making (Kasprzyk et al.; 2013) and related approaches, namely that actions optimized to certain scenarios may perform poorly in others. In the case of dynamic planning, a concern is that an adaptation policy could overfit to the particular sequence of flood and drought events in a training scenario, when those events are the result of natural variability rather than a transient signal (Herman et al.; 2020). This challenge is amplified when considering irreversible infrastructure investments, where a false positive decision would incur high regret (Stephens et al.; 2018; Raso et al.; 2019). Operational changes are a reversible and less costly alternative, although their performance gains are limited by infrastructure capacity (Culley et al.; 2016; Cohen et al.; 2021). Another option is planned incremental investments, in which increased marginal cost is accepted to allow more flexibility in implementation, one goal of Engineering Options Analysis (Jeuland and Whittington; 2014; de Neufville and Smet; 2019). The long-term planning problem ultimately involves joint infrastructure sizing and operation (Bertoni et al.; 2019; Geressu and Harou; 2019), as well as the selection of reliable indicator variables to reduce the risk of over- and under-adaptation.

The challenge of simultaneously optimizing indicators, actions, thresholds, timing, and policy structure for long-term adaptation can leverage ideas from recent work in short-term reservoir control. Approaches vary depending on the functional form of the policy. For instance, Optimizing policy structure for reservoir control has been investigated with policy tree optimization (Herman and Giuliani; 2018; Nayak et al.; 2018) and neuroevolution

(Zaniolo et al.; 2021). The former adds interpretability, but has less flexibility to represent nonlinear relationships. These studies have used a set of policy input variables common to reservoir control problems, typically the storage, time of year, and forecast information, extended to also include the climate state (Giuliani et al.; 2019). The concept of optimizing the choice of indicator variables within a heuristic policy search shares a similar goal to methods for input variable selection (Galelli et al.; 2014; Giuliani et al.; 2015), though does not guarantee that a globally optimal set of inputs has been found. While prior studies provide a foundation for these ideas, they have not been extended to policy design for long-term adaptation problems, which introduces new challenges in terms of the information and actions that are available, as well as the uncertainties that planners face.

3.3 Model and study area

3.3.1 Study Area

A large, complex system of water resources infrastructure has been built in California to support anthropogenic growth amid intense intra- and inter-annual variability in hydrology. Reservoirs in the foothills of the Sierra Nevada range capture winter and spring flood flows and snowmelt to be delivered for agriculture and municipal supply during summer. The State Water Project (SWP) and federal Central Valley Project (CVP) consist of a number of reservoirs and aqueducts throughout the Sacramento-San Joaquin river basin. In the Sacramento River Basin, three of the largest Sierra foothill reservoirs by volume (Shasta, Oroville, and Folsom), in parallel, contain 9 million acre-feet (11.1 km^3) of storage. These reservoirs play a major role in balancing the state's human and environmental water needs. Carryover storage in these reservoirs, measured at the end of the water year (September 30th), is a strong indicator of overall system performance and potential economic vulnerabilities (Draper and Lund; 2004) and has been found to be a reliable descriptor of system's vulnerabilities to climate change (Medellín-Azuara et al.; 2008). The terminal Delta of this

system is the site of pumped water exports, which support agriculture and municipal supply in the southern portion of the state. Environmental requirements related to the salinity of outflows from the Delta are a major constraint on these exports. Delta exports are a key metric for water supply reliability in the state and have been found to be vulnerable to climate change, due to both changes in precipitation levels and seasonal runoff timing (Anderson et al.; 2008; Ray et al.; 2020; Cohen et al.; 2020).

Adaptation to climate change in California water resources is a topic that has been studied extensively (Vicuna and Dracup; 2007). A number of studies show that projected hydroclimatic changes could have damaging affects to water supply in the region, (e.g., Lettenmaier and Sheer; 1991; VanRheenen et al.; 2004; Vicuna et al.; 2007; Brekke et al.; 2009; Ray et al.; 2020)). Those that have considered adaptations to reservoir operations (Yao and Georgakakos; 2001; Tanaka et al.; 2006; Georgakakos et al.; 2012; Cohen et al.; 2020, 2021) consider only stationary policy changes, where actions are not changed over time. There remains an opportunity to consider dynamic adaptation in the region conditioned on observed changes, combining a mix of reversible and irreversible adaptations.

3.3.2 Model

To demonstrate the new dynamic planning framework, we use the open source model Operation of Reservoirs in California (ORCA) to simulate the northern California reservoir system (Cohen et al.; 2020). The model simulates the major components of the California system north of the Delta, including the Shasta, Oroville, and Folsom Reservoirs, and Delta water supply exports via the Harvey O. Banks (SWP) and Tracy (CVP) pumping plants (Figure 3.1a,b). The pure simulation model allows for flexible adjustments to infrastructure and operating rules, as well as straightforward evaluation of alternative hydrologic scenarios.

ORCA is driven by a basic mass balance update for each of the three reservoirs. Based on timestep t, storage S_t^r in reservoir r is updated based on inflows Q_r^t, evaporative losses L_t^r, and a release R_t^r:

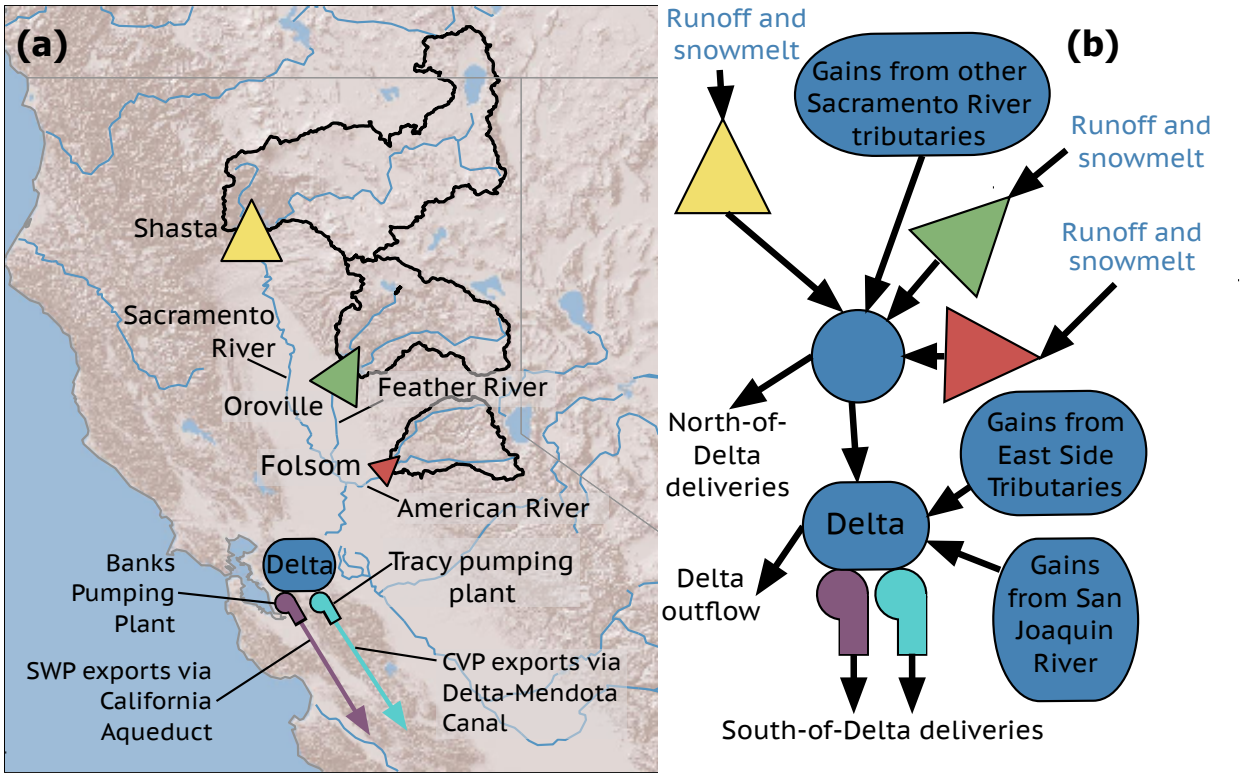

Figure 3.1: (a) Map of northern California reservoir system modeled in ORCA, adapted from Cohen et al. (2020). (b) Model schematic showing primary storage and pumping operations.

$$S_t^k = S_{t-1}^k + Q_t^k - R_t^k - L_t^k \tag{3.1}$$

The release is determined by the maximum of three operating requirements for each reservoir:

$$R_t^k = \max(R_{t,\text{environment}}^k, R_{t,\text{flood}}^k, R_{t,\text{demand}}^k) \times c_t^k \tag{3.2}$$

The first is an environmental flow constraint $u_{t,\text{environment}}^r$ that varies based on the time of year and water year type. The second is a flood control release constraint $u_{t,\text{flood}}^r$. The flood control release is determined by a flood control index based on the previous day's precipitation, and current reservoir storage. Finally, a demand release constraint consists of demands north of the Delta, south of Delta demands to be delivered by Banks and Tracy pumping plants, and a Delta outflow demand for environmental benefits and salinity control.

For further details on the baseline configuration of the model, see Cohen et al. (2020).

3.4 Methods

We introduce a new approach to the design and testing of dynamic adaptation policies under uncertainty in climate, land use, water demand, and economic factors. The method consists of four components, described in the following subsections: (1) Development of forcing scenarios from a combination of climate and land-use data; (2) Generation of adaptation policies using indicators, actions, and multi-objective optimization; (3) Testing policy robustness and tracking policy dynamics over time; and (4) Analysis of the actions and indicators most commonly used in the set of robust policies, including timing of implementation and sensitivity to cost values (Figure 3.2). These steps follow the general framework for decision making under uncertainty proposed by Herman et al. (2015) and Kwakkel and Haasnoot (2019), but extended for the specific case of dynamic policy structure optimization in several key ways. First, the choice of policy architecture is embedded in the policy generation step. Second, the policy analysis steps (3-4) focuses on identifying thresholds of indicator variables used to trigger specific actions, which extends the concept of vulnerability analysis to consider the policy response.

3.4.1 Forcing scenarios

3.4.1.1 Climate projections

For a dynamic adaptation problem, it is crucial to consider transient scenarios with physically consistent nonstationary trends. We obtain GCM-driven hydroclimate projections from downscaled CMIP5 Climate and Hydrology Projections, developed by the United States Bureau of Reclamation (USBR) (Brekke et al.; 2014). These consist of 31 GCMs run for various emissions scenarios to generate 97 scenarios of precipitation and temperature on a daily timestep through 2100 (see supplemental material for additional information). The

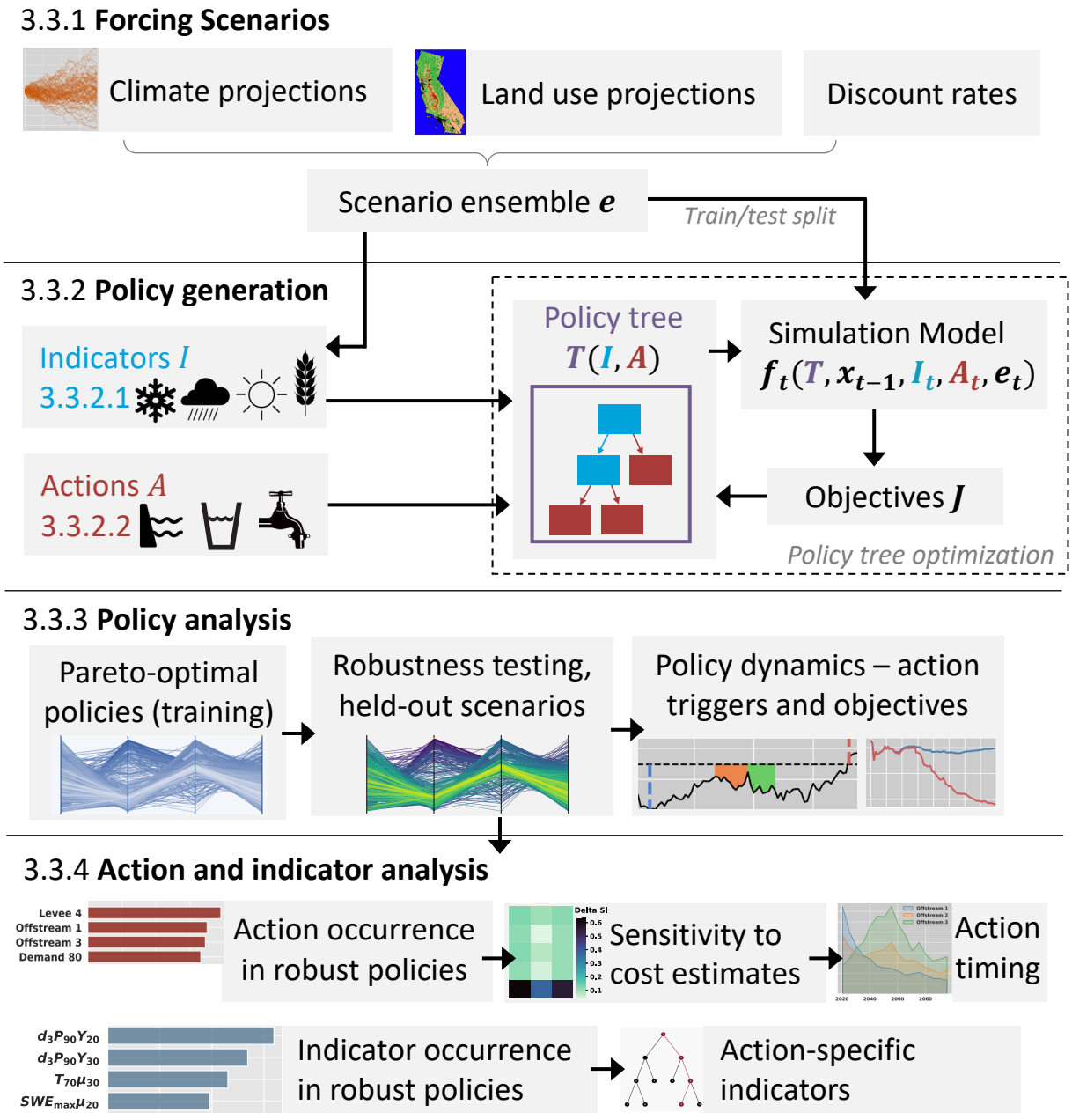

Figure 3.2: Overview of methods

projections include runoff and snow-water equivalent (SWE), obtained by routing precipitation and temperature through the Variable Infiltration Capacity (VIC) model (Liang et al.; 1994) calibrated for each basin. In cases where inputs to the reservoir simulation model, such as SWE and precipitation, were averaged across multiple stations, the relevant information was extracted from the gridded VIC output to produce basin-wide spatial averages. Physically consistent hydrologic variables including streamflow, temperature, precipitation, and snowpack are necessary to simulate the reservoir system. These hydro-climatic variables exhibit transient trends, with significant variation in the magnitude and direction of change across scenarios (Cohen et al.; 2020).

3.4.1.2 Water demand and land use projections

We use a total of 10 land use scenarios to develop projections of water demand, obtained from three existing sets of model projections. These include: three separate business-as-usual (BAU) scenarios from the USGS LUCAS model used for California's 4th Climate Assessment (Sleeter et al.; 2017), 3 scenarios from the Department of Energy's GCAM model (West and Le Page; 2014), and 4 scenarios from the USGS FORE-SCE model (Sohl et al.; 2014). From these land use projections, we estimate water demand values for 2020-2100 using crop water intensity estimates from Mall and Herman (2019), and urban water demand estimates from Christian-Smith et al. (2012). All of these scenarios show a trend toward increased water demand relative to the historical case, though with uncertainty in the magnitude of these changes.

3.4.1.3 Scenario ensemble

There are a total of 970 possible future scenarios to be created by combining each hydroclimate scenario (97) with each demand scenario (10). We split the scenarios into two sets: training and testing, consisting of 50 and 47 hydroclimatic scenarios, respectively. The water demand projections are also split evenly into training and testing sets. This yields 250

training scenarios and 235 testing scenarios. The total number of scenarios is therefore only half of the possible total, because the climate and demand projections are only combined within the training and testing sets, not across them, to ensure that the information remains separate. To each scenario we also assign a randomly sampled discount rate ranging from 2-8%.

3.4.2 Policy Generation

We extend the recently developed policy tree optimization method (Herman and Giuliani; 2018) to solve the dynamic climate adaptation problem. This method uses genetic programming to optimize the structure and thresholds of binary trees, rather than optimizing a vector of values as in traditional heuristic optimization approaches. As a result, the optimization does not need to prespecify the indicators, actions, timing, or threshold values, and can instead determine the policy structure during the optimization. Previous studies using policy tree optimization have focused on short-term reservoir control for a single reservoir with one objective function (Herman and Giuliani; 2018; Nayak et al.; 2018). Here we extend the approach to optimize long-term dynamic adaptations for a multi-reservoir system, combining operational actions with irreversible infrastructure investments. We also extend the approach to include multiple objectives.

3.4.2.1 Indicators

Indicators represent a hydroclimatic or socioeconomic variable at a certain timescale, aggregated over a moving window using a statistical transformation. In this study, we use many combinations of these options to generate indicator variables on an annual timestep, $\boldsymbol{I_t}$ (Table 3.1). Each of the variables is measured at different locations in the system (Figure 3.1). Streamflow represents the sum of unregulated flows into the Sacramento basin from its tributaries. Snow-water equivalent (SWE) uses the annual maximum value of the spatial average in the four river basins. Air temperature is measured at the three reservoir sites

and averaged. A seasonal timing indicator represents the day of the water year at which some percentage of the total annual streamflow has occurred. Demand indicators representing percent of historical demand. Certain combinations of the options in Table 1 are not considered: for example, SWE is not monitored on sub-annual timescales. Demand is only monitored the annual timescale and does not use the standard deviation or percentile statistics. Temperature is only monitored on the annual timescale, representing the average annual temperature. Given these, and several other omissions of certain possibilities of transformation combinations, a total of 540 indicator variable timeseries are created. These represent the vector $\boldsymbol{I}$ input to the optimization. The upper and lower bounds for each indicator, required for the optimization, were assigned as the 10% and 90% percentiles of the specific indicator's values at 2100, across all scenarios.

Variable	**Timescale**	**Window**	**Statistic**
Streamflow (Q)	Annual (A)	5 years	Mean (μ)
Snow-water equivalent (SWE)	Monthly (M)	10 years	Standard deviation (σ)
Air temperature (TAS)	3 Months ($3M$)	20 years	Percentile, P (10,30,50,70,90)
Seasonal timing (T)	Daily (d)	30 years	Difference, Δ (5,10,20,30,50 years)
Water demand (D)	3 Days ($3d$)	50 years	

Table 3.1: Descriptions of indicators and their notation.

The notation for each indicators containing the mean or standard deviation statistics, given a variable V, timescale t, window w, and statistic s, is $\{V_t\, s_w\}$. For variables which only consider annual timescales (SWE, TAS, D), the annual timescale does not need to be specified in the notation. In cases where a difference (Δ) is used, an additional statistic must also be applied, and Δ_w is appended to the end of the notation. When the percentile statistic is used, two variables are added: p, denoting the percentile value, and Y a placeholder for the window over which the percentile is taken. In these cases, the notation becomes $\{V_t\, s_p Y_w\}$. Lastly, when the seasonal timing T variable is used, a percentile p, rather than timescale, is the subscript for T. Thus, the notation for seasonal timing indicators becomes $\{V_p\, s_w\}$Some examples of indicators created from combinations of the above transformations and notation include:

- $Q_A\sigma_{50}$: annual streamflow, 50-year rolling standard deviation
- $Q_A\mu_{20}\Delta_5$: annual streamflow, 5-year difference in 20-year rolling average
- $Q_{3M}P_{30}Y_{20}$: 3-month streamflow, 30th percentile over 20-year moving window
- $Q_{1d}P_{90}Y_{30}$: 1-day streamflow, 90th percentile over 30-year moving window
- $Q_{3d}P_{30}Y_{50}$: 3-day streamflow, 30th percentile over 50-year moving window
- $T_{70}\mu_{30}$: seasonal timing, day of water year at which 70% of total annual streamflow has occurred, 30-year rolling average
- $T_{10}\sigma_{20}$: seasonal timing, day of water year at which 10% of total annual streamflow streamflow has occurred, 20-year rolling standard deviation
- $T_{50}\mu_{20}\Delta_5$: seasonal timing, day of water year at which 50% of total annual streamflow streamflow has occurred, 5-year difference in 20-year rolling average
- $SWE\mu_{30}$: maximum annual SWE, 30-year rolling average
- $SWE\sigma_{20}\Delta_{30}$: maximum annual SWE, 30-year difference in 20-year rolling standard deviation
- $TAS\mu_{10}$: average annual air temperature, 10-year rolling average
- $D\mu_5$: percent of historical annual demand, 5-year rolling average
- $D\mu_{20}\Delta 10$: percent of historical annual demand 10-year difference in 20-year rolling average

3.4.2.2 Actions

Six categories of actions are proposed in this study, each with multiple levels of implementation, making up a total of 22 options in vector $\boldsymbol{A}$. Actions can either be reversible or

irreversible (Table 3.2). These infrastructure expansions are informed by agency planning documents (Table 3.2), but our analysis does not explicitly consider the physical or political feasibility of implementing them. Additionally, we recognize that the costs used for these actions are uncertain, and therefore we perform a sensitivity analysis, specifically examining the sensitivity of results to cost estimates, in Section 3.2.6.2. The first of the reversible actions is demand curtailment (Demand 70,80,90), which reduces deliveries to 70-90% of the current demand, resulting in a curtailment cost C_C for each unit of shortage. These costs are estimated based on agency documents, including those from the California Public Utilities Commission (CPUC; 2016) and US Bureau of Reclamation (USBR; 2018, 2020). The curtailment is reversed at the next timestep when the policy selects a non-demand action. The second reversible action is changes to the reservoir operating policy, including two hedging rules (Hedging A,B) for the multi-reservoir system determined by Cohen et al. (2021). Once these hedging rules are triggered, they will continue until the policy selects the standard operating rule (Standard policy) action. For this study, there is no cost attributed to changing operating policies. However, we acknowledge that this is not always the case, as significant funding and resources is often be invested in developing new reservoir operating policies. The last reversible action is groundwater recharge (GW 0.1, 0.5, 1, 2, 4), which is not included in the baseline model but can be implemented when any of the five groundwater recharge actions are triggered. This action enables conjunctive use in each sub-basin when surplus flows occur, and groundwater pumping use when there are supply shortages. A cost C_{GW}, estimated from Medellín-Azuara et al. (2015), is applied for each unit of groundwater recharged or pumped. Once a groundwater action is triggered, groundwater use cannot be eliminated in the model, but can be reduced if a lower recharge action is later triggered. See supplemental materials for further details on operations of demand curtailments, operating policy changes, and groundwater recharge actions as they relate to existing system operations.

Additionally, three categories of irreversible actions are considered: reservoir capacity

Action A	**Type**	**Description**	**Cost**	**Source(s)**
Demand 70, 80, 90	Reversible	Curtail water supply demands to 70, 80, or 90% of current	$C_c = 235,000\$$ (per TAF curtailed)	CPUC (2016) USBR (2018) USBR (2020)
Hedging A,B & Standard policy	Reversible	Hedging policies A,B Standard operating policy	None	Cohen et al. (2021)
GW 0.1, 0.5, 1, 2, 4	Reversible	Expand each basin's GW pumping capacity to 0.1, 0.5, 1, 2, 4 (TAF/day)	$C_{GW} = \$22,500$ (per TAF)	Medellín-Azuara et al. (2015) Dogan et al. (2018)
Dam 1,2,3	Irreversible	Expand reservoir storage to: t4828, 5015, 5207 TAF (Shasta) 1020, 1075, 1130 TAF (Folsom)	$C_E = 50 + 3\Delta K^{0.7}$ $C_M = 0.272\Delta K$ (per year) ($ million, ΔK in TAF/day)	USBR (2015) USACE (2016)
Levee 1,2,3,4,5	Irreversible	Expand current downstream levee capacity by 10, 20, 30, 40, 50%	$C_E = 20 + \Delta K^{0.7}$ $C_M = 0.085\Delta K$ (per year) ($ million, ΔK in TAF/day)	CADWR (2017a) CADWR (2017b)
Offstream 1, 2, 3	Irreversible	Implement/expand offstream reservoir storage to 1500, 1800, 3000 TAF	$C_E = 100 + 2\Delta K^{0.7}$ $C_M = 0.5\Delta K$ (per year) ($ million, ΔK in TAF/day)	USBR (2017)

Table 3.2: Descriptions of actions and their notation, costs, and sources.

expansion (Dam 1,2,3), levee capacity expansion (Levee 1,2,3,4,5), and offstream storage implementation (Offstream 1,2,3). Each of these has the potential for incremental expansion if an action node denoting a larger capacity is later triggered. A one-time expansion cost C_E is applied each time a new expansion is triggered (Table 3.2). The cost equation for each expansion contains a base value and an expansion coefficient, ΔK, which represents the total change in infrastructure capacity, and is assigned the exponent 0.7 to represent decreasing marginal cost of expansion. After the action is initially triggered, a lag occurs before the expansion is fully implemented: we assume 5-year lags for levee expansions, and 10-year lags for dam expansion and offstream storage. Additional maintenance costs C_M for expanded infrastructure are applied each year after the action is fully implemented. Reservoir capacity expansion (Dam 1, 2, 3) involves increasing the maximum capacities of Shasta and Folsom reservoirs from 4552 and 975 TAF, respectively, to the values shown in Table 2, with the cost equations applied for each reservoir. Sizing, implementation cost, and maintenance costs for the dam actions are obtained from the Shasta Lake Water Resources Investigation Feasibility Report (USBR; 2015) and Folsom Dam Raise Draft Environmental Impact Statement/Environmental Impact Report (USACE; 2016).

Levee capacity expansions (Levee 1, 2, 3, 4, 5) occur downstream of each reservoir by a given percentage of baseline capacity, enabling both flood reduction and increased maximum allowed outflow from each reservoir. Sizing, implementation cost, and maintenance costs for the levee actions are obtained from the Central Valley Flood Protection Plan 2017 Update (CADWR; 2017a) and Flood System Long-Term Operations, Maintenance, Repair, Rehabilitation, and Replacement Cost Evaluation Technical Addendum (CADWR; 2017b). Lastly, the offstream storage action (Offstream 1, 2, 3) involves implementing a new reservoir linked to the Sacramento River downstream of Shasta reservoir. This action is based on the proposed Sites reservoir, to intake surplus flood flows in the winter, and release to mitigate potential shortages in Delta outflow and water supply deliveries. Proposed operations, sizing, construction costs, and maintenance costs for the offstream action are obtained from the

North-of-the-Delta Offstream Storage Investigation Feasibility Report (USBR; 2017). See supplemental materials for further details on operations of offstream storage in the context of overall system operations.

It is also important to note the institutional relations of these actions. For instance the offstream actions would be constructed and operated by the United States Bureau of Reclamation (USBR) and California Department of Water Resources (CADWR). The dam expansion actions would be implemented by USBR and the United Stated Army Corps of Engineers (USACE), while the levee expansions would be implemented by CADWR and USACE. Regarding reversible actions, all three of these agencies would also be involved in hedging decisions. Demand curtailments would affect USBR through the CVP and CADWR through the SWP. Additionally, these demand and delivery curtailments would be crucial for water and irrigation districts. Lastly, groundwater use would be beneficial to these districts, as well as individual water users. Based in the plethora of institutions that would be involved in these implementations, it clear that the many different actors must be involved in any large-scale climate adaptation planning process. Additionally, the policy-tree dynamic climate adaptation approach would be applicable on a smaller scale, where a single institution develops an individual adaptation plan based on a set of proposed actions.

3.4.2.3 Multi-objective optimization

In this heuristic optimal control problem, the state variables are represented by the vector of indicators $\boldsymbol{I}$, where I_s is the set of indicators specific to scenario s. Policy tree $T(\bar{A}, \bar{I})$ represents the control policy, where $\bar{I}$ and $\bar{A}$ are the subset of indicators and actions, respectively, used by the policy T. Using multi-objective policy tree optimization, we aim to solve the optimal control problem defined by Equations 3.3 - 3.9, subject to operations in the simulation model. This aims to determine the Pareto-optimal set of control policies T^* such that:

$$T^*(\bar{I}, \bar{A}) = \text{argmin}_T \boldsymbol{J}(T, \boldsymbol{I}, \boldsymbol{A}, \boldsymbol{e}) \tag{3.3}$$

subject to:

$$\boldsymbol{x}_{t+1} = f_t(T, \mathbf{x}_{t-1}, \boldsymbol{I}_t, \boldsymbol{A}_t, \boldsymbol{e}_t) \quad \forall t \tag{3.4}$$

Where:

$$\boldsymbol{J} = \begin{bmatrix} J_{\text{cost}} \\ J_{\text{reliability}} \\ J_{\text{carryover}} \\ J_{\text{flood}} \end{bmatrix} \tag{3.5}$$

$$J_{\text{cost}} = \sum_{y=1}^{Y} \left[\left(\frac{1}{1+r} \right)^{y} C(y)) \right] \tag{3.6}$$

$$J_{\text{reliability}} = \frac{\sum_{p \in P} Rel_p D_p}{\sum_{p \in P} D_p}, \qquad P = \{\text{SWP}, \text{CVP}\} \tag{3.7}$$

$$J_{\text{carryover}} = \sum_{y=1}^{Y} \left\{ \begin{array}{ll} 1, & \sum_{r=1}^{3} Cr_y^k \leq 5000 \text{ TAF} \\ 0, & \sum_{r=1}^{3} Cr_y^k > 5000 \text{ TAF} \end{array} \right\} \tag{3.8}$$

$$J_{\text{flood}} = \sum_{t=1}^{n} \sum_{r=1}^{3} \max \left(R_t^k - DQ^k, 0 \right)^2 \tag{3.9}$$

where f_t represents state transition equations with policy tree T given state x_t at timestep t, including scenario-specific state variables $\boldsymbol{x}$ and forcing variables $\boldsymbol{e}$, yielding the next state variable $\boldsymbol{x}_{t+1}$.

In Equation 3.6, the discounted cost J_{cost} is the sum of the annual $C(y)$, the overall cost of all actions implemented in a given year. These costs only reflect the actions given in Table 3.2, and do not include the cost of water shortage or flood damages, which are represented by the other objectives in this formulation.

$$C(y) = C_E(A_y) + C_M(A_{y-1}, A_{y-2}, ..., A_1) + C_c(y) + C_{GW}(y) \tag{3.10}$$

In Equation 3.7, the reliability objective $J_{\text{reliability}}$ is the weighted reliability between the State

Water Project and Central Valley Project, including both of their north-of-Delta deliveries and Delta exports. The reliability is represented by Equation 3.11, where $S_{t,p}$ is the supply or deliveries for project p at timestep t, and $D_{t,p}$ is the demand.

$$Rel_p = \frac{\sum_{t=1}^{n} \max(S_{t,p} - D_{t,p}, 0)}{\sum_{t=1}^{n} D_{t,p}} \tag{3.11}$$

In Equation 3.8, the carryover objective $J_{\text{carryover}}$ is a count of the number of years in the scenario in which the combined carryover storage Cr_y^k for the three reservoirs falls below 5000 TAF, a realistic value denoting vulnerable storage levels. For the flooding objective J_{flood} in Equation 3.9, DQ^k is the downstream levee capacity of reservoir k. Thus, this equation represents the total volume of water which exceeds this levee capacity over the time horizon.

The optimization is performed over the set of 250 training scenarios. In a function evaluation, the objective calculations are performed over model results from all training scenarios as a whole. A total of 10 random seeds are run with 50,000 function evaluations each, using the HPC1 cluster at the UC Davis College of Engineering. The algorithm parameters for the policy tree optimization include: a population size of 96, parent-selection value of 20, 0.7 crossover probability, maximum depth of 7, minimum depth of 4, and epsilon values roughly equal to 1% of the objective values. The ten random seeds are Pareto-sorted together to obtain the set of non-dominated solutions in training.

3.4.3 Policy analysis

3.4.3.1 Robustness testing

We test the policies from the optimization over the 235 held-out scenarios, and perform a Pareto sort to remove any policies that are dominated in testing. We then evaluate the robustness of the remaining policies based on their performance with respect to that of the baseline no-action policy, $\boldsymbol{J}_B$. For each testing scenario, we determine if the policy T^* outperforms the no-action case in the reliability, carryover, and flooding objectives. The cost

objective is not included in this comparison because it only reflects the costs of the actions, and therefore is zero (optimal) for the no-action case. If T^* outperforms the three remaining objectives for testing scenario s, a score of 1 is assigned for that scenario. The robustness score is then computed as the sum of these scores divided by the total number of testing scenarios, resulting in a value between 0 and 1 (Equation 3.12):

$$R(T^*) = \frac{1}{n}\sum_{s=1}^{n}\left\{\begin{array}{ll} 1, & J(T^*, s) < J_B, \quad \forall J \in (J_{\text{reliability}}, J_{\text{carryover}}, J_{\text{flood}}) \\ 0, & \text{otherwise} \end{array}\right\} \tag{3.12}$$

Among commonly used robustness metrics, this is most similar to satisficing. The satisficing metric represents the percentage of scenarios for which some performance target is satisfied (Starr; 1963; Schneller and Sphicas; 1983). In this case, we weight each scenario as equal, rather than using weighted probabilities. This ensures that the influence scenarios with more extreme trajectories remains present. This approach differs slightly in that the target performance metrics change in each scenario, as they are represented by the performance of the no-action case. We consider this a minimum standard of performance for the climate adaptation problem, that an optimized policy should outperform the no-action case in a scenario it has not seen before. A policy is considered robust if $R \geq 0.8$, and non-robust otherwise. This is a subjective choice, and a different threshold value could be chosen. The purpose of distinguishing two sets of policies is to identify any key differences in policy structure in the robust set, described in Section 3.3.4. We

3.4.3.2 Policy dynamics

We select one robust policy to illustrate the dynamics of the policy tree optimization approach by visualizing the indicators and actions taken over time in a particular testing scenario (i.e., a combination of climate and land use projections). Additionally, we analyze the dynamics of the cumulative objective function values over time to understand how the policy improves performance relative to the no-action case. This step provides details into

how a single policy works, beyond the objective values given in the Pareto front, and helps to understand why certain actions and indicators are selected from among the larger set of possible options in the search.

3.4.4 Action and indicator analysis

While the previous steps analyze one particular policy and scenario, there are many other such combinations that can provide insight into general patterns in policy structure. We therefore propose methods for action and indicator analysis using all of the nondominated policies and testing scenarios to explore the overall effectiveness and structure of policies as a group. A major goal of this step is to determine the differences between robust and non-robust policies, and more broadly to understand the dynamic responses needed to mitigate vulnerabilities to climate and land use change as they unfold.

3.4.4.1 Action occurrence in robust policies

Each policy tree in the Pareto set contains actions $\bar{A}$ selected from the full set of possible actions, $\boldsymbol{A}$, described in Table 2. We develop a metric for the occurrence of each action A based on the fraction of the total number of action nodes that it represents across a set of policy trees. This is done separately for the robust policies and the non-robust policies, for both individual actions and their more general categories. An example finding would be that the action "Levee 4" makes up 8% of all action nodes in the set of robust policies, and this could be compared to other actions as well as its occurrence in the non-robust set. This metric only depends on the structure of the policies and does not account for how frequently the actions are triggered in the test scenarios.

3.4.4.2 Sensitivity to cost estimates

The occurrence of certain actions in the set of nondominated, robust policies will depend on the cost estimates for each action (Table 3.2), which are uncertain. While these uncer-

tainties are not included in the development of the scenario ensemble, the sensitivity of the action occurrence to their costs can be determined separately. To do this, we assign five cost multipliers for the levee, dam, offstream, demand, and groundwater actions, each ranging from 0.5 to 1.5, i.e. -/+ 50% of the cost estimates in Table 2. We use the Delta moment-independent sensitivity analysis (Borgonovo; 2007) with 1000 Latin Hypercube samples of the cost multiplier parameters. For each sample, we repeat steps 3.3.3.1 and 3.3.4.1 by simulating the policies over all testing scenarios. We then perform a Pareto sort on the resulting performance metrics and identify robust policies with $R \geq 0.8$. While the policies are not re-trained, they will shift between the robust and non-robust groups depending on how they are affected by the different cost parameters. The Delta method determines the sensitivity of the action occurrence metric in the robust policy set to the five cost multipliers, identifying the influence of infrastructure cost assumptions on the optimal actions to use in the dynamic policies.

3.4.4.3 Action timing

For each possible action A, we can calculate the distribution of its implementation over the time horizon, considering the set of all robust policies in all testing scenarios. For irreversible actions, we count the year in which it is first implemented in the time horizon. We do not count cases in which an action node is triggered after a larger capacity action of the same category has already been implemented. For reversible actions, we count instances in which the action is implemented after having not been implemented in the previous year. We can then visualize the distribution of each action over the period 2020-2100 to understand which decisions are common across all policies and scenarios.

3.4.4.4 Indicator occurrence in robust policies

Each policy tree in the Pareto set contains indicators $\bar{I}$ selected from the full set of possible indicators, $\boldsymbol{I}$, described in Table 3.1. We develop a metric for the occurrence of each

indicator I based on the fraction of the total number of indicator nodes that it represents across a set of policy trees. Similar to the action occurrence metric, this is done separately for the robust policies and the non-robust policies to identify what information distinguishes the robust policies. An example finding would be that the 30-year mean annual streamflow indicator makes up 10% of all indicator nodes in the set of robust policies, which could then be compared to other variables. We additionally examine the set of indicators responsible for triggering each type of action. These are identified as the indicators present in the branch leading to each action node in a policy (i.e., ancestor nodes in the tree). For each action type, we calculate the occurrence of these action-specific indicator nodes for the robust policy set.

3.5 Results and Discussion

3.5.1 Policy analysis

3.5.1.1 Robustness testing and tradeoffs

The multi-objective optimization yields a total of 1176 policies. We re-evaluate these in the testing scenarios and perform a Pareto sort to remove 146 policies that are dominated, leaving a total of 1030 policy trees for the remainder of the study. Figure 3.3 shows a parallel axis plot describing the performance of each policy over all testing scenarios. The policy color is determined by the robustness score. The primary tradeoffs are between cost and the three other objectives, as well as between carryover storage and water supply reliability. The robust policies appear in yellow, while the non-robust policies are blue/green; notably, many policies fail to outperform the no-action case in the testing scenarios. However, the most robust policies clearly outperform the no-action case in the three non-cost objectives, and offer a range of cost options similar to the non-robust policies. This finding indicates that while investment does have some relationship with improved system performance, the robustness of this improvement is not always proportional to the total investment cost. Instead, it is

likely that the timing and threshold triggers of the actions are different between the robust and non-robust policies. This motivates a more detailed analysis of policy structure.

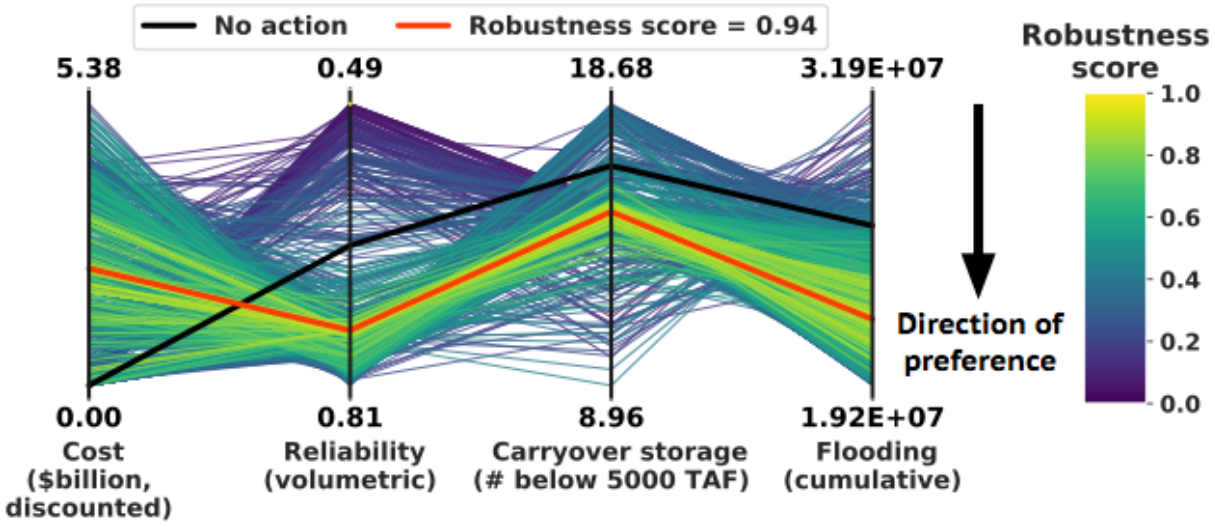

Figure 3.3: Parallel axis plot displaying the performance of the 1031 nondominated solutions in the testing scenarios. The reliability values denote the policies' total reliability over all testing scenarios. The other three objectives display the average values of the objective across the same testing scenarios.

The relationships between the objectives and robustness score can also be described with a nonlinear correlation analysis (see supplemental material). For example, we find that there is no significant correlation between cost and robustness, supporting the visual findings above. Among the non-cost performance metrics, there are few statistically significant trade-offs (i.e., negative correlation values). However, the performance objectives are positively correlated with the investment cost. In general, the objectives in Figure 3.3 are much improved relative to prior studies in which only static operational adaptations were considered (Cohen et al.; 2021). Because this optimization gives solutions where all non-cost objectives can be improved simultaneously, it gives the potential for combining dynamic infrastructure investment with operational changes.

3.5.1.2 Robust policy dynamics

To illustrate the dynamics of an example policy tree under a future scenario, we select one robust policy ($R = 0.94$) highlighted in red in Figure 3.3. This policy was chosen as a compromise across the four objectives, and is shown in Figure 3.4. The indicator variables are defined in the figure description following the notation in Table 1. The policy structure shows the conditions under which each action is triggered. However, these conditions may occur at different times, or not at all, depending on the scenario. For this reason we analyze the dynamic policy decisions in a particular scenario in Figures 3.5 - 3.6, which show the indicator variables and cumulative objectives, respectively, over the time horizon.

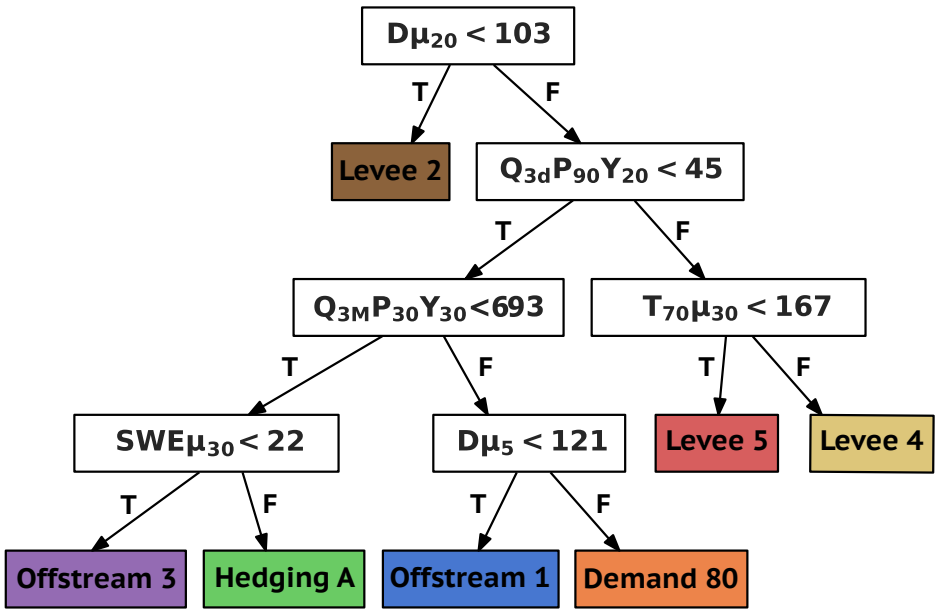

Figure 3.4: Compromise policy selected from Figure 3.3 with $R = 0.94$. $D\mu_{20}$ is the 20-year rolling average demand, percent of historical. $Q_{3d}P_{90}Y_{20}$ is the 3-month flow, 30th percentile over a 20-year moving window. $T_{70}\mu_{30}$ is the day of water year at 70/% of annual flow, 30-year rolling average. $Q_{3M}P_{30}Y_{20}$ is the 3-month flow, 30th percentile over 20-year moving window. $D\mu_5$ is the 5-year rolling average demand. Lastly, $SWE\mu_{30}$ is 30-year rolling average of maximum annual snow water equivalent.

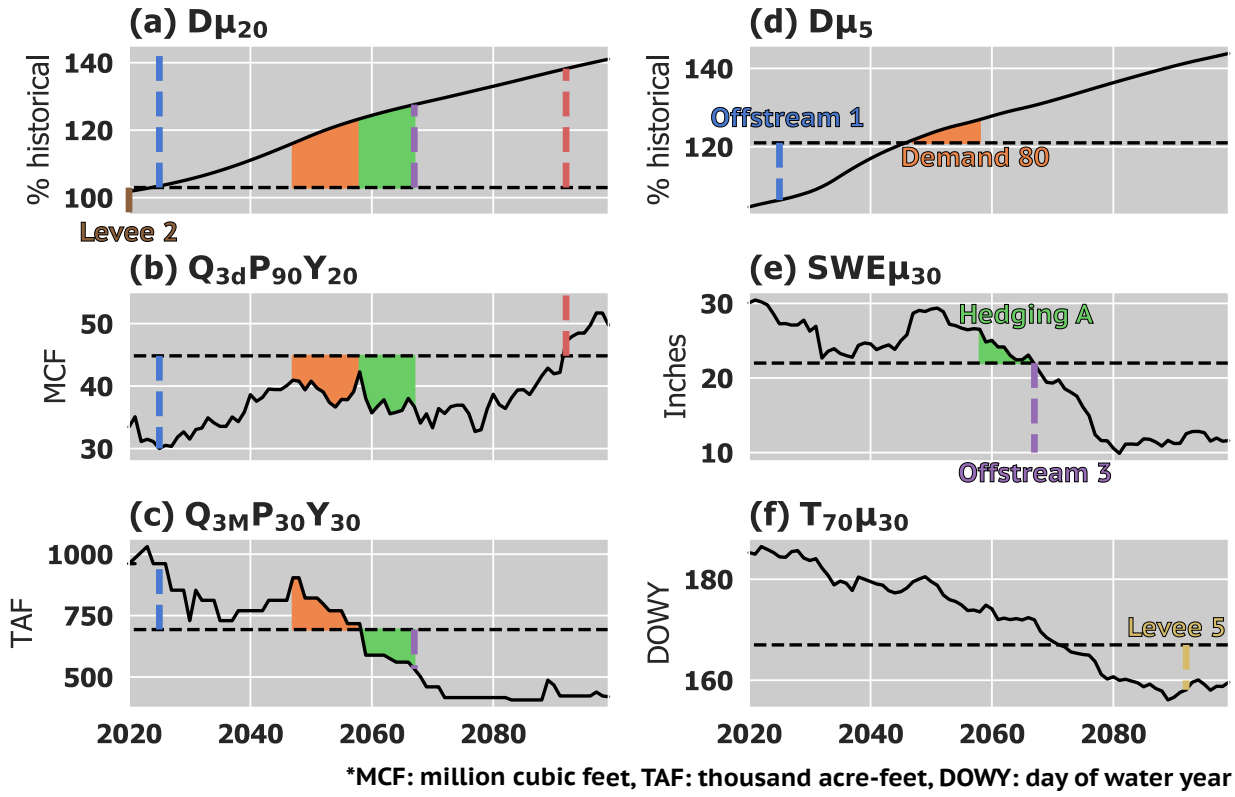

Figure 3.5: Time series of each indicator and timing of each action implementation when the robust policy tree is applied to the CESM RCP 8.5 climate projection coupled with the LUCAS BAU high land use projection. Actions are labeled in the subplots where they are implemented in part due to that indicator variable.

Figure 3.5a shows that the first action to be implemented is a moderate levee expansion (Levee 2), which is triggered at the start of the time horizon by the indicator $D\mu_{20}$, the head node of the policy in Figure 3.4. This condition will occur near the beginning of every scenario ($D = 100\%$) and does not reflect a specific benefit from levee expansion when demand is lower. Instead, it implies that a moderate expansion of the levees will improve system performance, specifically for the flood objective, in any scenario. Figure 3.6d shows the impact of this expansion on the flooding objective. After the five-year construction lag, this investment consistently reduces the cumulative flooding for a moderate cost relative to a higher capacity expansion. The process of analyzing this first action also highlights that the indicator variables may serve as a proxy for timing, which was not explicitly included as an indicator in the search but would be an apt replacement in this case.

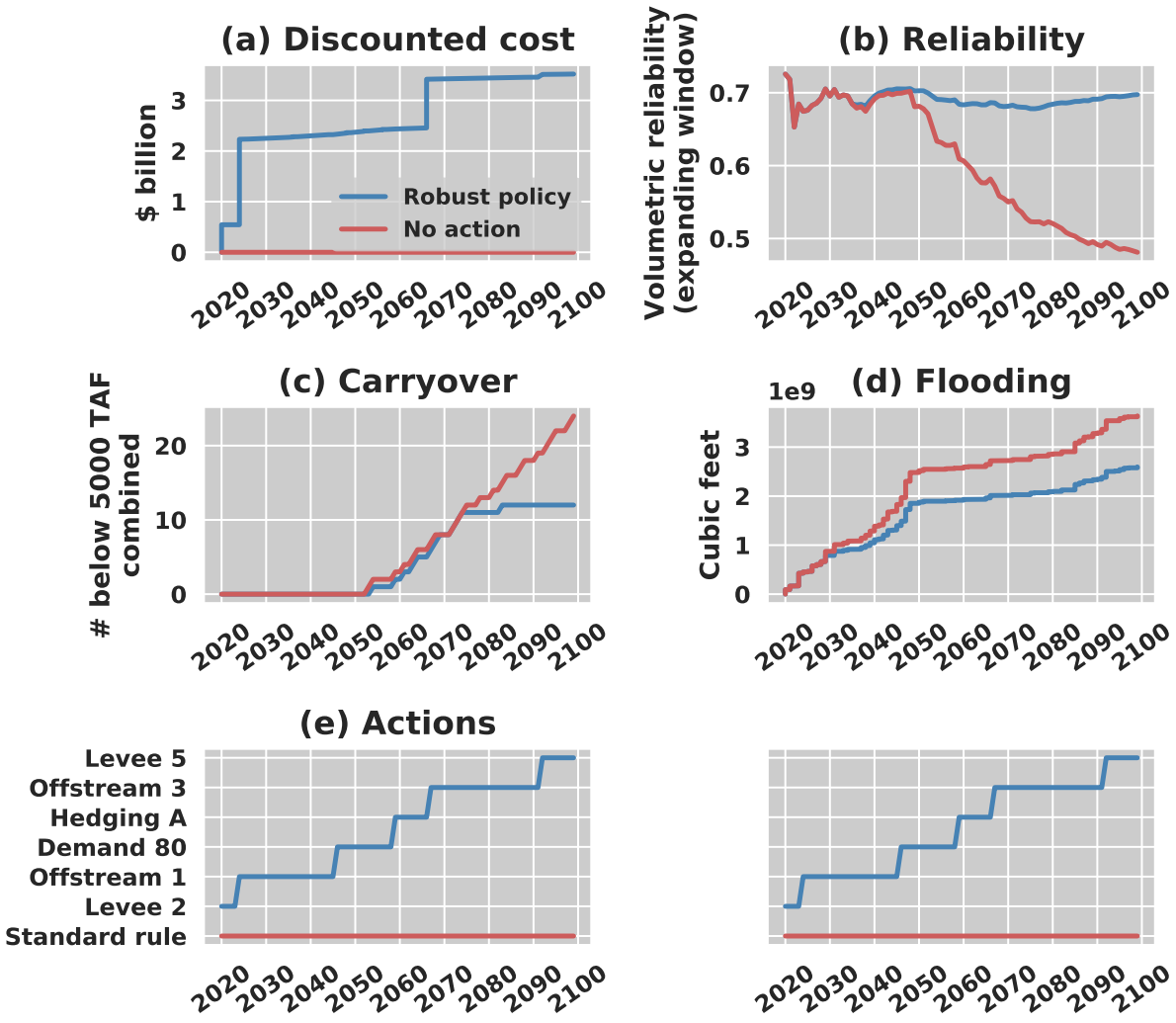

Figure 3.6: Timeseries of objective values from the robust policy tested on the CESM RCP 8.5/LUCAS BAU high scenario: (a) cumulative present value of discounted cost, (b) reliability with an expanding window starting at 2020, (c) cumulative vulnerable carryover storage years, and (d) cumulative flooding volume. (e) represents the path of actions triggered when the policy tree is simulated over the future scenario, repeated on the right to align with the timing of the objective plots.

When the demand threshold is exceeded, which happens early in almost all scenarios, the next indicator to be tracked is $Q_{3d}P_{90}Y_{20}$, which denotes high percentiles of three-day peak flows and can generally be associated with the level flood risk. The threshold value is high relative to the starting point for this scenario, so it is exceeded only much later in the century (Figure 3.5b). In the meantime, the next indicator tracked is $Q_{3M}P_{30}Y_{30}$, which represents the lower tail of the distribution of seasonal flows. A decrease would denote

lower summer flows, which have the potential to negatively impact water supply in this system (Cohen et al.; 2020). As a result, all four action nodes below this indicator aim to improve water supply, regardless of whether the specific threshold is crossed. The indicator begins the scenario well above this threshold, and crosses below it roughly mid-century. The four water supply actions are conditioned on the 5-year rolling mean demand ($D\mu_5$, Figure 3.5d) and the 30-year mean snowpack ($SWE\mu_{30}$, Figure 3.5e. These each trigger water supply actions, including reversible conservation and hedging actions (orange and green, respectively), over the following decades. The benefits are visible in the reliability and carryover storage objectives (Figure 3.6b-c).

The last action triggered for this scenario is the Levee 5, the largest expansion (note Levee 4 is not implemented in this scenario). This occurs as flood risk increases, with high percentile peak flows crossing above the 45 MCF threshold near the end of the century. The decision then depends on the flow timing indicator $T_{70}\mu_{30}$ denoting the day of the water year at which 70% of the annual flow has occurred. This value decreases consistently throughout the time horizon, an expected outcome of snowpack decline and more rain-on-snow events, which leads to heightened flood risk (e.g. McCabe et al.; 2007; Surfleet and Tullos; 2013; Huang et al.; 2018). This explains why the larger levee capacity action is triggered when the timing indicator is lower, denoting earlier streamflow and potentially more frequent and severe flood events. The effects of the second levee expansion are not significant in this analysis, as it is implemented close to the end of time horizon in 2097 after the 5-year lag. Other scenarios may reach a threshold in which the Levee 5 action is triggered earlier in the time horizon, so the action could still be beneficial in other testing and training scenarios.

We also consider the influence of each action on the discounted cost objective in Figure 3.6a, aligned with the action timing in Figure 3.6e. Note that the Levee 2 implementation at the start of the time horizon has a low cost relative to the other actions. This again highlights that it contributes to a reduction in flood volumes in all scenarios, while maintaining lower regret if a scenario turns out to have less flood risk. The costliest action is the offstream

storage expansion in 2025, as it is a large irreversible investment early in the time horizon. The demand curtailments contribute gradual cost increases during the middle of the century. The second offstream storage expansion contributes less cost than the first due to discounting. This suggests that the policies may favor implementation of larger irreversible actions only under more extreme conditions that may not occur until later in the time horizon. Lastly, implementation of Levee 5 has a low cost influence, since it is triggered late in the time horizon.

3.5.2 Action occurrence metrics

As a result of the flexible policy generation, certain actions may be used in the policy structures more often than others, suggesting that they are more effective adaptations under uncertainty. The action occurrence metrics for the sets of robust and non-robust policies are compared in Figure 3.7. For the robust set (Figure 3.7a), the most common action is Levee 4, followed by Offstream 1, 3. This is due to the clear benefits that levees give for the flooding objective, and that that offstream storage gives for water supply reliability. Changes to operating rules and demand curtailments are also common in the robust policies. On the other hand, the categories of actions that are implemented least frequently are groundwater use and reservoir expansion (Figure 3.7a,c). This suggests that these actions give little benefit in improving system performance relative to their estimated costs. In addition, the flexibility attained by the system when offstream expansion is implemented is more effective in improving upon all objectives than expanding existing reservoirs. Overall, identifying the occurrence of actions over the full robust set shows the ability of policy optimization to choose the most effective adaptations from a suite of options.

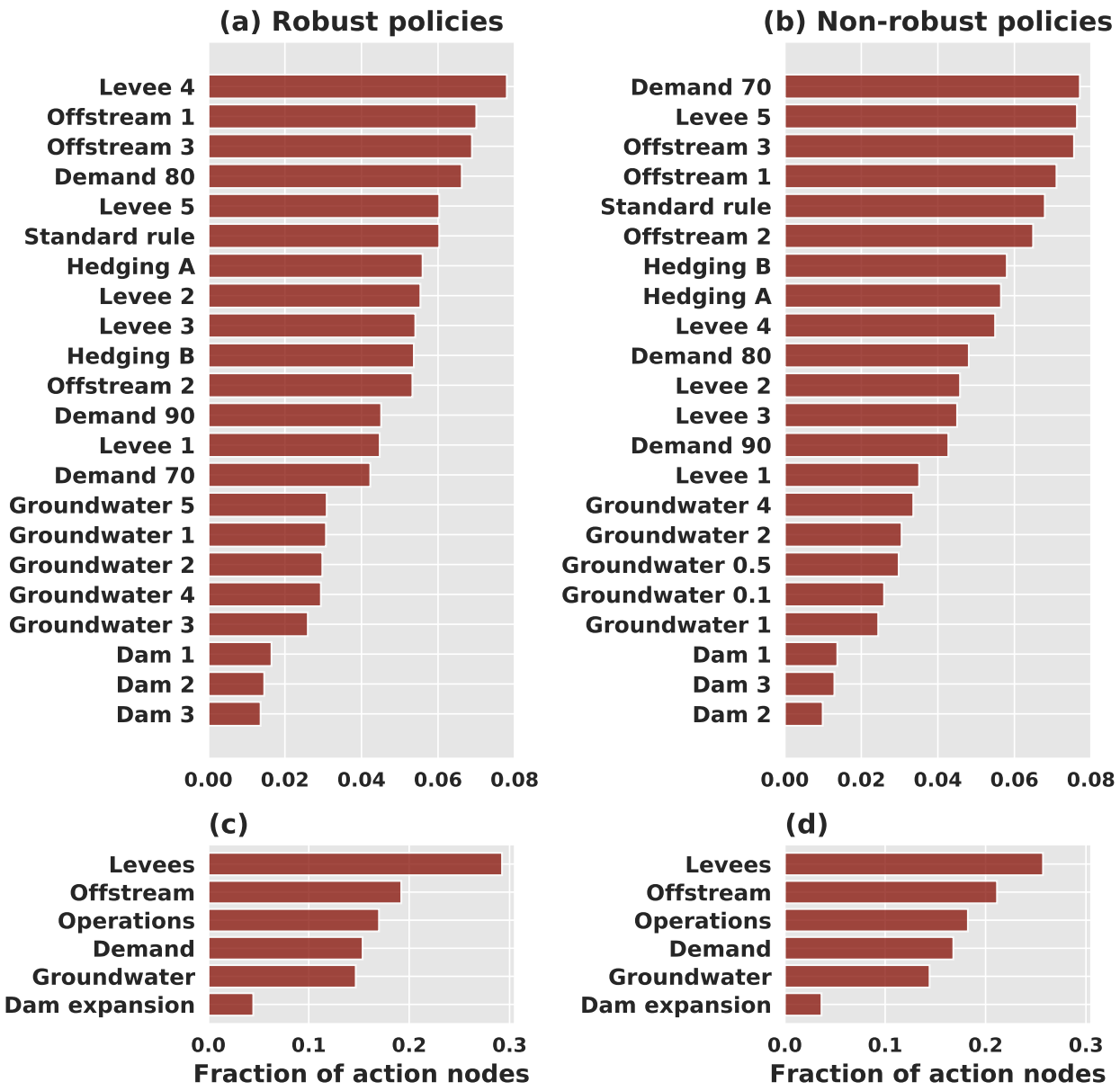

Figure 3.7: Action occurrence metrics for policies in the (a) robust set and (b) non-robust set. (c-d) Metrics grouped by action category. The occurrence metric is the fraction of all action nodes in the set of policies.

We also compare the actions taken between the robust and non-robust sets Figure 3.7b. Some of the most commonly used actions differ in the non-robust set. For instance, the most severe demand reduction (Demand 70) is the most common, followed by Levee 5, which is used notably more than in the robust set. In general, the non-robust policies rely on larger infrastructure expansions and demand reductions. This implies that effective implementation

of incremental actions, based on policy structure, may lead to less regret and more robust and flexible solutions. However, the general distribution of actions in the non-robust set is roughly the same as in the robust set, only with a different order (Figure 3.7c,d). This suggests that the policy robustness is determined more by the structure of indicators and action timing rather than the specific actions used, which underscores the importance of a dynamic rather than static approach to robustness. In other words, there is a major difference in robustness between some of these policies. If a policy is robust, it is either because the actions are always needed, and the timing is irrelevant, or because it picks up valid signals of long-term change and times the actions in a reasonable way. We find that the latter is true because the same actions are used in both the robust and non-robust policies.

3.5.3 Cost sensitivity

The action occurrence metrics in Figure 3.7 depend on the cost estimates for the actions, which are uncertain. This ancillary sensitivity analysis determines which of the cost estimates most strongly influence the occurrence of each action type in the set of nondominated policies. Results show that the occurrence of all actions is most sensitive to groundwater cost (Figure 3.8). The sensitivity of other cost multipliers is negligible by comparison. Groundwater use is among the least-used actions (Figure 3.7c). This is counter-intuitive, as several studies have shown the benefits of conjunctive use in this system, where groundwater provides inexpensive interannual storage (e.g. Harou and Lund; 2008; Medellín-Azuara et al.; 2015; Kourakos et al.; 2019). The sensitivity results imply that a lower estimate for groundwater recharge and pumping costs would encourage more frequent use. This would also change the frequencies of the other actions, which explains their similarly high sensitivity to groundwater cost. Overall, this sensitivity analysis suggests that groundwater use could be more useful relative to other adaptations, with a more detailed study of its costs. Conversely, the low sensitivity of all other cost multipliers indicates that the occurrence of those actions in the robust set is determined more by their performance in the non-cost objectives.

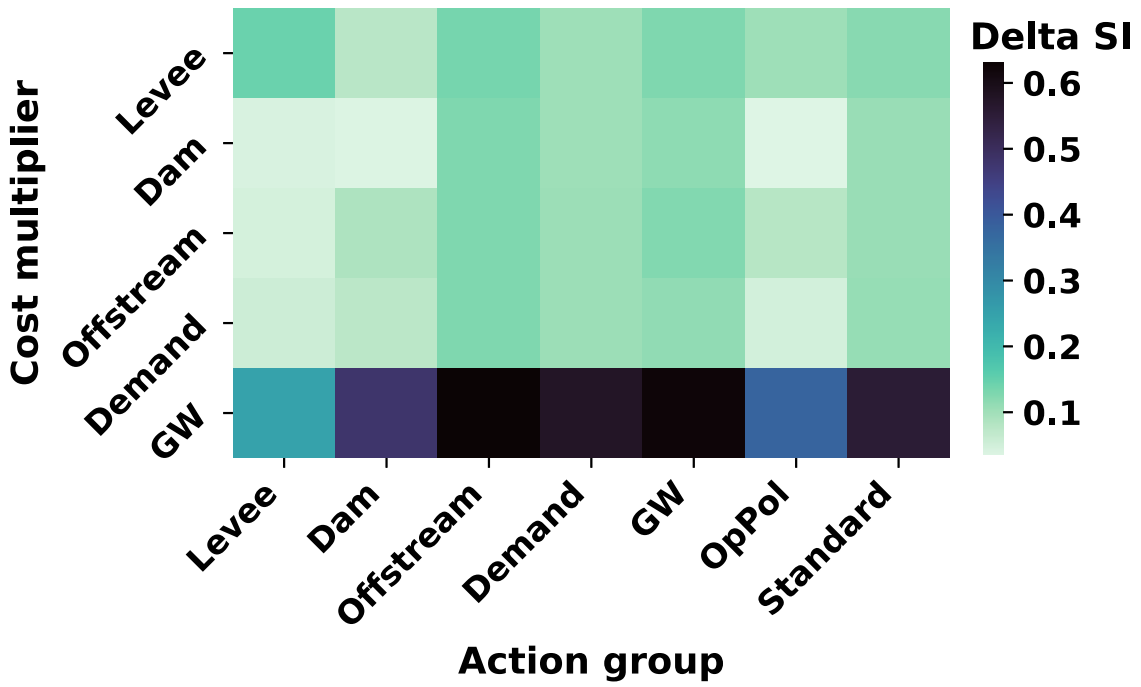

Figure 3.8: Sensitivity of action occurrence metrics in the set of robust policies (x-axis) to the cost multiplier parameters (y-axis), using the Delta moment-independent method.

3.5.4 Action timing

Figure 3.9 displays the distribution of action timing from all policies in the robust set, simulated over all testing scenarios. For the irreversible infrastructure actions (Figure 3.9 a-b), it is clear that higher capacity expansions are implemented later in the time horizon. By contrast, the reversible demand and hedging actions tend to be distributed throughout the time horizon, as they can be triggered on and off multiple times. This more general pattern confirms the specific finding in the policy dynamics example. The benefits of incremental infrastructure expansion have been widely recognized (e.g. Jeuland and Whittington; 2014; de Neufville and Smet; 2019; Fletcher et al.; 2019), but in this case the strategy was discovered by the policy search and was not a required constraint. There are four interrelated explanations for this shift in the distribution of timing for larger-capacity infrastructure actions: the ability for incremental expansion, potential for lower regret, extreme value thresholds, and discounted cost.

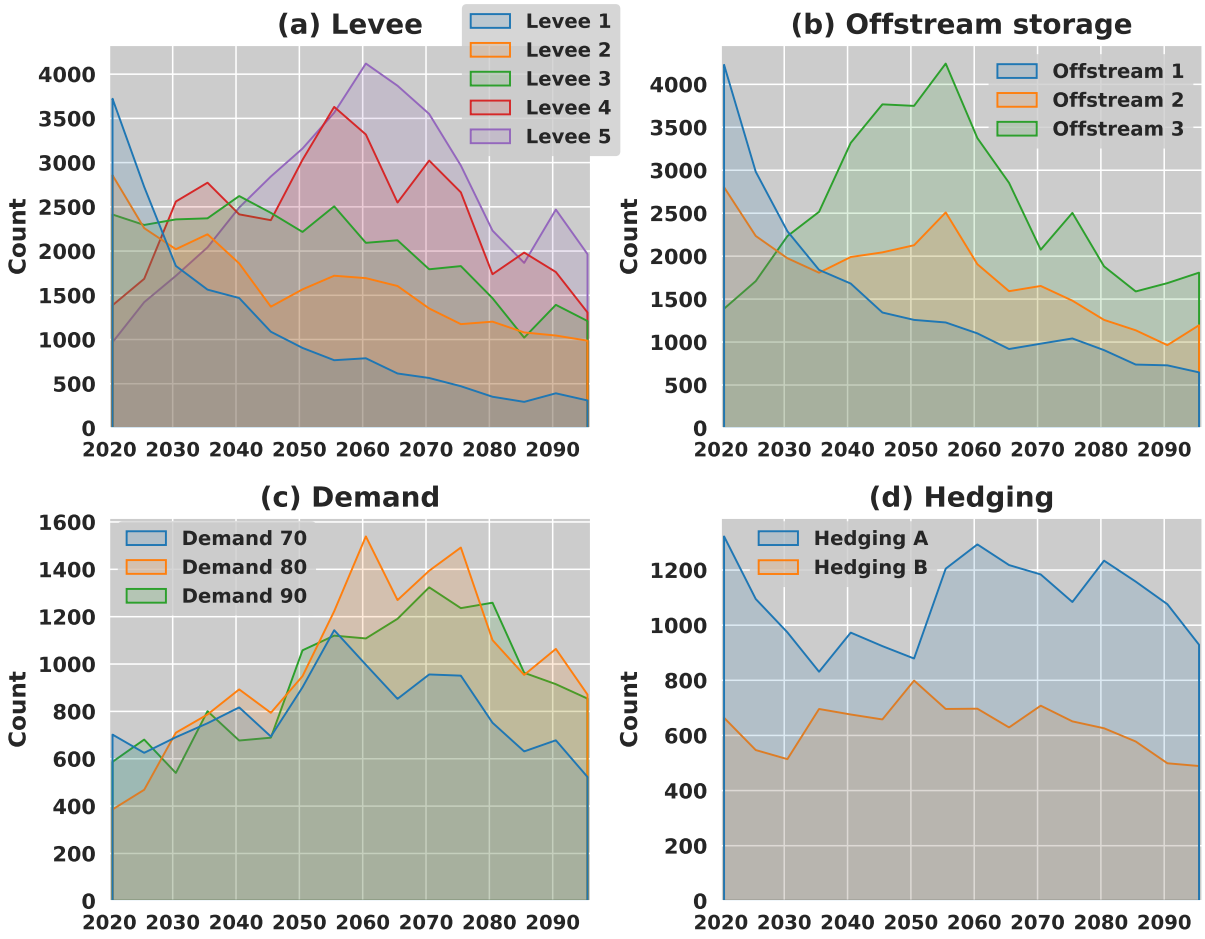

Figure 3.9: Timing of actions when each policy in the robust set (203 policies) is simulated over each of the testing scenarios (235). Therefore, in each timestep there are a total of 47705 possible occurrences of each action

Incremental infrastructure expansion allows for smaller initial investments, followed by later expansion. This behavior is highlighted by Figure 3.9a,b, where the smaller capacity infrastructure is more common earlier, while larger capacity occurs later in the time horizon. Policies ensure this by associating large capacity expansions with indicator variables using more extreme threshold values and/or longer aggregation windows, investigated in the following subsections. Additionally, implementing moderate infrastructure expansions may sufficiently improve performance in some scenarios, holding off further investment for the conditions or scenarios where they are needed, reducing the regret of over-investment. This behavior was exhibited by the sequences of levee and offstream storage investments in Figure

3.5. Finally, discounting also incentivizes this behavior, as a greater present value will be attributed to infrastructure built earlier in the time horizon.

3.5.5 Indicator occurrence metrics

Just as with action nodes, we compare the occurrence of indicators across all policies in the robust and non-robust sets. Figure 3.10 displays only the most common 16 indicators in each set, out of the total possible 540 indicators. These follow the notation defined in Table 1. In the robust policies, the most common indicator is $Q_A\mu_{20}\Delta_5$, the 5-year difference in the 20-year mean streamflow. This indicator represents very recent changes in the overall average streamflow, which could reflect both prolonged droughts and wet periods and therefore inform the implementation of various actions serving multiple objectives. Other common indicators in the robust set include high percentiles of three-day peak streamflow ($Q_{3d}P_{90}Y_{20}$, $Q_{3d}P_{90}Y_{30}$) and low percentiles of three-month streamflow ($Q_{3M}P_{30}Y_{20}$, $Q_{3M}P_{30}Y_{30}$) (Figure 3.10a). As shown in the policy dynamics analysis, these indicators represent changes in extreme events under climate change: increases in flood magnitude, and decreases in spring-summer flow volumes, respectively. However, the percentile values (30, 90) are not as extreme as they could be—other percentiles such as the 5th and 95th were omitted by the algorithm. This suggests that the 30th and 90th percentiles are effective transformations to reliably track signals for adaption, while the more extreme values likely contain more noise due to natural variability. These indicators also identify effective timescales for monitoring: 3-day for floods, and 3-month for low flows. These both differ greatly from the timescales in the non-robust set, where several 1-day and 1-month flow indicators are present in the policies that overfit to the training scenarios (Figure 3.10b).

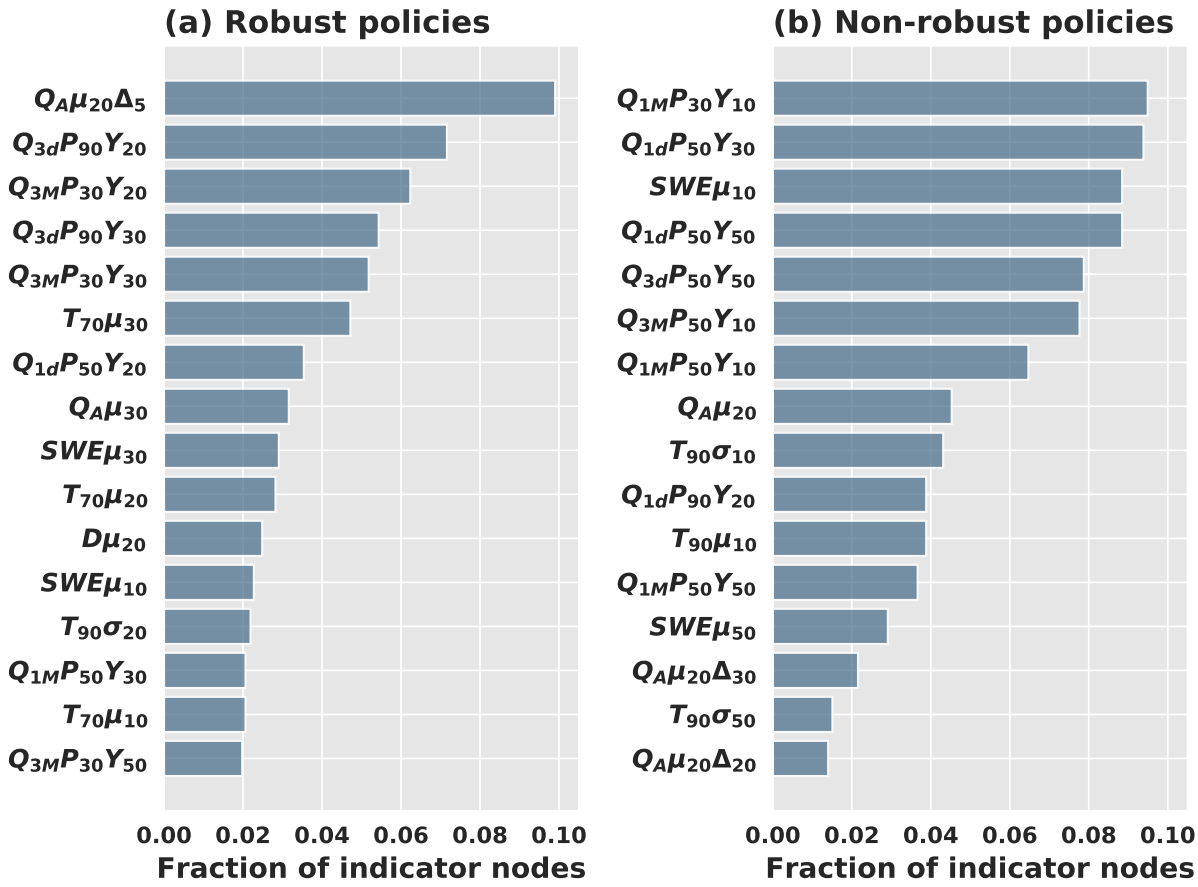

Figure 3.10: Indicator occurrence metrics (fractions) for all policy trees in the (a) robust set and (b) non-robust set. Indicator notation is defined in Table 3.1

Other common indicators in the robust policies are rolling averages of seasonal flow timing; both $T_{70}\mu_{30}$ and $T_{70}\mu_{10}$ denote the day of the water year at which 70% of annual flow has occurred. This is expected to move earlier in the year as snowpack decreases with rising temperatures, making a reliable signal for adaptations to mitigate the associated vulnerabilities in water supply and flood risk. The most notable indicator not present in Figure 3.10 is the air temperature, TAS. Most likely, the best predicted hydrologic impacts of rising temperature appear in the form of maximum annual snow water equivalent and flow timing, rather than temperature itself. The trends in these hydrologic variables are likely more effective in determining the conditions in which adaptations are necessary. Additionally, the presence of demand-related indicators such as $D\mu_{20}$ denotes that increasing demands (in

this case, due to land use change) are also a reliable signal to track for action implementation.

An important distinction between the robust and non-robust policies is the aggregation windows selected for the indicator variables. Among the most common indicators in the robust set, all but four are based on either 20-year or 30-year rolling windows, chosen from options ranging from 5 to 50 years. By contrast, he non-robust set contains many indicators with either 10-year or 50-year aggregation windows (Figure 3.10b). While these may provide optimal performance on the training scenarios, they do not generalize to the testing scenarios because they will either over-adapt to noise, in the case of the 10-year window, or wait too long to adapt, in the case of the 50-year window. The robust policies identify a compromise between adapting quickly versus adapting correctly. In general, this approach to optimizing policy structure allows the indicators to be selected during the search rather than constrained a priori—but the most useful indicators can only be identified by testing policies on held-out scenarios.

3.5.6 Action-specific indicators

The occurrence of certain indicators in the robust policies is informative, but is a few steps removed from explaining the performance of the policies. Therefore, we next examine which indicators are present in the "path" (i.e., ancestor nodes in the policy tree) to trigger certain actions. This provides detail about relationships between policy structure, indicators, and actions. Figure 3.11 shows the four most common indicators in the robust policy set that lead to the four most common actions. First, the action Levee 4 is primarily triggered by indicators $Q_{3d}P_{90}Y_{20}$ and $Q_{3d}P_{90}Y_{30}$. These high percentile peak flow indicators are an intuitive trigger for levee expansion. The remaining indicators are flow timing ($T_{70}\mu_{30}$) and snowpack ($SWE\mu_{30}$ conditions. This is also justified, as earlier timing and snowpack decline will lead to more severe flooding.

For the action Offstream 3 (Figure 3.11b), the most common indicator is a lower percentile of 3-month flows, $Q_{3M}P_{30}Y_{20}$, which can indicate drought conditions or lower summer

flows. This leads to water supply vulnerabilities, which the offstream storage actions help to mitigate. As the snowpack and flow timing indicators ($SWE\mu_{30}$ and $T_{90}\mu_{50}$) decrease, it indicates a shift in the precipitation phase from snow to rain. This shift is projected to impact water supply, so they also denote effective indicators for conditions to trigger offstream storage expansion. Additionally, the 50th percentile peak flow indicator, $Q_{1d}P_{50}Y_{20}$ likely represents general increases or decreases in the total flow volume, given that it is a median rather than extreme percentile.

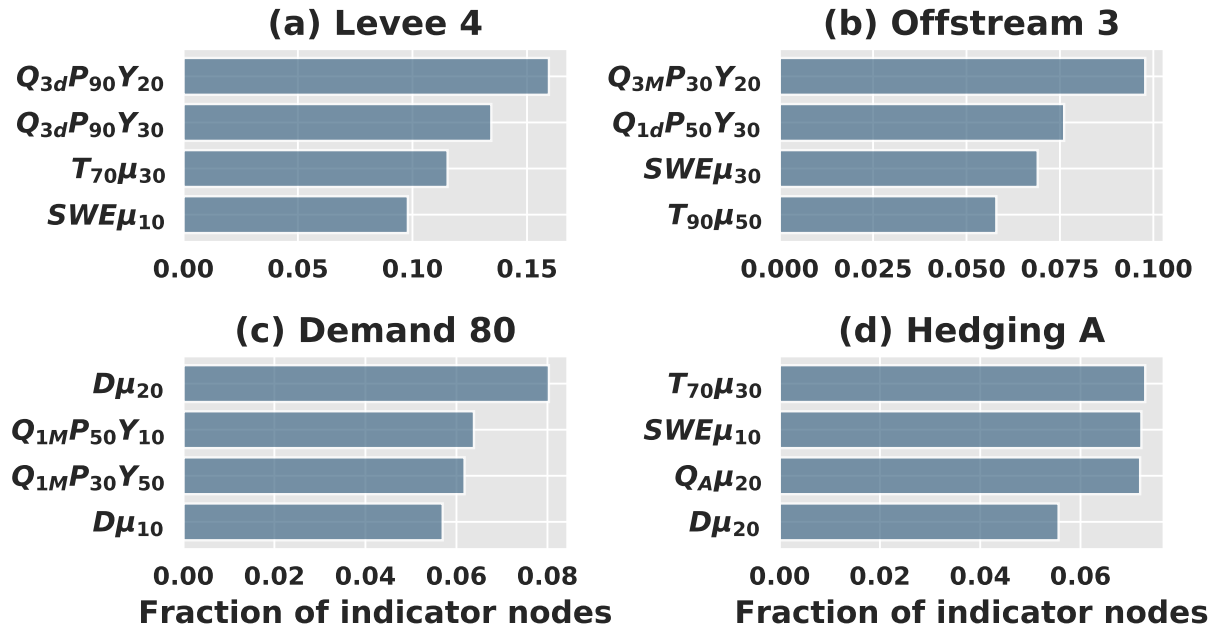

Figure 3.11: Fraction of indicator nodes in the robust set which are ancestor nodes to the actions (a) Levee 4, (b) Offstream 3, (c) Demand 80, and (d) Hedging A. The indicator notation is defined in Table 3.1

For the action Demand 80 (Figure 3.11c), the most common indicators include the 20-year rolling demand $D\mu_{20}$, and also $D\mu_{10}$. This suggests that as water demand increases throughout the century due to land use change, it may be necessary to activate conservation measures to ensure high reliability and carryover storage. Additional indicators for this action include transformations of monthly streamflow, $Q_{1M}P_{50}Y_{10}$ and $Q_{1M}P_{30}Y_{50}$, which likely represents the onset of drought conditions under which conservation measures would

be required. These indicators are helpful for the demand action based on its reversibility. They can detect drought conditions in which it should be triggered, while also being able to turn off the demand action when conditions become more stable

Finally, we consider the most common indicators for the Hedging A action (Figure 3.11d), which was designed in prior work based on static adaptations to the impacts of snowpack decline (Cohen et al.; 2020). Here, these impacts can be tracked dynamically by the two most common indicators associated with this action, the flow timing $T_{70}\mu_{30}$ and snowpack $SWE\mu_{10}$. Additionally, the action is able to increase the intra-annual flexibility of reservoir releases, which would be required in dry futures or under intense demand growth. This justifies the next two most common indicators, the mean streamflow $Q_A\mu_{20}$ and demand $D\mu_{20}$. In general, for a multi-objective problem with many adaptation options, each candidate action will be best informed by different indicators, which can be discovered during the policy search.

3.5.7 Limitations and future work

Several improvements can be made to this dynamic adaptation approach based on policy tree optimization. First, a probabilistic adaptation method could be developed in which the indicators are derived through Bayesian updating to reduce uncertainty in scenarios over time (Hui et al.; 2018; Fletcher et al.; 2019). Second, operating policy adaptations could be implemented with a nested optimization, where the policy triggers a sub-optimization based on recent hydro-climatic conditions to generate new operating policies. These improvements could be supported by synthetic scenarios to increase sample size and create policies robust to different characterizations of uncertainty, instead of different realizations of the same uncertainty characterization as in this study. However, this would pose a challenge in generating transient forcing scenarios and indicators that are physically consistent. One approach would be to make land use an endogenous response to climate and other changes, or use combinations of exogenous *and* endogenous land use scenarios, rather than solely exogenous

scenarios (Giuliani et al.; 2016; Ekblad and Herman; 2020).

Further advances are needed in the approach to analyzing the frequency with which indicator variables are triggered. The results of this study contain many indicators with a strong physical justification, but also some that are artifacts of the heuristic optimization, or that are only a proxy for timing. This challenge could be reduced by removing redundant information from the set of indicators, through a principal components analysis or related methods, similar to recent work in reservoir control (Giuliani et al.; 2019), though this would sacrifice some interpretability in the process. Additionally, the indicators and actions selected by the policies should be more strongly linked to the performance of the system in different types of scenarios, perhaps by developing an importance metric in which the components are removed from the policy one-at-a-time and its performance is re-evaluated. The approach could be coupled with an uncertainty decomposition to understand the relative influence of climate models, emissions scenarios, and natural variability in the indicator variables, following (Hawkins and Sutton; 2009; Lehner et al.; 2020). This study provides a first step to make this connection by separating the indicators and actions that occur in the robust policies, a discrete representation of system performance under uncertainty that ensures that the policies are adapting to a signal of long-term change and are not overly influenced by natural variability.

3.6 Conclusions

This study advances dynamic adaptation under uncertainty by introducing a multi-objective policy tree optimization approach. In this problem formulation, the policy structure, indicators, actions, and thresholds are identified simultaneously during the search, enabling flexible policy generation with an interpretable structure. The method is demonstrated for a case study of northern California, where a mix of infrastructure and operational adaptations are considered over time in response to an ensemble of nonstationary hydrology,

water demand, and economic conditions. Several key findings emerge. First, in contrast to a static planning problem, the cost of adaptation does not fully determine the robustness of a policy. Instead, robustness depends on the dynamic sequence of actions and the information used to trigger them. Second, the robust policies are not differentiated by the actions they select, but rather by substantial differences in their indicator variables. The indicators can be interpreted in the context of physically-based trends, and the robust policies identify statistical transformations of indicator variables that balance the risks of over- and under-investment. For example, robust policies rely on aggregation windows of 20-30 years, while non-robust policies include indicators with 10 and 50-year windows, suggesting that these policies adapt either too quickly or too slowly in a way that does not generalize to the testing scenarios.

3.7 Appendix

3.7.1 Multi-objective tradeoffs

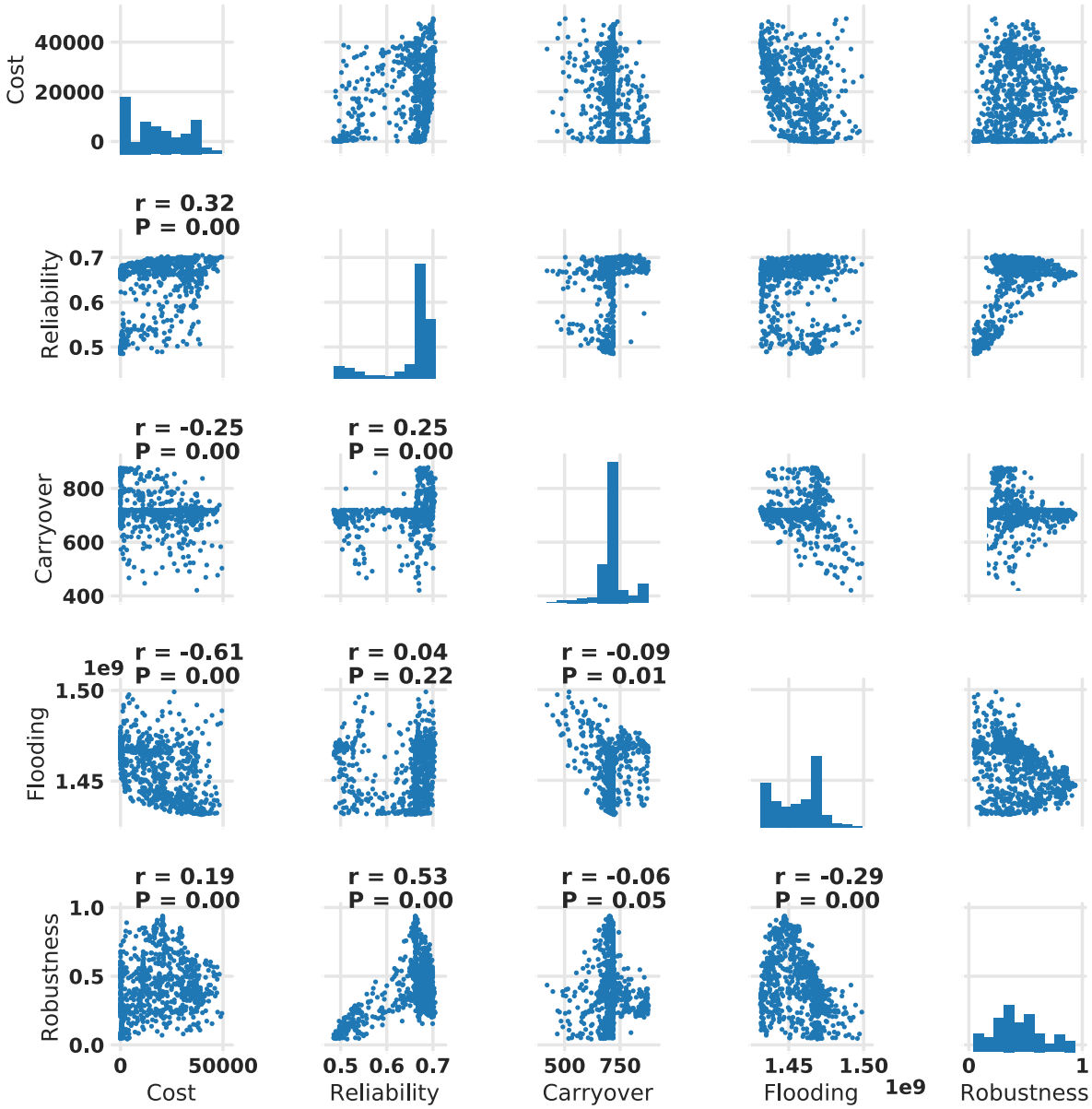

Figure 3.12: Pairplot displaying pearson-r coefficients and P values between objectives from the parallel axis plot in Figure 3.3

As discussed in Section 3.5.1.2, relationships between the objectives and robustness score in Figure 3.3 can be described with a nonlinear correlation analysis.As noted, there is no

significant correlation between cost and robustness. This highlights the fact the that the policy robustness is determined more by the structure of indicators and action timing rather than the specific actions used, underscoring the importance of a dynamic rather than static approach to robustness (Section 3.5.2). Among the non-cost performance metrics, there do not exist many statistically significant tradeoffs (i.e., negative correlation values). However, the performance objectives are positively correlated with the investment cost.

3.7.2 Action details

In this section, each of the six action categories are described in further detail.

3.7.2.1 Demand curtailments

For each demand curtailment action we reduce the demand values, regardless of WYT, to either 70, 80, or 90% of that in the baseline simulations. A multiplier is applied to the north-of-Delta (NOD^k) and south-of-Delta (SOD^k) demands for each reservoir k. These demands vary by month and WYT (see Cohen et al. (2020)). A total of $235,000 is added to the cost objective for each TAF curtailed. These simplified costs are estimated based on agency documents, including those from the California Public Utilities Commission (CPUC; 2016) and US Bureau of Reclamation (USBR; 2018, 2020). In future work, it could be advantageous to use cost estimate that are non-linear, or vary by water year type.

3.7.2.2 Hedging A,B & standard policy

Hedging policies A and B are equivalent to policies A in B in Cohen et al. (2021), respectively. The standard operating policy is equivalent to that modeled in Cohen et al. (2020).

3.7.2.3 Groundwater use

Groundwater pumping and recharge operations are modeled based on Nayak et al. (2018). Pumping costs and groundwater storage capacities are based on those in the CALVIN model, specifically those obtained from Medellín-Azuara et al. (2015) and the repository from Dogan et al. (2018). We assign groundwater storage capacities of 10, 8, 6 MAF (million acre-feet) downstream of Shasta on the Sacramento River, Oroville on the Feather River, and Folsom on the American River, respectively. Pumping costs are assigned as 22,500$ per TAF.

In for each timestep t, a groundwater flow target is $QT^t_{gw_t}$ is calculated for each reservoir's downstream basin k. The sign of $QT^t_{gw_t}$ denotes whether groundwater recharhge (positive sign) or pumpung (negative sign) occurs at timestep t. The sign and value for the groundwater target is determined based on R^k_t the release from reservoir k and the environmental and demand release constraints: $u^k_{t,env}$ and $u^k_{t,demand}$, respectively. If the release is greater than the larger of the two constraints, this denotes that flood releases are occurring, and artificial groundwater recharge should occur. The target groundwater recharge level $QT^t_{gw_t}$ is determined to be either the difference between the reservoir release and the greater of the two constraints, or the maximum groundwater recharge capacity if the difference between the release and demand exceed the recharge capacity. This is represented by the first piece-wise portion in Equation 3.13. The groundwater recharge or pumping capacity $QT^t_{gw_t}$ is equal to that specified by the groundwater action has been initiated for the time period represented. In this case, the quantity of diversions for groundwater recharge is guaranteed to be no larger that the combined north-of-Delta demands, south-of-Delta demands, and Delta outflow demands for the corresponding reservoir.

For the groundwater pumping case, the reservoir release R^k_t will not satisfy the the demand constraint, indicating that a north-of-Delta release curtailment has. We assume that groundwater in these basins is not used to replenish environmental flows, Delta outflow, or south-of-Delta demands. In this case, we note that groundwater pumping target is a negative value. It is equal to the maximum of $-QT^t_{gw_t}$ or the negative north-of-Delta supply

shortage, denoted by the negative shortage (see second piece-wise of Equation 3.13). The shortage is defined as the difference between the north of Delta demand, NOD_t^k the north-of-Delta diversions $Q_{NOD,t}^k$.

$$QT_{gw_t}^k = \left\{ \begin{array}{ll} \min\left[QT_{gw_t\,\max}^k, R_t^k - \max(u_{t,env}^k, u_{t,demand}^k)\right], & \text{if } R_t^k \geq \max(u_{t,env}^k, u_{t,demand}^k) \\ \max\left[-QT_{gw_t}, -(NOD_t^k - Q_{NOD,t}^k)\right], & \text{if } R_t^k < u_{t,demand}^k \end{array} \right\} \tag{3.13}$$

In the case that $QT_{gw_t}^k$ is positive, groundwater recharge occurs. The groundwater recharge value at time t is denoted by $Qin_{gw_t}^k$. If the target groundwater recharge is greater than the difference between the maximum groundwater reservoir capacity $S_{gw_{\max}}^k$ and groundwater storage $S_{gw_t}^k$, than the recharge at the timestep is equal to this difference (Equation 3.14). Otherwise the rechage inflow remains the same:

$$\text{if } QT_{gw}^k \geq 0: \quad Qin_{gw_t}^k = \left\{ \begin{array}{ll} QT_{gw_t}^k, & \text{if } S_{gw_{\max}}^k - S_{gw_t}^k \leq QT_{gw_t}^k \\ S_{gw_{\max}}^k - S_{gw_t}^k, & \text{if } QT_{gw_t}^k S_{gw_{\max}}^k - S_{gw_t}^k > QT_{gw_t}^k \end{array} \right\} \tag{3.14}$$

The groundwater storage is then updated based on the storage at the previous timestep and groundwater inflow value:

$$S_{gw_t}^k = S_{gw_{t-1}}^k + Qin_{gw_t} \tag{3.15}$$

In the case that $QT_{gw_t}^k$ is positive, groundwater pumping occurs (t.e. water is withdrawn from the aquifer). This is constrained to ensure that the groundwater pumped from the aquifer is no more than the current storage. Therefore, the groundwater outflow value $Qout_{gw_t}^k$ is determined by Equation 3.16:

$$\text{if } -QT_{gw}^k < 0: \quad Qout_{gw_t}^k = \left\{ \begin{array}{ll} QT_{gw_t}^k, & \text{if } -QT_{gw_t}^k - S_{gw_t}^k \geq 0 \\ -QT_{gw_t}^k - S_{gw_t}^k, & \text{if } -QT_{gw_t}^k - S_{gw_t}^k < 0 \end{array} \right\} \tag{3.16}$$

The groundwater storage is then updated based on the storage at the previous timestep and

groundwater outflow value:

$$S^k_{gw_t} = S^k_{gw_{t-1}} - Qout_{gw_t} \tag{3.17}$$

3.7.2.4 Dam expansion

Dam expansion capacities, expansion costs, and maintenance costs for Shasta are based off the Shasta Lake Water Resources Investigation Feasibility Report (USBR; 2015). For Folsom these values are obtained from Folsom Dam Raise Draft Environmental Impact Statement/Environmental Impact Report (USACE; 2016). The dam expansion in the model leads to increased floodpool and maximum capacity values for these reservoirs. There are no other explicit operational changes when the expansions occur.

3.7.2.5 Levee expansion

Levee expansions increase the threshold DQ^k, equivalent to the downstream levee capacity, used in the flooding objective (Equation 3.9). In addition, they increase the maximum allowed reservoir outflow that is allowed by increasing the allowed emergency spillage, which would otherwise exceed the downstream levee capacity. Levee expansion and maintenance costs are obtained from CADWR (2017a) and CADWR (2017b).

3.7.2.6 Offstream storage

Proposed operations, sizing, construction costs, and maintenance costs for the offstream action are obtained from the North-of-the-Delta Offstream Storage Investigation Feasibility Report (USBR; 2017). Here, we model a simplification of the proposed offstream reservoir operations from this feasibility report. There are three conditions that must be satisfied for intakes to the offstream reservoir to occur. The first is that the flow on the Sacramento River at Red Bluff must be greater than 11000 cfs. Additionally, for each month from October through May, there must be a one week period pulse protection period for which the the 3-day rolling average of flow of the Sacramento at Bend Bridge is between 15000 and 25000

CFS, and during which no diversions to the offstream reservoir occur. This pulse protection period minimizes the entrainment and impingement of juvenile salmonoids, because the first storm event is crucial for their migration (**?**). Lastly, intakes can only occur in October through May. If the month falls in October through May, the timestep occurs when there is no pulse protection, and flow at Red Bluff is greater than 11000, the intake $Q^o_{in,t}$ to the offstream reservoir is defined as:

$$Q^o_{in,t} = \max(\min(Q^o_{in,t}, S^o_{\max} - S^o_t), Q_{tributaries,t}) \tag{3.18}$$

Where $Q^o_{in,t}$ is the maximum allowable intake, $S^o_{\max}$ is the maximum capacity for the offstream reservoir, S^o_t is the storage in the offstream reservoir at timestep t, and $Q_{tributaries,t}$ is the total flow of all tributaries to the Sacramento River upstream of its confluence with the Feather River and downstream of Shasta Dam. $Q^o_{in,t}$, the maximum allowable intake, varies based on the size of the offstream reservoir. For Offstream 1, it is 8600 cfs, for Offstream 2- 10600 cfs, and for Offstream 3- 15000 cfs.

In the months June through September, if there is a daily inflow deficit d_t at the Delta greater than 15 TAF, an outflow from the offstream reservoir,Q^o_{out} occurs to help lower the deficit (Equation 3.19). It is subject to the Delta inflow deficit d_t, maxmimum outflow $Q_{out,\max}$ (equal to the maximum inflow), and current offstream storage S^o_t at timestep t.

$$Q^o_{\text{out}} = \min(d_t, Q_{out,\max}, S^o_t) \tag{3.19}$$

3.7.3 CMIP5 modeling centers

Modeling Center (or Group)	Institute ID	Model Name
Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	CSIRO-BOM	ACCESS1.0
Beijing Climate Center, China Meteorological Administration	BCC	BCC-CSM1.1 BCC-CSM1.1(m)
Canadian Centre for Climate Modelling and Analysis	CCCMA	CanESM2
National Center for Atmospheric Research	NCAR	CCSM4
Community Earth System Model Contributors	NSF-DOE-NCAR	CESM1(BGC) CESM1(CAM5)
Euro-Mediterranean Center on Climate Change	CMCC	CMCC-CM
Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-QCCCE	CSIRO-Mk3.6.0
LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and CESS,Tsinghua University	LASG-CESS	FGOALS-g2
The First Institute of Oceanography, SOA, China	FIO	FIO-ESM
NASA Global Modeling and Assimilation Office	NASA GMAO	GEOS-5
NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	GFDL-CM3 GFDL-ESM2G GFDL-ESM2M
NASA Goddard Institute for Space Studies	NASA GISS	GISS-E2-H-CC GISS-E2-R GISS-E2-R-CC
National Institute of Meteorological Research/Korea Meteorological Administration	NIMR/KMA	HadGEM2-AO
Met Office Hadley Centre	MOHC	HadGEM2-CC HadGEM2-ES
Institute for Numerical Mathematics	INM	INM-CM4
Institute Pierre-Simon Laplace	IPSL	IPSL-CM5A-MR IPSL-CM5B-LR
Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies	MIROC	MIROC-ESM MIROC-ESM-CHEM
Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	MIROC	MIROC5
Max Planck Institute for Meteorology	MPI-M	MPI-ESM-MR MPI-ESM-LR
Meteorological Research Institute	MRI	MRI-CGCM3
Norwegian Climate Centre	NCC	NorESM1-M

Table 3.3: CMIP5 modeling information

3.8 Data availability

Some or all data, models, or code generated or used during the study are available in a repository online:

ORCA:

https://github.com/jscohen4/orca

ORCA CMIP5 inputs:

URL: https://github.com/jscohen4/orca_cmip5_inputs

3.9 Acknowledgements

This work was partially supported by the U.S. National Science Foundation grants CBET-2041826 and CNS-1639268. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views or policies of the NSF. We further acknowledge the World Climate Research Program's Working Group on Coupled Modeling and the climate modeling groups listed in the supplement of this paper for producing and making available their model output.

Bibliography

Anderson, J., Chung, F., Anderson, M., Brekke, L., Easton, D., Ejeta, M., Peterson, R. and Snyder, R. (2008). Progress on incorporating climate change into management of California's water resources, *Climatic Change* **87**(1): 91–108.

Asadieh, B. and Krakauer, N. Y. (2017). Global change in streamflow extremes under climate change over the 21st century, *Hydrology and Earth System Sciences* **21**(11): 5863.

Beh, E. H., Dandy, G. C., Maier, H. R. and Paton, F. L. (2014). Optimal sequencing of water supply options at the regional scale incorporating alternative water supply sources and multiple objectives, *Environmental Modelling & Software* **53**: 137–153.

Beh, E. H., Maier, H. R. and Dandy, G. C. (2015). Adaptive, multiobjective optimal sequencing approach for urban water supply augmentation under deep uncertainty, *Water Resources Research* **51**(3): 1529–1551.

Beh, E. H., Zheng, F., Dandy, G. C., Maier, H. R. and Kapelan, Z. (2017). Robust optimization of water infrastructure planning under deep uncertainty using metamodels, *Environmental modelling & software* **93**: 92–105.

Bertoni, F., Castelletti, A., Giuliani, M. and Reed, P. (2019). Discovering dependencies, trade-offs, and robustness in joint dam design and operation: An ex-post assessment of the Kariba Dam, *Earth's Future* **7**(12): 1367–1390.

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

Borgomeo, E., Mortazavi-Naeini, M., Hall, J. W., O'Sullivan, M. J. and Watson, T. (2016). Trading-off tolerable risk with climate change adaptation costs in water supply systems, *Water Resources Research* **52**(2): 622–643.

Borgonovo, E. (2007). A new uncertainty importance measure, *Reliability Engineering & System Safety* **92**(6): 771–784.

Brekke, L. D., Maurer, E. P., , J. D., Dettinger, M. D., Townsley, E. S., Harrison, A. and Pruitt, T. (2009). Assessing reservoir operations risk under climate change, *Water Resources Research* **45**(4).

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

CADWR (2017a). Central Valley flood protection plan 2017 update, https://cawaterlibrary.net/wp-content/uploads/2017/10/2017CVFPPUpdate-Final-20170828.pdf.

CADWR (2017b). Flood system long-term operations, maintenance, repair, rehabilitation, and replacement cost evaluation technical addendum.

Ceres, R. L., Forest, C. E. and Keller, K. (2017). Understanding the detectability of potential changes to the 100-year peak storm surge, *Climatic Change* **145**(1): 221–235.

Christian-Smith, J., Matthew, H. and Allen, L. (2012). Urban water demand in California to 2100: Incorporating climate change.

Clarke, L., Bennett, N., Hejazi, M. I., Horing, J., Janetos, A., Lee, K., Lewis, K., Mach, K. J., Mastrandrea, M., Nichols, L. et al. (2018). National climate assessment iv, Chapter 17: Sector interactions, multiple stressors, and complex systems, *U.S. Global Change Research Program, Washington, DC, USA*, p. 638–668.

Cohen, J. S., Zeff, H. B. and Herman, J. D. (2020). Adaptation of multiobjective reservoir operations to snowpack decline in the western United States, *Journal of Water Resources Planning and Management* **146**(12): 04020091.

Cohen, J. S., Zeff, H. B. and Herman, J. D. (2021). How do the properties of training scenarios influence the robustness of reservoir operating policies to climate uncertainty?, *Environmental Modelling & Software* p. 105047.

CPUC (2016). What will be the cost of future sources of water for California?, https://www.cpuc.ca.gov/uploadedFiles/CPUC_Public_Website/Content/About_Us/Organization/Divisions/Policy_and_Planning/PPD_Work/PPD_Work_Products_(2014_forward)/PPD%20-%20Production%20costs%20for%20new%20water.pdf.

Culley, S., Maier, H., Westra, S. and Bennett, B. (2021). Identifying critical climate conditions for use in scenario-neutral climate impact assessments, *Environmental Modelling & Software* **136**: 104948.

Culley, S., Noble, S., Yates, A., Timbs, M., Westra, S., Maier, H., Giuliani, M. and Castelletti, A. (2016). A bottom-up approach to identifying the maximum operational adaptive capacity of water resource systems to a changing climate, *Water Resources Research* **52**(9): 6751–6768.

De Neufville, R. and Scholtes, S. (2011). *Flexibility in engineering design*, MIT Press.

de Neufville, R. and Smet, K. (2019). Engineering Options Analysis (EOA), *Decision Making under Deep Uncertainty*, Springer, pp. 117–132.

Dogan, M. S., Fefer, M. A., Herman, J. D., Hart, Q. J., Merz, J. R., Medellín-Azuara, J. and Lund, J. R. (2018). An open-source python implementation of california's hydroeconomic optimization model, *Environmental Modelling & Software* **108**: 8–13.

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A., Mongelli, I., Frieler, K., Betts, R. A. et al. (2018). Increased human and economic losses from river flooding with anthropogenic warming, *Nature Climate Change* **8**(9): 781–786.

Draper, A. J. and Lund, J. R. (2004). Optimal hedging and carryover storage value, *Journal of Water Resources Planning and Management* **130**(1): 83–87.

Ekblad, L. and Herman, J. D. (2020). Toward data-driven generation and evaluation of model structure for integrated representations of human behavior in water resources systems, *Water Resources Research* p. e2020WR028148.

Erfani, T., Pachos, K. and Harou, J. J. (2018). Real-options water supply planning: Multi-stage scenario trees for adaptive and flexible capacity expansion under probabilistic climate change uncertainty, *Water Resources Research* **54**(7): 5069–5087.

Fletcher, S., Lickley, M. and Strzepek, K. (2019). Learning about climate change uncertainty enables flexible water infrastructure planning, *Nature communications* **10**(1): 1782.

Galelli, S., Humphrey, G. B., Maier, H. R., Castelletti, A., Dandy, G. C. and Gibbs, M. S. (2014). An evaluation framework for input variable selection algorithms for environmental data-driven models, *Environmental Modelling & Software* **62**: 33–51.

Georgakakos, A., Yao, H., Kistenmacher, M., Georgakakos, K., Graham, N., Cheng, F.-Y., Spencer, C. and Shamir, E. (2012). Value of adaptive water resources management in northern California under climatic variability and change: Reservoir management, *Journal of Hydrology* **412**: 34–46.

Geressu, R. T. and Harou, J. J. (2019). Reservoir system expansion scheduling under conflicting interests, *Environmental Modelling & Software* **118**: 201–210.

Giuliani, M., Li, Y., Castelletti, A. and Gandolfi, C. (2016). A coupled human-natural systems analysis of irrigated agriculture under changing climate, *Water Resources Research* **52**(9): 6928–6947.

Giuliani, M., Pianosi, F. and Castelletti, A. (2015). Making the most of data: an information selection and assessment framework to improve water systems operations, *Water Resources Research* **51**(11): 9073–9093.

Giuliani, M., Zaniolo, M., Castelletti, A., Davoli, G. and Block, P. (2019). Detecting the state of the climate system via artificial intelligence to improve seasonal forecasts and inform reservoir operations, *Water Resources Research* **55**(11): 9133–9147.

Gold, D., Reed, P., Trindade, B. and Characklis, G. (2019). Identifying actionable compromises: Navigating multi-city robustness conflicts to discover cooperative safe operating spaces for regional water supply portfolios, *Water Resources Research* **55**(11): 9024–9050.

Haasnoot, M., Kwakkel, J. H., Walker, W. E. and ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world, *Global environmental change* **23**(2): 485–498.

Haasnoot, M., Schellekens, J., Beersma, J., Middelkoop, H. and Kwadijk, J. C. J. (2015). Transient scenarios for robust climate change adaptation illustrated for water management in the netherlands, *Environmental Research Letters* **10**(10): 105008.

Haasnoot, M., van't Klooster, S. and Van Alphen, J. (2018). Designing a monitoring system to detect signals to adapt to uncertain climate change, *Global environmental change* **52**: 273–285.

Harou, J. J. and Lund, J. R. (2008). Ending groundwater overdraft in hydrologic-economic systems, *Hydrogeology Journal* **16**(6): 1039.

Hawkins, E. and Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions, *Bulletin of the American Meteorological Society* **90**(8): 1095–1108.

Hegerl, G. and Zwiers, F. (2011). Use of models in detection and attribution of climate change, *Wiley interdisciplinary reviews: climate change* **2**(4): 570–591.

Herman, J. D. and Giuliani, M. (2018). Policy tree optimization for threshold-based water resources management over multiple timescales, *Environmental modelling & software* **99**: 39–51.

Herman, J. D., Quinn, J. D., Steinschneider, S., Giuliani, M. and Fletcher, S. (2020). Climate adaptation as a control problem: Review and perspectives on dynamic water resources planning under uncertainty, *Water Resources Research* p. e24389.

Herman, J. D., Reed, P. M., Zeff, H. B. and Characklis, G. W. (2015). How should robustness be defined for water systems planning under change?, *Journal of Water Resources Planning and Management* **141**(10): 04015012.

Huang, X., Hall, A. D. and Berg, N. (2018). Anthropogenic warming impacts on today's Sierra Nevada snowpack and flood risk, *Geophysical Research Letters* **45**(12): 6215–6222.

Hui, R., Herman, J., Lund, J. and Madani, K. (2018). Adaptive water infrastructure planning for nonstationary hydrology, *Advances in water resources* **118**: 83–94.

Jeuland, M. and Whittington, D. (2014). Water resources planning under climate change: Assessing the robustness of real options for the Blue Nile, *Water Resources Research* **50**(3): 2086–2107.

Kasprzyk, J. R., Nataraj, S., Reed, P. M. and Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change, *Environmental Modelling & Software* **42**: 55–71.

Kenney, M. A., Janetos, A. C. and Gerst, M. D. (2018). A framework for national climate indicators, *Climatic Change* pp. 1–14.

Kourakos, G., Dahlke, H. E. and Harter, T. (2019). Increasing groundwater availability and seasonal base flow through agricultural managed aquifer recharge in an irrigated basin, *Water Resources Research* **55**(9): 7464–7492.

Kwakkel, J. H. and Haasnoot, M. (2019). Supporting DMDU: A taxonomy of approaches and tools, *Decision Making under Deep Uncertainty*, Springer, Cham, pp. 355–374.

Kwakkel, J. H., Haasnoot, M. and Walker, W. E. (2015). Developing dynamic adaptive policy pathways: a computer-assisted approach for developing adaptive strategies for a deeply uncertain world, *Climatic Change* **132**(3): 373–386.

Kwakkel, J. H., Haasnoot, M. and Walker, W. E. (2016). Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty, *Environmental Modelling & Software* **86**: 168–183.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., Knutti, R. and Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, *Earth System Dynamics* **11**(2): 491–508.

Lettenmaier, D. P. and Sheer, D. P. (1991). Climatic sensitivity of California water resources, *Journal of Water Resources Planning and Management* **117**(1): 108–125.

Liang, X., Lettenmaier, D. P., Wood, E. F. and Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres* **99**(D7): 14415–14428.

Lopez-Cantu, T., Prein, A. F. and Samaras, C. (2020). Uncertainties in future us extreme precipitation from downscaled climate projections, *Geophysical Research Letters* **47**(9): e2019GL086797.

Mall, N. K. and Herman, J. D. (2019). Water shortage risks from perennial crop expansion in California's Central Valley, *Environmental Research Letters* **14**(10): 104014.

McCabe, G. J., Clark, M. P. and Hay, L. E. (2007). Rain-on-snow events in the western United States, *Bulletin of the American Meteorological Society* **88**(3): 319–328.

McPhail, C., Maier, H., Westra, S., Kwakkel, J. and van der Linden, L. (2020). Impact of scenario selection on robustness, *Water Resources Research* **56**(9): e2019WR026515.

Medellín-Azuara, J., Harou, J. J., Olivares, M. A., Madani, K., Lund, J. R., Howitt, R. E., Tanaka, S. K., Jenkins, M. W. and Zhu, T. (2008). Adaptability and adaptations of California's water supply system to dry climate warming, *Climatic Change* **87**(1): 75–90.

Medellín-Azuara, J., MacEwan, D., Howitt, R. E., Koruakos, G., Dogrul, E. C., Brush, C. F., Kadir, T. N., Harter, T., Melton, F. and Lund, J. R. (2015). Hydro-economic analysis of groundwater pumping for irrigated agriculture in California's Central Valley, usa, *Hydrogeology journal* **23**(6): 1205–1216.

Mortazavi-Naeini, M., Kuczera, G., Kiem, A. S., Cui, L., Henley, B., Berghout, B. and Turner, E. (2015). Robust optimization to secure urban bulk water supply against extreme drought and uncertain climate change, *Environmental Modelling & Software* **69**: 437–451.

Nayak, M. A., Herman, J. D. and Steinschneider, S. (2018). Balancing flood risk and water supply in California: Policy search integrating short-term forecast ensembles with conjunctive use, *Water Resources Research* **54**(10): 7557–7576.

Pielke Sr, R. A., Wilby, R., Niyogi, D., Hossain, F., Dairuku, K., Adegoke, J., Kallos, G., Seastedt, T. and Suding, K. (2012). Dealing with complexity and extreme events using a bottom-up, resource-based vulnerability perspective, *Extreme events and natural hazards: The complexity perspective* **196**: 345–359.

Quinn, J., Hadjimichael, A., Reed, P. and Steinschneider, S. (2020). Can exploratory modeling of water scarcity vulnerabilities and robustness be scenario neutral?, *Earth's Future* **8**(11): e2020EF001650.

Quinn, J., Reed, P., Giuliani, M. and Castelletti, A. (2019). What is controlling our control rules? opening the black box of multi-reservoir operating policies using time-varying sensitivity analysis, *Water Resources Research* .

Raso, L., Barbier, B. and Bader, J.-C. (2019). Modeling dynamics and adaptation at operational and structural scales for the ex-ante economic evaluation of large dams in an African context, *Water Resources and Economics* **26**: 100125.

Ray, P., Wi, S., Schwarz, A., Correa, M., He, M. and Brown, C. (2020). Vulnerability and risk: climate change and water supply from California's Central Valley water system, *Climatic Change* pp. 1–23.

Robinson, B. and Herman, J. D. (2019). A framework for testing dynamic classification of vulnerable scenarios in ensemble water supply projections, *Climatic Change* **152**(3-4): 431–448.

Schneller, G. and Sphicas, G. (1983). Decision making under uncertainty: Starr's domain criterion, *Theory and Decision* **15**(4): 321–336.

Shortridge, J. E. and Zaitchik, B. F. (2018). Characterizing climate change risks by linking robust decision frameworks and uncertain probabilistic projections, *Climatic change* **151**(3-4): 525–539.

Sleeter, B., Wilson, T. and Sherba, J. (2017). Land use and land cover projections for California's 4th climate assessment: US geological survey data release.

Sohl, T. L., Sayler, K. L., Bouchard, M. A., Reker, R. R., Friesz, A. M., Bennett, S. L., Sleeter, B. M., Sleeter, R. R., Wilson, T., Soulard, C. et al. (2014). Spatially explicit modeling of 1992–2100 land cover and forest stand age for the conterminous United States, *Ecological Applications* **24**(5): 1015–1036.

Starr, M. K. (1963). *Product design and decision theory*, Prentice-Hall.

Stephens, S. A., Bell, R. G. and Lawrence, J. (2018). Developing signals to trigger adaptation to sea-level rise, *Environmental Research Letters* **13**(10): 104004.

Surfleet, C. G. and Tullos, D. (2013). Variability in effect of climate change on rain-on-snow peak flow events in a temperate climate, *Journal of Hydrology* **479**: 24–34.

Tanaka, S. K., Zhu, T., Lund, J. R., Howitt, R. E., Jenkins, M. W., Pulido, M. A., Tauber, M., Ritzema, R. S. and Ferreira, I. C. (2006). Climate warming and water management adaptation for California, *Climatic Change* **76**(3-4): 361–387.

Taner, M. Ü., Ray, P. and Brown, C. (2019). Incorporating multidimensional probabilistic information into robustness-based water systems planning, *Water Resources Research* .

Trindade, B., Gold, D., Reed, P., Zeff, H. and Characklis, G. (2020). Water pathways: An open source stochastic simulation system for integrated water supply portfolio management and infrastructure investment planning, *Environmental Modelling & Software* **132**: 104772.

Trindade, B., Reed, P., Herman, J., Zeff, H. and Characklis, G. (2017). Reducing regional drought vulnerabilities and multi-city robustness conflicts using many-objective optimization under deep uncertainty, *Advances in Water Resources* **104**: 195–209.

USACE (2016). Folsom Dam raise draft environmental impact statement/environmental impact report, https://www.spk.usace.army.mil/Portals/12/documents/civil_works/JFP/Raise/Folsom%20Dam%20Raise%20Draft%20SEIS-SEIR%20-%20July%202016.pdf.

USBR (2015). Shasta Lake water resources investigation feasibility report, https://www.usbr.gov/mp/ncao/slwri/docs/feasability/slwri-final-fr-full.pdf.

USBR (2017). North-of-the-Delta offstream storage investigation feasibility report, https://sitesproject.org/resources/feasibilty-report/.

USBR (2018). Addendum to the coordinated operation agreement Central Valley Project/State Water Project, https://www.usbr.gov/mp/nepa/includes/documentShow.php?Doc_ID=36503.

USBR (2020). Central Valley Project final cost allocation study, https://www.usbr.gov/mp/cvp/docs/cvp-final-cost-allocation-study-2020.pdf.

VanRheenen, N. T., Wood, A. W., Palmer, R. N. and Lettenmaier, D. P. (2004). Potential implications of PCM climate change scenarios for Sacramento–San Joaquin River basin hydrology and water resources, *Climatic change* **62**(1-3): 257–281.

Vicuna, S. and Dracup, J. (2007). The evolution of climate change impact studies on hydrology and water resources in California, *Climatic Change* **82**(3-4): 327–350.

Vicuna, S., Maurer, E. P., Joyce, B., Dracup, J. A. and Purkey, D. (2007). The sensitivity of California water resources to climate change scenarios, *JAWRA Journal of the American Water Resources Association* **43**(2): 482–498.

Walker, W., Haasnoot, M. and Kwakkel, J. (2013). Adapt or perish: a review of planning approaches for adaptation under deep uncertainty, *Sustainability* **5**(3): 955–979.

West, T. and Le Page, Y. (2014). NACP CMS: Land cover projections (5.6-km) from GCAM v3.1, conterminous US, 2005-2095.
URL: *http://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id* = 1216

Wilby, R. L. and Dessai, S. (2010). Robust adaptation to climate change, *Weather* **65**(7): 180–185.

Yao, H. and Georgakakos, A. (2001). Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. reservoir management, *Journal of Hydrology* **249**(1-4): 176–196.

Zaniolo, M., Giuliani, M. and Castelletti, A. (2021). Neuro-evolutionary direct policy search for multiobjective optimal control, *IEEE Transactions on Neural Networks and Learning Systems* .

Zeff, H. B., Herman, J. D., Reed, P. M. and Characklis, G. W. (2016). Cooperative drought adaptation: Integrating infrastructure development, conservation, and water transfers into adaptive policy pathways, *Water Resources Research* **52**(9): 7327–7346.

Conclusion

This dissertation has presented three new frameworks for adapting water resources systems to the many challenges faced by climate uncertainty. Chapter 1 introduces a top-down approach to isolate and adapt to specific physical impacts of climate change projected with confidence, namely snowpack decline, while also designing adaptations that respond to more uncertain impacts in total water availability. These methods contribute to the literature on top-down climate change adaptation in water resources systems while also providing adaptation policies generalizable to snowmelt-dominated systems in the Western U.S. and elsewhere. Chapter 2 advances the design and testing of robust control policies as an adaptation to uncertainty in hydroclimatic conditions, contributing an experimental design to better understand the influence of the forcing scenario properties and baseline regret of training scenarios on the robustness of resulting policies. A key finding is that that policies trained to scenario sets with high baseline regret tend to outperform those generated with other training sets in both wetter and drier futures. Additionally, the policies adapted under these conditions develop an intra-annual hedging strategy to mitigate the effects of snowpack decline under rising temperatures. The approach highlights the general importance of considering the specific properties of training scenarios in the design of robust control policies to adapt to climate uncertainty. Chapter 3 develops a new dynamic planning that addresses some shortcomings of previous adaptive planning frameworks in water resources system. Multi-objective policy tree optimization yields dynamic climate adaptation policies which do not require prespecification of actions, indicators, thresholds, timing, or structure. Analysis of a non-dominated set of policies identifies the most relevant indicators and ac-

tions. Policies also identify strategies for incremental expansion of infrastructure, in order to minimize regret. The framework introduced in Chapter 3 identifies relevant signals for adaptation among the many interrelated impacts of climate and demand changes, and the policy search assigns actions to these signals without a prespecified structure, making it a transferable problem formulation to advance dynamic planning under uncertainty for water resources systems and related fields. Overall, these three studies allow for a synthesized robust-dynamic approach for climate adaptation. The frameworks build upon each other, combining scenario selection with policy design and vulnerability assessment, which in the end are used to yield a new, flexible approach to dynamic policy design via policy tree optimization. In total, these contributions provide a range of strategies which can be used in tandem for adaptation of water resources systems to climate uncertainty.