CHARACTERIZING AND MANAGING DEEPLY UNCERTAIN RISKS IN COUPLED HUMAN-NATURAL SYSTEMS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Julianne Dorothy Quinn

August 2017

CHARACTERIZING AND MANAGING DEEPLY UNCERTAIN RISKS IN COUPLED HUMAN-NATURAL SYSTEMS

Julianne Dorothy Quinn, Ph.D.

Cornell University 2017

Coupled human-natural systems are complex systems composed of interacting human and natural components. Managing these systems requires careful characterization of which system uncertainties drive their dynamics and how human actions interact with the natural system to create feedbacks. This dissertation advances exploratory modeling techniques to discover interactions and dependencies between elements of the human and natural systems to better characterize risks to each component. These techniques are illustrated on two socio-ecological systems serving multiple objectives: a managed lake and a multi-reservoir system. These case studies illustrate ways in which the coupled dynamics in these systems can differ under alternative human control strategies due to complex interactions between the two components, and their conclusions have important implications for managing several common challenges in socio-ecological systems, namely: tipping points, problem formulation uncertainty and risk characterization. The first case study on managed lakes shows that state-dependent control rules describing a town's pollutant discharge policy are more robust to deep uncertainties in lake model parameters than static, temporal control rules, reducing the probability of the lake's water quality crossing an irreversible tipping point. Furthermore, adaptive state-dependent control rules can be readily coupled with statistical learning techniques to better navigate deeply uncertain lake parameterizations. The second case study illustrates

how uncertainty in how to formulate a socio-ecological management problem, specifically a multi-objective, multi-reservoir operating problem, strongly influences the resulting human control strategies found to be optimal, and consequently how those strategies impact the system dynamics. This underlines the importance of exploring rival framings of how to formulate socio-ecological management problems to discover unintended consequences of different formulations. Finally, further work on the same multi-reservoir problem analyzing the impacts of plausible changes in monsoonal dynamics and sectoral water demands highlights the importance of sampling a broad range of potential drivers of change to characterize the most important risks to coupled human-natural systems, as failure modes may result from mixtures of complex factors. In summary, this work advances exploratory modeling techniques to yield a greater understanding of the dynamics of coupled human-natural systems that can be used to inform adaptive management strategies for building more robust and resilient systems.

BIOGRAPHICAL SKETCH

Julianne Quinn grew up in Bedford, NH. She graduated from Manchester High School West in 2007 and then attended Columbia University's School of Engineering and Applied Science, where she received a B.S. in Earth and Environmental Engineering with a concentration in Water Resources and Climate Risk in 2011. She then began a Ph.D. in Environmental and Water Resources Systems in Cornell University's Department of Civil and Environmental Engineering. She will begin a postdoctoral research position in the same department with Professor Patrick Reed in August 2017.

To my parents, Mary and Michael Quinn,

who always let me choose my own path,

and made sure every path was available to me.

ACKNOWLEDGEMENTS

Completing a Ph.D. requires a great support network and I am fortunate to have had countless mentors and friends guide and encourage me throughout the process. First and foremost, I would like to thank my Ph.D. committee for their technical and professional guidance and feedback on my research. I would especially like to thank my adviser, Prof. Patrick Reed, for his undying commitment to my (and all of his students') success. Working with him has been a pleasure and I look forward to continuing to learn from him as a postdoctoral researcher next year. I would also like to thank Prof. Jery Stedinger for his concern, encouragement, and selfless support throughout my Ph.D., and for his exemplary commitment to excellence in scholarship. Finally, I would like to thank Prof. Todd Walter for always expressing an active interest in my work, and for bringing a lighthearted spirit and sense of humor to every situation.

I've also greatly appreciated the opportunity to collaborate with other researchers at Cornell and elsewhere, broadening my knowledge in a wide range of topics. I'd like to thank Prof. Ellen McCullough (Georgia), Prof. Andrew Simons (Fordham), Dr. Leslie Verteramo Chiu (Cornell) and Prof. Josh Woodard (Cornell) for sharing their knowledge and ideas about agricultural economics and finance, and for inviting me to collaborate with them on development work in Africa. I'd also like to thank Prof. Andrea Castelletti and Prof. Matteo Giuliani of Politecnico di Milano for introducing me to multi-objective control and for always providing constructive feedback on my work. I'm also grateful to many researchers at Penn State. I'd like to thank Dr. Jared Oyler and Dr. Rob Nicholas for sharing their expertise in statistical climate downscaling to enhance my research, and Prof. Klaus Keller and Prof. Riddhi Singh (IIT Hyderabad) for teaching me about statistical learning techniques and inviting me to work

with them on collaborative research projects. Finally, I'd like to thank Prof. Jon Herman (UC Davis) for his friendship and encouragement, as well as his great research ideas and constant willingness to help.

In addition to my committee and research collaborators, I've been fortunate to have the support and friendship of many fellow graduate students at Cornell: Dr. Taimoor Akhtar, Liang Chen, Kwang-Bae Choi, Dr. Antoine Espinet, Dave Gold, Prof. Jon Herman, Prof. Jon Lamontagne, Dr. Min Pang, Dr. Rebecca Potash, Kyle Perline, Jared Smith, Dr. Sue Nee Tan, Dr. Danielle Toupo, Bernardo Trindade, Tori Ward, Dr. Tom Wild, Dr. Josh Woodbury, Candice Yu, Jazmin Zatarain Salazar, and Dr. Ying Zhang. I've also greatly appreciated the camaraderie of countless runners in the Ithaca area willing to geek out about running or research, as well as listen to me vent about my academic frustrations in exchange for listening to them vent about theirs: Chelsea Benson, Kristian Clemens, Dr. Anne Elise Creamer, Adam Engst, Jullien Flynn, Dr. Ved Gund, Dr. Matt Hall, Justin Hendrick, Rachel Hilliard, Dr. Matt Holden, Jonathan Klus, Kristina Kronauer, Dr. Michael Lam, Alex Looi, Rob MacDonald, Sean McCauley, Cassidy Nagle, Eric Obeng, Aaron Proujansky, Scott Wehrwein, Henry Wells and many others.

This research was funded by the U.S. National Science Foundation (NSF) through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the Penn State Center for Climate Risk Management. High-performance computing resources were provided by the Texas Advanced Computing Center (TACC) and Blue Waters. I also received funding from the Cornell School of Civil & Environmental Engineering Olin Fellowship, the Cornell University Bernard Meyers Academic Award and an NSF Integrative Graduate Education and Research Traineeship

(IGERT) on Food Systems and Poverty Reduction. I gratefully acknowledge the support of these programs.

Finally, I could have never gotten to where I am without the love and support of my parents, who have always encouraged me to pursue my goals and have provided me with the foundation to do so. And meeting those goals would be an uphill battle without the love and compassion of my partner, Alex Looi. His support and understanding help pull me through tough times, while his humor and incessant teasing make the good times even better.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Current Challenges to Managing Deep Uncertainties in Coupled Human-Natural Systems

Coupled human-natural systems are complex systems composed of interacting human and natural components [Liu et al., 2007]. Designing management strategies for the human component of these systems that will yield anthropologic benefits without unduly harming the natural system poses several challenges due to the complex interactions between the two components. These include nonlinear dynamics, thresholds, reciprocal feedback loops, time lags, heterogeneity and unpredictability [Liu et al., 2007]. Consequently, studying only the human or natural component misses important interactions and feedbacks, a realization that has bred a call for coupled modeling of the two systems from researchers in both disciplines, spawning the fields of socio-hydrology [Sivapalan et al., 2012], social-ecological systems [Anderies et al., 2006], and socio-environmental systems [Filatova et al., 2016]. This dissertation contributes to these burgeoning fields by better characterizing deep uncertainties impacting interactions between the two components in order to inform more robust control strategies for the human component that balance often conflicting societal and environmental objectives. "Deep," or Knightian [Knight, 1921], uncertainties refer to those for which there is "no scientific basis on which to form any calculable probability whatever" [Keynes, 1937]. The work that follows specifically focuses on tackling three major challenges associated with managing complex human-natural systems in the presence of deep uncertainty: tipping points,

problem formulation uncertainty, and risk characterization.

1.1.1 Balancing Conflicting Objectives in Socio-Ecological Systems with Deeply Uncertain Tipping Points

Many socio-ecological systems exhibit tipping points that, when crossed, result in a regime shift between opposing equilibria [Filatova et al., 2016; Liu et al., 2007]. Horan et al. [2011] note that most research related to managing tipping points in these systems has either assumed no institutional constraints, making the system robust to shocks that cause tipping, or rigid constraints with no flexibility to adapt. We strike a balance between these two extremes in our study of managing tipping points in socio-ecological systems, described in Chapter 3. Introducing a control strategy termed Direct Policy Search (DPS; Rosenstein & Barto [2001]) to the socio-ecological systems literature, we show how a town managing phosphorus emissions to a lake exhibiting threshold behavior can formulate adaptive pollutant control strategies, limited by some institutional constraints, that result in more stable and robust management plans compared to static temporal control strategies. These findings show promise for managing other socio-ecolocial systems with tipping points, such as harvested fish populations, grasslands consumed by cattle on rangelands and, most importantly, the Earth system [Carpenter et al., 2015; Anderies et al., 2013].

1.1.2 Tackling Problem Formulation Uncertainty

The first step in designing policies for managing socio-ecological systems is to define the problem. Most research focuses on solving an assumed problem, though, not defining it. However, the definition of the problem itself may be deeply uncertain. In particular, when the problem is to optimize the system performance with respect to several objectives, it is not clear what these objectives should be, or how they should be quantified. Chapter 4 tackles these often ignored uncertainties on a multi-objective, multi-reservoir control problem using a rival framings optimization framework. Under this framework, several candidate performance measures for system objectives are tested against one another to discover a problem formulation that effectively characterizes stakeholders' risk preferences with respect to different system objectives. This framework extends several constructive decision aiding approaches promoted in the literature [Tsoukias, 2008; Walker et al., 2003; Roy, 1990; Zeleny, 1989; Maass, 1967] using the same flexible optimization method utilized in Chapter 3, DPS. Importantly, we highlight how improved computational power enables us to more formally implement these approaches, and how advanced computational methods allowing for complex mixtures of objective functions to be optimized simultaneously necessitate the exploration of how these objectives interact when combined in a single problem formulation.

1.1.3 Characterizing Hydrologic and Socioeconomic Risks to Inform Adaptive Water Systems Control Policies

Central to effective management of socio-ecological systems is their ability to adapt to changes in external stressors. Anticipating how to adapt requires knowledge of which stressors are most important to system performance. In Chapter 5 of this dissertation, we illustrate how exploratory modeling approaches [Bankes, 1993] in which management strategies are re-evaluated over a broad range of system conditions can be used to characterize the greatest risks to the system and discover the most important drivers of system performance to inform adaptive management. In particular, we illustrate the need to explore a range of potential drivers rather than a subset assumed to be most important *a priori* in order to capture the complex system dynamics that emerge in coupled human-natural systems [Pruyt & Islam, 2016]. Chapter 5 specifically focuses on exploring the effects of changes in monsoonal dynamics on the robustness of alternative multi-reservoir management plans, finding that hydrologic conditions beyond mean streamflow can interact with other drivers to cause system failures that a limited sampling of factors would miss.

1.2 Scope and Organization

1.2.1 Chapter 2: Methodological Components

Chapter 2 provides a brief overview of common computational methods utilized in Chapters 3-5 of this dissertation. These methods include Multi-

Objective Evolutionary Optimization with the Borg MOEA and optimal control with Evolutionary Many Objective Direct Policy Search (EMODPS). The goal of these methods is to find alternative control policies for socio-ecological and multi-reservoir systems that span key tradeoffs.

1.2.2 Chapter 3: Direct Policy Search for Robust Multi-Objective Management of Deeply Uncertain Socio-Ecological Tipping Points

Chapter 3 introduces a closed looped control method called Direct Policy Search (DPS) to the socio-ecological systems literature, illustrating how the method's state-based adaptivity offers several advantages over open loop control strategies often used to design environmental and climate policies. These two solution techniques are compared on a stylized problem from ecological economics in which a town situated on a lake must balance the economic benefits it receives from polluting the lake with phosphorus (P) against the associated ecological consequences [Carpenter et al., 1999]. Managing this conflict is especially challenging because the lake exhibits threshold behavior where it is clean and clear at low P levels, but then quickly transitions to a permanently polluted state upon crossing an irreversible tipping point, dubbed the "critical P threshold." Furthermore, the location of this tipping point may be deeply uncertain, meaning managers cannot agree on prior probability distributions for its value [Lempert & Collins, 2007].

Despite these challenges, we find that using DPS to define P release policies

for the town as a function of the current P concentration in the lake enables the discovery of solutions that increase economic benefits relative to solutions designed using an open loop control strategy without degrading performance on ecological objectives (Section 3.6.1). Moreover, DPS is much more computationally efficient, finding these policies after a limited number of policy evaluations (Section 3.6.1). DPS also finds more stable policies that do not exhibit undesirable end-of-time behavior like open-loop control strategies (Section 3.6.2). Finally, it discovers solutions that are more robust to deep uncertainties, satisfying minimum performance levels across objectives under a greater range of plausible lake model parameters (Section 3.6.3). However, despite these benefits, even the most robust DPS policies are still sensitive to assumptions about the lake's critical P threshold (Section 3.6.3). The chapter concludes with a discussion of how this sensitivity could be reduced by including time-evolving estimates of model parameters and their associated uncertainty into the DPS control policies themselves. Chapter 3 is drawn from Quinn et al. [2017], published in *Environmental Modelling & Software*.

1.2.3 Chapter 4: Rival Framings Framework for Discovering how Problem Formulation Uncertainties Shape Risk Management Tradeoffs in Water Resources Systems

Chapter 4 promotes a formal evaluation of rival framings of multi-objective water resources optimization problems to discover the most effective combinations of performance measures for capturing stakeholder risk preferences across multiple system objectives. In particular, the goal is to guard against unintended

consequences of particular formulations that might go undiscovered without formally evaluating several alternatives and diagnosing reasons for their failure or success. This is particularly important when designing control policies with EMODPS, as it allows for the formulation of complex mixtures of objective functions, which may enable a better representation of stakeholder objectives if they are formulated effectively, or result in unpredictable adverse interactions if they are formulated poorly. While no single formulation will be without consequences, comparing rival framings will enable stakeholders to not only choose preferred solutions, but preferred formulations whose consequences they are most willing to accept.

In Chapter 4, this framework is illustrated on the Red River basin in Vietnam, where operating policies for the system's four largest reservoirs must serve several objectives: flood protection to the capital city of Hanoi, hydropower production, and water supply to multiple sectors, agriculture being the most dominant. In evaluating four rival framings of the Red River multi-objective optimization problem, several objective formulations commonly used in the water resources literature are found to suffer severe unintended consequences. In particular, expectation objectives are found to be prone to overfitting, standard deviation objectives to result in excessive sacrifices in mean performance, min-max objectives to provide a poor representation of system tradeoffs because of their instability, and expected damages objectives to mask extreme flood events, inadvertently favoring hydropower production. These consequences accentuate the need to compare competing problem formulations when designing water resources management plans. Chapter 4 is drawn from Quinn et al. [Accepted], recently accepted for publication in *Water Resources Research*.

1.2.4 Chapter 5: Exploring How Changing Monsoonal Dynamics and Human Pressures Challenge Multi-Reservoir Management of Food-Energy-Water Tradeoffs

Chapter 5 builds off of the analyses in Chapter 4 by evaluating the robustness of alternative Red River reservoir operating policies to deep uncertainties in hydrologic and socioeconomic conditions. Since there is great uncertainty in how climate change, urbanization and agricultural technology will impact current operating conditions, it is valuable to re-evaluate alternative operating policies across a broad range of potential system conditions to discover which policies are most robust, i.e. less sensitive to these uncertainties. Not only can this re-evaluation help identify robust policies, but it can also discover which uncertainties are most important in driving system performance. This information can be used to inform more adaptive operating policies for the future.

In order to determine what the most influential factors are, it is important to re-evaluate policies over a wide range of potential changes in system conditions and external stressors to effectively capture the range of system dynamics that may result from operating with different policies under different conditions. In this study, we contribute a new sampling scheme to generate changing monsoonal dynamics in addition to changes in the first two moments of annual flow and sectoral water demand. Applying this approach to the Red River basin, we find that the log-space mean annual flow, annual standard deviation, and amplitude of the monsoonal cycle emerge as the most important drivers of system performance across the system's food-energy-water objectives (Section 5.5.2), highlighting the importance of sampling a complex mix of factors. Frighten-

ingly, under all of the evaluated reservoir operating policies, the Red River basin is sitting at the edge of failure: the combined effects of the three most important uncertain factors result in the most robust solution across all system objectives only meeting minimum performance criteria under a corridor of system conditions. Furthermore, downscaled climate projections quickly move to regions outside of this corridor, in particular to regions that cause severe flooding. Consequently, it is crucial that operating policies for the future be designed to adapt as the world moves outside the current safe operating space. Future work will explore how the factors deemed most important to performance can be used to inform the design of more adaptive policies. Chapter 5 is drawn from Quinn et al. [In Prep].

1.2.5 Chapter 6: Contributions & Future Work

Finally, Chapter 6 summarizes the contributions of this dissertation to the field of decision making for coupled human-natural systems under deep uncertainty. Also discussed are several avenues of future research building off of the contributions made here. These include information selection and policy formulation, learning and adaptation, and designing more resilient socio-ecological systems.

1.2.6 Author Contributions for Collaborative Work

Chapter 3: Julianne Quinn conceived the study and led the modeling, data analysis and writing. Patrick Reed and Klaus Keller supervised the experiments and contributed to the data analysis and writing.

Chapter 4: Julianne Quinn conceived the study and led the optimization, data analysis and writing. Patrick Reed supervised the experiments and contributed to the data analysis and writing. Matteo Giuliani built the system model and contributed to the data analysis and writing. Andrea Castelletti supervised the experiments and contributed to the data analysis and writing.

Chapter 5: Julianne Quinn conceived the study and led the computational experimentation, data analysis and writing. Patrick Reed supervised the experiments and contributed to the data analysis and writing. Matteo Giuliani built the system model and contributed to the data analysis. Andrea Castelletti supervised the experiments and contributed to the data analysis. Jared Oyler performed the statistical climate downscaling and contributed to the writing. Robert Nicholas supervised the statistical climate downscaling.

CHAPTER 2

METHODOLOGICAL COMPONENTS

2.1 Multi-Objective Evolutionary Optimization

Multi-objective evolutionary optimization is the heuristic optimization of multi-objective problems using Multi-Objective Evolutionary Algorithms (MOEAs). A defining feature of MOEAs that has led to their growing popularity is that they do not discover a single optimal solution, but a set of non-dominated solutions also called the Pareto approximate set [Pareto, 1896]. Within this set, performance in any component objective can only be improved by degrading performance in one or more of the remaining objectives. This set is called "approximate" because the search process is limited by a user-defined number of function evaluations (NFE), or model runs, reducing the likelihood of discovering the true optimal set for non-trivial applications. MOEAs have emerged as dominant solution tools for complex environmental and water resources problems because they are effective at solving nonlinear, nonconvex, multi-modal and discrete multi-objective problems where derivative-based algorithms often fail [Nicklow et al., 2010]. They rely on an iterative search process in which a number of different probabilistic operators for mating, mutation, selection, and archiving are used to modify and evolve a population of candidate solutions [Reed et al., 2013]. In this thesis, the Borg MOEA [Hadka & Reed, 2013] is utilized to find multi-objective control policies for socio-ecological and multi-reservoir systems because of its proven reliability on challenging multi-objective environmental problems [Hadka & Reed, 2013; Reed et al., 2013; Ward et al., 2015; Zatarain Salazar et al., 2016]. Section 2.1.1 provides a brief description of

the algorithm.

2.1.1 The Borg MOEA

The Borg MOEA is a hyper-heuristic optimization algorithm, meaning it uses feedbacks from its search progress to adapt the underlying strategies it employs for improving its performance on a given problem [Burke et al., 2003]. The Borg MOEA does this through the use of multiple self-adaptive search operators that have shown success on a range of multi-objective optimization problems. These operators generate new candidate solutions by manipulating existing solutions in one of two ways: 1) by "mutating" them to promote random diversification, and 2) by recombining them with one or more other existing solutions to translate search through the decision space. The operators included in the Borg MOEA are simulated binary crossover (SBX; Deb & Agrawal [1994]), differential evolution (DE; Storn & Price [1997]), polynomial mutation (PM) and uniform mutation (UM) applied to each of the D objectives with probability $1/D$ [Hadka & Reed, 2013], parent-centric crossover (PCX; Deb et al. [2002]), unimodal normal distribution crossover (UNDX; Kita et al. [1999]) and simplex crossover (SPX; Tsutsui et al. [1999]). Figure 2.1, reproduced from Hadka & Reed [2013], illustrates the distribution of solutions generated by these operators.

The Borg MOEA adapts the probability of using these operators throughout the search based on their recent success in generating non-dominated solutions. It also adapts the population size throughout the search [Kollat & Reed, 2007] and utilizes an easily parallelized steady-state algorithm structure [Deb et al., 2005] in which new solutions are generated one-at-a-time (i.e., asynchronous

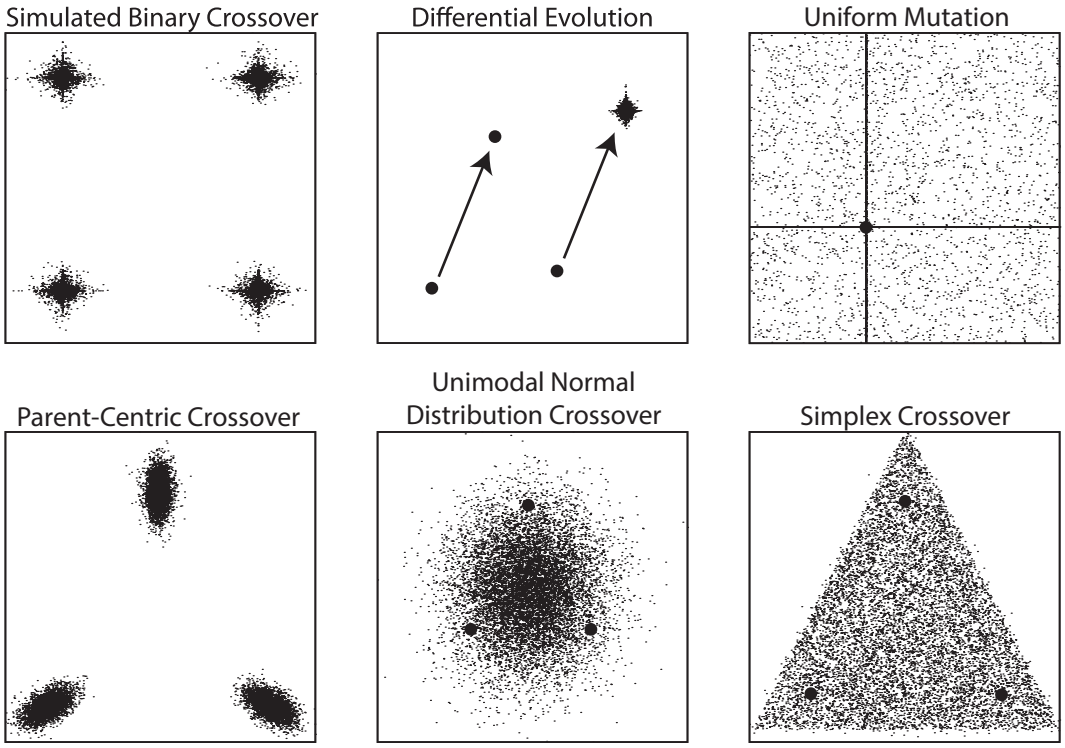

Figure 2.1: Example solutions generated by the algorithmic operators used by the Borg MOEA for a problem with two decision variables, reproduced from Hadka & Reed [2013]. The large circles represent the initial ("parent") solutions located in the (x,y) plane at the value of their respective decisions variables, while the small circles represent sample new solutions ("children") generated by the operators.

search), as opposed to an entire population at a time in generational search. This adaptability as well as ε-dominance archiving and ε-progress have been credited as the reasons for the Borg MOEA's consistent success on challenging multi-objective optimization problems [Hadka & Reed, 2013; Reed et al., 2013; Ward et al., 2015; Zatarain Salazar et al., 2016]. A recent, parallel extension of the Borg MOEA called Multi-Master Borg has further improved the algorithm's reliability [Hadka & Reed, 2015]. These concepts are described below.

ε-dominance archiving and ε-progress

Under traditional point dominance, a solution is added to the non-dominated set, or "archive," throughout the search if it improves upon all other evaluated solutions in at least one objective. The ε-dominance, or block dominance, concept distinguishes itself from point dominance by instead calculating dominance over ε-"boxes." These multi-dimensional boxes are created based on user-specified precision levels, ε, in each dimension [Hadka & Reed, 2013]. Dominance is then calculated with respect to the boxes, meaning that if a solution resides within a box that is dominated by another box in which another solution resides, it is not added to the non-dominated set. Additionally, if multiple solutions lie within the same ε-box, only the solution closest to the ideal point of that box is stored in the Pareto-approximate set. This thins the number of solutions stored in the archive, preventing a computationally burdensome growth in the number of solutions that must be sorted each iteration [Hadka & Reed, 2013; Laumanns et al., 2002]. Consequently, ε-dominance speeds up algorithmic search processes relative to traditional point dominance while also providing a mathematical proof of convergence and guaranteeing a diverse representation of problems' tradeoffs [Laumanns et al., 2002].

Figure 2.2, reproduced from Kasprzyk [2013], provides an illustration of the ε-dominance concept on a two-objective minimization problem. In this figure, all of the solutions are non-dominated when using point dominance, as each solution performs better than any other in one objective, but degrades with respect to the other objective. However, using the ε-dominance concept, solution A would be eliminated from the set because it resides within an ε-box that is dominated by the ε-box in which solution B resides. Additionally, solution D

would not be included because it falls within the same ε-box as solution C, but solution C is closer to the ideal point within that box.

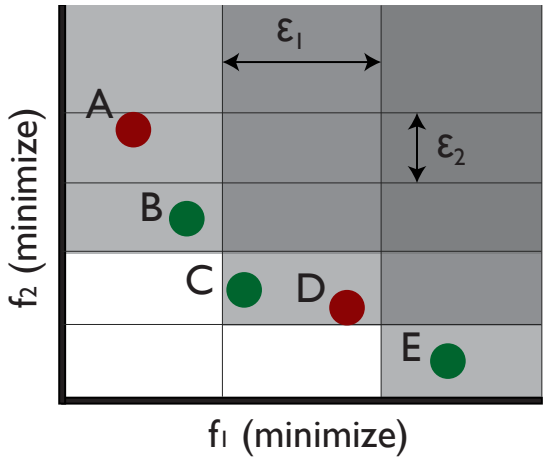

Figure 2.2: Illustration of ε-dominance on a two-objective minimization problem, reproduced from Kasprzyk [2013]. ε_1 indicates the precision on objective 1, while ε_2 indicates the precision on objective 2. Since solution B lies within an ε-box that dominates the ε-box in which solution A resides, solution A is not included in the ε-dominance Pareto set. Similarly, since solution D lies within the same ε-box as solution C but is further from that box's ideal point, solution D is not included in the ε-dominance Pareto set. This leaves only the green solutions: B, C and E.

In the Borg MOEA, ε-progress measures how frequently a solution is added to the archive that improves upon a previous solution by $> \varepsilon$, meaning the new solution lies within an ε-box that dominates the ε-box in which the previous solution resided [Hadka & Reed, 2013]. Borg increases the probability of using algorithm operators that frequently contribute to ε-progress and decreases the probability of using operators that do not frequently contribute to ε-progress. This makes the algorithm largely insensitive to the initial values of its parameters, unlike many other algorithms [Hadka & Reed, 2013; Reed et al., 2013; Ward et al., 2015].

Multi-Master Borg

Multi-Master Borg is a hierarchical parallelization of the Borg MOEA that has been shown to improve the reliability of attaining high-quality approximations to the Pareto optimal set for challenging real-world problems [Hadka & Reed, 2015; Giuliani et al., In Press]. The Multi-Master Borg consists of multiple master-worker implementations of the Borg MOEA. Each of these implementations, called islands, co-evolves through the aid of a controller that keeps a global archive of the best solutions across all of the islands. Figure 2.3 illustrates how the parallelization works. On each island, a master node allocates jobs to the workers as in the Borg MOEA, keeping a local archive of its best solutions. However, if search stagnates at that island, the controller injects solutions from the global archive into that island's local archive. This reduces the probability of poor search performance, increasing the algorithm's reliability across random seeds [Hadka & Reed, 2015].

2.2 Optimal Control

Optimal control refers to the mathematical optimization of control policies [Bertsekas, 1995]. Control policies can fall into one of two categories: open loop or closed loop. In open loop control, all decisions are determined based only on the initial state and the model. In closed loop control, information on the system state feeds back to the control to inform future decisions. Figure 2.4 illustrates these concepts on a sample problem in which the control decision is how much anthropogenic phosphorus (P) to release to a lake at time t, denoted a_t, and the system state is the current P concentration in the lake at time t, de-

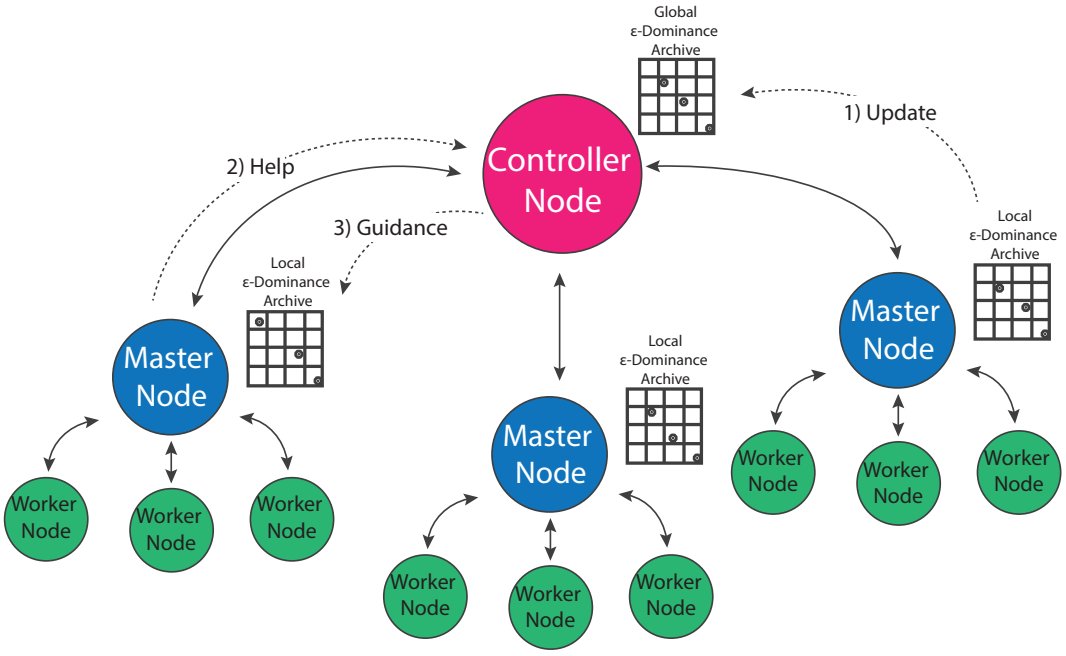

Figure 2.3: Illustration of the hierarchical structure and communication of the Multi-Master Borg MOEA, adapted from Hadka & Reed [2015]. Each island, composed of a master node shown in blue and several worker nodes shown in green, begins with its own set of solutions that initially evolve independently as in the serial Borg algorithm. However, if ε-progress at any of the master node's local archives stalls, that master seeks help from the controller node shown in pink. The controller node maintains a global ε-dominance archive across the islands and injects solutions from the global archive into the local archive of any master node seeking help.

noted X_t. The system also receive P inputs from the sediment, $\frac{X_t^q}{1+X_t^q}$, as well as stochastic P inputs from the environment, Y_t.

In this thesis, a multi-objective, closed loop control strategy called Evolutionary Many Objective Direct Policy Search (EMODPS; Giuliani et al. [2016b]) is employed to design operating policies for socio-ecological and multi-reservoir systems. This method is described in Section 2.2.1.

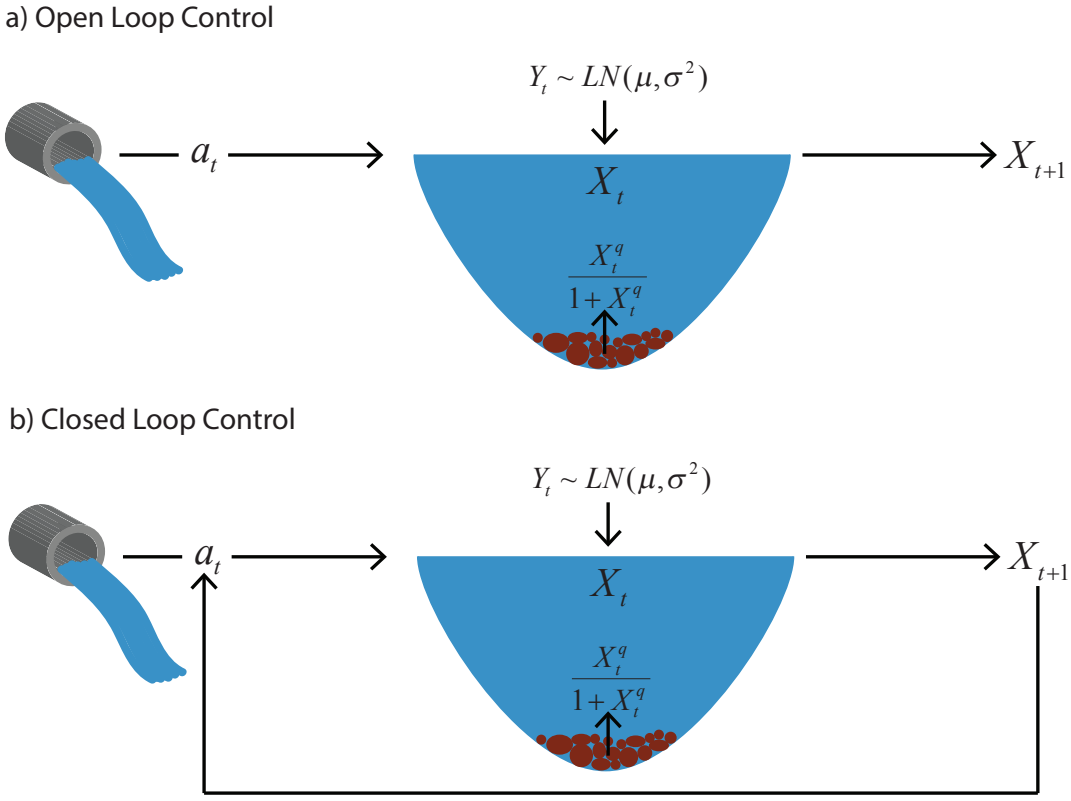

Figure 2.4: Example of open vs. closed loop control. In open loop control (panel a), the control decision is made without any feedback from the system state. In this example, the control decision is how much anthropogenic phosphorus (P) to release to a lake, a_t, while the system state is the P concentration in the lake, X_t. In closed loop control (panel b), the lake P concentration resulting from the control decision, X_{t+1}, feeds back into the control decision for the next time step, a_{t+1}.

2.2.1 Evolutionary Many Objective Direct Policy Search (EMODPS)

EMODPS is a many objective extension of DPS [Rosenstein & Barto, 2001], a closed loop control strategy that has proven to be a simple and computationally efficient but effective method for solving challenging control problems [Giu-

liani et al., 2016b]. Also called parameterization-simulation-optimization in the water resources literature [Koutsoyiannis & Economou, 2003], the EMODPS approach involves parameterizing control policies (e.g., reservoir release policies) within a given family of functions (e.g., piecewise linear functions, radial basis functions, etc.), simulating them over a series of stochastic inputs, and then optimizing their parameters to improve system performance over multiple objectives computed in the simulation. EMODPS utilizes non-linear universal approximators to parameterize candidate operating policies and MOEAs to optimize their performance over the problem's conflicting objectives.

This process is illustrated in Figure 2.5 for a reservoir optimization problem in which a time series of inflows are fed to a reservoir simulation model along with a policy, p_θ, defined by an initial set of parameters θ_0. The simulation yields a set of system trajectories, τ, over which objectives are calculated. The MOEA then generates a new set of policy parameters, θ_k for iteration k, attempting to improve the objective values computed over the simulation. This process repeats until a user-specified maximum NFE.

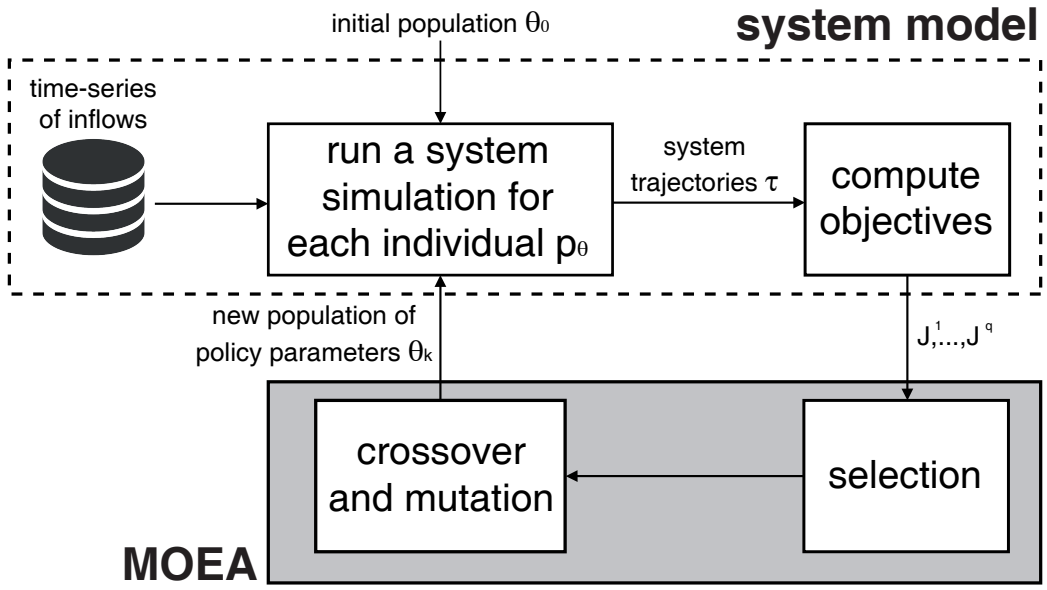

Figure 2.5: Illustration of Evolutionary Many Objective Direct Policy Search (EMODPS) on a multi-objective reservoir optimization problem. Stochastic inputs (in this case, streamflows), as well as control policies, p_θ (in this case, reservoir release policies defined by an initial set of parameters, θ_0) are fed to a system simulation model. This yields a set of system trajectories, τ, over which obejctives are calculated. An MOEA generates new parameter sets, θ_k, for each iteration k, attempting to improve the objective values computed over the simulation. The process repeats until a maximum number of function evaluations have been performed.

CHAPTER 3

DIRECT POLICY SEARCH FOR ROBUST MULTI-OBJECTIVE MANAGEMENT OF DEEPLY UNCERTAIN SOCIO-ECOLOGICAL TIPPING POINTS

This chapter is drawn from the following peer-reviewed journal article:

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 and Software, 92, 125–141.

This work was partially supported by the U.S. National Science Foundation (NSF) through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the Penn State Center for Climate Risk Management. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding entities.

3.1 Abstract

Managing socio-ecological systems is a challenge wrought by competing societal objectives, deep uncertainties, and potentially irreversible tipping points. A classic, didactic example is the shallow lake problem in which a hypothetical town situated on a lake must develop pollution control strategies to maximize its economic benefits while minimizing the probability of the lake crossing a critical phosphorus (P) threshold, above which it irreversibly transitions into a eutrophic state. Here, we explore the use of direct policy search (DPS) to design robust pollution control rules for the town that account for deeply uncer-

tain system characteristics and conflicting objectives. The closed loop control formulation of DPS improves the quality and robustness of key management tradeoffs, while dramatically reducing the computational complexity of solving the multi-objective pollution control problem relative to open loop control strategies. These insights suggest DPS is a promising tool for managing socio-ecological systems with deeply uncertain tipping points.

3.2 Introduction

As economic development continues globally, severe ecological consequences of human actions are manifesting themselves in many forms. Altered nutrient cycling, shifting biomes, and decreased biodiversity are just a few examples of the repercussions of anthropogenic activities [Parry, 2007]. More responsible socio-ecological management will require balancing conflicting objectives, some of which exhibit uncertain and precarious threshold behavior (e.g., Werners et al. 2013; Keller et al. 2008). For example, we are currently balancing a severe tradeoff between increasing energy production using fossil fuels and avoiding potentially irreversible ecological damages from crossing a threshold atmospheric CO_2 concentration [Solomon et al., 2009]. In fact, Lenton et al. [2008] highlight eight components of the Earth System that could reach catastrophic tipping points as a result of global warming, with the areal extent of Artic summer sea-ice and the Greenland ice sheet facing the most imminent threat.

In environmental systems with thresholds, balancing conflicts in societal values or objectives is further complicated by severe uncertainties associated with identifying thresholds as well as the consequences of crossing them [Lenton,

2013; Keller & McInerney, 2008]. These uncertainties are often considered "deep" or Knightian uncertainties, meaning planners cannot agree on prior probability density functions to describe the parameters of the system model, or even on the model itself [Lempert & Collins, 2007; Knight, 1921]. In these cases, it is desirable to find robust management plans that perform well across a broad range of possible system conditions [Herman et al., 2015; Kwakkel et al., 2016b].

Since its seminal inception [Bankes, 1993; Lempert et al., 2002; Walker et al., 2003; Lempert et al., 2010], the field of decision making under deep uncertainty has emphasized a transition from classical "predict then act" risk management frameworks to exploratory modeling frameworks (e.g., see Dessai et al. 2009). These methods move beyond planning for a single expected future and instead emphasize investigating the response of system management plans to a wide range of deeply uncertain states-of-the-world (SOWs) in order to discover robust actions for avoiding unacceptable outcomes [Bryant & Lempert, 2010; Lempert et al., 2006; Hall et al., 2012]. In their recent review, Herman et al. [2015] highlight the rapid growth in new methodologies and applications of decision analysis frameworks focused on robustness or deep uncertainty, such as robust decision making (RDM) [Lempert et al., 2006], dynamic adaptive policy pathways [Haasnoot et al., 2013], many-objective robust decision making (MORDM) [Kasprzyk et al., 2013], and decision scaling [Brown et al., 2012]. Despite the growing diversity of robustness-focused frameworks, the taxonomy of methods presented by Herman et al. [2015] emphasizes the commonalities between them and the importance of bridging their capabilities to advance the field. These approaches share four core methodological components: (1) eliciting or searching for alternative management actions, (2) using exploratory modeling to broadly

sample possible SOWs that could impact the performance of alternative policies or actions, (3) eliciting robustness measures that distinguish SOWs of concern, and (4) potentially using sensitivity analysis to clarify the key factors that most strongly influence robustness for subsequent monitoring [Herman et al., 2015].

This study advances the MORDM framework [Kasprzyk et al., 2013] with a specific focus on two technical contributions: (1) demonstrating the value and use of direct policy search (DPS) [Rosenstein & Barto, 2001] for identifying adaptive robust operational control strategies for socio-ecological systems and (2) demonstrating how nonlinear environmental thresholds, or tipping points, pose fundamental challenges for balancing economic benefits and their consequent risks to socio-ecological systems. As initially developed by Kasprzyk et al. [2013], the MORDM framework focuses on aiding decision makers and stakeholders in learning how to frame complex, ill-defined environmental planning problems and in discovering robust decisions that perform well across a broad array of possible SOWs. A distinguishing feature of MORDM relative to other frameworks is its use of many-objective evolutionary optimization to identify approximately Pareto optimal management decisions. Pareto optimal, or non-dominated, decisions represent those management actions for which improvement in one objective is only possible with degrading performance in one or more other objectives [Pareto, 1896]. These solutions are first discovered through multi-objective optimization to one's best estimate of the true SOW. The solutions are then re-evaluated in alternative SOWs to determine how robust they are to uncertainties in system parameters. At its core, the MORDM framework provides *a posteriori* decision support, meaning it first presents explicit representations of key system tradeoffs and robustness challenges and then elicits stakeholder preferences in selecting management actions

(i.e., generate first, choose later, as classified by Cohon & Marks [1975]).

In this study, we demonstrate the value of exploiting DPS in the MORDM framework using the classical shallow lake problem [Carpenter et al., 1999]. In this didactic example, a hypothetical town situated on a lake attempts to balance the economic benefits it receives from discharging phosphorus (P) into the lake with the environmental costs of irreversibly tipping the lake into a eutrophic state. The behavior of this stylized model of lake eutrophication is representative of many socio-ecological systems with tipping points, such as harvested fish populations, grasslands consumed by cattle on rangelands, and global carbon cycle dynamics [Carpenter et al., 2015; Anderies et al., 2013]. Early work on the lake problem (e.g., Carpenter et al. 1999; Lempert & Collins 2007) has focused on optimizing the town's pollution control policy to maximize the expected net present value of a utility function which rewards economic benefits and penalizes pollution using a monetary valuation of displaced ecological benefits. Collapsing these objectives into a single expected utility function poses several problems. First, it assumes *a priori* knowledge of stakeholders' values, and agreement among stakeholders on those values. Monetizing environmental benefits and costs to find a single "optimal" solution can fail to capture the full range of achievable objective values, which would better embody the range of preferences among different stakeholders (see, e.g., Admiraal et al. 2013). Second, maximizing the expected value of a utility function requires agreement on the probability distribution of stochastic inputs, which poses severe challenges for systems with deeply uncertain characteristics [Lempert & Collins, 2007; Knight, 1921].

Recent many-objective extensions of the lake problem have sought to explic-

itly capture the tradeoffs between economic and environmental objectives (e.g., Singh et al. 2015; Hadka et al. 2015; Ward et al. 2015), as well as deep uncertainty in the lake model parameters (e.g., Singh et al. 2015; Hadka et al. 2015). Ward et al. [2015] find that when optimizing pollution control strategies for the town as a time series of P release decisions, several state-of-the-art multi-objective evolutionary algorithms (MOEAs) fail to find effective policies due to the high dimensional decision space for candidate pollution control action, weak system responses to late period decisions (i.e., temporal salience structure as discussed by Thierens et al. [1998]), and the non-linear pollution threshold. One way to potentially overcome these challenges is to employ a closed loop control method in which knowledge of the system state is used as a feedback control to inform the decision at each time step [Bertsekas, 1995]. Not only can the additional information provided at each time step improve the signal of the late-period decisions, but it can also allow for a different set of P release decisions under different realizations of stochastic P inflows. The open loop intertemporal pollution control strategy employed by Ward et al. [2015], however, only finds one vector of pollution control decisions that perform best in expectation, and is reflective of the methodologies used in many environmental policy studies (e.g., Nordhaus 2013).

Here, we employ a closed loop control strategy called direct policy search (DPS), that has proven to be a simple and computationally efficient but effective method for solving challenging reservoir control problems [Giuliani et al., 2016b]. In DPS, the parameters of a function mapping the system state(s) to decisions are optimized rather than the decisions themselves. What results from optimization with DPS is therefore not a sequence of decisions, but a policy that one can operate. DPS is also called parameterization-simulation-optimization in

the water resources literature [Koutsoyiannis & Economou, 2003], as it involves parameterizing policies, simulating them, and then optimizing their parameters so that they perform best under simulation. In constrast to stochastic dynamic programming (SDP), a commonly used closed loop stochastic control method, DPS can more easily include multiple state variables in the optimized policies, and simulate them over a range of stochastic inputs during the optimization without building an explicit transition probability model. This drastically reduces the curse of dimensionality and allows direct use of system simulations [Giuliani et al., 2016b], a shortcoming of SDP. Additionally, the DPS simulation model can be coupled with an MOEA to directly optimize multiple objectives. The computational efficiency of DPS makes the method an effective technique for optimizing single and multi-objective control problems for a variety of applications [Guariso et al., 1986; Koutsoyiannis & Economou, 2003; Rosenstein & Barto, 2001].

Here we demonstrate that the built-in control feedback of DPS, whereby the policies use updated knowledge of the system state to inform decisions, also results in control rules that are robust to deep uncertainties in the system characteristics. Carpenter et al. [1999] show that even in lakes with long records of nutrient inputs and concentrations, major uncertainties remain in the values of parameters describing nutrient dynamics in the lake. Furthermore, uncontrolled nutrient inputs from non-point sources are likely to change with deeply uncertain changes in land use. For these reasons, developing adaptive lake management strategies that are robust to deep uncertainty is a vital objective itself. Given the similarity of the lake problem to many socio-ecological systems, DPS holds promise as an effective method for developing robust multi-objective management plans across a variety of problems. This study demonstrates DPS

as valuable method for advancing the field of decision making under deep uncertainty, especially when confronting socio-ecological systems with severe tipping points.

3.3 Lake Problem Description

In this study we consider a pollution control problem in which a town must develop a P emissions strategy that balances its economic benefits and the quality of the lake. We model the P dynamics in the lake using a theoretical model of lake nutrient dynamics developed by Carpenter et al. [1999] in which water quality can transition between two states: (i) an oligotrophic equilibrium and (ii) a eutrophic equilibrium. Transitions between these states have been observed in many shallow lakes, such as Tomahawk Lagoon in New Zealand and Lakes Takern and Krakesjon in Sweden [Scheffer et al., 1993]. In this study, we adapt a dimensionless version of the Carpenter et al. [1999] model used to abstract phosphorus (P) dynamics in Lake Mendota, Wisconsin:

$$X_{t+1} = X_t + a_t + Y_t + \frac{X_t^q}{1 + X_t^q} - bX_t, \tag{3.1}$$

where

X is the normalized concentration of P in the lake,

a are anthropogenic (controlled, point source) P inputs,

$Y \sim LN(\mu,\sigma^2)$ are natural (uncontrolled, non-point source) P inputs modeled as coming from a log-normal distribution with mean μ and variance σ^2,

q is a parameter controlling the rate at which P is recycled from the sediment,

b is a parameter controlling the rate at which P is lost from the lake, and

$t \in \{0,1,2,...\}$ is the time index in years.

The term $\frac{X^q}{1+X^q}$ represents the inputs of P recycled from the sediment as a function of the current P level, X. These inputs follow a sigmoid curve whose maximum steepness is controlled by q. Larger values of q result in steeper curves and therefore more sudden changes in recycling rates as a function of lake P concentration. The term bX represents the losses of P through, for example, outflow or sediment adsorption. Losses are assumed to be a linear function of the lake P concentration, with larger values of b resulting in greater losses [Carpenter et al., 1999].

A graphical depiction of the lake model for different values of b and q is shown in Figure 3.1, where the temporal fluxes of P into and out of the lake are plotted as a function of the current concentration of P in the lake, X_t. The black line in Figure 3.1(a) shows the rate of losses for $b = 0.4$, while the colored curves show the rate of P inputs from recycling under different values of q. In Figure 3.1(b), the black curve shows inputs from recycled P when $q = 2.5$, while the colored lines show the losses of P for different values of b. Inputs from anthropogenic and natural P are assumed to be 0 in this figure; adding them would vertically translate the recycling curves.

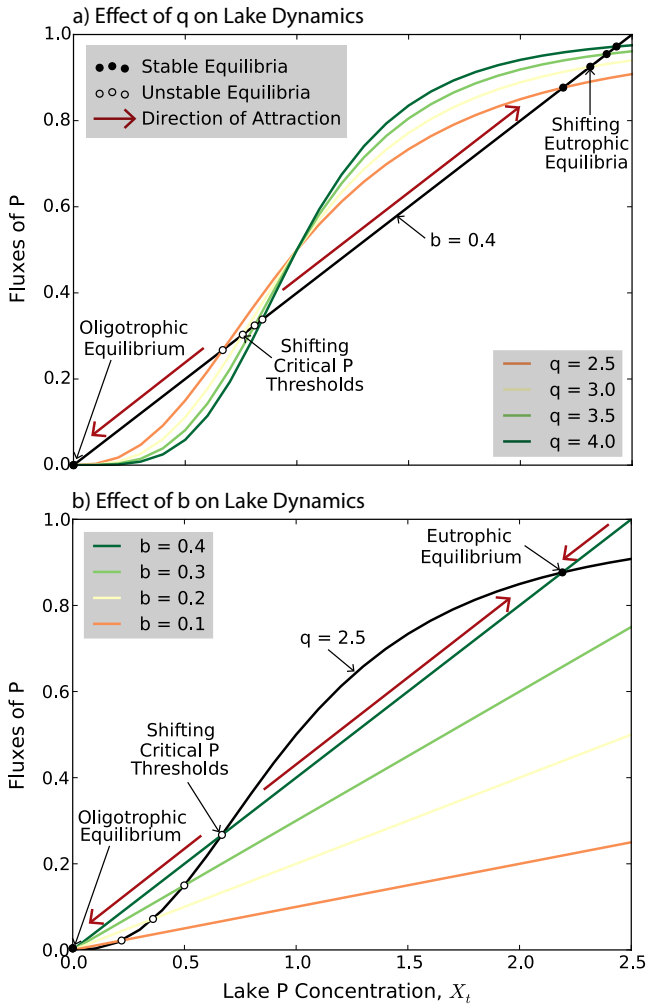

Figure 3.1: Non-linear dynamics of the irreversible lake model. Curves indicate inputs of P recycled from the sediment and lines indicate P sinks. The system is in equilibrium when the P inputs and sinks are equal, which occurs at three locations for each combination of inputs and sinks functions. Two of these intersections are stable equilibria, and the third in the middle is unstable. The unstable equilibrium is referred to as the "critical P threshold," as crossing it moves the system to an irreversibly eutrophic state for the parameters of b and q investigated in this study. Panel (a) shows the effect of q on the location of the critical P threshold, while panel (b) shows the effect of b. Higher values of both b and q result in higher critical P thresholds, making it easier to avoid moving the system from a stable oligotrophic equilibrium to a stable eutrophic equilibrium.

Note that for the alternate values of b and q plotted in Figures 3.1(a) and (b), each of the curves representing inputs of recycled P and the lines representing outputs from P sinks intersect at three different P concentrations. At each of the points of intersection, the inputs are equal to the outputs and the net flux of P is 0, meaning the lake's P concentration is in equilibrium. When the sources are greater than the sinks, the P concentration tends to increase, and when the sinks are greater, it tends to decrease. These dynamics result in two stable equilibria shown in black that are separated by an unstable equilibrium shown in white. The stable equilibrium that occurs at a low P concentration represents a healthy, oligotrophic equilibrium, while that at a high P concentration represents an unhealthy, eutrophic equilibrium.

If the concentration of P in the lake exceeds the unstable equilibrium in any of the cases plotted above, it will be impossible for the lake to return to an oligotrophic equilibrium unless P is actively removed from the system. Reducing P inputs alone will not be sufficient, as this figure already shows the system with 0 anthropogenic and natural P inputs. These sample lake model parameterizations consequently possess "irreversible" P thresholds, or tipping points, that pose severe management challenges. Such lakes tend to be shallow, located in P-rich regions, or receive sustained high P inputs [Carpenter et al., 1999]. Because crossing the unstable equilibrium results in a permanently polluted lake, we call the P concentration at this point the "critical P threshold."

As Figure 3.1 shows, different lake parameterizations can have profound implications for how one manages the system, as the value of the critical P threshold can vary greatly for different values of b and q. For a given loss rate, b, higher values of q result in higher critical P thresholds, allowing managers to

emit more P before the lake enters a permanently polluted state. However, the slope of the recycling curve is steeper for larger values of q, making the transition across this threshold more abrupt. For a given value of q, the same is true; larger values of b result in higher critical P thresholds but often with a faster transition across them.

Furthermore, the parameters describing the lake's dynamics can be difficult to estimate from empirical data. Carpenter et al. [1999] try to do so with 21 years of data on P input rate and mass in Lake Mendota. After generating point estimates of the model parameters, they generate bootstrapped parameter estimates by adding back randomly selected residuals and find that reasonable estimates for the characteristics of the lake dynamics span the three main possibilities of reversible, hysteretic, and irreversible. This uncertainty severely challenges the development of effective management strategies. Even in modeled settings, reliable estimates of model parameters often cannot be made until a threshold has already been crossed (see, e.g., Lempert & Collins 2007), at which point it is too late to inform management if the system is irreversible.

Here, we investigate how to formulate effective management policies for a system with assumed values of model parameters related to lake dynamics and economic discounting, but then test how robust these policies are to parametric uncertainty in the case that the assumed values are incorrect, applying a robustness method called Multi-Objective Robust Decision Making (MORDM; Kasprzyk 2013). While more robust solutions may potentially be discovered by initially optimizing to the deeply uncertain SOWs, this improved robustness may result in large regrets if the system never confronts the most extreme SOWs. In MORDM, the sampled ranges of uncertain parameters are subjectively cho-

sen to be wide enough to discover failures and it could be that some of the sampled SOWs are extremely unlikely. Following Kasprzyk [2013], we implement MORDM to establish an initial baseline of performance of alternative management plans. In discovering robust baseline solutions, we implicitly minimize regret to an assumed SOW representing the best current state of knowledge. Once key uncertainties driving system performance are discovered, more robust policies can be developed by monitoring these uncertainties and adapting the policies as more information becomes available.

3.4 Optimization Problem

3.4.1 Lake Management Objectives

The general challenge we consider in this study is how much anthropogenic P, a_t, to emit over time in order to simultaneously optimize multiple, conflicting objectives with respect to environmental and economic goals. Because the objectives are conflicting, tradeoffs exist between different lake management plans. We find the set of "non-dominated" management plans, meaning that among the solutions in the set, no solution does better than another in all objectives. This set is called the Pareto set.

Following Ward et al. [2015], we consider four objectives listed in Table 3.1 and described below.

1. **Maximize expected economic benefits:** In each of N simulations of T years of random natural P inflows, the discounted economic benefits are

Table 3.1: Performance objectives of lake pollution control problem

Objective	Description	Min/Max	Epsilon
Economic Benefits	Discounted economic benefits, assumed proportional to discounted P emissions	Max	0.01
Lake P Concentration	Measure of the lake quality, where lower P concentrations correspond to clearer lakes	Min	0.01
Policy Inertia	Measure of the stability of the control policy, where stable policies are favored	Max	0.0001
Reliability	Percentage of time that the lake P concentration is below the critical P threshold, X^{crit}	Max	0.0001

calculated as $\sum_{t=0}^{T-1} \alpha a_{t,i} \delta^t$, where α is a dimensionless parameter representing the town's willingness to pay for pollution, δ is the discount factor used to convert future benefits to present benefits, and $a_{t,i}$ is the anthropogenic P release in the t^{th} year of the i^{th} simulation. The expected economic benefits, O_1, are the average economic benefits across the N simulations of T years:

$$O_1 = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=0}^{T-1}\alpha a_{t,i}\delta^t\right). \tag{3.2}$$

As in Ward et al. [2015], we use $\alpha = 0.4$ and $\delta = 0.98$.

2. Minimize worst case average P concentration: In each of T years, the average P concentration across N simulations is calculated. For water quality purposes, the lake manager seeks to minimize the maximium of these T averages, O_2:

$$O_2 = \max_{t\in(1,\ldots,T)} \frac{1}{N}\sum_{i=1}^{N} X_{t,i}, \tag{3.3}$$

where $X_{t,i}$ is the lake P concentration in the t^{th} year of the i^{th} simulation.

3. Maximize average inertia of P control policy: Since rapid P reductions require large investments in infrastructure such as tertiary treatment, the lake manager would like to best maintain policy inertia, or stability. Here inertia, O_3, is defined as the average fraction of $T-1$ time steps across N simulations that require reductions of less than some limit, I_{limit}:

$$O_3 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T-1}\sum_{t=1}^{T-1}\phi_{t,i}\right) \text{where } \phi_{t,i} = \begin{cases} 1, & a_{t-1,i} - a_{t,i} < I_{limit} \\ 0, & a_{t-1,i} - a_{t,i} \geq I_{limit} \end{cases}. \tag{3.4}$$

As in Ward et al. [2015], we use a value of 0.02 for I_{limit}. In a real system, the value of I_{limit} should be elicited from stakeholders, who would have to evaluate the costs and benefits associated with these reductions.

4. Reliability of policy in staying below the lake's critical P threshold: Because the lake is irreversible and crossing its critical P threshold, X^{crit}, results in permanent eutrophication, the lake manager seeks to maximize the average fraction of time the lake is below this threshold. This is defined as the policy's reliability, O_4:

$$O_4 = \frac{1}{NT}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\theta_{t,i}\right) \text{where } \theta_{t,i} = \begin{cases} 1, & X_{t,i} < X^{crit} \\ 0, & X_{t,i} \geq X^{crit} \end{cases}. \tag{3.5}$$

The reliability is constrained to be at least 85%. It was challenging for several MOEAs to find solutions meeting this constraint in the prior work by Ward et al. [2015].

In this study, we compare two solution strategies for optimizing multi-objective lake management control policies: 1) open-loop control via intertemporal optimization and 2) closed loop control via direct policy search.

3.4.2 Solution Strategies

Intertemporal Open Loop Control

Following prior work on multi-objective optimization of the lake problem Singh et al. [2015]; Hadka et al. [2015]; Ward et al. [2015] and common practice in the integrated assessment literature (e.g., Nordhaus [2013]), we use an open loop intertemporal solution strategy in which there are T decision variables, a_t, representing the anthropogenic discharges at each time step $t \in (0, 1, ..., T-1)$.

Drawing on the objectives defined in equations (3.2)–(3.5), the intertemporal optimization problem is thus defined as:

$$\text{Minimize } F(a) = (-O_1, O_2, -O_3, -O_4) \tag{3.6}$$

$$a = (a_0, a_1, ..., a_{T-1}) \tag{3.7}$$

Subject to:

$$O_4 \geq 0.85 \tag{3.8}$$

$$0.01 \leq a_t \leq 0.1 \ \forall t \tag{3.9}$$

Under this solution strategy, T anthropogenic P emissions, a_t, are optimized to maximize expected economic benefits, minimize maximum average lake P concentration, maximize inertia, and maximize reliability. The same T releases are made in each of the N simulations, such that $a_{t,i} = a_t \ \forall i$. The only constraints are that the reliability be at least 85% (Equation 3.8), and that P emissions range between 0.01 and 0.1 so that economic activity is relatively stable (Equation 3.9).

Direct Policy Search Closed Loop Control

The second solution strategy, termed direct policy search (DPS) [Rosenstein & Barto, 2001], is a closed loop control method that does not optimize the anthropogenic P releases, a_t, themselves, but rather the parameters of a state-aware control rule (see Ch 2 Section 2.2.1). The present study builds on prior work with DPS in the water resources literature [Guariso et al., 1986; Koutsoyiannis & Economou, 2003] by introducing the method into the socio-ecological tipping points literature [Singh et al., 2015; Carpenter et al., 2015; Ward et al., 2015]

and demonstrating the value of its state-based feedback control for navigating deeply uncertain threshold management tradeoffs [Singh et al., 2015; Hadka et al., 2015].

As applied to the lake problem, candidate DPS control rules are represented as functions mapping the current state of the lake's P concentration, X_t, to a P release decision, a_t. Figure 3.2 illustrates an example of such a control policy function. At low P concentrations, one would expect to be able to release high amounts of P without worrying about crossing the critical P threshold. However, as one approaches this threshold (shown in red), the P release should decrease to ensure high reliability. If the P concentration is above the critical P threshold, however, reliability can no longer be improved since the lake is irreversible, so it may be best to increase P releases again to maximize further economic benefits.

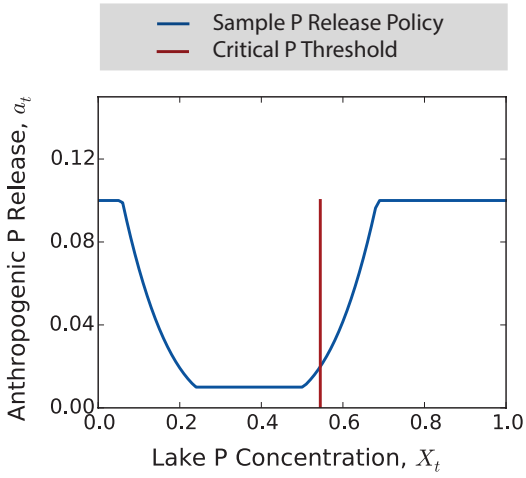

Figure 3.2: Sample P release policy. The blue curve indicates the amount of anthropogenic P, a_t, released as a function of the current lake P concentration, X_t. The red line is the critical P threshold, X^{crit}, above which the system is irreversibly eutrophic.

In this study, we use cubic radial basis functions to define the function map-

ping lake P concentrations to P release decisions, as many studies have shown them to be effective universal approximators for DPS, generalizing well when operating a system on out-of-sample inputs (e.g., Buşoniu et al. [2011]; Giuliani et al. [2016b]). This optimization problem is formulated as:

$$\text{Minimize } F(a) = (-O_1, O_2, -O_3, -O_4) \tag{3.10}$$

$$a = (a_{0,i}, a_{1,i}, ..., a_{T-1,i}) \tag{3.11}$$

$$a_{t,i} = \min\left(\max\left(\sum_{j=1}^{n} w_j \left|\frac{X_{t,i} - c_j}{r_j}\right|^3, 0.01\right), 0.1\right) \forall t, i \tag{3.12}$$

Subject to:

$$O_4 \geq 0.85 \tag{3.13}$$

$$-2 \leq c_j \leq 2 \tag{3.14}$$

$$0 \leq r_j \leq 2 \tag{3.15}$$

$$0 \leq w_j \leq 1 \tag{3.16}$$

$$\sum_{j=1}^{n} w_j = 1, \tag{3.17}$$

where c_j, r_j and w_j are the centers, radii and weights of n cubic radial basis functions. The decision variables are these $3n$ parameters, rather than the T decision variables a_t in the intertemporal optimization. For this problem, we use $n = 2$, resulting in 6 decision variables compared to $T = 100$ decision variables using the intertemporal solution strategy. Notice that with this solution strategy, different P release decisions, $a_{t,i}$, can be made in each of the N simulations, as the decisions are informed by the different resulting lake P concentrations, $X_{t,i}$.

DPS has proven to be a powerful and computationally efficient method for solving many-objective stochastic control problems [Giuliani et al., 2016b]. Un-

like the open loop intertemporal strategy, DPS is able to incorporate information about the lake's state in each stochastic scenario to guide more adaptive and responsive P release decisions. Additionally, one could easily incorporate more information into the DPS policies. For example, the above formulation could include the lake P concentrations in both the current and previous time steps. For simplicity, we demonstrate the method with just a single input. Note that this simple formulation assumes that one can perfectly measure the lake P concentration before each release decision and that the presence of an irreversible tipping point is known. In future extensions, we will explore more advanced data assimilation strategies that can account for observational error and more complex learning (e.g., about the model parameters).

3.4.3 Optimization Algorithm

We use the intertemporal and DPS solution strategies to solve the multi-objective formulation of the lake problem for $T = 100$ years and $N = 100$ realizations of randomly generated natural P inflows, initializing the lake concentration at $X_0 = 0$. Values of the lake model parameters are given in Table 3.2. The slope of the losses curve, b, is set to 0.42, and the shape parameter of the recycling curve, q is set to 2.0. These parameters correspond to a lake with a critical P threshold, X^{crit}, of 0.54. Natural P inflows are simulated from the same distribution as in Scenario 2 from Ward et al. [2015]: a log-normal distribution with a real-space mean of $\mu = 0.03$ and variance of $\sigma^2 = 10^{-5}$. Ward et al. [2015] showed that this scenario poses a severe challenge to six state-of-the-art MOEAs in finding Pareto-approximate lake management strategies. Of these six MOEAs, only the Borg MOEA (Hadka & Reed [2013]; see Ch 2 Section 2.1.1)

was able to consistently find high quality tradeoff solutions. Ward et al. [2015] attributed Borg's success to its use of ε-dominance archiving, ε-progress and multiple self-adaptive search operators. For these reasons, we also use the Borg MOEA as our solver of choice.

3.4.4 Evaluation of Solution Strategies

We use the hypervolume metric to compare the intertemporal and DPS solution strategies in terms of algorithmic efficiency and effectiveness in attaining high quality approximations of the problem's Pareto frontier. Hypervolume is a measure of the multi-dimensional "volume" of space dominated by a set of Pareto-approximate solutions [Zitzler et al., 2003]. One can calculate the hypervolume of the best-known approximation of the Pareto set, or "reference set," after different numbers of function evaluations (NFE) throughout the search process and observe how quickly the different solution strategies are able to find good solutions. Additionally, the hypervolume metric is useful for comparing the final performance of different solution strategies. Here, we use it to distinguish and compare the relative quality of the best-known Pareto sets of the intertemporal and DPS solution strategies.

Figure 3.3 illustrates how hypervolume (HV) can be used for such a comparison. This figure shows sample reference sets from two solution strategies, A and B, on a two-objective minimization problem. The HV of each reference set is therefore an area calculated with respect to a reference point. One can see that two solutions in reference set B dominate one solution in reference set A, while the third solution in reference set B is dominated by three points in reference set

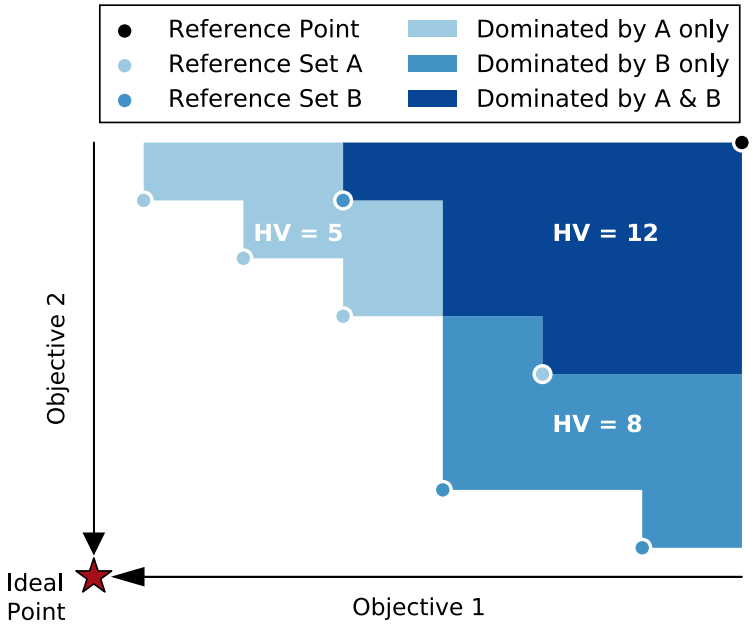

Figure 3.3: Sample hypervolume comparison. Light blue points represent solutions found by method A, while dark blue points represent solutions found by method B. The solutions found by method B dominate a greater area, or hypervolume, indicating that this method performs better than method A.

A. Looking at the hypervolume metric, though, one can see that the reference set from method B dominates a larger space (HV=8+12=20) than the reference set from method A (HV=5+12=17), indicating that method B is able to find a better approximation to the true Pareto front and is therefore the preferred method.

3.5 Computational Experiment

3.5.1 Random Seed Analysis

Since the Borg MOEA is a stochastic optimization algorithm whose search depends on the random seed used to initialize the population and generate new solutions, we solve the lake problem using 50 different random seed runs with Borg's default initial parameter values [Hadka & Reed, 2013]. Performing a random seed analysis allows one to control for variability and evaluate the consistency in search performance. In this study, we set the NFE for each seed to 200,000 to provide sufficient time for the Borg MOEA to attain high quality approximations of the tradeoffs using both the intertemporal and DPS solution strategies (see Figure 3.5 in Section 3.6.1).

To evaluate the search efficiency of each solution strategy, we calculate the hypervolume of the archive every 1,000 function evaluations for each seed. We then normalize these values by dividing by the best final hypervolume achieved across all seeds and solution strategies. To compare the final performance of each strategy, we find a reference set from each method by aggregating the terminal set of solutions found by each seed from that method and re-sorting them into one Pareto-approximate set per method. We plot these two reference sets on the same axes to visualize the amount of space dominated by the solutions found from each method (see Figure 3.4 in Section 3.6.1).

The results of the computational experiment are given in Section 3.6.1, and the quality of the final solutions analyzed in Section 3.6.2.

3.5.2 Robustness Analysis

As discussed in our introduction of the lake problem's dynamics in Section 3.3, it is an important and realistic concern that assumptions related to the lake model's parameters could fundamentally change the performance of alternative pollution control strategies. Many systems like managed lakes are subject to deep parametric uncertainty, meaning decision makers do not know, or cannot agree on, prior probability distributions to describe the potential values of the systems' parameters [Knight, 1921; Lempert et al., 2002; Kwakkel et al., 2016b]. In such cases, it is desirable to find "robust" policies that perform well under a broad range of potential system characteristics. Here, we seek to compare the performance of the lake management plans derived by each method under the assumed lake characteristics to which they were optimized, as well as under other deeply uncertain states of the world (SOWs) to which they may actually be applied.

Building on initial work by Kasprzyk et al. [2013] and several recent extensions [Herman et al., 2014; Singh et al., 2015; Hadka et al., 2015], we use the many-objective robust decision making (MORDM) framework to evaluate the robustness of the control policies derived from the intertemporal and DPS solution strategies. As mentioned in the introduction, MORDM, like several other methods used to assess robustness, follows four key steps outlined by Herman et al. [2015]. The first step is to generate alternative management plans for the robustness analysis. Some methods pre-specify such plans, while MORDM searches for them using multi-objective optimization. In this study, we generate alternative lake management plans using multi-objective optimization with the intertemporal and DPS solution strategies.

The next step in any robustness analysis is to re-evaluate the management alternatives across a broad range of SOWs generated by globally sampling the deeply uncertain factors. The third step is to quantify how well the different management plans generalize across these uncertain SOWs. There are many different measures of robustness used in the literature. Herman et al. [2015] highlight four: two satisficing metrics [Starr, 1962; Hine & Hall, 2010] and two regret metrics [Lempert & Collins, 2007; Savage, 1951]. In comparing these four metrics on a water portfolio planning problem, Herman et al. [2015] find that the domain criterion satisficing measure [Starr, 1962], which quantifies the fraction of SOWs in which desired performance levels are met, is the only metric that prioritizes solutions meeting the stakeholders' minimum performance criteria. For this reason, we also use the domain criterion satisficing metric to quantify robustness in this study. However, in a real system we recommend calculating a variety of robustness metrics considering a range of thresholds of acceptability, as stakeholders may favor different metrics or thresholds based on their risk attitudes. Furthermore, the choice of metric or threshold level may influence which solutions are considered most robust, introducing more conflicts that stakeholders should consider in evaluating alternative management plans [Herman et al., 2015]. On the contrary, different metrics or thresholds may ease tension if similar conclusions are drawn despite being derived under different assumptions of what it means to be robust [Hall et al., 2012]. In both cases, the additional robustness analysis can aid stakeholders in coming to consensus on candidate actions even if their conclusions emerge for different reasons.

Regardless of which metrics are ultimately used to quantify robustness, the final step employed by several robustness analysis methods including MORDM is a sensitivity analysis to determine which parameters are most important for

ensuring acceptable system performance. This is also referred to as scenario discovery, and can be achieved through factor mapping [Bryant & Lempert, 2010; Friedman & Fisher, 1999], or factor prioritization [Sobol, 2001; Saltelli, 2002]. Factor mapping techniques identify parameter ranges or combinations of parameter ranges that lead to poor performance, while factor prioritization ranks the uncertain parameters by importance to system performance. Ideally, managers could use the results of such sensitivity analyses to focus their attention on controlling or monitoring the key parameters driving system performance.

Evaluation of Alternatives

For this study, we generate 1,000 alternative SOWs from a Latin hypercube sample across the ranges of all uncertain parameters. All parameter ranges are sampled uniformly. We assume the same deeply uncertain parameters and ranges used by Hadka et al. [2015] for a robustness analysis of the same four-objective version of the lake problem, considering only irreversible lakes. We also add one uncertain parameter: the initial P concentration in the lake, X_0. Table 3.2 lists the uncertain parameters, their base values in the SOW to which the policies are optimized and the ranges over which they are sampled for the robustness analysis. The results of the re-evaluation are given in Section 3.6.3.

Calculation of Robustness

As stated earlier, we quantify the robustness of each lake management plan in terms of the fraction of sampled SOWs in which certain performance criteria are met. For real-world decision-support applications, these criteria are typically

Table 3.2: Ranges of sampled uncertain parameters and their base values. Lake management plans are optimized to a lake modeled by the base values and re-evaluated on 1,000 alternative SOWs generated from a Latin hypercube sample across the parameter ranges given in the table. All parameter ranges are sampled uniformly.

Parameter	Description	Base Value	Minimum	Maximum
b	Linear P loss parameter	0.42	0.1	0.45
q	Shape parameter of sigmoid curve modeling P recycling	2.0	2.0	4.5
μ	Mean natural P inflow	0.03	0.01	0.05
σ^2	Real-space variance of natural P inflow	10^{-5}	$(0.001)^2$	$(0.005)^2$
X_0	Initial lake P concentration	0.0	0.0	0.3
δ	Discount parameter for calculating economic benefits	0.98	0.93	0.99

elicited from stakeholders (see e.g., Lempert & Groves 2010; Moody & Brown 2012; Herman et al. 2014). For illustrative purposes on this stylized model, we simply specify three possible sets of criteria a stakeholder might demand:

1. Average Economic Benefits > 0.2
2. Average Reliability > 95%
3. Average Economic Benefits > 0.2 and Average Reliability > 95%

For each solution found by DPS and the intertemporal optimization, we calculate the percent of sampled SOWs in which each of the above criteria are met. The results of these robustness calculations are given in Section 3.6.3.

Sensitivity Analysis

The final step in MORDM is to use scenario discovery [Bryant & Lempert, 2010] to determine what ranges of model parameters lead to poor performance. Two commonly used factor mapping approaches include the Patient Rule Induction Method, or PRIM [Friedman & Fisher, 1999], and Classification and Regression Trees, or CART [Breiman et al., 1984]. Both of these methods assume poor performance can be explained by independent combinations of input parameter ranges, and therefore fail to capture non-linear interactions between model parameters that can lead to unacceptable performance. Recent approaches have improved upon PRIM by first performing Principal Components Analysis (PCA) and then applying PRIM to the rotated datasets [Dalal et al., 2013]. However, this modification only addresses the orthogonality constraint of PRIM, not its linearity constraint. For this reason, we perform a visual factor mapping sensitivity analysis to identify which joint ranges of lake model parameters cause particular management plans to fail. This approach enables a better treatment of the nonlinear nature of the lake model's irreversible tipping point. The results of this analysis are given in Section 3.6.3.

3.6 Results and Discussion

3.6.1 Tradeoffs and Problem Difficulty

The results of our computational experiment indicate that DPS is able to find more dominant lake management policies than the intertemporal method. Figure 3.4 shows the final reference sets found using both solution strategies. Here, we obtain the solutions in each method's reference set by combining the results from the 50 random seed runs of the Borg MOEA (see Section 3.5.1). DPS found 260 non-dominated solutions and the intertemporal optimization found 86. In Figure 3.4, the arrows designate the directions of increasing preference for the three primary axes. Larger points designate higher levels of reliability. The star designates the ideal solution, thus larger points closer to the star are favorable. The DPS solutions in blue dominate many of the intertemporal solutions in red, as they perform better in every objective. That is, the DPS solutions are able to achieve greater economic benefits at much lower maximum average P concentrations while maintaining high levels of reliability and inertia. Figure 3.4 clearly highlights that the final hypervolume of the management plans found using DPS exceeds that of the management plans found using the intertemporal approach.

In addition to finding better final solutions, the DPS method is able to discover these solutions much faster. Figure 3.5 shows the relative computational efficiency of each method by plotting the relative hypervolume achieved by each seed versus the number of function evaluations in the search. Relative hypervolume is calculated as a fraction of the space dominated by the best reference set achieved by a single seed. The figure shows that all of the DPS seeds

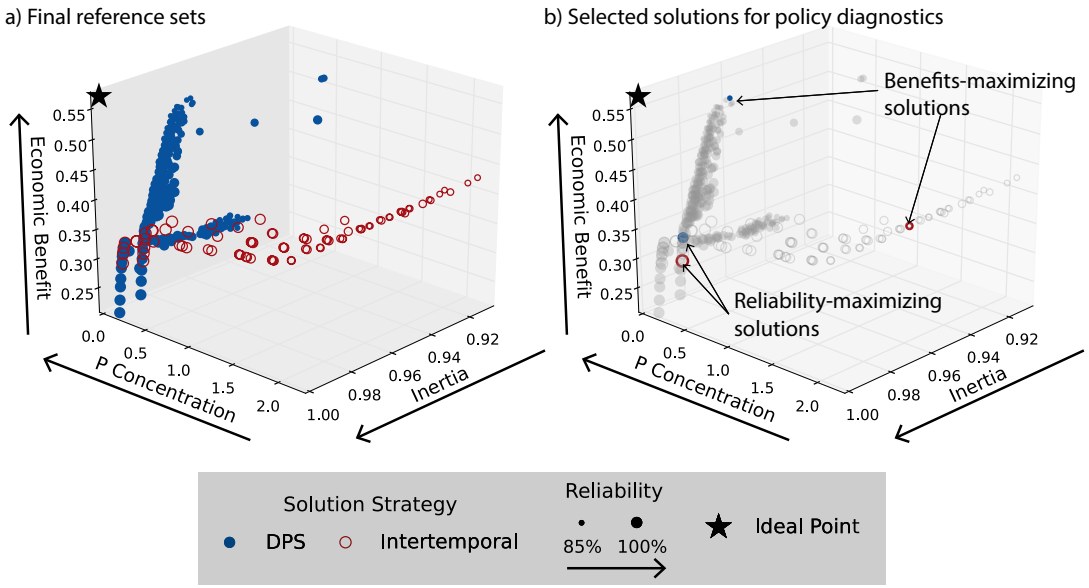

Figure 3.4: Pareto-approximate reference sets obtained using the intertemporal (red) and DPS (blue) solution strategies. Arrows indicate direction of preference along each axis. The size of the points represents their reliability, with larger circles preferred. Panel (a) shows that most of the DPS solutions dominate the intertemporal solutions, performing better in every objective. Panel (b) highlights the solutions from each method with the greatest reliability and economic benefits. The decisions resulting from these solutions are examined in further analyses.

in blue converge to near-perfect relative hypervolume. Additionally, these solutions are found very quickly, some within only a few thousand function evaluations and all before 100,000. Using the intertemporal solution strategy, however, the Borg MOEA is unable to generate any hypervolume until about 20,000 evaluations, and while most of the seeds also converge by 100,000 evaluations, they do so to a much lower relative hypervolume (roughly half) with a greater spread across seeds. While it may not be surprising that the DPS solution strategy converges faster than the intertemporal solution strategy since it has far less decision variables, it is noteworthy how much faster it converges, and to what degree it improves upon the attainable system performance. DPS converges so

fast that even the initial population of 100 solutions generated from a uniform random sample by the Borg MOEA often includes Pareto-approximate solutions. It was not obvious *a priori* that such a simple policy rule with only six decision variables would be able to adequately model such effective policies.

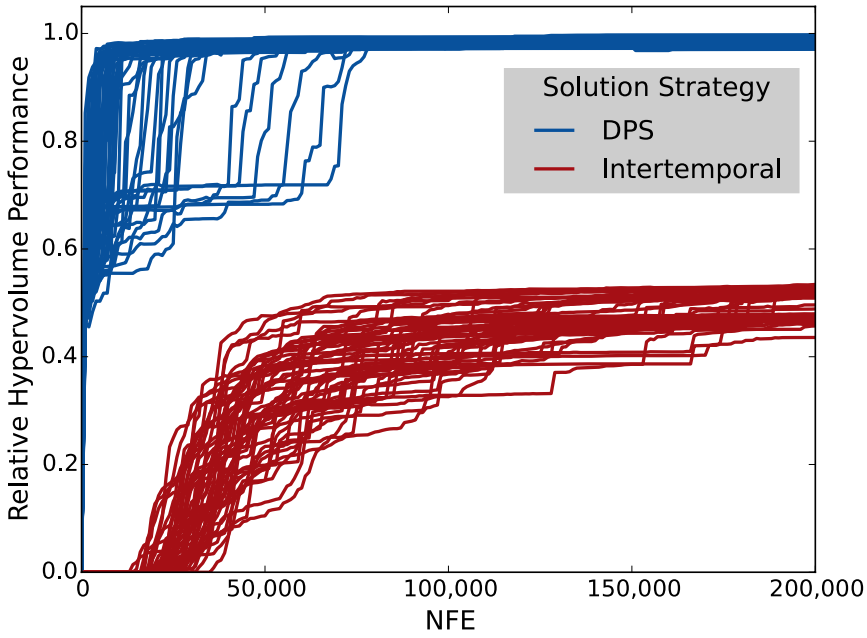

Figure 3.5: Search convergence using intertemporal (red) vs. DPS (blue) solution strategies. Each line depicts the relative hypervolume vs. the number of function evaluations (NFE) for a different random seed of the Borg MOEA, where hypervolume is given as a fraction of the best-known Pareto-approximation set. Greater hypervolume values are preferred. Across all seeds, the DPS solution strategy is able to generate much greater hypervolume significantly faster.

The difficulty of solving the intertemporal version of the lake problem illustrated in Figure 3.5 is consistent with the prior published MOEA benchmarking results by Ward et al. [2015]. Two core challenges and limitations explain the slower convergence and reduced hypervolume performance of the intertemporal solution strategy relative to DPS: temporal salience structure and open loop

control. The temporal salience structure of the lake problem refers to the differential effects of P release decisions over time [Thierens et al., 1998], with earlier decisions having a greater influence on the economic benefits and reliability of the lake management plans due to the lake's irreversible dynamics and the discounting of economic benefits. Since the intertemporal solution strategy attempts to optimize 100 decision variables representing the P release decisions in each control period, the temporal salience structure of the lake problem results in sequential convergence of the release decisions from early to late stages, termed domino convergence [Thierens et al., 1998]. Consequently, the least important decisions at the end of time do not strongly influence the hypervolume progress of the Borg MOEA, making it difficult for the algorithm's selection operators to optimize the P releases for these time steps. This results in slow convergence of the late-term decisions, termed drift stall [Thierens et al., 1998].

The open loop control of the intertemporal strategy limits its final hypervolume performance due to its lack of feedback control compared to the closed loop DPS approach. Even though there are only 6 decision variables defining each DPS-derived policy function, when operated, the policy's feedback control enables different P release decisions in each realization of 100 years of random natural P inflows, as the release decisions respond to the different resulting trajectories of lake P concentrations. The open loop intertemporal strategy, however, finds only one set of 100 release decisions that must perform well on average across the 100 realizations of natural P inflows.

3.6.2 Performance Comparison of the Different Control Strategies

To analyze the control policies derived from the intertemporal and DPS solution strategies, we highlight a few solutions for further investigation. For each method, we select two solutions: that maximizing reliability and that maximizing economic benefits. Figure 3.4(b) shows the locations of these solutions in the Pareto-approximate sets and Table 3.3 summarizes their objective values.

Table 3.3: Objective values of selected solutions from the intertemporal and DPS solution strategies. Low values are preferred for O_2, while high values are preferred for the other three.

Method	Solution	O_1: Expected Economic Benefits	O_2: Maximum Average P Concentration	O_3: Average Inertia	O_4: Average Reliability
Intertemporal	Highest Reliability	0.32	0.21	0.99	1.0
	Highest Benefits	0.43	2.25	0.97	0.86
DPS	Highest Reliability	0.34	0.22	0.99	1.0
	Highest Benefits	0.57	0.57	0.98	0.85

Policies vs. Intertemporal Pathways

Figure 3.6(a) shows the shapes of the policies corresponding to the two DPS solutions, with the most reliable solution in blue and the benefits-maximizing solution in green. Also shown in red is the critical P threshold beyond which the lake is irreversibly eutrophic. As one would expect, both solutions map low P concentrations, X_t, to high P releases, a_t, and prescribe decreased releases as the lake P concentration increases. The most reliable policy is much more conservative, decreasing P discharges at much lower lake P concentrations than the benefits-maximizing solution. Both policies subsequently increase P discharges beyond some lake P concentration, with the benefits-maximizing solution doing so at much lower values. This is because tipping becomes inevitable as the lake P concentration approaches the critical threshold, and once the threshold is crossed, reliability can no longer be improved, so it is best to further maximize economic benefits. As is shown later, though, these policies are able to find oligotrophic equilibria below the critical P threshold such that the increasing region of the policy is rarely operated, especially under the most reliable policy.

Figures 3.6(b) and (c) compare the two selected DPS policies with simulations of the corresponding intertemporal control strategies. Given that the intertemporal strategies are optimized release decisions at specific times, the time series of releases must be simulated to infer the corresponding lake P concentration when each release is made. For this reason, the decisions do not form a continuous function and have been plotted as points. As can be seen in Figure 3.6(b), the most reliable intertemporal policy only observes a small range of lake P concentrations since it is able to find an oligotrophic equilibrium about which the concentration fluctuates. If a manager wants to know how to oper-

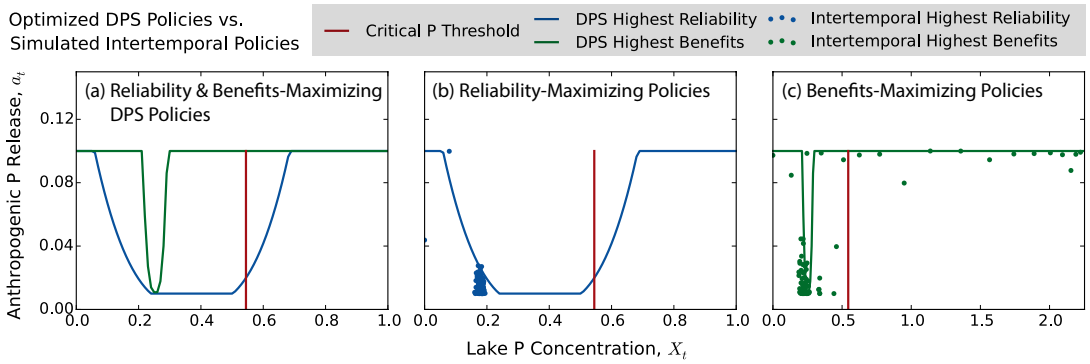

Figure 3.6: Optimized DPS policies vs. simulated intertemporal policies, where P release is plotted against the lake P concentration. Panel (a) compares the reliability-maximizing (blue) and benefits-maximizing (green) DPS policies, panel (b) compares the reliability-maximizing solutions from DPS (solid line) and the intertemporal solution strategy (points), and panel (c) compares the benefits-maximizing solutions from the two methods. The DPS policies are obtained directly from optimization, while proxy policies are obtained for the intertemporal policies by simulating them over their time horizon to determine the lake P concentration at the time of each P release. Consequently, the intertemporal policies are not continuous.

ate this solution beyond the simulated horizon, though, the simulated policy is not a helpful decision tool due to its limited domain of applicability and the somewhat random corresponding decisions within that domain. The benefits-maximizing policy forms a more complete policy, although the randomness of this solution's decisions is also evident. The randomness in the intertemporal release decisions of both policies is largely driven by the large number of decision variables in the optimization, and the low sensitivity to later term decisions resulting in algorithmic drift stall [Thierens et al., 1998], as discussed in Section 3.6.1.

Time Series Comparison

In addition to viewing the selected solutions in the policy space, it is informative to compare the time series of actual release decisions and corresponding lake P concentrations brought about by operating these policies. These time series are shown in Figure 3.7. As in Figure 3.6, the most reliable solutions are shown in blue, the benefits-maximizing in green, and the critical P threshold in red. The DPS solutions are represented by solid lines and the intertemporal solutions by dashed lines.

In the first row of this figure, the lake P concentration is plotted vs. time and one can see that all of the selected solutions quickly find a steady state P concentration at which the lake remains for most of the simulation. In the second row of Figure 3.7, where the P releases are plotted versus time, it becomes clear that these steady state P concentrations are quickly achieved by first releasing the maximum allowable P emission and then settling around a stable mean release. This is the same behavior Carpenter et al. [1999] find to be optimal when managing the lake to maximize an economic utility function which penalizes crossing the critical P threshold. This strategy is also called "bang-bang" control [Bellman et al., 1955], referring to the abrupt switch in control decisions from one extreme to another. In this case, bang-bang control is achieved by first releasing as much as possible in the first time step to reach a target steady state P concentration, and then drastically reducing emissions to a constant, lesser amount in all subsequent time steps to maintain equilibrium.

While both the bang-bang strategy and the DPS solutions exhibit this behavior, the Carpenter et al. [1999] optimization method only generates one solution that equilibrates to one P level found to be optimal according to a single utility

function. Our multi-objective approach finds several solutions which equilibrate to different P levels depending on one's preference. In Figure 3.7(a), it can be seen that the benefits-maximizing DPS solution equilibrates to a slightly higher steady state P concentration than the most reliable DPS solution. The closer proximity of this steady state P concentration to the critical P threshold demonstrates the greater implicit risk of this policy.

Figures 3.7(b) and (c) compare the reliability and benefits-maximizing solutions found by each solution strategy, respectively. In panel 3.7(b), one can see that the selected solutions from both methods result in similar lake dynamics, with the selected intertemporal solution being slightly more conservative than the selected DPS solution, as it has a lower steady state P concentration. In panel 3.7(c), however, the selected solutions from each method result in very different lake dynamics. Operation of the benefits-maximizing solution from the intertemporal optimization initially results in a fairly stable lake P concentration. However, after almost 80 years, the lake P concentration escapes this oligotrophic equilibrium and eventually crosses the critical P threshold, entering a eutrophic state.

The decisions resulting in this behavior are shown in panel 3.7(f). Here one can see that after almost 80 years, the P release decisions from the benefits-maximizing intertemporal solution suddenly increase dramatically. This solution prescribes conservative behavior initially to ensure that the lake P concentration stays below the critical P threshold for 85 years, meeting the reliability constraint. It then prescribes myopic behavior to maximize benefits, as if the end of the simulation were the end of time.

It should be noted that while these P releases at the end of time are much

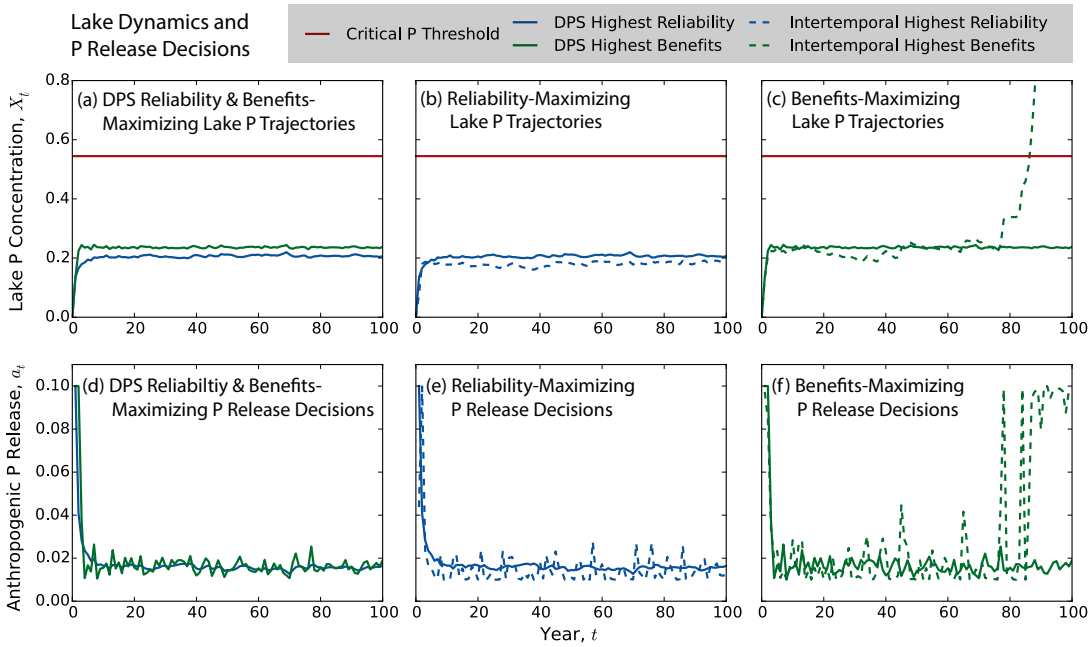

Figure 3.7: Time series of lake P concentration and anthropogenic P releases for select DPS and intertemporal solutions under one simulation of random natural P inflows. Panels (a)–(c) show that all solutions quickly reach a stable oligotrophic equilibrium, but that the benefits-maximizing intertemporal policy crosses the critical P threshold near the end of the simulation, moving toward a eutrophic equilibrium. Panels (d)–(f) indicate that DPS release decisions, shown by solid lines, are more temporally stable than their intertemporal counterparts, shown by dashed lines. They also do not exhibit myopic end-of-time behavior, like the benefits-maximizing intertemporal solution. Additionally, the most reliable DPS solution is more stable than the benefits-maximizing DPS solution.

greater for the intertemporal solution than for the DPS solution shown in the same panel, the average economic benefits across 100 simulations of 100 years are actually greater for the DPS solution than the intertemporal solution (see Table 3.3). There are two reasons for this. The first is because the economic benefits are discounted, such that earlier P releases are much more important to the overall economic benefits than later P releases. The DPS solution is able

to achieve greater average emissions early on without crossing the critical P threshold. The second reason is because the simulation shown in panel 3.7(c) is from one realization of random natural P inflows. In other realizations, there are large enough natural P inflows to increase the lake P concentration to the rising limb of the benefits-maximizing DPS policy. In these realizations, perfect reliability is not achieved (note this solution's average reliability of 85% in Table 3.3) and economic benefits are greater than in the plotted realization.

While there are realizations in which P emissions from the benefits-maximizing DPS solution will bring the lake over the critical P threshold, the reason it does not always exhibit the end-of-time behavior observed by the benefits-maximizing intertemporal solution is because the P release decisions from the policy are conditioned solely on the current state of the lake, not on time. One could add time as an input to the DPS policy, and the objective values could improve from this additional information, but doing so would result in similar end-of-time behavior. This stresses the importance of carefully considering what one includes in the DPS policy, or how one formulates objectives.

In addition to preventing this undesirable behavior by conditioning the release decision solely on the state, the releases from the DPS solutions are also more stable in time than the releases from the intertemporal solutions, as seen in panels 3.7(e) and 3.7(f). This is again due to the difficulty for the MOEA to optimize later-term release decisions. It is also important to note from panel 3.7(d) that the most reliable DPS policy is much more stable than the benefits-maximizing DPS solution. This is because the benefits-maximizing DPS policy is much steeper about its observed oligotrophic equilibrium of just over 0.2, as seen in Figure 3.6. A small change in the lake P concentration about this point

demands a much greater change in the P release decision compared to the most reliable DPS policy. This is an additional, unforeseen tradeoff between economic benefits and reliability; in order to maximize economic benefits, one must also make a sacrifice in terms of temporal release stability.

3.6.3 Robustness Analysis

Sections 3.6.1 and 3.6.2 illustrate the computational and performance benefits of using DPS for many objective optimization of control problems, but one potential concern is that the optimal policies are too specific to the system to which they were optimized. If the true system characteristics are different, the solutions may no longer perform well. Using the MORDM framework [Kasprzyk et al., 2013] described in Section 3.5.2, we re-evaluate the performance of the solutions found by each solution strategy on alternative plausible lake systems to determine how robust they are to commonly deep uncertainties in characterizing environmental thresholds and economic discounting.

Evaluation of Alternatives and Calculation of Robustness

As explained in Section 3.5.2, we re-evaluate the Pareto-approximate solutions in 1,000 alternative SOWs and, for each solution, calculate robustness as the percent of SOWs in which three criteria are met: 1) economic benefits > 0.2, 2) average reliability > 95% and 3) both of these. In Figure 3.8, we have sorted the Pareto-approximate solutions found by each method by robustness under each of these three criteria, plotting the satisficing metric for the most robust solution on the left and the least robust on the right. Figure 3.9 shows the reference sets

from the original optimization, with each point shaded by its robustness on the third criterion combining economic and reliability performance thresholds.

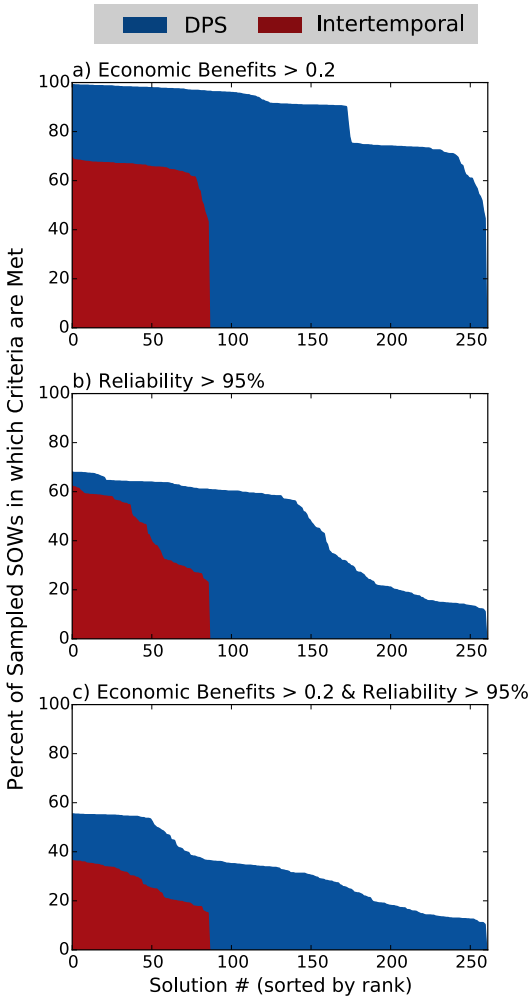

Figure 3.8: Robustness of DPS and intertemporal solutions according to the domain-satisficing criterion. Solutions from the reference sets of the DPS and intertemporal solution strategies are re-evaluated on 1,000 alternative SOWs and sorted by robustness. Robustness is quantified here as the percent of SOWs in which certain criteria are met: (a) economic benefits > 0.2, (b) reliability > 95% and (c) economic benefits > 0.2 and reliability > 95%. The DPS solutions are able to meet each of these criteria in a greater percent of the SOWs than the intertemporal solutions.

From Figure 3.8, it is clear that for all criteria, the DPS solutions in blue are more robust than the intertemporal solutions in red. However, robustness drops dramatically from one criterion to the next. Even the most robust DPS solution on criterion 3 satisfies the economic and reliability thresholds in less than 60% of sampled SOWs. Decision makers managing the lake would likely be concerned by this result.

To determine which parameter combinations are preventing these solutions from meeting criterion 3, we further analyze the most robust solutions from each solution strategy. These points are enlarged in Figure 3.9. Note the complex mapping between objective values and robustness of the DPS solutions that would make it difficult to predict the most robust solutions *a priori*, strongly limiting the value of classical multi-criterion decision making methods [Chankong & Haimes, 1983]. The DPS and intertemporal solutions with low economic benefits in the world to which they were optimized generally perform poorly on the robustness measure, likely failing to meet the economic criterion. Similarly, the intertemporal solutions with high maximum average P concentrations are not robust, likely failing to meet the reliability criterion. However, because many of the high benefits DPS solutions are able to achieve these benefits with low maximum average P concentrations, there is not an obvious failure mechanism on the reliability criterion.

This complex mapping of objective values to robustness highlights the importance of performing multi-objective search initially in order to find a diverse set of policy options, as it is often impossible to know *a priori* whether a single-objective optimization method will find a robust solution. Since a solution's robustness may influence a decision maker's policy preference, it is better to

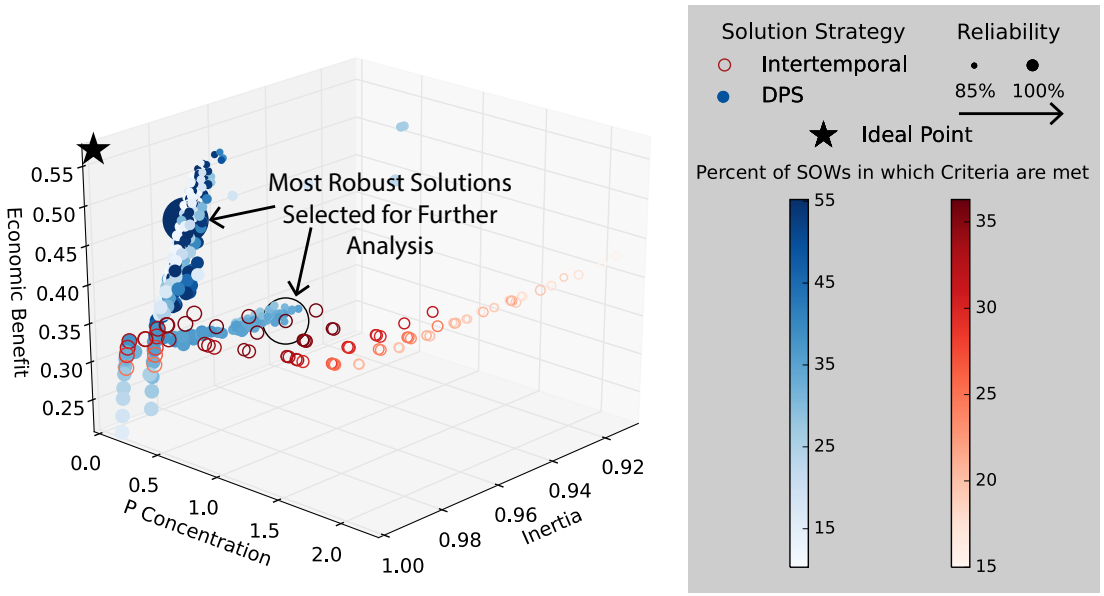

Figure 3.9: DPS and intertemporal reference sets shaded by robustness, with the most robust solutions enlarged. Robustness is defined as the percent of SOWs in which solutions have economic benefits > 0.2 and reliability > 95%. The complex spatial mapping to robustness, particularly of the DPS solutions, highlights the importance of *a posteriori* decision making; it would be impossible to predict in advance what weighted combination of objectives would yield robust solutions. The most robust solutions from each method, enlarged in the figure, are selected for further sensitivity analysis.

perform multi-objective optimization to find a range of Pareto optimal policies and then re-evaluate them on alternative SOWs to determine how robust each solution is. Only once these robust solutions have been discovered can one determine which parameter uncertainties are driving its failures.

Sensitivity Analysis of Uncertain Parameters

In Figure 3.10, we have plotted 2-D projections of the sampled SOWs, with each point representing one of the 1,000 samples. The y-axis gives the value of q in

each point's SOW, while the x-axis gives the value of b for the plots in the first column, and δ for the plots in the second. In the top row, each of these points has been colored green if the most robust DPS solution meets criterion 3 in that sampled SOW, and gray if it does not. In the bottom row, the same is done for the most robust intertemporal solution.

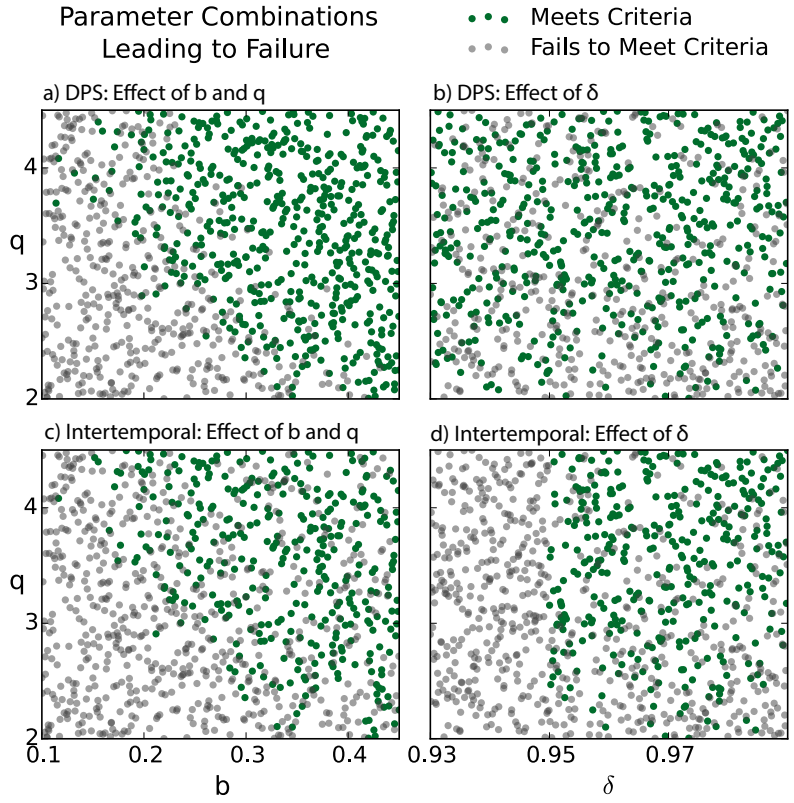

Figure 3.10: Combinations of uncertain parameter values leading to failure. Each dot represents a different SOW for a selected solution, with the most robust DPS solution shown in panels (a) and (b), and the most robust intertemporal solution in panels (c) and (d). Each dot is shaded green if the solution has economic benefits > 0.2 and reliability > 95% in that SOW, or gray if it does not. Panels (a) and (c) show that there is a nonlinear combination of small values of b and q which always result in failure. Panels (b) and (d) show that there is a threshold value of the discount parameter, δ, below which the most robust intertemporal solution always fails but the most robust DPS solution does not.

In the first column, one can see that both the selected intertemporal and DPS solutions fail for a non-linear combination of low b and low q values. In the second column, however, differences emerge. The DPS solution is able to achieve criterion 3 for any of the sampled values of the discount parameter, δ, used to calculate economic benefits. For the intertemporal policy, however, the selected solution is unable to achieve the economic threshold of criterion 3 for values of δ below about 0.95. Because the intertemporal solution prescribes the same time series of P release decisions in every realization, regardless of the other lake parameters, its economic performance is sensitive only to the discount parameter. While the economic benefits of the DPS solution also depend on this parameter, the P release decisions from this policy change depending on the state of the lake, and the lake P concentration follows a different trajectory under different system parameters. By adapting its P release decisions to the state of the lake, the selected DPS solution becomes insensitive to the discount parameter and is able to reach high economic benefits regardless of that parameter's value.

Effect of the critical P threshold on policy failure

While the insensitivity of the most robust DPS solution to the discount parameter was an unforeseen benefit of the method, the remaining sensitivity to b and q is concerning. Figure 3.11 shows that the non-linear combination of b and q causing these failures is actually driven primarily by the critical P threshold associated with these parameter combinations. This figure shows a contour map of the critical P threshold laid atop the dot plot in Figure 3.10(a), but with successes now plotted black instead of green. The critical P threshold below which the policies almost always fail lies around the 0.5 contour. The critical P thresh-

old in the world to which the DPS policies were optimized is 0.54, indicating a high sensitivity of the policies to this threshold. Just a small movement to a lower threshold results in it being crossed, likely due to the inexistence of an oligotrophic equilibrium when the policy is operated on a lake system with this threshold.

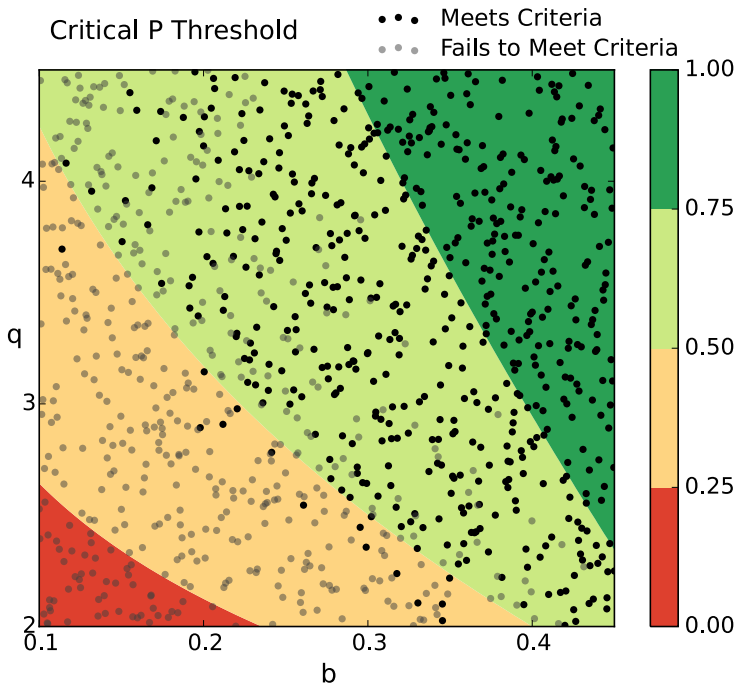

Figure 3.11: Effect of critical P threshold on policy failure. This is the same dot plot as shown in Figure 3.10(b), but atop a contour plot of the critical P threshold as a function of b and q. One can see that below a critical P threshold of about 0.5, the most robust DPS solution always fails to meet the satisficing criteria. This is very close to the critical threshold of 0.54 in the SOW to which the policy was optimized, indicating high sensitivity to the assumed model parameters.

Furthermore, there is a littering of failures even just above this threshold, with the failure rate increasing for higher values of q. In this region, the impact of the other uncertain parameters becomes important in determining whether or not the policy will succeed. For example, high values of q result in more abrupt

transitions across the threshold due to steeper recycling curves, while greater values of μ result in natural P inputs eliminating the system's oligotrophic equilibrium.

Clearly, one approach to improve robustness is to improve the understanding of the uncertain parameters and, in response, adjust the policy so that it maintains an oligotrophic equilibrium. Fortunately, the ability of DPS to incorporate additional information into the policy could enable the formation of more adaptive policies. For example, the DPS policy could include the mean and variance of one's estimate of the critical P threshold, and change adaptively as these estimates are refined through endogenous learning [Lempert & Collins, 2007; Keller & McInerney, 2008]. Alternatively, one could directly include estimates of the critical P threshold rather than both b and q. Prior work on tipping points has found that systems tend to slow down as they approach tipping points [Scheffer et al., 2009], so the rate of change in lake P concentration could potentially be a DPS indicator as well. We are pursuing these avenues in further research.

3.7 Conclusions

Managing socio-ecological systems for multiple objectives poses several challenges, particularly if the systems possess non-linear tipping points and deeply uncertain system dynamics. In this study, we use the classical shallow lake problem to illustrate how DPS can alleviate many of the difficulties associated with identifying robust environmental control policies for these systems, as well as their inherent tradeoffs. By directly parameterizing operating policies for the

system and optimizing the parameters of those policies with an MOEA, DPS can exploit the information available in detailed system simulations. Our results show that the DPS-based closed loop control strategy can drastically reduce the computational complexities of solving a benchmark many-objective control problem compared to optimizing open loop control policies. Additionally, the control feedback inherent to DPS policies results in more flexible system operations that can react to different observations of stochastic inputs. While the actual use of an intertemporal open loop control policy without re-optimization requires that one take a particular action regardless of the resulting system state, operation of a closed loop control policy found with DPS enables different decisions based on the system's state trajectory without having to re-optimize the policies at every time step. This adaptability makes DPS policies more robust to well-characterized uncertainty in stochastic inputs such as natural P loads to a lake, enabling better performance across a stochastic ensemble of potential inputs.

The adaptability of DPS policies not only yields improved performance under well-characterized uncertainty in stochastic inputs, but also improved robustness to deep uncertainties. In this study, we find that DPS policies optimized to an assumed SOW generalize better when subjected to alternative SOWs than intertemporal control policies. In particular, the DPS policies are able to meet minimum environmental and economic performance criteria in a larger fraction of alternative SOWs than the intertemporal policies. In fact, the most robust DPS policy is able to adapt its decisions well enough in alternative SOWs that its ability to meet the criteria is insensitive to the value of the discount parameter used to calculate economic benefits, which is not true of the most robust intertemporal policy.

Finally, the ability of DPS to include multiple sources of information into the optimized control policies allows one to create even more flexible policies that include estimates of the system parameters and their associated uncertainty. This information could prevent policies optimized to an assumed SOW from crossing a catastrophic tipping point below what was assumed, or conversely from behaving too conservatively in systems with tipping points above what was assumed. DPS could therefore be a simple adaptive management tool for identifying robust control policies in socio-ecological systems like the lake problem that exhibit deeply uncertain threshold behavior. The most obvious extension is to the Earth system and its many components that could be on the brink of a tipping point induced by anthropogenic climate change. Future work in this area should focus on developing state-dependent rather than intertemporal CO_2 emissions policies and adaptively changing these policies as one gains new information on system conditions critical to tipping.

CHAPTER 4

RIVAL FRAMINGS FRAMEWORK FOR DISCOVERING HOW PROBLEM FORMULATION UNCERTAINTIES SHAPE RISK MANAGEMENT TRADEOFFS IN WATER RESOURCES SYSTEMS

This chapter is drawn from the following peer-reviewed journal article:

Quinn, J.D., Reed, P.M., Giuliani, M. and Castelletti, A. (Accepted). Rival framings: a framework for discovering how problem formulation uncertainties shape risk management tradeoffs in water resources systems. Water Resources Research. doi: 10.1002/2017WR020524

Section 4.5.3 has been added to include material from the article's supporting information.

This work was partially supported by the U.S. National Science Foundation (NSF) through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the Penn State Center for Climate Risk Management. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding entities.

4.1 Abstract

Managing water resources systems requires coordinated operation of system infrastructure to mitigate the impacts of hydrologic extremes while balancing conflicting multi-sectoral demands. Traditionally, recommended management strategies are derived by optimizing system operations under a single problem

framing that is assumed to accurately represent the system objectives, tacitly ignoring the myriad of effects that could arise from simplifications and mathematical assumptions made when formulating the problem. This study illustrates the benefits of a rival framings framework in which analysts instead interrogate multiple competing hypotheses of how complex water management problems should be formulated. Analyzing rival framings helps discover unintended consequences resulting from inherent biases of alternative problem formulations. We illustrate this on the monsoonal Red River basin in Vietnam by optimizing operations of the system's four largest reservoirs under several different multi-objective problem framings. In each rival framing, we specify different quantitative representations of the system's objectives related to hydropower production, agricultural water supply and flood protection of the capital city of Hanoi. We find that some formulations result in counterintuitive behavior. In particular, policies designed to minimize expected flood damages inadvertently increase the risk of catastrophic flood events in favor of hydropower production, while min-max objectives commonly used in robust optimization provide poor representations of system tradeoffs due to their instability. This study highlights the importance of carefully formulating and evaluating alternative mathematical abstractions of stakeholder objectives describing the multi-sectoral water demands and risks associated with hydrologic extremes.

4.2 Introduction

Managing both intra-annual and inter-annual hydrologic variability has posed a continual challenge to human societies. This challenge is especially difficult for low income countries whose economies depend largely on agriculture, but

lack the institutional and infrastructure capacity to adapt to variable hydrologic conditions [Hall et al., 2014]. Climate change is only expected to exacerbate this issue, as greater warming should increase both evaporation and precipitable water, paradoxically leading to both longer, more severe droughts and more intense flooding [Trenberth, 2011]. Recent observations indicate intensification of the hydrologic cycle has already begun [Huntington, 2006], with more frequent heat and precipitation extremes observed over the last half century [Coumou & Rahmstorf, 2012]. Again, these impacts are felt most by the disadvantaged, deepening poverty in low income, climate-dependent economies [Olsson et al., 2014; Hallegatte et al., 2015; World Bank, 2016]. Furthermore, as these economies grow and diversify out of agriculture, competition for water resources across their developing sectors will increase. In order to reduce, and if possible overcome, the negative impacts and water conflicts associated with hydrologic extremes, it is of paramount importance that innovative water management policies be discovered [Tanaka et al., 2006; Giuliani et al., 2016a; World Bank, 2016].

Conventionally, water resources managers have attempted to reduce the multi-sectoral impacts of hydrologic variability through optimized reservoir operations. Given that most river basins now contain multiple reservoirs, optimizing operations is mathematically challenging just considering the competing objectives and the high-dimensional and stochastic nature of the multi-reservoir control problem [Giuliani et al., 2016b; Zatarain Salazar et al., 2016]. While addressing these challenges, this study also confronts the often-ignored epistemic uncertainties surrounding how to formulate the control problem itself. In classical decision theory, the problem formulation is designed to conform to the chosen modeling approach [Tsoukias, 2008], disregarding the fact that the cho-

sen procedure will affect the predictions of the consequences of alternative solutions [Majone & Quade, 1980], and consequently which solutions are considered "optimal" [Roy, 1990]. Kasprzyk et al. [2009] and Zeff et al. [2014] illustrate this on separate multi-objective water supply portfolio planning problems in which the attainable system performance depends heavily on which objectives, constraints and decisions are included in the optimization. More specifically, Kasprzyk et al. [2009] find that different families of solutions emerge from different formulations, with some formulations missing entire regions of decision relevant tradeoff solutions. As such, how one frames a problem can wield an inadvertent influence of power on the outcome [Stirling, 2008]. This has led natural resources managers to advocate for the exploration of alternative problem structures [Hoppe, 2011] in participatory planning processes in order to discover tensions between competing framings formulated under different perceptions of stakeholder values [Bosomworth et al., 2017].

Acknowledging that the most appropriate problem formulation is itself uncertain, in this study we explore alternative problem structures using what Tsoukias [2008] dubs a "constructive" decision aiding approach in which the problem formulation itself is constructed, not just the optimal solutions. Within the water resources literature, similar methods were developed in the 1960s under the Harvard Water Program through which Maass et al. [1962] proposed a four-step process for designing water resources systems: 1) identifying the objectives, 2) translating the objectives into design criteria, 3) using these criteria to design water resources development and management plans, and 4) evaluating the consequences of the plans that have been developed, in particular, by quantifying regrets associated with using one objective over another. Emerging from the early origins of behavioral economics, this approach has inspired new

decision theories such as the version concept of Roy [2010], the rival problem framings concept of Walker et al. [2003], and de novo programming of Zeleny [1981], which Kasprzyk et al. [2012] expand on to explicitly capture multiple objectives. This approach is perhaps best described by Zeleny [1989]:

> *Making decisions does not mean finding our ways through a fixed maze (problem solving) – decision making refers to the very construction of that maze – ordering of nature so that we ourselves can find our way through it.*

In this study, we highlight the importance of evaluating alternative constructions of the maze that is the multi-objective, multi-reservoir control problem through Vietnam's Red River basin, where operations of the four largest reservoirs must balance agricultural water demands for food and energy production, while also reducing flood risks to the capital city of Hanoi. Similar in concept to the approaches of Kasprzyk et al. [2009] and Zeff et al. [2014], we build and evaluate four rival problem framings of the Red River control problem; however, we not only vary the objectives and constraints included in each formulation, but also the mathematical quantification of those objectives. Building on insights from Giuliani & Castelletti [2016] in highlighting the importance of capturing multiple risk attitudes in problem framings for water management applications, we construct several alternative management objectives representing a range of stakeholder risk preferences from highly risk-averse (e.g. min-max objectives) to risk-neutral (e.g. expectation objectives). By optimizing Red River reservoir operations to alternative formulations encompassing a gradient of risk attitudes and re-evaluating the resulting solutions from each formulation on the objectives from each of the other competing formulations, we seek to mitigate the unintended systematic biases that Majone & Quade [1980] caution analysts

to guard against. We also perform visual diagnostics of the Pareto-approximate operating policies discovered under each problem formulation to better understand the effects of preference and framing on the resulting system behavior.

One of the difficulties of applying this constructive decision aiding approach is that traditional optimization methods may be limited in the type and scale of problems they can solve. For example, linear programming methods can only solve problems with linear objectives and constraints, while linear quadratic programming methods can only solve problems for linear systems with quadratic objective functions [Yeh, 1985]. In terms of scale, Giuliani et al. [2016b] note that commonly used stochastic dynamic programming methods are limited by several dimensional curses that confine the number of reservoirs whose operations can feasibly be optimized simultaneously, the number of exogenous variables, such as streamflow and precipitation, that can be used to condition reservoir release decisions, and the number of Pareto-optimal solutions that can be discovered. Additionally, the system states must be discretized, the objective functions and constraints must be time-separable and the disturbance process must be uncorrelated in time [Castelletti et al., 2012b]. These latter constraints severely limit the type of objectives that can be formulated by traditional methods. For example, time-separability constraints make it impossible to reflect different risk attitudes with respect to different objectives within a single problem formulation, e.g. by calculating some objectives in expectation and others using a min-max formulation.

Fortunately, Giuliani et al. [2016b] show that these restrictions can be overcome with Evolutionary Multi-Objective Direct Policy Search (EMODPS), a simulation-optimization approach in which multi-objective evolutionary algo-

rithms (MOEAs) are used to optimize the performance of multi-reservoir operating policies simulated over stochastic streamflows. Importantly, the objectives calculated over this simulation need not be time-separable, and the operating policies can be flexibly formulated to approximate any mathematical form. For example, non-linear approximators can be used to describe the operating policies, allowing for adaptable, state-dependent operating rules. Consequently, EMODPS allows us to formulate complex problem formulations, improving our ability to accurately assess system performance. This added layer of problem complexity further motivates the need to test competing formulations of water resources optimization problems: first, because different combinations of performance measures may more effectively capture the stakeholders' objectives, and second, because multiple nonlinear objectives may interact in unpredictable ways, increasing the risk of unintended consequences. Fortunately, recent computational advancements in our ability to solve complex, multi-objective control problems [Reed & Hadka, 2014] have enabled a formal implementation of a rival framings approach to better account for problem formulation uncertainty.

Building off of foundational work by the Harvard Water Program [Maass et al., 1962] and others in highlighting the importance of utilizing multiple performance measures to evaluate system performance, in this study we exploit the computational power of EMODPS to optimize multi-reservoir operating policies for the Red River basin, described in Section 4.3, under multiple problem formulations outlined in Section 4.4. While many uncertainties surround reservoir operations, such as model, demand and climate uncertainty, as well as non-stationarity in risk-preferences, we focus our analyses on stationary problem formulation uncertainty to isolate its effects. In Section 4.5 we use visual diagnostics to assess the importance of this uncertainty by illustrating how policy

operations designed under different formulations impact system performance. Through this analysis, we find several unintended consequences and unforeseen benefits of particular framings that have important implications for how the system can better manage extremes and conflicting multi-sectoral demands. Unfortunately, it is not always possible to know *a priori* what the effects of alternative objectives will be. For this reason, we conclude in Section 4.6 with a discussion of the importance of applying constructive decision aiding processes to effectively manage the negative impacts and water conflicts associated with hydro-climatic variability and change. Future work will explore how problem formulation uncertainty compares with other sources of uncertainty in influencing overall system performance.

4.3 Red River Context

4.3.1 Basin Description

From its source in southern China to its mouth in the South China Sea, the Red River basin spans 169,000 km^2, fifty-one percent of which lies in Vietnam. As the second largest river basin in Vietnam, the Red River serves as a vital agricultural and economic resource to the developing nation. Recent reservoir construction in the system has significantly contributed to Vietnam's energy growth, with hydropower currently representing 46% of the country's total installed electric power capacity [Asian Development Bank, 2016]. These reservoirs have also enabled more secure and stable food production through irrigable agriculture, a key component in poverty alleviation, as 70% of the Vietnamese population is

employed in agriculture, 76% of which is irrigated [Nguyen et al., 2002]. With cultivation and fisheries representing 58% and 29% of the basin's average water demand in the delta, respectively (Figure 4.1b), managing droughts is vital for Vietnam's food security.

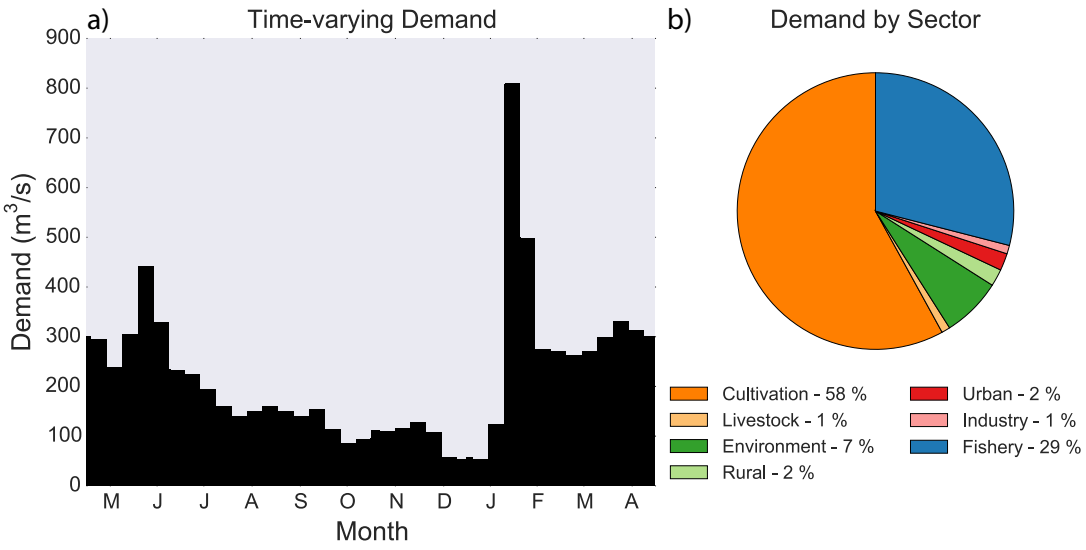

Figure 4.1: Average water demand over time in the Red River delta (panel a) and its distribution across sectors (panel b), obtained from the Vietnamese Institute of Water Resources Planning (IWRP). There is a large spike in demand at the beginning of February for field flooding at the time of planting, illustrating why agriculture represents the largest source of demand (58%). The next most important sector is fisheries (29%), further highlighting the importance of water supply for the region's food security.

Yet, while drought concerns during the dry season threaten the region's ability to provide sufficient water supply for agriculture and hydropower, large scale floods during the monsoon season endanger the basin's infrastructure. The rapidly urbanizing Vietnamese capital of Hanoi lies in the Red River delta, where average annual flood damages have been estimated at 130 million USD [Hansson & Ekenberg, 2002]. Seeking to reduce the impacts of severe and fre-

quent flooding, the Red River's second largest reservoir, Hoa Binh, was specifically designed to reduce the maximum observed flood peak at Hanoi from 14.8 m to 13.3 m, just below the 13.4 m dike height [Le Ngo et al., 2007]. Flood protection requires maintaining low storage at Hoa Binh and the other system reservoirs during monsoonal months to ensure that there is sufficient storage capacity to capture large flood events. However, maintaining low storage in the reservoirs reduces hydropower production and the ability to supply water for irrigation. In this study, we investigate if improved multi-reservoir operations in Vietnam's Red River basin can better balance the multi-sectoral demands of agricultural water supply, energy production, and flood protection.

4.3.2 Model Description

Figure 4.2a shows the locations of the four largest reservoirs within the Red River basin whose operations we optimize. Figure 4.2b, reproduced from Giuliani et al. [In Press], provides a more detailed schematic of how flows are simulated through the system. The two largest reservoirs, Son La (SL) and Hoa Binh (HB), are located in series along the Da River, which provides roughly half of the total system flow. Hoa Binh is the most important reservoir for flood protection since it is the last reservoir before Hanoi along the largest tributary. Parallel to Son La and Hoa Binh are the Thac Ba (TB) reservoir on the Chay River and the Tuyen Quang (TQ) reservoir on the Gam River. These reservoirs are much smaller in terms of storage and power capacity. Altogether, the four modeled reservoirs have a storage capacity of 22.67 billion m^3 and power capacity of 4782 MW. Table 4.1 lists the storage and power capacities of each reservoir individually.

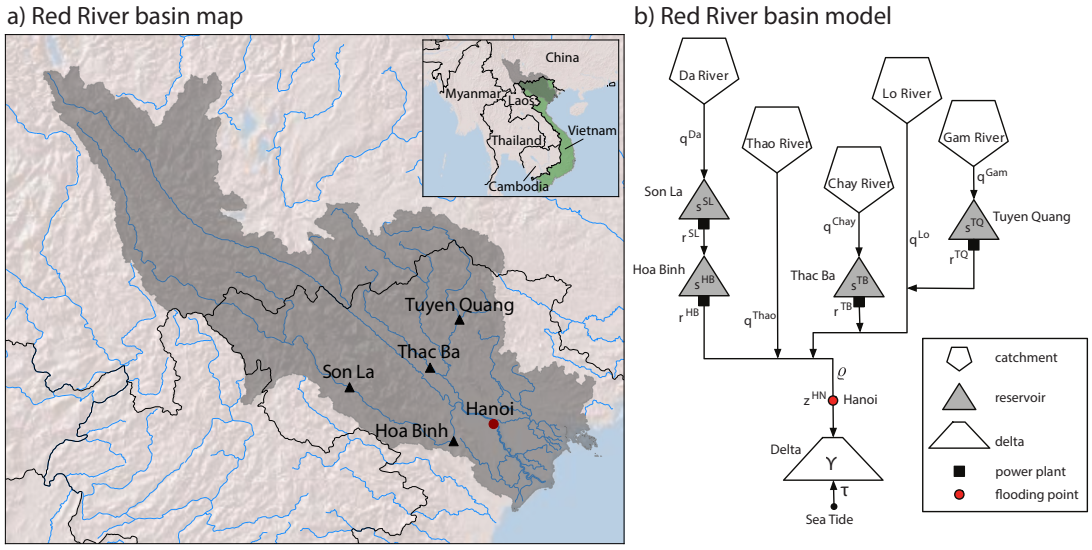

Figure 4.2: Map of the Red River basin (panel a) and schematization of the main components of the Red River basin model (panel b, reproduced from Giuliani et al. [In Press]). The inflows shown in panel b are generated synthetically, the releases at each of the reservoirs are determined by the optimized operating policies, and subsequent flows through the delta are modeled by a dynamic emulator of a MIKE 11 simulation of the downstream hydraulics.

We simulate flows through the Red River system using two sub-models: 1) flows through the reservoirs and power plants and 2) flows through the delta. All data used to build the model are from the Ministry of Agriculture and Rural Development (MARD) of Vietnam and were collected during the Integrated and sustainable water Management of Red Thai Binh Rivers system in changing climate (IMRR) project (http://xake.elet.polimi.it/imrr/). In the first sub-model, we estimate the volume of storage, s_t^k, in the k-th reservoir at time t using simple mass balance equations:

Table 4.1: Storage and power capacities of reservoirs in the Red River basin whose operations are optimized.

Reservoir	Storage in Bm^3 (% of total)	Maximum Power Capacity (MW)
Son La	9.58 (42.3%)	2400
Hoa Binh	8.38 (37.0%)	1920
Thac Ba	2.81 (12.4%)	120
Tuyen Quang	1.90 (8.4%)	342

$$s_t^{SL} = s_{t-1}^{SL} + q_t^{Da} - r_t^{SL} - e_t^{SL} S(s_{t-1}^{SL}) \tag{4.1}$$

$$s_t^{HB} = s_{t-1}^{HB} + q_t^{Da,lat} + r_t^{SL} - r_t^{HB} - e_t^{HB} S(s_{t-1}^{HB}) \tag{4.2}$$

$$s_t^{TB} = s_{t-1}^{TB} + q_t^{Chay} - r_t^{TB} - e_t^{TB} S(s_{t-1}^{TB}) \tag{4.3}$$

$$s_t^{TQ} = s_{t-1}^{TQ} + q_t^{Gam} - r_t^{TQ} - e_t^{TQ} S(s_{t-1}^{TQ}) \tag{4.4}$$

where r_t^k is the actual release from the k-th reservoir in the time interval $[t-1, t)$; q_t^{Da}, q_t^{Chay} and q_t^{Gam} are the water volumes from the Da, Chay and Gam rivers flowing into the Son La, Thac Ba and Tuyen Quang reservoirs, respectively, in this time interval; $q_t^{Da,lat}$ is the lateral inflow to the Da River between Son La and Hoa Binh during this time interval; e_t^k is the average volume of water evaporated from the k-th reservoir during this time interval and $S(s_{t-1}^k)$ is the surface area of k-th reservoir at time $t-1$ as a function of its storage level at time $t-1$. For each reservoir k, deterministic rates of e_t^k are assumed for each calendar day based on a 10-day moving average of the average historical evaporation rates from the nearest meteorological station over the period 1959-2011 (see Bernardi et al. [2014] for more details). In our notation, the value of the subscript indicates the

time step at which each variable's value is deterministically known. The time step at which the release decision is made is one day; however, this volume of water is allocated hourly in the model assuming the operator optimally engages the turbines to maximize daily energy production. Following this assumption, we estimate the daily hydropower produced by the k-th reservoir's hydropower plant using the function $\eta_t^k = f(s_{t-1}^k, r_t^k)$, which is described by an artificial neural network (ANN; see Giuliani et al. [2016a] for more details).

Because it is unrealistic and unsafe to assume no future streamflows will lie outside of those which have already been observed [Thomas, Jr. & Fiering, 1962], we run the first sub-model with synthetically generated hydrology. Compared to the limited 51-year historical record (1960-2010), these synthetic streamflows expand the range of hydrologic scenarios to which reservoir operations are optimized. In this study, we assume hydrologic stationarity in generating synthetic flows for the model simulations such that optimized operating policies represent baseline tradeoffs under our best perception of the current state of the world. Consequently, this study focuses solely on problem formulation uncertainty. In future work, we will explore the effects of uncertainty in the distribution of future hydrologic flows on the performance of these optimized policies.

Here we synthetically generate correlated monthly streamflows on the five tributaries, q_t^{Da}, q_t^{Thao}, q_t^{Chay}, q_t^{Lo} and q_t^{Gam} using the method of Kirsch et al. [2013]. This method uses Cholesky decomposition to preserve auto-correlation, and a simultaneous resampling of historical flows at each site to preserve spatial correlation. We then disaggregate the synthetic monthly flows to daily flows using the method of Nowak et al. [2010], which proportionally scales historical daily

flows at each site from a probabilistically selected month of the historical record such that the synthetic monthly total is preserved. Finally, we scale the lateral inflow between the Son La and Hoa Binh reservoirs, $q_t^{Da,lat}$, from q_t^{Da} assuming a constant flow per unit drainage area. Readers interested in a more detailed discussion and statistical validation of the synthetic streamflows can reference Appendix B.

For the delta sub-model, we use a meta-model developed by Dinh [2015] to approximate a 1D hydrodynamic model (MIKE 11) of the flow routing from the reservoirs to Hanoi and the irrigation districts. The meta-model employs an ANN to approximate the water volume in the irrigation canals, Υ_t , the water level at Hanoi, z_t^{HN}, and the supply deficit, D_t:

$$\Upsilon_t = f(\Upsilon_{t-1}, \varrho_t, W_t, \tau_{t-1}) \tag{4.5}$$

$$z_t^{HN} = f(z_{t-1}^{HN}, \varrho_t, \tau_{t-1}) \tag{4.6}$$

$$D_t = f(\varrho_t, W_t, \tau_{t-1}, \Upsilon_t) \tag{4.7}$$

where $\varrho_t = r_{t-1}^{HB} + r_{t-1}^{TB} + r_{t-1}^{TQ} + q_{t-1}^{Thao} + q_{t-1}^{Lo}$ is the total inflow to the canals assuming a one-day travel time from the reservoirs and streamflow gauges of the Thao and Lo rivers to the delta; W_t is the time-dependent water demand; and τ_{t-1} is the previous day's tide. Using this meta-model reduces the computational demands of simulating 20 years of operations from a few days to a few seconds, making optimization computationally feasible. See Dinh [2015] for more details.

4.4 Methods

In this study, we evaluate four competing problem formulations of the Red River control problem. Because many of the problem formulations we explore are mathematically complex, we need a flexible optimization approach that does not require a specific problem structure. The Evolutionary Multi-Objective Direct Policy Search (EMODPS) framework [Giuliani et al., 2016b] provides this flexibility (see Ch 2 Section 2.2.1). Earlier work in the Red River basin by Castelletti et al. [2012a] found that EMODPS was able to discover operating policies for HoaBinh that outperformed historical operations on every objective, and more recent work by Giuliani et al. [2016a] indicated it was capable of converging on a more challenging three-reservoir version of the model.

Here, we advance the EMODPS framework with an additional diagnostic verification step. This step includes re-evaluating the optimized policies on an out-of-sample set of stochastic inputs to ensure that they generalize well, and analyzing the policies themselves to understand how they operate on the system to achieve the given objectives. In summary, the primary steps in the EMODPS framework presented in this study are: 1) formulation of the system objectives, 2) formulation of reservoir operating policies as functions whose parameters are to be optimized, 3) multi-objective optimization of the policies, and 4) diagnostic verification of the optimized policies. We describe each of these steps in detail below.

4.4.1 Formulation of Objectives

A core goal and contribution of this work is to better understand the nature of the operational tradeoffs across the Red River system's three primary functions: flood protection, hydropower production and agricultural water supply. However, as noted in the introduction, translating each of the system objectives into quantitative performance measures is not straightforward. Consequently, we explore four rival problem framings that capture a range of stakeholder attitudes toward risk from highly risk-averse to risk-neutral. We take a multi-objective optimization approach since collapsing these objectives into a single economic performance measure weighting the different objectives may lead to a single objective unexpectedly dominating the system performance [Arrow, 1950; Kasprzyk et al., 2015]. This could be especially concerning if the estimates of the costs and benefits associated with flood damages, hydropower revenue, and agricultural losses during drought are highly uncertain and non-stationary [Dittrich et al., 2016; Smith et al., 2017]. For these reasons, in each formulation we quantify the three key stakeholder objectives using un-monetized measures of performance, but recommend that economic estimates be incorporated *a posteriori* to aid decision makers in choosing among alternative operating policies.

The four candidate formulations we explore in this study are as follows: (1) Worst Case (WC), (2) Worst First Percentile (WP1), (3) Expected Value (EV), and (4) Expected Value & Standard Deviation of Hydropower (EV&SD_H). In each formulation, operating policies are simulated over N ensemble members of T years of synthetically generated streamflows, with N and T varying by formulation. In all formulations, each T-year simulation begins on May 1^{st}, the first day of the monsoon season, and initial conditions for April 31^{st} must be specified for

the storages at the four reservoirs, $\{s_0^{SL}, s_0^{HB}, s_0^{TB}, s_0^{TQ}\}$, the water level at Hanoi, z_0^{HN}, the water volume in the canals, Υ_0, the flow to the delta, ϱ_0, and the total system inflow, q_0^{TOT}. In the WC formulation, objectives are calculated over N=50 ensemble members of length T=20 years (i.e., 50 unique 20 year streamflow records). Since this is a fairly long simulation length, performance is relatively insensitive to initial conditions, so constant, typical values for April 31st are assumed. In the WP1, EV and EV&SD$_H$ formulations, however, objectives are calculated over N=1000 ensemble members of length T=1 year (i.e., 1000 unique 1 year streamflow records). In order to better sample inter-annual variability under these shorter simulations, initial conditions are randomized for each ensemble member by sampling joint conditions on April 31st from 10,000-year simulations of the optimal policies from the WC formulation. Table 4.2 summarizes the characteristics of the simulations over which policies are optimized for each formulation.

When formulating candidate objective functions for optimization, a single performance statistic across the N ensembles of T-year simulations must be quantified mathematically. We calculate the d-th objective, J_d, according to Equation 4.8:

$$J_d = \Psi_{i\in(1,\ldots,N)}\left[\Phi_{t\in(1,\ldots,365T)}[g_d(t,i)]\right] \tag{4.8}$$

where $g_d(t,i)$ is the value of the d-th objective on day t of the i-th ensemble member, Φ is an operator for the aggregation of $g_d(t,i)$ over time, such as the sum ($\sum$), and Ψ is a statistic used to filter the noise across ensemble members, such as the expected value ($\mathbb{E}$). It is through these key variables, $g_d(t,i)$, Φ, Ψ, N and T, that the problem formulation can vary to reflect different risk attitudes, e.g.

Table 4.2: Characteristics of simulations over which policies are optimized for each formulation.

Formulation	Initial Conditions	Ensemble Size, N	Years per Ensemble, T
Worst Case (WC)	Constant, average conditions	50	20
Worst First Percentile (WP1)	Randomly sampled from simulation of WC policies over 10,000 years	1000	1
Expected Value (EV)			
Expected Value & Standard Deviation of Hydropower ($EV\&SD_H$)			

by changing how objectives are aggregated over time and filtered across noise, as well as the time horizon over which they are calculated [Soncini-Sessa et al., 2007]. While these variables change across the four formulations we explore, the general form of the highest dimensional multi-objective optimization problem can be summarized by equations 4.9–4.11 below:

$$\theta^* = \text{argmin}_\theta J(\theta) \tag{4.9}$$

where

$$J = \begin{bmatrix} -J_{Hydro}(\theta) \\ J_{Deficit^2}(\theta) \\ J_{Flood}(\theta) \\ J_{Recovery}(\theta) \\ J_{Hydro\ Std}(\theta) \end{bmatrix} \tag{4.10}$$

subject to

$$J_{Flood}(\theta) \leq C \tag{4.11}$$

where θ is a vector of decision variables describing the operating policies defined in Section 4.4.2 and C is a constraint on the flooding objective defined for each formulation in Section 4.4.1. Table 4.3 summarizes which objectives and constraints are included in each formulation. As indicated by the superscripts, some of the objectives are the same across formulations, while others are not. Mathematical descriptions of each of the objectives under each formulation, including $g_d(t, i)$, Φ and Ψ, are provided in Appendix A, while summary text descriptions and our rationale for each candidate formulation are provided in Section 4.4.1 below.

Worst Case (WC) Formulation

The WC formulation assumes a highly risk-averse operator concerned with minimizing the worst case performance of the hydropower, flooding, and water supply objectives across an ensemble of potential conditions, similar to prior published studies by Orlovski et al. [1984] and Soncini-Sessa et al. [1990]. This

Table 4.3: Objectives and constraints included in optimization under each problem formulation. See Section 4.4.1 for further explanation.

Formulation	Objectives	Constraints
Worst Case (WC)	J^{WC}_{Hydro},$J^{WC}_{Deficit^2}$,J^{WC}_{Flood}	–
Worst First Percentile (WP1)	J^{WP1}_{Hydro},$J^{WP1}_{Deficit^2}$,J^{WP1}_{Flood},$J^{WP1}_{Recovery}$	$J^{WP1}_{Flood} \leq 2.15$ m
Expected Value (EV)	J^{EV}_{Hydro},$J^{EV}_{Deficit^2}$,J^{WP1}_{Flood},$J^{EV}_{Recovery}$	$J^{WP1}_{Flood} \leq 2.15$ m
Expected Value & Standard Deviation of Hydropower (EV&SD$_H$)	J^{EV}_{Hydro},$J^{EV}_{Deficit^2}$,J^{WP1}_{Flood},$J^{EV}_{Recovery}$,$J^{EV\&SD_H}_{Hydro\ Std}$	$J^{WP1}_{Flood} \leq 2.15$ m

formulation was designed based on the desire of the MARD to formulate conservative operating policies, particularly with respect to flooding.

In the WC formulation, the first objective, J^{WC}_{Hydro}, seeks to maximize hydropower production based on the desire of the Vietnamese Ministry of Industry and Trade (MOIT) to generate as much hydropower as possible in order to minimize costs of production from thermal plants and import. Here, we calculate average hydropower production over the synthetically generated 20-year streamflow sequences ($\Phi = \mathbb{E}_{365T}$) and minimize the worst case average production across the 50-member ensemble ($\Psi = \min_N$). Simulations of 20 years are chosen to provide estimates of production over a typical planning period, while the worst case across 50 simulations of 20 years is minimized to ensure reasonable performance in even the worst case potential planning period. In the Red River system, we optimize hydropower production rather than revenue because the Vietnamese electricity market is regulated by the Government and energy is sold at a fixed rate. Since the price is fixed, maximizing production is equiva-

lent to maximizing the revenue from production [Castelletti et al., 2012a]. While unexplored here, uncertainty in how best to formulate this objective could be considered in an additional rival framing.

The second objective, $J^{WC}_{Deficit^2}$, seeks to minimize the squared water supply deficit. As with hydropower, the average daily squared deficit is calculated over every 20-year ensemble member ($\Phi = \mathbb{E}_{365T}$) and we minimize the maximum of these averages across the 50 ensemble members ($\Psi = \max_N$). The daily deficit is squared to numerically favor several small deficits over a small number of large deficits. This objective was accepted by the MARD through the IMRR project in 2013.

The final objective of this formulation, J^{WC}_{Flood}, seeks to minimize flood damages. In each of the 20-yr simulations, we approximate expected flood damages by the penalty function shown in Figure 4.3 and minimize the maximum expected damages across the 50-member ensemble ($\Phi = \mathbb{E}_{365T}$ and $\Psi = \max_N$). Based on alarm levels elicited from stakeholders, water levels between 6 m and 11.25 m are penalized minimally by a linearly increasing function of depth, while water levels above 11.25 m are more harshly penalized by a fourth order polynomial to reduce the probability of overtopping the dikes at 13.4 m [Giuliani et al., 2016a]. This damage function was suggested by the Vietnamese Central Committee for Flood and Storm Control (CCFSC).

Minimizing flooding damages is a common approach to optimizing reservoir operations for flood control [Windsor, 1973; Needham et al., 2000; Lund, 2002; Malekmohammadi et al., 2009]. In past studies of flooding in the Red River basin, Vinh Hung et al. [2007] estimated damages using a 2-D hydrodynamic model of the delta mapping water levels to inundated area, while De Kort &

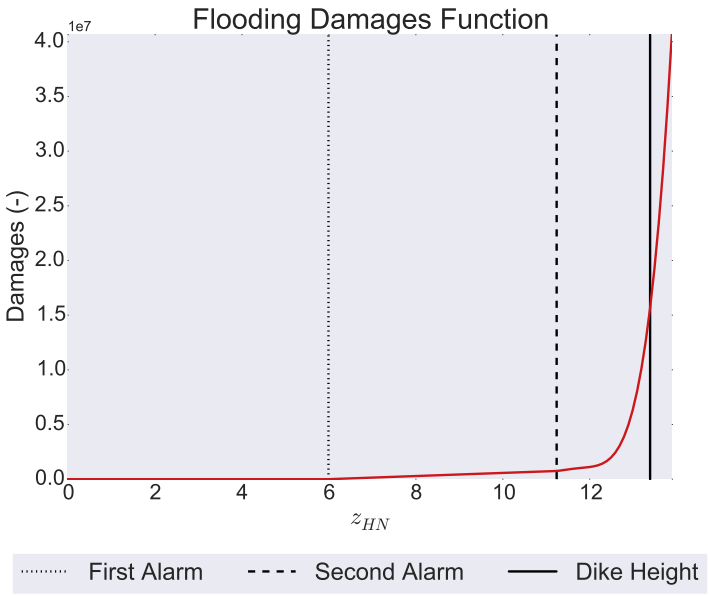

Figure 4.3: Flood penalty function used to approximate damages at Hanoi. Below 6 m (dotted black line) there are assumed to be no damages. Between 6 m and 11.25 m (dashed black line), damages are assumed to be minor and linearly increasing with depth, but above 11.25 m they become severe and are modeled by a fourth-order polynomial. This shape is intended to keep water levels from breaching the dikes at 13.4 m (solid black line). The shape of the damage function and the alarm levels were elicited from stakeholders.

Booij [2007] used past flood recovery costs and flow rates to estimate a damage curve. In contrast, Castelletti et al. [2012a] concluded that minimizing estimated damages was not appropriate in the Red River because the delta is constantly changing as a result of dike breaching events and urban development.

When estimating actual damages is difficult or they are non-stationary as is the case here, it is common to instead create a flood penalty function that harshly penalizes high water levels (see e.g. Orlovski et al. [1984]; Needham et al. [2000]). This is the intent of the above function. Castelletti et al. [2012a]

take a similar approach by minimizing the average squared excess of 9.5 m at Hanoi, an alarm level chosen from Hansson & Ekenberg [2002]. The excesses are squared to reduce the total force on the levee, the driver of collapse, which increases with the square of the water level. Similarly, Le Ngo et al. [2007] minimize a weighted sum of squared maximum water levels at Hanoi and squared deviations of the HoaBinh reservoir level from its maximum each flood season. The piecewise fourth-order polynomial used here is intended to be extremely risk-averse.

Worst First Percentile (WP1) Formulation

In prior work in the Red River Basin, Giuliani et al. [In Press] solved the WC formulation of the problem as both a challenging computational benchmark application and to provide an initial understanding of the system's multi-sectoral tradeoffs. Subsequent to this effort, the authors re-evaluated the policies derived from the WC formulation on a larger set of out-of-sample streamflows and observed that the policies did not generalize well. For this reason, we explore an alternative risk-averse formulation here in which we minimize the worst first percentile across a 1000-member ensemble of 1-yr simulations rather than the absolute worst across a 50-member ensemble of 20-yr simulations. The motivations for this are two-fold: 1) the worst first percentile should be more stable than the worst case, as the worst case has a higher sampling variance and may be unbounded [Stedinger et al., 1993], and 2) aggregating objectives over 1-yr simulations and minimizing the worst first percentile across a 1000-member ensemble may better capture inter-annual variability. Aggregating objectives over 20-yr simulations and minimizing the worst case across a 50 member ensemble

as in the WC formulation may mask particularly bad years if several good years are also included.

In the WP1 formulation, we compute hydropower production and the squared deficit in expectation within each ensemble member's 1-yr simulation ($\Phi = \mathbb{E}_{365T}$), just as in the 20-yr simulations of the WC formulation. However, we then calculate J^{WP1}_{Hydro} and $J^{WP1}_{Deficit^2}$ as the worst first percentile of these averages across the 1000-member ensemble rather than the absolute worst ($\Psi = \text{quantile}_N\{\Phi, 0.01\}$ for J^{WP1}_{Hydro} and $\Psi = \text{quantile}_N\{\Phi, 0.99\}$ for $J^{WP1}_{Deficit^2}$). In the WP1 formulation, we also reframe how the flooding objective is calculated within each 1-yr simulation so that the objective values are more semantically meaningful. While the fourth order polynomial is intended to be very conservative with respect to flooding, the value of the damage function is hard to comprehend since it maps water levels to a dimensionless number, not a monetary value. If stakeholders are trying to weigh the tradeoff between two solutions, it is unclear how much better or worse one solution does with respect to the other in terms of flooding based on their objective values. Motivated by Hashimoto et al. [1982], we partition the flooding objective into two components: resilience, $J^{WP1}_{Recovery}$, and vulnerability, J^{WP1}_{Flood}.

Hashimoto et al. [1982] argue that objectives related to the mean and variance of benefits commonly used in engineering do not capture other important aspects of system performance during drought, peak demands or extreme weather. They propose that system failures instead be evaluated in terms of their frequency (reliability), duration (resilience) and severity (vulnerability). These objectives have since been widely adopted in the water resources literature, primarily with respect to failures in meeting demand [Moy et al., 1986;

Bayazit & Ünal, 1990; Maier et al., 2001; Fowler et al., 2003; Kjeldsen & Rosbjerg, 2004; Asefa et al., 2014]. Several of these studies have also found synergies between reliability and resilience, but strong tradeoffs between each of these and vulnerability Hashimoto et al. [1982]; Moy et al. [1986]; Bayazit & Ünal [1990], suggesting it is worthwhile to consider vulnerability and either reliability or resilience in a multi-objective optimization, as we do here.

In this study, we quantify flood resilience within each 1-yr simulation using its inverse, measured as the average time to recovery after the water level at Hanoi exceeds 6 m ($\Phi = \mathbb{E}_{365T}$). We quantify flood vulnerability as the maximum annual water level in excess of 11.25 m ($\Phi = \max_{365T}$). These two thresholds are based on the same cutoffs used to define the piecewise polynomial function used to estimate flood damages in the worst case formulation (see Figure 4.3). Since the WP1 formulation minimizes the worst first percentile across a 1000-member ensemble of 1-yr simulations ($\Psi = \text{quantile}_N\{\Phi, 0.99\}$ for both J^{WP1}_{Flood} and $J^{WP1}_{Recovery}$), the flood vulnerability objective, J^{WP1}_{Flood}, is equivalent to minimizing the amount by which the 100-yr flood exceeds 11.25 m. Unlike the WC flood damages objective, which we do not constrain because it is unclear what is an acceptable level of dimensionless damages, we constrain the WP1 flood vulnerability objective to be less than 2.15 m (the difference between the second alarm level and the dike height) under the assumption that stakeholders would like to be protected to at least the 100-yr flood level. The resilience objective, $J^{WP1}_{Recovery}$, is intended to keep water levels at Hanoi persistently low in order to reduce the sustained pressure on the dikes. It is expected that these two flood objectives will conflict; maintaining low water levels at Hanoi may require higher storages in the reservoirs, reducing their capacity to capture large floods, putting Hanoi at risk of higher maximum water levels.

Expected Value (EV) Formulation

The WC and WP1 formulations both assume a risk-averse operator who is concerned with the tails of the distribution of each objective. However, optimizing to the tails often requires sacrifices in the mean [Beyer & Sendhoff, 2007]. In the EV formulation, we assume a risk neutral operator who is primarily concerned with average performance, representing common practice in water resources optimization problems (for examples, see reviews by Yakowitz [1982]; Yeh [1985]; Labadie [2004] and sources cited therein). Under the EV formulation, we quantify J^{EV}_{Hydro}, $J^{EV}_{Deficit^2}$ and $J^{EV}_{Recovery}$ as the expected annual hydropower production, squared deficit and recovery time for water levels over 6 m at Hanoi, respectively, calculating their averages across a 1000-member ensemble of 1-yr simulations ($\Phi = \mathbb{E}_{365T}$ and $\Psi = \mathbb{E}_N$). We do not change the flood vulnerability objective and constraint from the WP1 formulation, though, as flooding is not a concern in the average year; it is only the extremes that put the city of Hanoi at risk and consequently need to be minimized.

Expected Value and Standard Deviation of Hydropower (EV&SD_H) Formulation

The final formulation we test can be viewed as a compromise between the risk-averse WP1 formulation and the risk-neutral EV formulation with a specific focus on the inter-annual variability of hydropower production. In the classical robust optimization literature, it has long been recognized that there is often a direct conflict between the mean and variance of stochastic performance measures [Taguchi, 1986]. Knowing this, operators may be willing to trade off exceptionally high years of hydropower production if operations can be dis-

covered that reduce their exposure to drought-driven losses in production. In the water resources literature, these concerns have been addressed by including measures of variability in addition to expectation, either as an additional objective in a multi-objective optimization problem [Kawachi & Maeda, 2004; Reed & Kasprzyk, 2009], as a constraint [Kasprzyk et al., 2012], or as part of a weighted single objective function [Watkins & McKinney, 1997; Ray et al., 2013]. Here we take the first approach and explicitly quantify the tradeoff between maximizing mean hydropower performance and minimizing the variability about that mean by adding an objective to the EV formulation, $J^{EV\&SD_H}_{Hydro\ Std}$, to minimize the standard deviation in average annual hydropower production ($\Phi = \mathbb{E}_{365T}$ and $\Psi = \text{std}_N$). All other objectives and constraints are the same as in the EV formulation.

Table 4.4 provides a summary of the objective calculations from each formulation. For a more detailed, mathematical description of the objectives from each formulation, see Appendix A.

4.4.2 Formulation of Operating Policies

In order to optimize the complex objective functions defined for each of the four rival framings described in Section 4.4.1, we need to specify an operating policy for each of the reservoirs. Several operating policies, or rule curves, have been proposed in the water resources literature. The most basic is a linear operating policy [Revelle et al., 1969; Hashimoto et al., 1982; Guariso et al., 1986; Oliveira & Loucks, 1997] in which releases are described by a linear function of inputs, often storage plus inflow. Recently, more flexible, non-linear functions have been used, such as artificial neural networks (ANNs) [Raman & Chan-

Table 4.4: Within-ensemble aggregators and across-ensemble noise filters for objective calculations under each formulation

Objective	Φ (within-ensemble aggregator)	Ψ (across-ensemble noise filter)
J^{WC}_{Hydro}	$\mathbb{E}_{365T}$	$\min_N$
J^{WP1}_{Hydro}		quantile$\{\Phi, 0.01\}$
J^{EV}_{Hydro}		$\mathbb{E}_N$
$J^{WC}_{Deficit^2}$	$\mathbb{E}_{365T}$	$\max_N$
$J^{WP1}_{Deficit^2}$		quantile$\{\Phi, 0.99\}$
$J^{EV}_{Deficit^2}$		$\mathbb{E}_N$
J^{WC}_{Flood}	$\mathbb{E}_{365T}$	$\max_N$
J^{WP1}_{Flood}	$\max_{365T}$	quantile$\{\Phi, 0.99\}$
$J^{WP1}_{Recovery}$	$\mathbb{E}_{365T}$	quantile$\{\Phi, 0.99\}$
$J^{EV}_{Recovery}$		$\mathbb{E}_N$
$J^{EV\&SD_H}_{HydroStd}$	$\mathbb{E}_{365T}$	std_N

dramouli, 1996; Neelakantan & Pundarikanthan, 2000; Cancelliere et al., 2002] and radial basis functions (RBFs) [Giuliani et al., 2016b,a; Giuliani & Castelletti, 2016] that do not require an assumed mathematical form but can universally approximate a variety of functional shapes. Giuliani et al. [2016b] compare operating policies optimized for Hoa Binh with EMODPS using ANNs and RBFs and find that ANNs tend to overfit to the stochastic simulations they are trained on, generalizing less well when re-evaluated out-of-sample. For this reason, we parameterize the operating policies of the four reservoirs with RBFs.

As shown in equation 4.12, the RBF-based representation of operational poli-

cies prescribe releases, u_t^k (normalized on [0,1]), from the k-th reservoir at time t as a function of B time-varying inputs, x_{t-1} (normalized on [0,1]):

$$u_t^k = \sum_{i=1}^{A} w_i^k \exp\left(- \sum_{j=1}^{B} \frac{((x_{t-1})_j - c_{j,i})^2}{b_{j,i}^2} \right) \tag{4.12}$$

where $(x_{t-1})_j$ is the normalized value of the j-th input at time $t-1$, A is the number of RBFs, w_i^k is the weight of the i-th RBF associated with the k-th reservoir, and $c_{j,i}$ and $b_{j,i}$ are the centers and radii, respectively, of the i-th RBF associated with the j-th input. Due to physical constraints, the actual release from reservoir k at time t, r_t^k, is not always the same as the un-normalized policy-prescribed release. If there is insufficient water to release the un-normalized value of u_t^k, only the available water is released, and if there is insufficient storage capacity, s_{cap}^k, to allow only releasing the un-normalized value of u_t^k, the excess is spilled.

The centers, radii and weights of the RBF policies compose the decision variables, θ, optimized by the MOEA (see Equation 4.9):

$$\theta = \begin{bmatrix} c_{i,j} \\ b_{i,j} \\ w_i^k \end{bmatrix} \text{with } i = \{1, ..., A\},\ j = \{1, ..., B\} \text{ and } k = \{1, ..., M\}. \tag{4.13}$$

where $c_{i,j} \in [-1, 1]$, $b_{i,j} \in [0, 1]$ and $w_i^k \in [0, 1]$ with $\sum_{i=1}^{A} w_i^k = 1\ \forall\ k$. For M outputs (reservoirs), this corresponds to $A(M + 2B)$ decision variables. We model the releases at M=4 reservoirs using A=11 RBFs and B=6 inputs, where the inputs are the storages at each reservoir on the previous day, the total system inflow on the previous day, and the day of the year: $x_{t-1} = \{s_{t-1}^{SL}, s_{t-1}^{HB}, s_{t-1}^{TB}, s_{t-1}^{TQ}, q_{t-1}^{TOT}, t\}$ where $q_{t-1}^{TOT} = q_{t-1}^{Da} + q_{t-1}^{Da,lat} + q_{t-1}^{Thao} + q_{t-1}^{Chay} + q_{t-1}^{Gam}$. This represents a total of 176 decision

variables.

4.4.3 Multi-Objective Optimization

We use the Multi-Master Borg MOEA [Hadka & Reed, 2015], described in Ch 2 Section 2.1.1, to optimize the operating policies of the four reservoirs in the Red River basin. We use Multi-Master Borg with 16 islands and run 5 random algorithm trials, or seeds, of 400,000 function evaluations per island. Visual inspection of search progress indicated that this was sufficient, as progress had reached an asymptote of diminishing returns with little variability across the 5 random seeds. The epsilon dominance archiving used in Borg requires that users specify levels of precision for each objective below which they are indifferent to differences in performance. Table 4.5 shows the values of epsilon (or significant precisions) used for each objective in this study. Giuliani et al. [In Press] optimized the worst case formulation on the Texas Advanced Computing Center (TACC) Stampede Cluster (https://www.tacc.utexas.edu/stampede/) using 512 cores per island and 400,000 computational hours, and we optimized the remaining formulations on the Blue Waters supercomputer (http://www.ncsa.illinois.edu/enabling/bluewaters) using 1024 cores per island and a total of 1.7 million computational hours. For each formulation, we obtained approximate Pareto sets by combining and re-sorting the best solution sets found by each seed.

Table 4.5: Epsilons used for multi-objective optimization under each problem formulation.

Objective	WC Formulation	WP1 Formulation	EV Formulation	EV&SD_H Formulation
J_{Hydro}	0.1	0.5	0.5	0.5
$J_{Deficit^2}$	5.0	25.0	25.0	25.0
J_{Flood}	275.0	0.05	0.05	0.05
$J_{Recovery}$	–	0.5	0.5	0.5
$J_{Hydro\ Std}$	–	–	–	0.05

4.4.4 Diagnostic Verification of Optimized Policies

The final step we have added to the EMODPS framework in this paper is the diagnostic verification of the optimized policies. We evaluate the performance of the control policies in two ways: 1) by re-evaluating their performance over out-of-sample streamflow ensembles, and 2) by analyzing their multi-reservoir management behavior and the system dynamics that result from operating with policies that emphasize different objective preferences from the rival problem formulations. In the first step, we re-evaluate all of the policies over a second ensemble of stochastic streamflows that is 100 times larger than the ensemble used during optimization. This corresponds to a 5000-member ensemble of 20-yr simulations for the WC formulation, and a 100,000-member ensemble of 1-yr simulations for all other formulations. If the solutions achieve similar objective values in the re-evaluation as in optimization, then both the objectives and policies are stable, so we can trust our representation of policy performance.

In this study, we also re-evaluate the solutions from each problem formu-

lation on the objectives from each of the other formulations using the set of streamflows from both the optimization and validation. This allows us to visualize the regrets associated with risk-averse vs. risk-neutral objectives to determine the costs in expectation of optimizing to the worst case and vice versa. We can also see if the stability of a particular objective depends on whether or not that objective was included in the optimization. This can be used to diagnose whether poor performances in re-evaluation are due to the policies being overfit, or to the objectives themselves being inherently unstable.

Finally, the second step of the diagnostic verification is to analyze the behavior and consequences of operating with the optimized policies. In this study we select the best solution on each objective within each formulation to see how operations differ as a function of both preference and formulation. Analyzing the operations and state behavior of the system opens the black box of the policy function, providing insights into how the policies are able to achieve the objective values that they do.

4.5 Results and Discussion

Here, we present the results of the EMODPS policy optimization and diagnostic verification. In Section 4.5.1, we show the multi-objective tradeoffs that emerge from the Pareto-approximate solutions discovered under each of the four candidate problem formulations (see Tables 4.2–4.4). In Section 4.5.2, we assess how well these solutions generalize on out of sample streamflows, and re-evaluate the competing formulations in each other's spaces (i.e., we re-simulate the control policies found under each problem formulation to calculate their perfor-

mance on the objectives from all of the other formulations). In Section 4.5.3, we investigate how preference and problem formulation shape the behavior of the optimized control policies. Lastly, in Section 4.5.4 we illustrate how these control policies affect downstream flood dynamics at Hanoi.

4.5.1 Rival Representations of Red River Tradeoffs

The best known approximations of the Pareto optimal sets discovered for each of the four candidate problem formulations of the Red River test case are shown using parallel axes plots in panels a-d of Figure 4.4. In these plots each shaded line corresponds to an operating policy for the system's four reservoirs that intersects each vertical axis at the value it achieves for the objective that axis represents. Solutions from the WC formulation are shown in red (panel a), from the WP1 formulation in blue (panel b), the EV formulation in green (panel c), and the EV&SD_H formulation in purple (panel d). We use this color scheme to distinguish the candidate Red River problem formulations in all subsequent figures. In Figures 4.4a–4.4d, all of the vertical axes have been oriented such that the optimal direction is downward. All lines have been shaded according to their performance on the hydropower objective, with darker shades representing greater production. Consequently, theoretical ideal solutions in each of the spaces plotted in Figure 4 would be dark shaded horizontal lines intersecting the bottom of each axis.

In parallel axes plots, intersecting lines between pairs of vertical axes designate tradeoffs between those two objectives, as superior performance in one objective comes at the expense of inferior performance in another. In Figures

4.4a–4.4d, one can also observe tradeoffs between the hydropower objective and objectives oriented on non-adjacent axes through shading. For visual clarity, we have thinned the four Pareto approximation sets illustrated in Figure 4.4 by re-sorting them with larger epsilons to attain representative sets of approximately 100 solutions in each panel that fully span the tradeoffs discovered in this study.

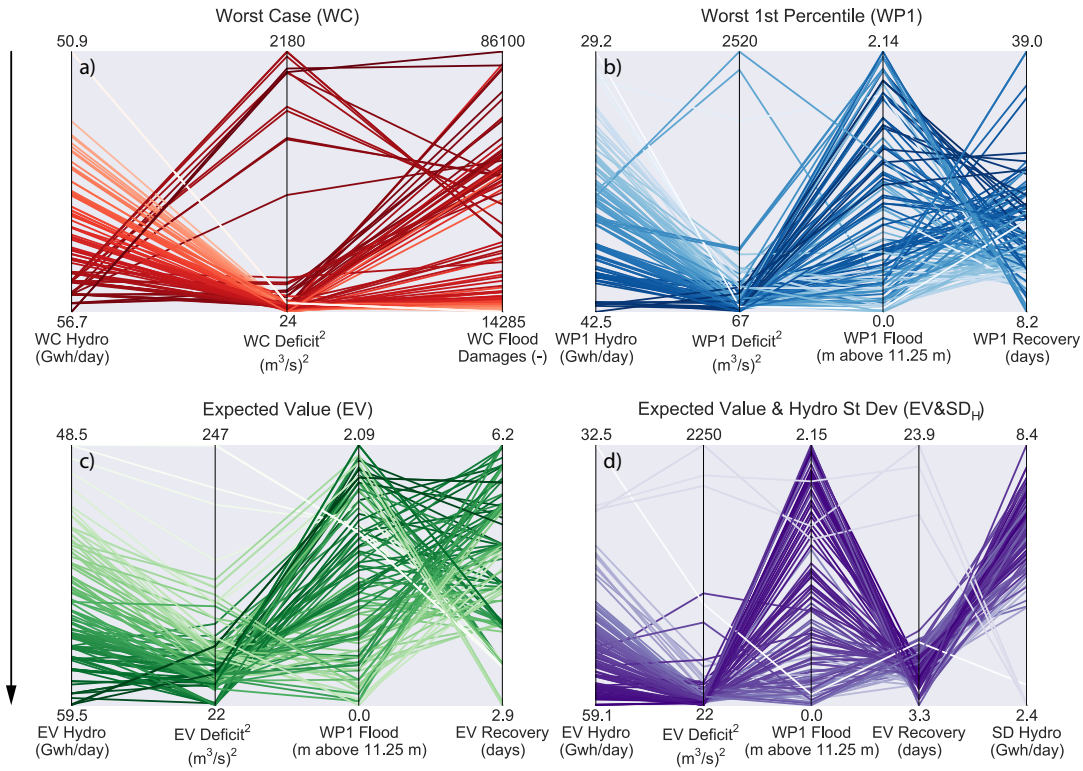

Figure 4.4: Approximate Pareto sets from the WC formulation (panel a), WP1 formulation (panel b), EV formulation (panel c) and EV&SD$_H$ formulation (panel d). Each axis represents a different objective from that formulation and each shaded line a solution in the approximate Pareto set. All lines are shaded by their performance on the hydropower objective, with darker shades representing better performance, and all axes are oriented such that the optimal direction is down. An ideal solution would therefore be a dark horizontal line across the bottom of the axes.

Across Figures 4.4a–4.4d, the major tradeoff of note is between hydropower and flooding, which can be seen by the inversion of the color gradients along these two axes. This conflict results because high storages favor hydropower production, while low storages favor flood protection. There is also a weak, but non-linear tradeoff between the squared water supply deficit and flooding, as well as between the squared water supply deficit and hydropower, which can be seen by the crossing diagonal lines between these adjacent axes across all formulations. For the formulations that include both the flood vulnerability and resilience objectives (WP1 in panel b, EV in panel c and EV&SD_H in panel d), crossing diagonal lines between these axes indicate that there is a strong tradeoff between these two objectives. This suggests that in order to reduce maximum flood levels, moderately high flood levels must be maintained, resulting in more sustained pressure on the dikes.

Interestingly, the shapes of the tradeoffs that emerge for the WP1 (panel b) and EV formulations (panel c) are similar, suggesting that these conflicts are not quantile-dependent. However, this does not imply that there is not a tradeoff between average performance and the stability of performance. As can be seen by the inversion of colors along the axes for expected hydropower and standard deviation of hydropower in the EV&SD_H formulation (panel d), these objectives strongly conflict. In particular, the solutions with the lowest standard deviation in annual hydropower production have similar average hydropower production to the worst first percentile hydropower production observed in the WP1 formulation. This severe degradation in average performance occurs with the squared deficit and recovery time objectives as well.

4.5.2 Verification of Control Policies

As summarized in Section 4.4.4, our first diagnostic verification step is to re-evaluate the solutions from each of the problem formulations in the objective spaces of all of the other formulations using both the synthetic streamflow ensembles to which they were optimized and an out-of-sample validation set with 100 times as many ensemble members. The results of the re-evaluation are shown in Figure 4.5, where each panel represents a different objective. The first row of Figure 4.5 shows the WC objectives, the second row the WP1 objectives, and the bottom row the EV and EV&SD_H objectives. Within each panel, points representing the multi-reservoir control policies are positioned along the x-axis at their objective values achieved in optimization, and along the y-axis at their objective values in validation. Each panel is oriented such that the lower left corner represents the most favorable direction. Solutions that achieve similar values in optimization and validation will fall near the 1:1 line, shown by a black dashed line. If solutions lie above this line, their performances degraded in re-evaluation, and if they lie below, their performance improved.

Figure 4.5 provides several insights into the stability of each of the problem formulations, as well as their inherent biases, some of which were intended and others of which were not. Beginning with the WC objectives in the first row (panels a-c), the most prominent observation is the instability of these objectives, as nearly all of the solutions lie above the dashed 1:1 line, indicating degrading performance in re-evaluation. The degradation is particularly bad on the flood damages objective (panel c) due to the fourth order polynomial used to approximate damages when the water level at Hanoi exceeds 11.25 m. The fact that all of the solutions degrade similarly whether or not they were

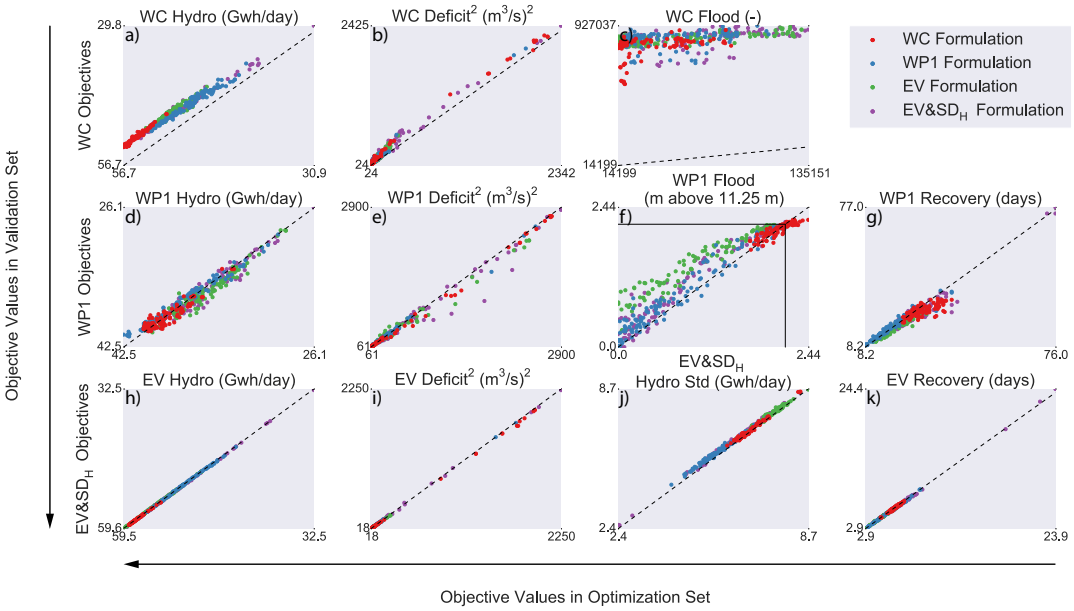

Figure 4.5: Validation of Pareto-approximate solutions from each formulation. Each dot represents a different control policy positioned along the x-axis at its objective value over the optimization set of streamflows, and along the y-axis at its value over an out-of-sample validation set with 100 times as many ensemble members. Solutions with stable performance between optimization and validation lie near the black dashed 1:1 line. WC solutions are shown in red and their objectives in the top row (panels a-c), WP1 solutions in blue and their objectives in the middle row (panels d-g), EV solutions in green and EV&SD$_H$ solutions in purple, with their objectives in the bottom row (panels h-k). All panels are arranged such that the optimal direction is toward the lower left corner. The black solid line in panel f represents the dike height at Hanoi.

optimized under this problem formulation indicates that the degradation is not due to overfitting the radial basis functions that define the control policies but instead due to the worst case formulation of the flood objective itself. This is not surprising, as the worst case is often unbounded and therefore likely to worsen as the sample size increases. This may not be problematic, for example, in the case of the hydropower objective (panel a) where the ordering of the so-

lutions from most favorable to least favorable does not considerably change in re-evaluation. However, Figure 4.5c indicates that this is not the case on the WC Flood objective, as some solutions that do relatively poorly over the optimization set do relatively well over the validation set, suggesting that the original representation of the tradeoffs from optimization may not be accurate.

This difference in the magnitude of degradation across objectives is likely due to their distributions. Hydropower has a bounded minimum production, and the squared deficit a bounded maximum. Flooding, however, is unbounded, and the fourth order polynomial used to estimate damages in this study results in an extremely fat tail. Optimizing the worst case of an unbounded, non-linear objective is particularly difficult, as its value degrades severely with increasing sample sizes, resulting in greater noise than signal. Compounding this difficulty is the non-uniqueness of calculating damages in expectation; a solution with frequent small floods may have similar expected damages to a solution with infrequent large floods. As a result, the performance of the optimized operating policies is highly sensitive to the streamflows they are optimized to, making it difficult to reliably compare alternative solutions. These results call into question the effectiveness of using min-max objectives for robust optimization, as is often recommended in the literature [Wald, 1992; Beyer & Sendhoff, 2007], at least when the objective's performance is noisy and unbounded.

Fortunately, the second row of Figure 4.5 (panels d-g) shows that the WP1 objectives are much more stable, as the points cluster around the 1:1 line, with some solutions degrading in re-evaluation and others improving. Interestingly, though, unlike on the WC objectives for which instability is independent of for-

mulation, there is evidence of formulation-dependent instability on the WP1 Flood objective (panel f), as the WC and WP1 solutions do not systematically degrade, while the EV and EV&SD_H solutions do. Since the WC formulation is the only formulation that does not include the WP1 Flood objective, the stability of these solutions on this objective suggests that the formulation of the objective itself does not cause instability. The degradation of the EV and EV&SD_H solutions therefore must be due to overfitting of the control policies to the streamflows over which they were optimized. While the degradation of the EV&SD_H solutions on the WP1 Flood objective is less than for the EV solutions, it is still greater than for the WP1 solutions. This suggests that optimizing to the worst first percentile across all objectives results in fairly stable policies from year to year, while including expectation in the formulation results in more variable inter-annual performance on all objectives, not just the objectives optimized in expectation. Adding an objective related to inter-annual variability such as the standard deviation in hydropower production enables more stable performance on WP1 objectives, but not as stable as when optimizing to the worst first percentile across all objectives.

Another noteworthy observation from the second row of Figure 4.5 (panels d-g) is that the WC solutions lie far from the ideal point on both the WP1 flood resilience (WP1 Recovery, panel g) and flood vulnerability (WP1 Flood, panel c) objectives. Performance is particularly bad on the WP1 Flood objective, as several of the WC solutions lie above the black line drawn at 2.15 m, indicating that these solutions do not provide protection to the 100-yr flood. While the WC formulation of the flooding objective was intended to be especially risk averse by modeling damages above 11.25 m with a fourth-order polynomial and minimizing the worst case performance across the ensemble, the calculation of

expected damages over 20 years enables severe flood events to be masked by drier years with little to no damages. Instead of forcing the discovery of more conservative flood policies, the fourth-order polynomial only serves to make the objective values unstable, as shown in the row above (panel c). Ironically, the WC solutions perform well on the hydropower objective from every formulation (panels a, d and h), though, indicating that despite the harsh flood penalty this formulation actually favors optimizing hydropower production over flood protection.

On the contrary, the WP1 formulation of the flood objective, which focuses solely on large events, allows for the discovery of policies that provide protection at the 100-yr level under both the optimization and validation streamflow ensembles. Unlike the non-unique damage function, the flood resilience and vulnerability objectives are able to distinguish between flooding caused by small, frequent events (captured by the resilience objective) and large, infrequent events (captured by the vulnerability objective). Additionally, despite the WP1 Flood objective only having the equivalent of a linear penalty on the maximum water level at Hanoi as opposed to the fourth order polynomial on damages, the WP1 solutions still obtain low flood damages according to the WC Flood objective (panel c). This linear penalty is able to reduce the noise in the tails of the flooding objective while the worst first percentile bounds its performance, resulting in a stable objective that is able to simultaneously minimize expected damages. This greater flood protection does come at a cost, however, as the WP1 solutions do not do well on the WC Hydro objective (panel a).

These results suggest that minimizing flood damages in expectation may be ill-advised because it is difficult to know *a priori* whether or not doing so will be

effective in reducing severe floods, especially when damages are uncertain or non-stationary and need to be approximated by a non-linear penalty function. However, damage functions may still be useful for comparing optimized solutions *a posteriori*. For example, one could re-simulate alternative non-dominated policies over a larger ensemble of streamflows to estimate the maximum water level and corresponding damages of more extreme events like the 500-yr flood, which one may not be able to estimate precisely over computationally tractable ensemble sizes for optimization. This can also provide stakeholders with a more realistic representation of the non-linear mapping of stage to damages without suffering the negative consequences of optimizing to a noisy, non-linear objective function.

The final row of Figure 4.5 (panels h-k) shows the performance of all of the solutions on the EV and EV&SD_H objectives. With most of the solutions lying nearly on the 1:1 line, these objectives are the most stable due to the smaller sampling variability of the mean and standard deviation than quantiles in the tails [Stedinger et al., 1993]. This row also highlights the regret associated with optimizing to the worst first percentile, as the solutions from the WP1 formulation do poorly on the EV Hydro objective (panel h). Regret in the opposite direction is not as severe, as the EV solutions do fairly well on the WP1 Hydro objective in the row above (panel d). However, while the WP1 solutions sacrifice EV Hydro performance, they do fairly well on the EV&SD_H Hydro Std objective (panel j) despite not explicitly including it in optimization. Additionally, while the greater stability in inter-annual hydropower production enabled by the WP1 formulation does degrade performance in EV Hydro, the degradation is not as severe as for the best Hydro Std solutions from the EV&SD_H formulation (panel h), suggesting that optimizing to the worst first percentile is

a more effective way to reduce variability in performance without excessively sacrificing average performance.

Ray et al. [2013] and Watkins & McKinney [1997] draw similar conclusions from a water supply optimization problem where including standard deviation in costs as part of the objective function led to more reliable and sustainable results with respect to shortages, but increased vulnerability. Noting that minimizing variance penalizes outcomes both above and below the mean, Ray et al. [2013] and Watkins & McKinney [1997] re-formulated their objective to incorporate a penalty for squared positive cost deviations from a target, a modification inspired by Takriti & Ahmed [2004]. Similarly, the worst first percentile only penalizes outcomes below the mean for maximization objectives and above the mean for minimization objectives. Both of these alternative objective formulations are able to achieve more stable policies without excessively compromising mean performance, as Ray et al. [2013] and Watkins & McKinney [1997] found that the cost deviations penalty resulted in less variable direct costs without increasing vulnerability with respect to shortages.

The conclusions from Figure 4.5 highlight the importance of evaluating rival framings of how stakeholder objectives should be translated into quantitative performance measures when designing water resources systems optimization problems. While engaging stakeholders in the problem formulation process is important for ascertaining their qualitative objectives and such participatory modeling has improved decision-making for water resources applications [Palmer et al., 1990], this process alone does not guarantee the design of effective policies, as it is not obvious *a priori* what the best mathematical characterization of stakeholder risk preferences will be. For example, if we had only translated

stakeholder objectives into the metrics utilized by the WC formulation, policies optimized to that formulation would seem acceptable. However, since multiple quantitative translations of stakeholder objectives and preferences have been tested here, it is clear that these policies may not provide sufficient flood protection, a consequence that would have otherwise gone undiscovered. Thanks to this rival framings analysis, stakeholders in the Red River choosing from alternative policies can now better see the hydropower production, deficit and flood levels they would expect both on average and once every 100 years under policies from different formulations. Further analysis illustrating simulated behavior with different policies could better illustrate when and how severe periods of flooding, drought and high or low hydropower production might be.

4.5.3 Impacts of Problem Framing and Preference on Control Policies

For the second step of the verification process, we have selected solutions from different preference regions within each formulation's Pareto approximate set to analyze more deeply. First, we examine the reservoir operations that result from different policies to determine how each solution is able to achieve its objective values. In Figure 4.6, we show these operations for Hoa Binh, the most important reservoir for flood protection of Hanoi. The first row of Figure 4.6 (panels a-d) shows the average annual storage trajectories and the second row (panels e-h) the average annual release trajectories for the solutions from each problem formulation that perform best on each of the objectives included in that formulation. All of the trajectories are drawn from the beginning of the monsoon in

May to the end of the dry season in April. To reduce the influence of noisy stochastic daily flows, we smoothed the average storage and release trajectories using 7-day and 30-day moving averages, respectively. Note that differences in releases look much smaller than differences in storage because of the units and scales of the axes; releases are plotted in m^3/s, so small visual differences in flow rate actually correspond to substantial volumetric differences over an entire day.

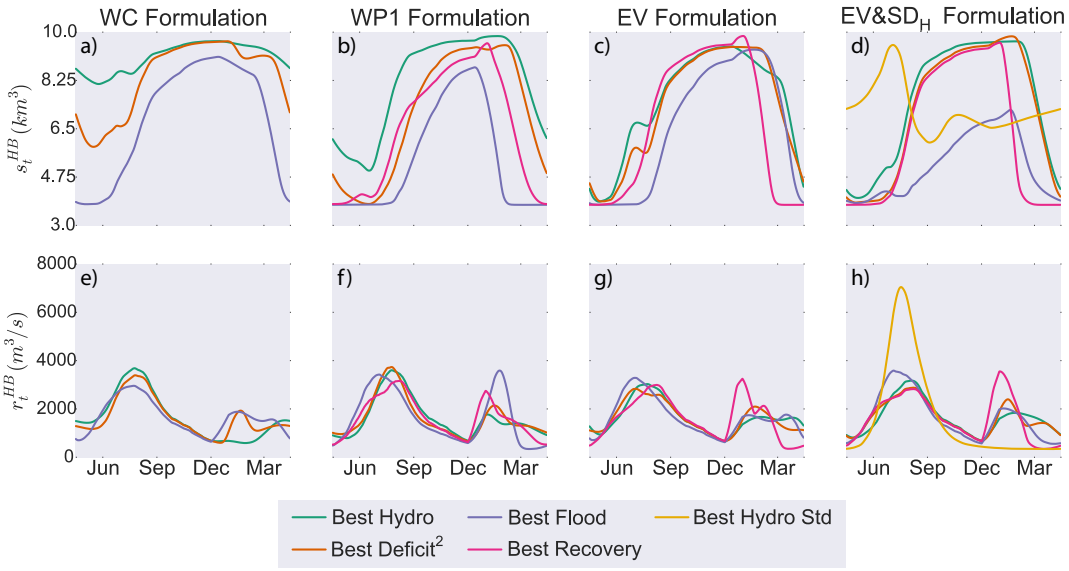

Figure 4.6: Average annual storage (top row, panels a-d) and release (bottom row, panels e-h) trajectories from May to April at Hoa Binh, the most important reservoir for flood protection. Solutions maximizing hydropower generally maintain the highest storage levels, those minimizing flooding the lowest, and those minimizing the deficit or flood recovery time, intermediate levels. All solutions increase releases during the monsoon when there is more water, and all but the solution with the lowest inter-annual variability in hydropower production increase releases again at the beginning of the calendar year to meet the agricultural water demand.

Across nearly all solutions and problem formulations, the annual cycle of

storages and releases is the same. At the beginning of the monsoon, storage is decreased or maintained at a constant low level to make capacity available for oncoming floods. This is achieved by increasing releases as the flood waters arrive. At the end of the monsoon in September and October, releases are decreased to increase storage for hydropower production during the dry season. Storage remains high throughout the dry season both to maximize hydropower production, and to ensure that there is enough water to meet the multi-sectoral water demand. At the beginning of the next calendar year, releases are increased again. These releases achieve two objectives simultaneously: they provide water to meet the spike in demand at the beginning of the planting season (see Figure 4.1a) and begin emptying the reservoir in anticipation of the oncoming monsoon.

The only policy that does not follow this shape is the solution with the lowest standard deviation in inter-annual hydropower production from the EV&SD_H formulation, shown in gold in panels d and h. This solution seeks to minimize inter-annual variability by maintaining moderately high storage year-round, never emptying the reservoir as much as the other solutions in case the next year is dry. This is achieved by letting the reservoir first fill when the monsoon rains arrive, while the other solutions try to maintain low storage in the reservoir at this time to make room for flood protection. Once the storage nears capacity, releases are sharply increased, maximizing hydropower production when it is guaranteed that water is available. At the end of the monsoon, releases are sharply decreased to prevent storage from dropping too low. This storage is then maintained throughout the dry season by releasing little to nothing, as water is hard to come by during the dry season, so stability is best maintained by releasing little. This is the reason this solution does so poorly with respect to

the squared water supply deficit and expected hydropower production.

For the other solutions though, there are consistencies across formulations in how the best solutions for each objective operate relative to one another. The best hydropower solutions, shown in green, maintain high storage levels year-round to maximize production. The high storages are maintained by releasing little water at the end of the dry season and waiting until later into the monsoon season to start increasing releases. On the contrary, the best flood solutions, shown in purple, maintain the lowest storage levels year-round, often being the last solutions to increase storage at the end of the monsoon and the first to increase releases in order to empty the reservoir at the end of the dry season. The earlier releases at the beginning of the monsoon season illustrate why there is a tradeoff between the flood resilience and vulnerability objectives; these releases increase the water level at Hanoi and increase the amount of time that it is above 6 m. However, they allow the reservoir to maintain lower storage and increase the probability of capturing large flood waves that might otherwise result in water levels above 11.25 m at Hanoi.

For the best recovery solutions, shown in pink, the opposite dynamic emerges. These policies increase releases later into the monsoon, reducing the amount of time that the water level is above 6 m at Hanoi. However, by increasing releases later, they reach higher average storages, decreasing their ability to capture large flood waves and increasing the probability of the water level at Hanoi exceeding 11.25 m. These policies result in average storage trajectories somewhere between the best hydro and best flood solutions. The storage trajectories of the best deficit solutions, shown in orange, also lie between these two, with the main difference from the best recovery solutions being that the

best deficit solutions increase releases (and consequently decrease storage) less sharply at the end of the dry season in anticipation of the monsoon. During this time of year, the storage trajectories of the best deficit solutions more closely track those of the best hydro solutions, maintaining higher storage longer to ensure enough water to meet demand.

Comparing solutions that perform best on the same objective from different formulations, the consequences of optimizing to one formulation over another emerge. Most notably, the solutions from the WC formulation (panels a and e) maintain higher storages year-round than their counterparts from the other formulations. This enables them to achieve greater hydropower production, but puts them at risk of not being able to capture large flood waves, explaining their strong performance across the hydropower objectives of every formulation, but poor performance on the WP1 Flood objective (see Figure 4.5).

Between the EV and WP1 formulations, there are only minor differences between corresponding solutions, with the exception of the best flood solution, which is much more conservative under the WP1 formulation than the EV formulation, maintaining lower storages year-round. It should be noted, though, that these are average trajectories. Figure 5 illustrates that the EV solutions exhibit much greater inter-annual variability than the WP1 solutions, suggesting that while these operations look similar on average, there may be greater differences from year to year. These figures also highlight the benefits of achieving stability through minimization of the worst first percentile rather than the standard deviation in hydropower production, as the WP1 policies are much more in line with typical operations than the solution with the best standard deviation from the EV&SD_H formulation.

While Figure 4.6 provides some useful insights into how different policies from each problem formulation operate the HoaBinh reservoir, it is useful to investigate joint operations of the system of four reservoirs to see how coordination is used to achieve the overall system objectives. In Figure 4.7, we show average annual storage trajectories at all four reservoirs for the best flood (left column) and best hydro (right column) solutions from each formulation. The left column therefore shows the trajectories of the solutions whose Hoa Binh trajectories are shown in purple in Figure 4.6, while the right column shows the trajectories of the solutions shown in green in Figure 4.6. One of the most striking observations from Figure 4.7 is that operations of the best flood solution from every formulation are nearly identical at Son La (panel a). Because Son La has the largest hydropower capacity and it is followed by Hoa Binh, the most important reservoir for flood protection, operations at Son La can afford to be less conservative with respect to flooding and can instead focus on maximizing its hydropower potential. Operations at Son La with the best hydro solutions (panel b) are less similar across formulations, but still much more similar than at the other reservoirs. Only the best hydro solution from the WC formulation distinguishes itself by drawing down the reservoir level at Son La later in the dry season, increasing the risk of flooding, but maximizing hydropower production longer.

Figure 4.7 also illustrates how different solutions coordinate operations in different ways. The best flood solution from the WP1 formulation maintains lower storages at Hoa Binh (panel c), the most important reservoir for flood protection, than the best flood solutions from the WC and EV formulations. To compensate for the hydropower potential lost by storing less water at Hoa Binh, this solution maintains higher storage levels at Tuyen Quang (panel g)

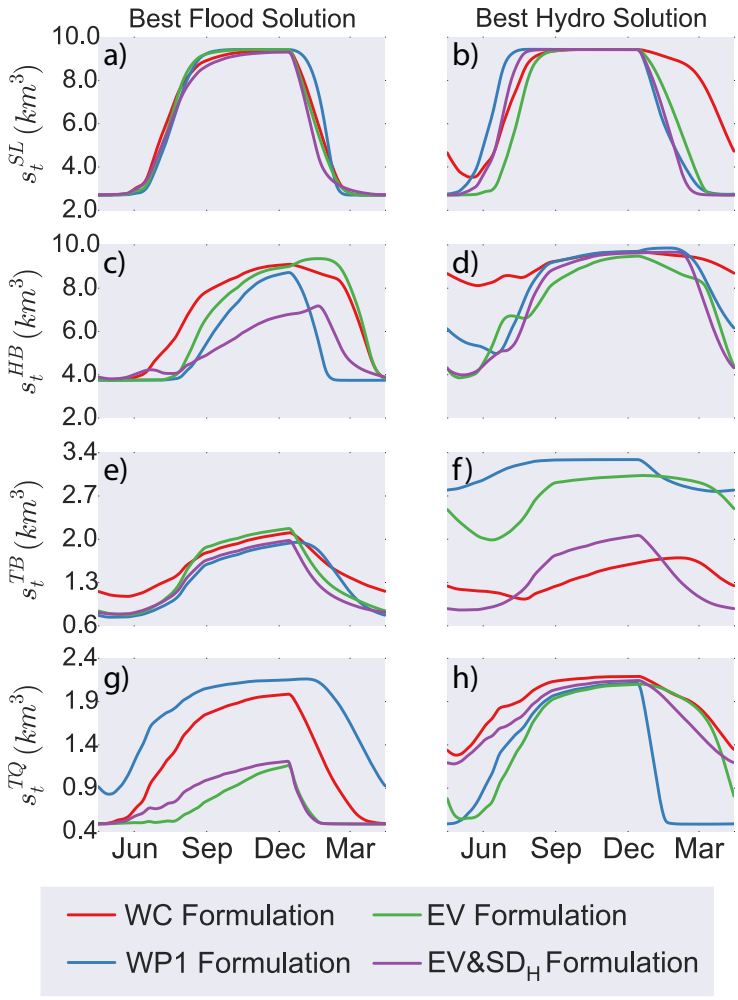

Figure 4.7: Storage trajectories at all four reservoirs when operating with the best flood (left column, panels a, c, e and g) and hydro solutions (right column, panels b, d, f and h) from each formulation. Panel a indicates that operations at Son La vary little across formulations, while the other panels show how coordination across reservoirs varies by formulation and preference.

than the best flood solutions from all other formulations. This is a much smaller reservoir, making its flood reduction potential minimal and the consequences of not being able to catch a flood wave there low. While this somewhat makes up for the hydropower lost at Hoa Binh, the smaller hydropower capacity of

Tuyen Quang is not nearly enough to fully compensate, explaining why the WP1 solutions do not do well on the EV and WC Hydro objectives.

On the opposite spectrum, the best hydro solution from the WC formulation maintains very high storage levels at Hoa Binh year-round (panel d). Because this leaves little room to catch any large flood waves, this solution attempts to make up for it by maintaining lower storage levels at Thac Ba (panel f), another small reservoir, than the best hydro solutions from the WP1 and EV formulations. However, because Thac Ba has much less storage capacity, this does not fully prevent high risk flood events, explaining the poor performance of the WC solutions on the WP1 Flood objective.

4.5.4 Impacts of Problem Framing and Preference on Flood Dynamics

In addition to examining the operations at each of the reservoirs, it is informative to visualize how these operations result in different responses downstream at Hanoi. For this analysis, we highlight solutions from the WC and WP1 formulations to distinguish the effects of these two variants of risk-averse optimization problems, and in particular two variants of flood control objectives. To illustrate a discrete number of policies spanning a wide range of preferences, we select the best hydro solution, best flood solution, and a compromise solution from each formulation. It has been long noted that multi-objective participatory planning in water resources systems is critical for discovering candidate compromise policies [Maass et al., 1962; Cohon & Marks, 1975; Matalas & Fiering, 1977; Haimes & Hall, 1977]. Recent advances in visual analytics have enhanced

the interactive and collaborative exploration of candidate compromise solutions using techniques such as brushing objectives on stakeholders' performance criteria [Basdekas, 2014; Dittrich et al., 2016; Herman et al., 2014; Huskova et al., 2016; Groves et al., 2016]. Analyzing the behavior of compromise solutions from different formulations and how they compare with extreme solutions, one can see how stakeholders will make better, more-informed decisions if choosing compromise solutions from a number of different problem formulations.

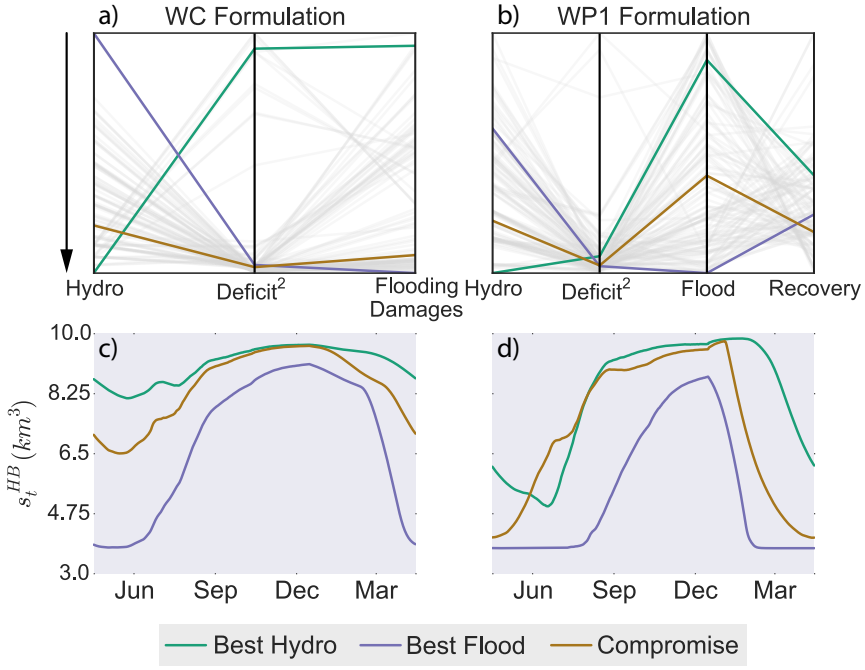

Figure 4.8: Best hydro solution (green), best flood solution (purple) and compromise solution (brown) from WC (left column, panels a and c) and WP1 (right column, panels b and d) problem formulations selected for further analysis. The top row (panels a and b) shows the location of these solutions in the objective space of their problem formulations, while the bottom row (panels c and d) shows their storage trajectories at Hoa Binh.

Figure 4.8 shows where the three selected solutions from each formulation lie with respect to each of the objectives on a parallel axis plot (panels a and b),

as well as how their storage trajectories differ at Hoa Binh (panels c and d). Generally speaking, the storage trajectories of the compromise solutions lie between those of the best hydro and best flood solutions. For each of these solutions, we use the simulations from the validation set of 100,000 years of synthetic inflows to estimate the probability density function (PDF) of the water level at Hanoi over time. Figure 4.9 shows these estimates in log space for each of the solutions, with high probabilities shaded red, moderate probabilities yellow, and low probabilities blue. A dotted line is drawn at the first alarm level of 6 m, a dashed line at the second alarm level of 11.25 m, and a solid line at the dike height of 13.4 m.

Across all solutions, the general shape of the time-varying PDFs is similar. The water level of the high probability density region in red increases from May to July due to the monsoonal rains. As the rains subside at the end of the season, the water levels begin to fall and low levels are maintained throughout the dry season until the beginning of the next calendar year. Then, in response to the high releases from the reservoirs to meet the agricultural water demand during the planting season (see Figure 4.1), water levels at Hanoi briefly spike and then fall again before the next monsoon season begins. The shape of the second spike varies by solution, illustrating the different ways that the water demand can be met and the reservoirs emptied in advance of the monsoon. The shape of the spike for the best WP1 flood solution is particularly interesting as there seems to be a bifurcation with two different release magnitudes depending on how much water needs to meet the agricultural water demand.

While the general shape of the time-varying PDFs is similar across solutions, there are some important differences. Comparing the three solutions from the

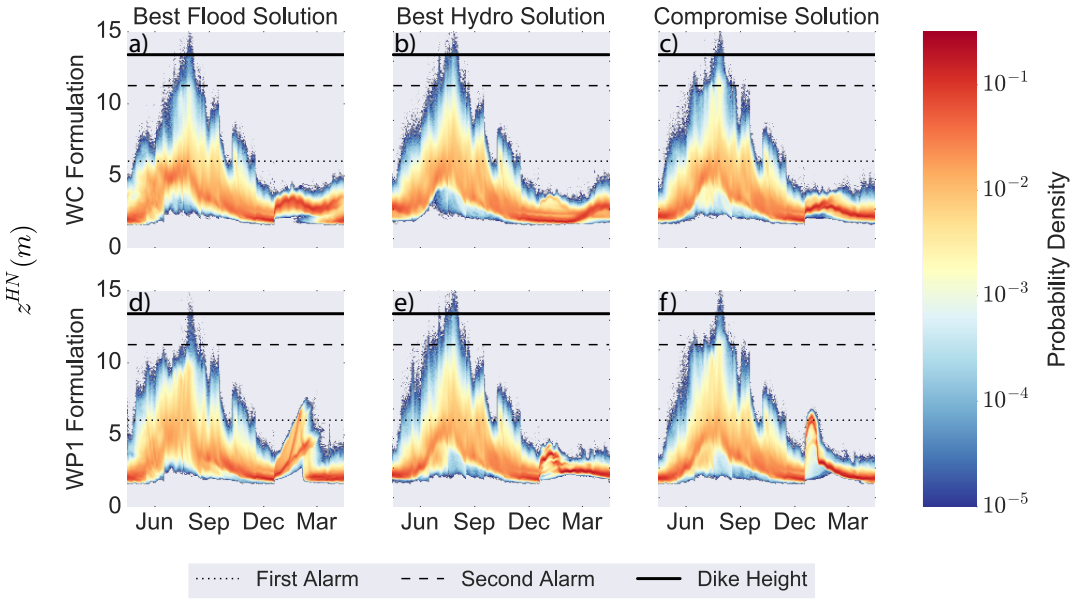

Figure 4.9: Probabilistic trajectories of the water level at Hanoi under three selected WC solutions (top row, panels a-c) and WP1 solutions (bottom row, panels d-f). The dotted line represents the first alarm level of 6 m, the dashed line the second alarm level of 11.25 m, and the solid line the dike height of 13.4 m. The best WP1 flood solution (panel d) has a higher probability of crossing 6 m than the other solutions, exhibiting less resilience, but a lower probability of crossing 11.25 m, exhibiting less vulnerability. The probabilistic behavior of the compromise WP1 solution (panel f) lies between that formulation's best flood (panel d) and best hydro (panel e) solutions, while differences are harder to see for the best WC solutions (panels a-c). The y-axis on each chart has been cut off at 15 m, which some simulations exceed in extremely rare cases.

WC formulation in the top row (panels a-c), these differences appear to be minor, but across the WP1 solutions in the bottom row (panels d-f), noticeable differences emerge. The best flood solution from the WP1 formulation (panel d) actively attempts to reduce the peak water level at Hanoi by maintaining moderately high water levels throughout the monsoon season. In contrast to the best flood solution from the WC formulation (panel a), it has a much greater prob-

ability of crossing 6 m, but a lower probability of crossing 11.25 m. This again highlights the tradeoff between flood resilience and vulnerability. It should be noted, though, that reducing the probability of crossing 11.25 m also appears to reduce the probability of overtopping the dikes at 13.4 m, as the best flood solution from the WC formulation (panel a) also overtops the dikes more often than the best flood solution from the WP1 formulation (panel d), although this is a rare event for both solutions.

Comparing the best WP1 flood solution (panel d) to the best WP1 hydro solution (panel e), one can see the probabilistic effects of different preferences on the water level at Hanoi over time. Under operations with the best WP1 hydro solution (panel e), the time-varying density of the water level at Hanoi more closely resembles that of the best WC hydro solution (panel b), with water levels exceeding 6 m less often earlier in the monsoon season than the best WP1 flood solution (panel d). However, this results in a greater probability of exceeding both the second alarm level of 11.25 m and the dike height of 13.4 m. The WP1 compromise solution (panel f), strikes a balance between the two, crossing 11.25 m more often than the best WP1 flood solution (panel d), but less often than the best WP1 hydro solution (panel e). The fact that these differences are less obvious between the WC solutions in panels a-c, whose probabilistic water level dynamics most closely resemble those of the WP1 hydro solution, highlights that the WC formulation is actually far less conservative with respect to flooding than intended, and instead maximizes hydropower production. Consequently, stakeholders simply choosing a compromise solution among a set of non-dominated policies from a single formulation would make a poor decision if only the WC formulation were used to design operating policies.

While the storage and release trajectories in concert with the time-varying PDFs in Figure 4.9 provide some understanding of how the operations lead to this coincident behavior, it is helpful to visualize these state trajectories jointly through a state space diagram [Nayfeh & Balachandran, 2008]. State space diagrams are useful for examining how a system evolves in time. For example, one can observe whether and when a system converges to a steady state, bifurcates into separate trajectories, or exhibits periodic orbital behavior [Nayfeh & Balachandran, 2008]. Applied here, the state space diagram may enhance our understanding of the stability of different operating policies, and how they achieve their objective values.

Figure 4.10 shows the probabilistic state space diagram for the compromise solutions from the WC (panel a) and WP1 (panel b) formulations, to further highlight the benefits of testing multiple problem formulations to guide stakeholders in discovering effective compromise solutions. Tracing the high probability density regions in dark red, one can see that for the WC compromise solution in panel a, total reservoir storage initially increases without significantly increasing the water level at Hanoi. This is because the reservoirs are not releasing much of what comes in, trying to maintain high storages for hydropower production. However, once the reservoirs reach maximum capacity, they are forced to increase releases and the water level at Hanoi quickly rises. Notice this occurs before the total storage capacity has been reached. This indicates that this policy is not making full use of all of the reservoirs. Recall from Figure 4.7 that the best hydro solution from the WC formulation maintains high storages at Hoa Binh and attempts to use the smaller Thac Ba reservoir for flood protection. The state space diagram suggests that this is insufficient, as the larger reservoirs fill to capacity and are forced to spill water downstream, while the smaller reservoirs

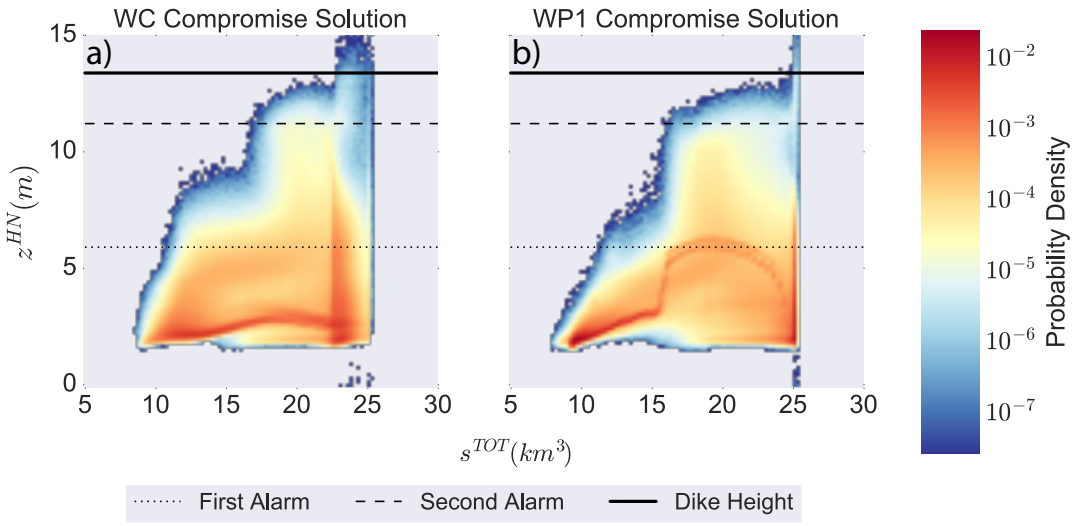

Figure 4.10: Joint probability density of total storage (x-axis) and water level at Hanoi (y-axis) under operations with the WC compromise solution (panel a) and WP1 compromise solution (panel b). The WC compromise solution initially fills the reservoirs without releasing much, maintaining low water levels at Hanoi until the largest reservoirs reach maximum storage and must spill, causing the water level at Hanoi to jump. The WP1 compromise solution fills the reservoirs more slowly, releasing at the same time, causing water level and storage to rise simultaneously. This behavior makes better use of the full system capacity, reducing the probability of reaching maximum storage and causing the water level at Hanoi to spike. The y-axis on each chart has again been cut off at 15 m, which some simulations exceed in extremely rare cases.

have unused capacity that is not being fully exploited for flood protection. Consequently, under the lower probability events in yellow and light blue, water levels exceed 11.25 m over a range of total storages from about 15-25 km^3 and overtop the dikes at storages between about 22-25 km^3.

The compromise solution from the WP1 formulation exhibits very different joint state behavior. Tracing again the highest probability streak in red, one

can see that storage levels increase in concert with the water level at Hanoi. This is because the reservoirs do not store up everything that comes in, but release some water to maintain more storage space for potential future flood events. Most of the time, the water level at Hanoi reaches the first alarm level of 6 m and levels off there. As the monsoon subsides, the water level at Hanoi drops while storage remains high. This is because little flow comes during the dry season, so the reservoirs store up as much water as possible to meet the agricultural demand. In the occasional wet year when the WP1 compromise solution reaches maximum storage, water levels also rise sharply as for the WC compromise solution, but this happens less often and only at maximum total system storage, indicating that this solution is better able to make use of the full system capacity for flood protection. As a result, the probability of crossing 11.25 m is lower, as seen by the greater blue shade above this line compared to the yellow shade above it for the WC compromise solution. Additionally, the WP1 compromise solution only ever results in overtopping when storage is at its maximum, while this occasionally occurs with the WC compromise solution before total storage capacity has been reached.

4.6 Conclusions

This study uses the multi-reservoir Red River system in Vietnam to illustrate that even modest changes in how objectives are quantified in a control problem can yield a surprising cascade of impacts on system performance. Consequently, it is important to test rival problem framings to determine the consequences of alternative quantitative abstractions of stakeholder objectives. In this system, where operating policies must balance the competing needs of

flood management, hydropower production and water supply for agriculture, we find that several commonly used problem framings can result in damaging unintended consequences. First, minimizing variance-based objectives often yields harsh consequences for expected performance while maximizing expected value objectives tends to expose systems to negative, high variance outcomes. In the context of reservoir controls, our results show that maximizing expected value objectives also has a tendency to yield over-fit control policies that do not generalize well out of sample. Finally, the Red River test case formulation with the greatest inherent negative consequences observed here is the worst case formulation commonly used in robust optimization [Wald, 1992; Beyer & Sendhoff, 2007] as it results in unstable policies that provide a poor representation of the system tradeoffs and unintended modes of failure.

While it is known that the worst case over a large ensemble will degrade compared to the worst case over a smaller ensemble, this is not always of consequence if policies that minimize performance near the tails of an objective's distribution simultaneously minimize even more extreme values of the distribution. In this study, that is the case for the worst case hydropower and squared deficit objectives, but not for the worst case flood damages objective due to the noisy, unbounded penalty function used to approximate damages in expectation. An important outcome of our results is that minimizing this non-unique, nonlinear functional abstraction of flood damages in expectation inadvertently maximizes hydropower production, yielding levee overtopping in Hanoi for the 100-year flood level. This is true even if minimizing the worst case expected damages across an ensemble of multi-year streamflows. Fortunately, we find that the consequences of minimizing non-unique, non-linear, worst case objectives can be overcome by formulating distinct, linear, worst first percentile ob-

jectives. In particular, minimizing the worst first percentile of the annual maximum water level solely targets large, infrequent events with a linear penalty, resulting in a stable objective that is able to simultaneously reduce expected damages and the probability of observing even larger flood events. Additionally, it is able to reduce inter-annual variability without compromising expected performance as significantly as when including variance minimization as an objective.

These conclusions have significant implications for how reservoir operations are optimized for flood protection. Minimizing expected flood damages is a common objective in reservoir operations, but it may not adequately reduce system hazard. However, minimizing the more effective worst first percentile maximum annual flood objective while also maximizing expected hydropower production is only possible with simulation optimization frameworks such as EMODPS. Furthermore, EMODPS facilitates coordinated control across multiple reservoirs without suffering from the curse of dimensionality. The high-dimensional, multi-objective Red River control problem explored in this study is representative of the contextual and mathematical challenges that will be faced in a broad range of global multi-reservoir systems. Our ability to discover and appropriately manage the water, energy, and food tradeoffs within these systems can be greatly advanced with the parameterization-simulation-optimization approach demonstrated here. Future work should focus on improving how information feedbacks and scalable control frameworks can be used to rigorously evaluate rival problem framings for managing complex river basins balancing evolving multi-sectoral demands, ecological impacts and changing hydrologic extremes.

CHAPTER 5

EXPLORING HOW CHANGING MONSOONAL DYNAMICS AND HUMAN PRESSURES CHALLENGE MULTI-RESERVOIR MANAGEMENT OF FOOD-ENERGY-WATER TRADEOFFS

This chapter is drawn from the following article currently in preparation:

Quinn, J.D., Reed, P.M., Giuliani, M., Castelletti, A., Oyler, J.W. and Nicholas, R.J. (In Prep). Exploring How Changing Monsoonal Dynamics and Human Pressures Challenge Multi-Reservoir Management of Food-Energy-Water Tradeoffs.

This work was partially supported by the U.S. National Science Foundation (NSF) through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507 and the Penn State Center for Climate Risk Management. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding entities.

5.1 Abstract

Multi-reservoir systems require robust and adaptive control policies capable of managing evolving hydroclimatic variability and human demands across a wide range of time scales. This is especially true for systems with high intra-annual and inter-annual variability, such as monsoonal river systems that need to buffer against seasonal droughts while also managing extreme floods. Moreover, the timing, intensity, duration, and frequency of these hydrologic extremes may be affected by deeply uncertain changes in socioeconomic and climatic pressures. This study contributes an innovative method for exploring

how possible changes in the timing and magnitude of monsoonal seasonal extremes impact the robustness of reservoir operating policies optimized to historical conditions assuming stationarity. We illustrate this analysis on the Red River basin in Vietnam, where reservoirs and dams serve as important sources of hydropower production, irrigable water supply, and flood protection for the capital city of Hanoi. Applying our scenario discovery approach, we find food-energy-water conflicts are exacerbated by potential hydrologic shifts, with wetter worlds threatening the ability of operating strategies to manage flood risk and drier worlds threatening their ability to provide sufficient water supply and hydropower production, especially if demands increase. Most notably, though, amplification of the within-year monsoonal cycle and increased inter-annual variability threaten all of the above. These findings highlight the importance of considering changes in both lower order moments of annual streamflow as well as intra-annual monsoonal behavior when evaluating the robustness of alternative water systems control strategies for managing deeply uncertain futures.

5.2 Introduction

Designing robust river basin management systems poses a severe challenge. Successful long-term performance in these systems requires a careful exploration of the capacity of alternative management plans to handle uncertain changes in climatic and human pressures over multiple time scales. From extreme precipitation events to prolonged drought, and rapid urbanization to intensive agriculture, a range of stochastic and potentially nonstationary climatic and anthropogenic factors influence how river basin systems should be managed both now and in the future [Vörösmarty et al., 2000; Bouwer, 2000; Field

et al., 2014]. These uncertainties are often considered "deep," meaning decision makers possess discordant beliefs about what are appropriate prior probability distributions for their occurrence [Knight, 1921; Lempert et al., 2002].

In deeply uncertain decision contexts, Dessai et al. [2009] highlight the deficiencies of classical "predict-then-act" risk-based assessments, since both likelihoods and consequences are poorly defined. Instead, they advocate for "bottom-up" approaches in which exploratory modeling techniques [Bankes, 1993] are employed to discover robust strategies that perform well across a broad range of possible system conditions and external forcings, regardless of their likelihood. These methods, reviewed in several recent papers [Herman et al., 2015; Maier et al., 2016; Dittrich et al., 2016] include decision scaling [Brown et al., 2012; Poff et al., 2015], information gap [Ben-Haim, 2004], robust decision making (RDM; Lempert et al. [2003]) and many-objective robust decision making (MORDM; Kasprzyk [2013]). Despite their methodological differences in implementation, these approaches share two common goals: 1) to discover management plans that are robust to deep uncertainties in system conditions and external forcings and 2) to determine conditions under which these plans can no longer satisfy system goals and therefore need to be redesigned.

Traditionally, these methods are used to evaluate and compare the robustness of either a pre-specified plan or set of plans (e.g. Brown et al. [2012]; Ben-Haim [2004]; Lempert et al. [2003]), or a non-dominated set of multi-objective management plans optimized to assumed conditions representing planners' best estimate of the current "state of the world" (SOW) (e.g. Kasprzyk [2013]). Recognizing that sampling the action space could discover more robust management plans, several recent studies have expanded on these approaches by

incorporating ideas from robust optimization, re-optimizing management policies over a broad sample of deeply uncertain conditions [Mortazavi-Naeini et al., 2015; Roach et al., 2016; Beh et al., 2017; Trindade et al., 2017; Watson & Kasprzyk, 2017]. The unifying principle underlying these studies is that decision makers may be willing to sacrifice optimality to assumed conditions in exchange for less sensitivity to these assumptions [Lempert & Collins, 2007]. However, as noted by Watson & Kasprzyk [2017], sub-optimal but robust policies may potentially result in large regrets, especially if some of the deeply uncertain worlds to which policies are re-optimized are extremely unlikely. Furthermore, which measure is used to quantify robustness in the optimization could result in the discovery of completely different strategies [Herman et al., 2015; Quinn et al., Accepted].

To bridge the differences between minimizing short-term regrets and maximizing potentially conflicting measures of long-term robustness, Walker et al. [2001] recommend formulating adaptive policies that evolve as observations and endogenous learning resolve system uncertainties. This idea helped spawn the Adaptive Policymaking [Kwakkel et al., 2010] and Adaptation Pathways [Haasnoot et al., 2012] design approaches, which were seminal to the ultimate formulation of the increasingly popular Dynamic Adaptive Policy Pathways planning framework (DAPP; Haasnoot et al. [2013]). Under DAPP, decision makers monitor *signposts* of change which *trigger* the adoption of different policies when some critical value, or *adaptation tipping point* [Kwadijk et al., 2010], has been observed.

Central to the concept of adaptive robustness is the design of the signposts (i.e., factors monitored to estimate the system trajectory) and tipping points (i.e.,

values of the signposts which trigger policy changes). Examples of signposts used in the literature to design adaptive water management and infrastructure planning strategies include the 30-year running average of the number of days in which river discharge is below some threshold [Kwakkel et al., 2016a], a probabilistic measure of the risk of falling below a threshold system storage level [Zeff et al., 2016], and the 50-year running average of the water-year centroid [Herman & Giuliani, In Review]. Clearly, a broad range of candidate factors may be suitable, but discovering effective signposts is a challenge (e.g., see discussions in Galelli & Castelletti [2013]; Galelli et al. [2014]; Hermans et al. [2017]). For example, Herman & Giuliani [In Review] find that even monotonically trending hydrologic conditions may not be informative for management. In designing adaptive reservoir operations for Folsom Dam in California, they find a 50-year moving average of annual streamflow is not helpful for controlling floods since floods are primarily driven by shorter time-scale events. Additionally, a 50-year moving average of the 100-year flood estimate from an LP3 distribution fit to the annual maxima is also ineffective since it is a lagging, not leading, indicator of change.

One promising way to design these signposts and triggers is by combining the DAPP approach with bottom-up exploratory modeling techniques such as RDM in which sensitivity analysis is performed to identify conditions under which current management plans fail. Groves et al. [2014] illustrate how these methods can be combined to design robust and adaptive water resources plans for the Metropolitan Water District of Southern California. However, in order to effectively combine these approaches, it is important to ensure that the scenarios sampled in the RDM analysis adequately capture the system dynamics that might emerge under alternative climatic and socioeconomic futures [Pruyt &

Islam, 2016]. In particular, as noted by Herman & Giuliani [In Review], simply sampling changes in mean hydroclimatic conditions may not be sufficient, especially given that changes in climate variability affect the frequency of extreme climate events more than changes in the mean [Katz & Brown, 1992], and have more severe consequences for agricultural production [Thornton et al., 2014]. Changes in intra-annual variability may also pose challenges, particularly for monsoonal systems for which water management strategies must balance competing concerns of flood protection in the wet season and drought management in the dry season.

In this study, we attempt to capture these complex hydroclimatic dynamics through an extensive sampling of how the timing and magnitude of monsoonal seasonal extremes may evolve under a changing climate using the Red River basin in East Asia as a case study. In this system, four major reservoirs on the Vietnamese side of the basin serve multiple system objectives related to food, energy and water. First, they provide flood protection to the capital city of Hanoi, situated in the delta. Second, they supply water to multiple economic sectors, the largest sector being 16 irrigation districts in the floodplain. Finally, they also supply the region with hydroelectric power, currently the primary source of electricity in Vietnam [Asian Development Bank, 2016].

However, the ability of the system's four reservoirs to meet all three of these objectives could be challenged by changing hydrologic and socioeconomic pressures. In the context of hydrologic pressures, streamflow in the Red River is dominated by the annual monsoon from May to October, which in six months provides nearly 80% of the total annual flow. Yet there is deep uncertainty in how the monsoon will be affected by climate change. A weakening of the East

Asian monsoon since the 1920s has been observed in wind speeds over China and sea level pressure (SLP) gradients [Guo et al., 2011; Jiang et al., 2010; Vautard et al., 2010; Zhou et al., 2009]; however, the IPCC Fifth Assessment Report [Hartmann et al., 2013] places low confidence in the projection that the monsoon will weaken, as Hsu et al. [2011] and Wang et al. [2012] observe increased precipitation since 1979 under an alternative definition of monsoon area. Furthermore, more than 85% of the Coupled Model Intercomparison Project 5 (CMIP5) models predict an increase in mean precipitation in the East Asian summer monsoons [Hijioka et al., 2014]. In the context of socioeconomic pressures, water demand in the basin is currently dominated by agriculture, followed by fisheries. However, total rice growing area in Vietnam has been declining recently due to rapid urbanization [Yu et al., 2010], suggesting there may be a shift in sectoral demand in the future.

Despite these uncertainties, prior work in the Red River basin has primarily focused on handling the challenge of managing conflicting multi-sectoral demands by optimizing reservoir operations to historical [Le Ngo et al., 2007; Castelletti et al., 2012a; Giuliani et al., 2016b] or stationary stochastic conditions [Giuliani et al., In Press; Quinn et al., Accepted]. While several studies have investigated the potential impacts of different climate change projections on agricultural production, hydropower production and flooding in the basin [Yu et al., 2010; Gebretsadik et al., 2012; Neumann et al., 2015], to our knowledge, only Giuliani et al. [2016a] have analyzed how reservoir operations could be re-optimized to mitigate these impacts. The work by Giuliani et al. [2016a] provides an informative benchmark of the maximum operational adaptive capacity [Culley et al., 2016; Whateley et al., 2014] of the system to the tested climate change projections. However, in practice it is impossible to know what

will actually happen and consequently, what the best adaptation strategy will be. Furthermore, a limited set of climate change projections may not span the full range of plausible climate futures, in particular since many of them share large portions of code, and consequently systematic structural biases [Pennell & Reichler, 2011; Steinschneider et al., 2015a]. Non-climatic uncertainties could also inhibit operational adaptive capacity, perhaps even more significantly than climate change (see e.g. Herman et al. [2014]).

The goal of this study is to demonstrate the utility of a newly developed sampling scheme for exploring how changes in both the first two moments of annual streamflow as well as intra-annual monsoonal behavior, combined with potential demand shifts, can help discover the most important drivers of system performance. In particular, we illustrate how this sampling method can determine under which combinations of interacting hydrologic and socioeconomic factors multi-reservoir operating policies in the Red River designed for stationary conditions fail to attain satisfactory performance. These combined factors can be used to design more robust operating policies in future analyses by serving as signposts for adaptive management plans.

Our paper is organized as follows. Section 5.3 provides a brief background of the Red River water systems management model. Section 5.4 describes our scenario discovery experiment for sampling potential changes in monsoonal dynamics and human pressures. Section 5.5 presents the results of this analysis. Finally, Section 5.6 concludes with a discussion of our key findings and opportunities for future work.

5.3 Multi-Objective Water Systems Management Model

This study builds off of work in the Red River basin described in Chapter 4. For a description of the basin and system model, see Section 4.3. Building off of insights from Chapter 4, reservoir operating policies for the Red River basin were re-optimized in this study to further improve upon the Worst First Percentile formulation described in Section 4.4. Reservoir policies were again optimized using Evolutionary Many-Objective Direct Policy Search (EMODPS) in which a Many-Objective Evolutionary Algorithm (MOEA) is used to optimize the parameters of multi-reservoir operating policies in order to improve the performance of multiple system objectives computed in simulation. Section 5.3.1 describes the formulation of the system objectives; Section 5.3.2 describes the formulation of the reservoir operating policies; and finally, Section 5.3.3 summarizes the experimental setup of the multi-objective optimization.

5.3.1 Formulation of Objectives

Quinn et al. [Accepted] compared four alternative multi-objective formulations of the Red River control problem in which stakeholder objectives were quantified in different ways. Two of the formulations were specifically intended to represent the goals of risk-averse stakeholders, as research partners at a number of Vietnamese governmental organizations stressed their desire to formulate conservative operating policies, particularly with respect to flooding. Quinn et al. [Accepted] found that minimizing tail measures of performance such as the worst first percentile across an ensemble of streamflows provided more accurate representations of system tradeoffs compared to min-max measures, and

that minimizing the 100-yr stage more effectively reduced the probability of severe flood events compared to minimizing expected damages. For this reason, we apply the same objective formulations here with improvements to how the policies are parameterized (described in Section 5.4.2) and simulated.

In this study, candidate operating policies are simulated over an ensemble of N stochastic T-year simulations in which the d-th objective, J_d, is quantified according to Equation 5.1:

$$J_d = \Psi_{i\in(1,\ldots,N)}\Big[\Phi_{t\in(1,\ldots,365T)}[g_d(t,i)]\Big] \tag{5.1}$$

where $g_d(t,i)$ is the value of the d-th objective on day t in the i-th ensemble member, Φ is an operator for the aggregation of $g_d(t,i)$ over T years (such as the sum, $\sum$), and Ψ is a statistic used to filter the noise across the N ensemble members (such as the expected value, $\mathbb{E}$). Here we simulate policies over a 1000-yr sequence of stochastic streamflows that we divide into $N = 1000$ consecutive ensemble members of $T = 1$ year simulations to compute objectives. This ensures that the distribution of initial conditions for the ensemble members is representative of those obtained by the control strategy being simulated.

The first system goal in this study is to minimize J_{Flood}, a measure of flooding. We quantify this objective as the amount by which the annual maximum water level at Hanoi exceeds 11.25 m in the worst first percentile year. This stage is an alarm level elicited from stakeholders. We constrain J_{Flood} to be less than 2.15 m, the difference between 11.25 m and the dike height of 13.4 m, ensuring protection to the 100-yr flood. This objective is therefore calculated by letting $g_{Flood}(t,i) = z^{HN}_{t,i}$ where $z^{HN}_{t,i}$ is the water level at Hanoi on day t of the i-th

ensemble member, $\Phi = \max_{365T}$ and $\Psi = \text{quantile}_N\{\Phi, 0.99\}$.

The second system goal is to maximize J_{Hydro}, a measure of hydropower production. We compute this objective as the average daily production within each ensemble member and again minimize the worst first percentile across the ensemble, so $g_{Hydro}(t,i) = \sum_k \eta^k_{t,i}$ where $\eta^k_{t,i}$ is the energy production from reservoir k on day t of the i-th ensemble member, $\Phi = \mathbb{E}_{365T}$ and $\Psi = \text{quantile}_N\{\Phi, 0.01\}$. We maximize production rather than revenue because energy is sold by the government at a fixed rate, so these measures are equivalent. Since markets and prices may change in the future, alternative formulations could be tested in subsequent analyses, but here we focus solely on analyzing the sensitivity of total production to uncertainties in system conditions and external forcings.

The final system goal is to minimize $J_{Deficit^2}$, a measure of the water supply deficit. For this objective, we compute the average daily squared deficit and minimize the worst first percentile across the ensemble, so $g_{Deficit^2}(t,i) = D^2_{t,i}$ where $D^2_{t,i}$ is the squared deficit on day t of the i-th ensemble member, $\Phi = \mathbb{E}_{365T}$ and $\Psi = \text{quantile}_N\{\Phi, 0.99\}$. Daily deficits are squared to numerically favor frequent but small deficits over less frequent, more severe deficits.

Combining all of these objectives, the goal of the multi-objective Red River control problem is to find non-dominated parameter sets θ^* minimizing the three objectives. This is defined formally by Equations 5.2-5.7:

$$\theta^* = \text{argmin}_\theta J(\theta) \tag{5.2}$$

where

$$J = \begin{bmatrix} -J_{Hydro}(\theta) \\ J_{Deficit^2}(\theta) \\ J_{Flood}(\theta) \end{bmatrix} \tag{5.3}$$

$$J_{Flood} = \text{quantile}_N\Big\{\max_{365T}(z_{t,i}^{HN}), 0.99\Big\} \tag{5.4}$$

$$J_{Deficit^2} = \text{quantile}_N\Big\{\mathbb{E}(D_{t,i}^2), 0.99\Big\} \tag{5.5}$$

$$J_{Hydro} = \text{quantile}_N\Big\{\mathbb{E}_{365T}\Big(\sum_k \eta_{t,i}^k\Big), 0.01\Big\} \tag{5.6}$$

subject to

$$J_{Flood}(\theta) \leq 2.15 \tag{5.7}$$

where θ is a vector of parameters describing the operating policies defined in Section 5.3.2. These parameters therefore serve as the decision variables optimized by the MOEA.

5.3.2 Formulation of Operating Policies

In this study, operating policies for each of the four reservoirs are determined by radial basis functions (RBFs) mapping a vector of system states to reservoir releases. RBFs have been shown to provide effective operating policies in this system, generalizing better to out-of-sample streamflows than policies approximated by ANNs [Giuliani et al., 2016b]. The RBF representation of daily release policies at each reservoir is given by Equation 5.8:

$$u_t^k = \sum_{i=1}^{A} w_i^k \exp\left(- \sum_{j=1}^{B} \frac{\left((x_{t-1})_j - c_{j,i}\right)^2}{b_{j,i}^2} \right) \tag{5.8}$$

where u_t^k is the policy-prescribed release from the k-th reservoir on day t (normalized on [0,1]), $(x_{t-1})_j$ is the value of the j-th input at time $t-1$ (normalized on [0,1]), A is the number of RBFs, B the number of inputs, and w_i^k , $c_{i,j}$ and $b_{i,j}$ are the weights, centers and radii, respectively, of the i-th RBF associated with the k-th reservoir and j-th input. The actual release at the k-th reservoir, r_t^k, is equal to the un-normalized value of u_t^k unless physical constraints prohibit it (e.g. if the prescribed release lies outside the minimum and maximum allowable releases, if there is insufficient water to meet the prescribed release, or if the prescribed release would result in the reservoir storage capacity being exceeded).

As defined in Equations 5.2-5.7 in Section 5.3.1, the goal of EMODPS is to find a non-dominated set of parameter vectors θ^* minimizing the system objectives. The parameter vector θ is composed of the weights, centers, and radii defining the RBF policies, i.e. $\theta = [w_i^k, c_{i,j}, b_{i,j}]$ where $i = \{1, ..., A\}$, $j = \{1, ..., B\}$ and $k = \{1, ..., K\}$. In this study, $K = 4$ for the four reservoirs, with $w_i^k \in [0, 1]$, $c_{i,j} \in [-1, 1]$, $b_{i,j} \in [0, 1]$. This results in a total of $A(K + 2B)$ parameters. We use A=11 RBFs and B=6 inputs: the previous day's storage at each of the four reservoirs and a cyclic representation of time with phase-shifted sin(·) and cos(·) functions of time, i.e. $\mathbf{x_{t-1}} = \left[s_{t-1}^{SL}, s_{t-1}^{HB}, s_{t-1}^{TB}, s_{t-1}^{TQ}, \sin\left(\frac{2\pi t}{365} - p_1\right), \cos\left(\frac{2\pi t}{365} - p_2\right) \right]$ where $q_{t-1}^{TOT} = q_{t-1}^{Da} + q_{t-1}^{Thao} + q_{t-1}^{Chay} + q_{t-1}^{Lo} + q_{t-1}^{Gam}$ and p_1 and p_2 are phase shifts on $[0, 2\pi]$. For the sin(·) and cos(·) functions, we also set their associated centers to 0 and radii to 1 since the intent of these parameters is to horizontally translate and scale the inputs, but outside of sin(·) and cos(·) functions they would be vertically translating them. We instead include phase shifts within the sin(·) and cos(·) functions

to serve as centers in horizontally translating the inputs, but do not include a scale parameter, as this would result in them no longer being full period over the year. Given the above changes, the RBF representation of the policies in this paper can be described by Equation 5.9:

$$u_t^k = \sum_{i=1}^{A} w_i^k \exp\left(- \sum_{j=1}^{B-2} \frac{\left((x_{t-1})_j - c_{j,i}\right)^2}{b_{j,i}^2} + (x_{t-1})_{B-1}^2 + (x_{t-1})_B^2 \right) \tag{5.9}$$

where $(x_{t-1})_{B-1} = \sin\left(\frac{2\pi t}{365} - p_1\right)$ and $(x_{t-1})_B = \cos\left(\frac{2\pi t}{365} - p_2\right)$. It should be noted that since all inputs are normalized on [0,1], squaring the $\sin(\cdot)$ and $\cos(\cdot)$ functions in Equation 5.9 does not create two separate cycles per year. The total number of parameters to be optimized in this formulation is $A(K + 2(B\text{-}2)) + 2 = 134$.

5.3.3 Multi-Objective Optimization

As stated in Section 2.2.1, EMODPS exploits MOEAs to optimize the parameter vector θ defining the multi-reservoir control policies in order to minimize multiple objectives computed over the system simulation. Given recent documented success of the Multi-Master Borg MOEA in optimizing operating policies for complex reservoir control problems [Giuliani et al., In Press; Zatarain Salazar et al., In Review], we use this algorithm to identify Pareto-approximate operating policies for the Red River's four largest reservoirs. We run the Multi-Master Borg with 5 seeds using a 16-master implementation with 400,000 function evaluations allocated to each master. The Borg algorithm requires that users specify "epsilons," or significant levels of precision below which they are impartial to differences in performance, for each objective. We use epsilons of 0.05 for J_{Flood},

25.0 for $J_{Deficit^2}$ and 0.5 for J_{Hydro}.

5.4 Methods

5.4.1 Robustness Analysis

The primary goals of this study are to 1) discover robust operating policies for reservoirs in the Red River basin and 2) determine the most important hydrologic and socioeconomic drivers of their performance in order to inform adaptive management strategies, for example by setting triggers for re-operations or by incentivizing demand management strategies. Many different strategies have been proposed in the literature for defining and evaluating robustness; however, as Herman et al. [2015] highlight in their taxonomy of robustness frameworks, these approaches all share four common components. The first component is the selection of decision alternatives, in this case, alternative reservoir operating policies. These may be pre-specified or discovered through optimization or a statistical design of experiments. Following the Many-Objective Robust Decision Making (MORDM) framework introduced by Kasprzyk [2013], we discover non-dominated operational alternatives through multi-objective optimization (described in Section 5.3), considering only well-characterized, stationary streamflow uncertainty in the optimization.

The second shared component of most robustness analyses is to evaluate the decision alternatives in different states of the world (SOWs) that could pose operational challenges in the future. These SOWs may be pre-specified through a limited sampling of a select subset of factors assumed to be most important (i.e.,

a priori scenario analyses), or by discovering which factors are most important through a larger design of experiments [Herman et al., 2015]. In this study, we discover the most important factors by generating alternative SOWs through a design of experiments, but focus specifically on hydrologic and socioeconomic drivers. A core contribution of this study includes the use of synthetic streamflow generation to capture a diverse suite of non-stationary monsoonal dynamics. Our approach for generating SOWs is described in detail in Section 5.4.2.

After re-evaluating the Red River system's candidate control policies in different SOWs, we utilize robustness measures to rank the alternatives and inform further vulnerability assessments using scenario discovery. Examples of robustness measures used in the literature include the expected value and standard deviation of an alternative's performance across the sampled SOWs [Kwakkel et al., 2016a], a satisficing measure of the percent of worlds in which some criteria are met [Herman et al., 2014; Moody & Brown, 2013], or a measure of regret in performance associated with applying a decision alternative in different SOWs compared to its baseline performance [Kasprzyk, 2013; Lempert & Collins, 2007]. Here, we calculate a satisficing metric [Starr, 1969] for each design alternative representing the percent of sampled worlds in which minimum performance levels across the three objectives defined in Section 5.3.1 are met. The final step in our robustness assessment employs sensitivity analysis to discover which conditions sampled in the scenario generation process control the robustness of different design alternatives. This is also called "scenario discovery" [Lempert et al., 2008] as the scenarios of concern are discovered rather than assumed. Here, we discover scenarios of concern using logistic regression, described in Section 5.4.3. This analysis could be used in future work to inform when policies should be re-designed [Groves et al., 2014] or when controlling

factors should be managed to mitigate their negative effects (see e.g. Herman et al. [2014]).

The final step we apply in our analysis is to evaluate the plausibility of the scenarios discovered to be of concern. While not mentioned as an explicit step in the robustness taxonomy of Herman et al. [2015], this step is often added to climate change studies to see where projections from general circulation models (GCMs) fall relative to the scenarios of concern [Brown et al., 2012; Whateley et al., 2014; Steinschneider et al., 2015b; Herman et al., 2016]. This scenario evaluation step is discussed in Section 5.4.4.

5.4.2 Sampling of Deep Uncertainties

In this study, we focus on analyzing the impacts of hydrologic and socioeconomic uncertainties on the robustness of alternative Red River control policies spanning key tradeoffs. Here, we outline how we explore monsoonal changes and evolving socioeconomic pressures to discover scenarios under which the Red River control alternatives identified with multi-objective optimization no longer attain satisfactory performance. This analysis provides a means of identifying tolerable windows of change as well as conditions that should trigger re-operation of the Red River reservoirs.

Hydrologic Uncertainties

In many scenario discovery analyses for water resources systems, hydrologic uncertainty is examined by first generating scenarios of alternative tempera-

ture and precipitation changes from current climatology and then feeding these scenarios through a physically based [Steinschneider et al., 2015b] or statistical hydrologic model [Brown et al., 2012; Moody & Brown, 2013]. However, this approach often requires extrapolating beyond conditions observed in the historical record to which the model was calibrated, assuming these relationships will hold in the future (see discussion in Wagener et al. [2010]). Recently, researchers have attempted to circumvent this problem by directly generating alternative streamflow conditions for scenario discovery [Nazemi et al., 2013; Borgomeo et al., 2015; Herman et al., 2016]. While using a hydrologic model may permit causal inferences on the effects of changes in precipitation, temperature, and in the case of some models, land use change on streamflow (assuming stationarity in parametric and structural model behavior), such inferences may not be of concern to water managers. Since streamflow is more directly tied to water management needs, simply discovering the streamflow characteristics that are of concern may be more informative, if not simpler. For these reasons, we directly generate synthetic streamflows in our scenario discovery analysis.

The main streamflow characteristics that we consider are the log-space mean annual flow, standard deviation of annual flow, and intra-annual flow variability tied to the monsoon. These characteristics are of concern because there is great uncertainty in the direction of mean changes in precipitation in East Asia, inter-annual variability is expected to increase globally, and it is unclear how the monsoon will be affected by climate change [Hijioka et al., 2014]. To alter these flow characteristics for our scenario discovery process, we first generate a series of synthetic streamflows built to stationary conditions and then rescale them to produce the desired effects. Although in this study we generate synthetic flows as described in Appendix B, our exploratory sampling approach

could be flexibly used with alternative synthetic streamflow generators.

The main goal of the rescaling process is to find a vector of monthly multipliers to apply to the stationary synthetic time series. For this process, consider two ensembles: $\mathbf{Q_S} \in \mathbb{R}^{N_S \times 12}$ of N_S years of synthetic monthly flows and $\mathbf{Q_H} \in \mathbb{R}^{N_H \times T}$ of N_H years of historical monthly flows. We first perform a log-transform of each of these ensembles to formulate time series of normally distributed monthly flows, $\mathbf{Y_S} = \ln(\mathbf{Q_S})$ and $\mathbf{Y_H} = \ln(\mathbf{Q_H})$. We then standardize $\mathbf{Y_S}$ to generate ensembles of synthetic standard normal monthly flows, $\mathbf{Z_S}$ as shown in Equation 5.10:

$$Z_{S_{i,j}} = \frac{Y_{S_{i,j}} - \hat{\mu_j}}{\hat{\sigma_j}} \tag{5.10}$$

where $\hat{\mu_j}$ and $\hat{\sigma_j}$ are the sample mean and sample standard deviation of the j-th month's log-transformed flows from $\mathbf{Y_H}$. To rescale the stationary synthetic monthly streamflows $\mathbf{Q_S}$ into alternative streamflow scenarios $\mathbf{Q'_S}$, vectors of monthly varying mean multipliers $\mathbf{M}_\mu = [M_{\mu,1}, ..., M_{\mu,12}]$ and standard deviation multipliers $\mathbf{M}_\sigma = [M_{\sigma,1}, ..., M_{\sigma,12}]$ are applied to $\hat{\mu_j}$ and $\hat{\sigma_j}$, respectively, when back-transforming Z_S, as shown in Equation 5.11:

$$Q'_{S_{i,j}} = \exp\left(M_{\mu,j}\hat{\mu, j} + M_{\sigma,j}\hat{\sigma_j} Z_{S_{i,j}}\right). \tag{5.11}$$

If only the log-space annual mean and standard deviation are changed, then all elements of $\mathbf{M}_\mu$ and $\mathbf{M}_\sigma$ are constant (i.e., $M_{\mu,i} = m_\mu \forall i$ and $M_{\sigma,i} = m_\sigma \forall i$, where m_μ and m_σ are constants). Once these multipliers have been applied to calculate $\mathbf{Q'_S}$, the adjusted monthly flows are disaggregated to daily flows according to the same proportions used to disaggregate the corresponding stationary monthly flows $\mathbf{Q_S}$ to daily flows.

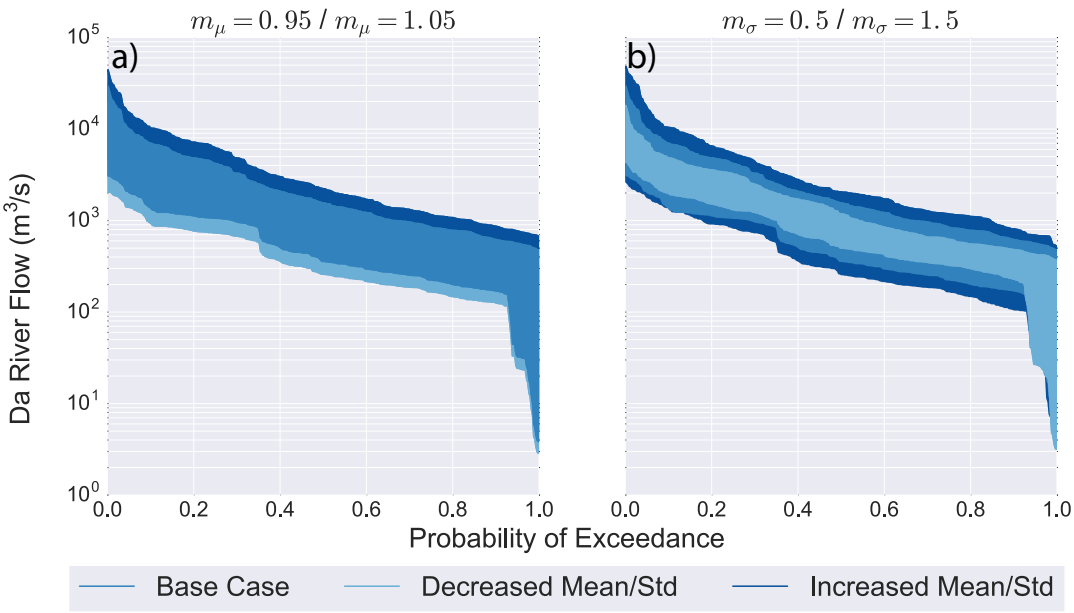

Figure 5.1: Ranges spanned by synthetically generated flow duration curves on the Da River under base case, i.e. historical, conditions (medium blue) as well as increased (dark blue) or decreased (light blue) log-space annual mean (panel a) and annual standard deviation (panel b). The multipliers displayed are those spanned in our scenario discovery experiment.

Figure 5.1 shows the effects of the constant multipliers m_μ and m_σ on the flow durations curves (FDCs) of daily flows in the Da River, which provides about half of the system's flow. In panel a, the range spanned by stationary synthetic flows is shown in medium blue, while the range spanned by rescaled flows generated with a log-space mean multiplier of $m_\mu = 1.05$ is shown in dark blue and $m_\mu = 0.95$ in light blue. Because the multipliers are applied in log-space, these 5% increases and decreases can yield fairly large differences in real-space flow magnitude. In panel b, the range of stationary synthetic flows is again shown in medium blue, while the range of rescaled flows generated with a log-space standard deviation multiplier of $m_\sigma = 1.5$ is shown in dark blue and $m_\sigma =$

0.5 in light blue. As one would expect, increasing the log-space annual standard deviation produces both wetter and drier years, while decreasing it reduces the range of generated flows. It should also be noted that increasing the log-space standard deviation also increases the real-space mean, greatly increasing flood risk.

To model solely monsoonal changes to the intra-annual distribution of flows, we vary the elements of $\mathbf{M}_\mu$ and $\mathbf{M}_\sigma$ cyclically in time, but with no change to the log-space mean flow. In order to achieve this, we fit Fourier series to the historical log-space monthly means and calculate time-varying multipliers by dividing the monthly means predicted by an adjusted harmonic, y_2, by those predicted by the historical fit, $\hat{y_1}$. We do not calculate different within-year multipliers for $\mathbf{M}_\sigma$, only $\mathbf{M}_\mu$, but future work could explore within-year changes to $\mathbf{M}_\sigma$ as well.

In this system, we found that the log-space mean monthly flows in each of the five Red River tributaries were modeled well by the first two harmonics (with R^2 values between 0.996 and 0.999 across the sites). That is, the log-space mean monthly flow, $\widehat{y_1(i)}$, at each site can be modeled by:

$$\widehat{y_1(i)} = \bar{y} + C_1 \cos\left(\frac{2\pi i}{12} - \phi_1\right) + C_2 \cos\left(\frac{2 \times 2\pi i}{12} - \phi_2\right) \tag{5.12}$$

where $\bar{y}$ is the mean of the historical time series of log-space monthly means, C_1 and ϕ_1 are the amplitude and phase, respectively, of the first harmonic (i.e. annual cycle), C_2 and ϕ_2 are the same for the second (i.e. semi-annual cycle), and i is the month of the year, from 1 for May (the beginning of the monsoon) to 12 for April (the end of the dry season). We create adjusted time series, y_2, for each site by applying multipliers to C_1 and/or C_2 and phase shifts

to ϕ_1 and/or ϕ_2. Increases in the amplitudes capture strengthening monsoons, while decreases capture weakening monsoons, and adjustments to the phase shifts capture changes in its timing. These parameters could also capture effects caused by upstream dam construction in China, again highlighting that the cause may not matter, as long as water managers can track the effect.

The amplitude multipliers, m_{C_1} and m_{C_2}, and phase shift deltas, d_{ϕ_1} and d_{ϕ_2}, are used to calculate a new cycle of mean monthly flows, y_2, according to Equation 5.13:

$$y_2(i) = \bar{y} + m_{C_1} C_1 \cos\left(\frac{2\pi i}{12} - (\phi_1 - d_{\phi_1})\right) + m_{C_2} C_2 \cos\left(\frac{2\pi i}{12} - (\phi_2 - d_{\phi_2})\right) \tag{5.13}$$

The i-th element of $\mathbf{M}_\mu$ and $\mathbf{M}_\sigma$ can then be calculated as $y_2(i)/\widehat{y_1(i)}$. In order to change the log-space mean annual flow and standard deviation simultaneous to the intra-annual distribution, the i-th element of $\mathbf{M}_\mu$ and $\mathbf{M}_\sigma$ can be calculated according to Equations 5.14 and 5.15:

$$M_{\mu,i} = [m_\mu y_2(i)/\widehat{y_1(i)}] \tag{5.14}$$

$$M_{\sigma,i} = [m_\sigma y_2(i)/\widehat{y_1(i)}] \tag{5.15}$$

Figure 5.2 illustrates the effects of m_{C_1}, m_{C_2}, d_{ϕ_1} and d_{ϕ_2} on the log-space mean hydrographs and FDCs of the Da River. Panel a shows the historical fit, $\hat{y_1}$, with a solid black line and the effects on y_2 of setting $m_{C_1} = 1.5$ or $m_{C_1} = 0.5$ with dashed black lines and of setting $m_{C_2} = 1.5$ or $m_{C_2} = 0.5$ with dotted black lines. In both cases, when the multiplier is greater than 1, the peaks are higher and

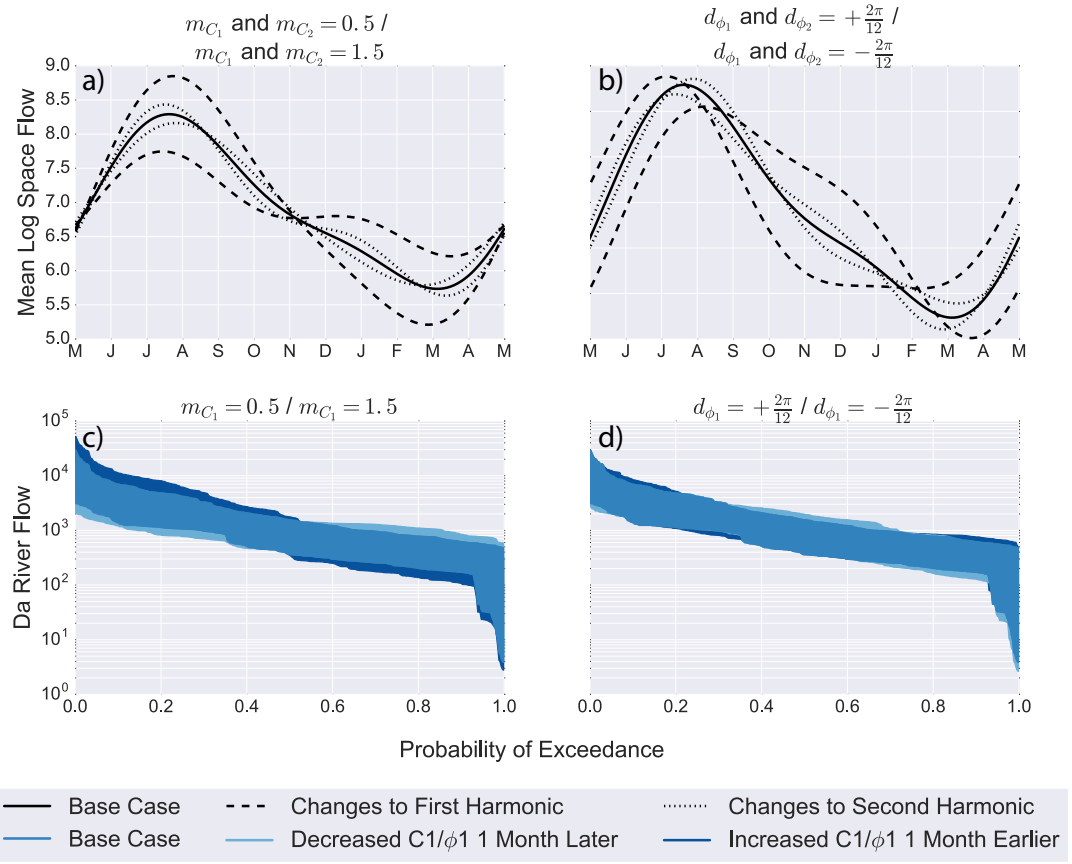

Figure 5.2: Effects of changes in the amplitude (panels a and c) and phase (panels b and d) of the first two harmonics on the log-space mean hydrographs (panels a and b) and flow duration curves (panels c and d) of the Da River. The multipliers displayed are those spanned in our scenario discovery experiment.

the troughs lower, and when the multiplier is less than one, the reverse is true. Panel a shows that m_{C_1} has a much greater influence on the annual cycle than m_{C_2}. Moving to panel c, the effect of m_{C_1} on the Da River FDCs is illustrated. The range of base case FDCs is colored medium blue, the range when m_{C_1} = 1.5 dark blue, and when m_{C_1} = 0.5 light blue. Again, one can see that the highest flows become larger when m_{C_1} = 1.5 and the lowest flows smaller, while the reverse is true when m_{C_1} = 0.5. However, it should be noted that increases

in monsoonal flows that occur without decreases in dry season flows can be modeled by combining changes to both the amplitude of the first harmonic and the log-space mean. It should also be noted that since these multipliers are applied in log-space, an increase in the amplitude of either harmonic will result in greater real-space increases in the monsoon season than decreases in the dry season, thereby increasing the real-space mean.

Panels b and d are similar to a and b, except they show the effects of d_{ϕ_1} and d_{ϕ_2} instead of m_{C_1} and m_{C_2}. In panel b, the historical mean fit, $\hat{y_1}$, is again shown with a solid black line, while the dashed lines show the effect of $d_{\phi_1} = \pm 2\pi/12$ radians (1 month shift) and the dotted lines the effect of $d_{\phi_2} = \pm 2\pi/12$ radians. Again, changes in the first harmonic have much greater effects than changes in the second. Panel d shows the effects of changes in d_{ϕ_1} on the Da River FDCs, with the range of historical flows shown in medium blue, the range of flows with a leftward phase shift in dark blue, and with a rightward phase shift in light blue. In panel b, it can be seen that the leftward phase shift of d_{ϕ_1} results in a slightly higher peak in July instead of August, and a higher trough in the dry season. These changes result in higher high and low flows in the FDCs in panel d. The rightward phase shift shows lower peaks and troughs in panel b, resulting in lower high and low flows in panel d, but higher moderate flows due to a slower die-off in the monsoon. However, these changes in the FDCs are much less significant than those observed as a result of m_{C_1} in panel c.

The final hydrologic factor we adjust in our scenario discovery analysis is the evaporation rate. This rate influences the amount of water evaporated from the reservoirs, which reduces the total system storage. Increased evaporation could therefore potentially help reduce floods, but at the cost of hydropower

production and water supply for irrigation. We apply a delta shift to historical evaporation rates of d_e mm/day in our analysis. Table 5.1 summarizes the ranges of all hydrologic factors explored in our study.

Table 5.1: Ranges of hydrologic factors sampled in our scenario discovery analysis. SOWs are generated from a Latin Hypercube sample across all hydrologic and socioeconomic factors, with each factor sampled uniformly within its bounds. However, it should be noted that since multipliers are applied in log-space, the samples are not uniform in real-space.

Deeply Uncertain Factor	Lower Bound	Upper Bound
Log-space Mean Multiplier, m_μ (−)	0.05	1.05
Log-space Std Multiplier, m_σ (−)	0.5	1.5
Log-space C1 Multiplier, m_{C_1} (−)	0.5	1.5
Log-space C2 Multiplier, m_{C_2} (−)	0.5	1.5
Log-space ϕ_1 Delta, d_{ϕ_1} (radians)	$-2\pi/12$	$+2\pi/12$
Log-space ϕ_2 Delta, d_{ϕ_2} (radians)	$-2\pi/12$	$+2\pi/12$
Evaporation Delta, d_e (mm/day)	-0.5	+1.0

Socioeconomic Uncertainties

In addition to generating alternative scenarios for how hydrologic characteristics might change in the future, we also generate alternative demand scenarios to capture how evolving socioeconomic conditions might influence the robustness of alternative reservoir operating policies. We break the demand down into three primary sectors: agriculture, aquaculture and other. We apply independent multipliers to each of these sectors: m_{ag} for agriculture, m_{aq} for aquaculture, and m_o for all other demands, sampling greater potential growth in other demands than agriculture and aquaculture due to urbanization. Figure 5.3a shows

the historical distribution of these demands over time, while Figure 5.3b shows the effect of simultaneously sampling the upper and lower limits of m_{ag}, m_{aq}, and m_o to illustrate the full range of sampled scenarios. We also sample a delta shift in the timing of total demand, d_D, from 30 days earlier to 30 days later, illustrated in Figure 5.3c. The ranges explored for each of these factor adjustments are given in Table 5.2. In our experimental design, we generate a total of 1000 alternative SOWs from a Latin Hypercube sample across the factor ranges given in Tables 5.1 and 5.2, sampling all factors uniformly within their bounds. While many of these deeply uncertain factors may be correlated in reality, the goal of this experiment is simply to fill the space of plausible futures to identify SOWs in which current policies fail. Subsequent analysis considering correlations among different uncertain factors can then be used to assess whether or not these failure states may be likely.

Table 5.2: Ranges of socioeconomic factors sampled in our scenario discovery analysis. SOWs are generated from a Latin Hypercube sample across all hydrologic and socioeconomic factors, with each factor sampled uniformly within its bounds.

Deeply Uncertain Factor	Lower Bound	Upper Bound
Agricultural Demand Multiplier, m_{ag} (–)	0.5	1.5
Aquaculture Demand Multiplier, m_{aq} (–)	0.5	2.0
Other Demand Multiplier, m_o (–)	0.5	5.0
Demand Delta, d_D (days)	-30	+30

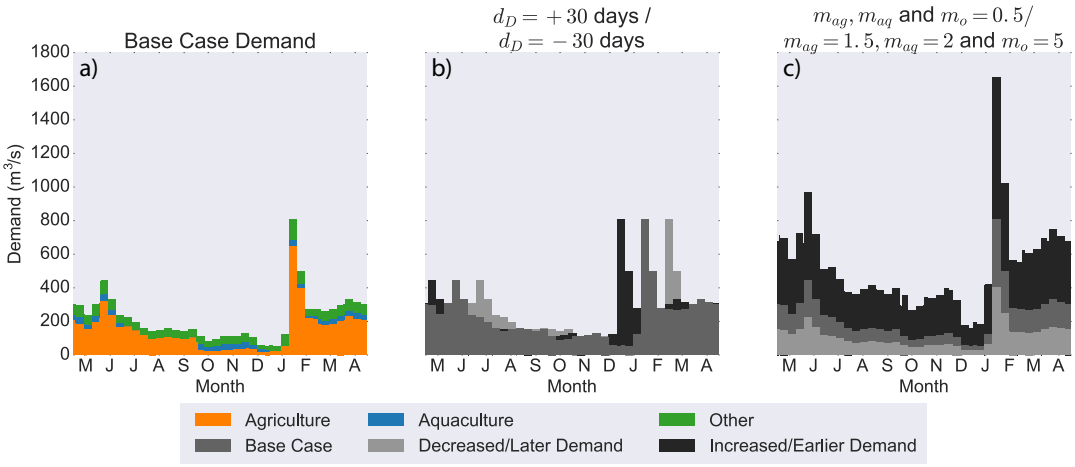

Figure 5.3: Base case distribution and timing of average water demand in the Red River basin across sectors (panel a), as well as the effects of changes in the timing (panel b) and amount (panel c) of sectoral demand. Panel a shows that most of the water demand is for agriculture (orange). The second largest sector is aquaculture, shown in blue. Demand from all other sectors, including industrial and municipal, are shown in green. In panels b and c, base case, i.e. average historical conditions, are shown in medium gray, increased and earlier demand in black, and decreased and later demand in light gray. The multipliers displayed are those spanned in our scenario discovery experiment.

5.4.3 Scenario Discovery

After re-evaluating the candidate operating policies generated from our multi-objective optimization on the alternative SOWs described in Section 5.4.2, we calculate the robustness of each policy using a satisficing metric quantified as the percent of worlds in which $J_{Flood} \leq 2.15$ m (providing protection to the 100-yr flood), $J_{Hydro} \geq 25$ Gwh/day, and the worst first percentile of the maximum daily deficit is less than 350 m^3/s. We choose this formulation of the deficit criterion, called J_{MaxDef}, because it is more intuitive than the worst first percentile of the average daily squared deficit. However we still optimize to $J_{Deficit^2}$ to mini-

mize both high deficits and persistent low deficits. The results of our robustness analysis are presented in Section 5.5.1.

After quantifying the robustness of each alternative operating policy, we select the most robust solution on each objective, as well as across objectives, for further analysis. We then use logistic regression to identify combinations of hydrologic and socioeconomic factors under which the selected policies fail to attain satisfactory performance. We also analyze why these conditions are unfavorable by visualizing their impacts on time series of downstream water levels, hydropower production and water supply deficits. Determining which factors most influence system performance, and visualizing in what way they influence performance, could aid water managers in deciding when to change management plans. The results of this scenario discovery analysis are given in Section 5.5.2.

5.4.4 Scenario Evaluation

The final step in our robustness analysis is to evaluate the plausibility of the hydrologic scenarios discovered to be of concern in Section 5.5.2. Typically, this is done by driving hydrologic or statistical models with projected precipitation and temperature series from GCMs. However, we note from Figure 5.4 that the amplitude (panel a) and phase (panel b) spectra of the historical precipitation series (shown in red) are similar in shape to those of the historical streamflow series (shown in blue), with the amplitude of the streamflow series tapering off at higher frequencies due to the landscape acting as a low pass filter. In particular, the shapes are similar around the frequencies of 1 and 2 cycles per year

(banded in green), corresponding to the first two harmonics. Assuming this relationship holds in the future, this suggests that projected changes in the amplitude and phase of the annual and semi-annual precipitation time series can be viewed as a proxy for the corresponding changes in streamflow time series. We can also assume changes in the mean and standard deviation of the precipitation series will be nearly the same as those for streamflow for mass balance purposes. For these reasons, we assess the plausibility of the hydrologic scenarios discovered to be of concern by tracking the deeply uncertain hydrologic factors throughout time in different climate projections. However, since the log-space magnitude of the precipitation spectrum is slightly larger than that of the streamflow spectrum, the real-space effect of a given percent change in the log-space precipitation spectrum will be larger than the corresponding real-space effect of the same percent change in the log-space streamflow spectrum. Consequently, the changes in precipitation projected by downscaled climate models can be considered conservative estimates of the corresponding changes to streamflow.

For this analysis, we track the deeply uncertain hydrologic factors in downscaled projections from 17 CMIP5 models run with multiple initial conditions across all four representative concentration pathways (RCPs). Before downscaling, we bias correct the CMIP5 projections using the APHRODITE observation dataset [Yatagai et al., 2012] and a variation of the modified equidistant cumulative distribution function matching (EDCDFm) algorithm [Li et al., 2010] developed by Pierce et al. [2015]. EDCDFm is a quantile mapping approach that adjusts the model CDF based on value differences by quantile between the model and observations over an historical calibration period. Unlike traditional quantile mapping approaches, EDCDFm preserves model-predicted me-

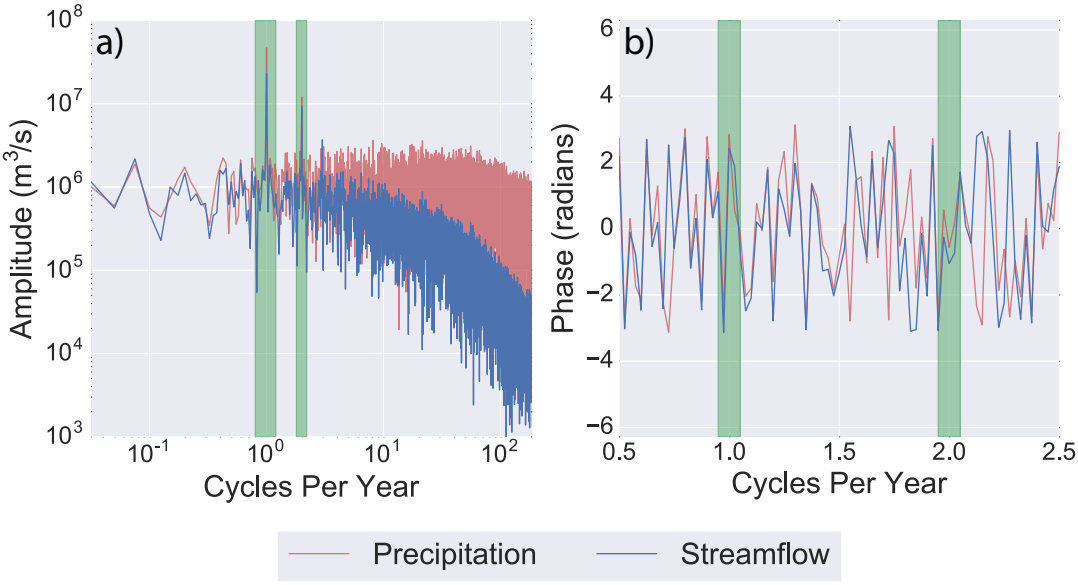

Figure 5.4: Amplitude (panel a) and phase (panel b) spectra of historical streamflow (blue) and basin-averaged precipitation (red) time series in the Red River basin from 1961-2000. Green bars surround the frequencies of 1 and 2 cycles per year, the frequencies of the first and second harmonics respectively. Panel a indicates that the shape of the streamflow and precipitation spectra is similar surrounding these two frequencies, while panel b indicates that the phase is also similar.

dian changes and accounts for future climate changes in distribution tails [Li et al., 2010; Pierce et al., 2015].

After bias correction, we statistically downscale the CMIP5 projections using a constructed analog approach similar to the method introduced by Pierce et al. [2014]. For a specific CMIP5 model and future day, we find the single best matching historical analog at the coarse spatial scale of the model. In traditional constructed analog methods [Hidalgo et al., 2008], multiple n best matching historical analogs are typically combined as a weighted average. Pierce et al. [2014] find that a single best analog reduces precipitation drizzle issues

caused by multi-analog averaging and provides a better representation of spatial dependence and extremes at the local scale. To decrease complexity, we do not use the localized analog selection procedure of Pierce et al. [2014] and instead select a best match analog using the entire Red River spatial domain. The final downscaled projections for the future day are the local high resolution APHRODITE observations that correspond to the historical coarse analog modified by a scaling factor that accounts for differences between the analog and future day [Pierce et al., 2014]. The results of our scenario evaluation over the downscaled climate projections are presented in Section 5.5.3.

5.5 Results

This section presents the results of our robustness analysis on the optimized multi-reservoir operating policies for the Red River basin. In Section 5.5.1, we show the multi-objective tradeoffs of the optimized policies under base case conditions, as well as their robustness across the deeply uncertain SOWs. In Section 5.5.2, we explore a subset of tradeoff solutions that capture a range of sectoral preferences to determine what ranges of the deeply uncertain factors cause these solutions to perform poorly on particular objectives. We also analyze how changes in the factors most influencing system performance affect extreme flooding, hydropower and deficit events in the basin. Finally, in Section 5.5.3 we evaluate whether or not the SOWs discovered to be of concern in Section 5.5.2 may plausibly occur under different climate projections. Together, these analyses could help reservoir operators set monitoring priorities and design triggers for re-optimization of system operations if the Red River system shows signs of evolving toward a state of concern.

5.5.1 Robustness of Tradeoff Solutions

Figure 5.5a displays the Pareto approximate Red River operating policies discovered through the multi-objective optimization described in Section 5.3 using a 2-D glyph plot in which each circle represents a different operating policy. The location of each policy along the x-axis corresponds to its average daily hydropower production in the worst first percentile year, its location on the y-axis to its height above 11.25 m under the 100-yr flood, and its size to its average daily squared deficit in the worst first percentile year, with smaller circles representing smaller deficits. The solutions have also been shaded by their performance on the hydropower objective, with darker shades corresponding to greater production. Arrows along the x and y axes indicate the directions of favorable performance, with a star located at the ideal point in the lower left corner. The arc of the solutions from the upper left corner to the bottom right shows the strong tradeoff between the hydropower and flooding objectives, since high storages favor production, while low storages favor flood protection. Despite this strong tradeoff, all of the solutions in the initial baseline SOW meet the 2.15 m flood constraint while generating an average daily production of at least 30 Gwh/day in the worst first percentile year. Resimulating the optimized policies to calculate their maximum daily deficit, all of the solutions have a maximum daily deficit of less than 350 m^3/s in the worst first percentile year.

Panels b, c and d of Figure 5.5 show how the values of the flood, hydropower, and deficit objectives change when the Red River control solutions in panel a are re-evaluated across the deeply uncertain SOWs generated as described in Section 5.4.2. Each line in these figures corresponds to a different solution from panel a, and shows the cumulative percent of SOWs in which that solution

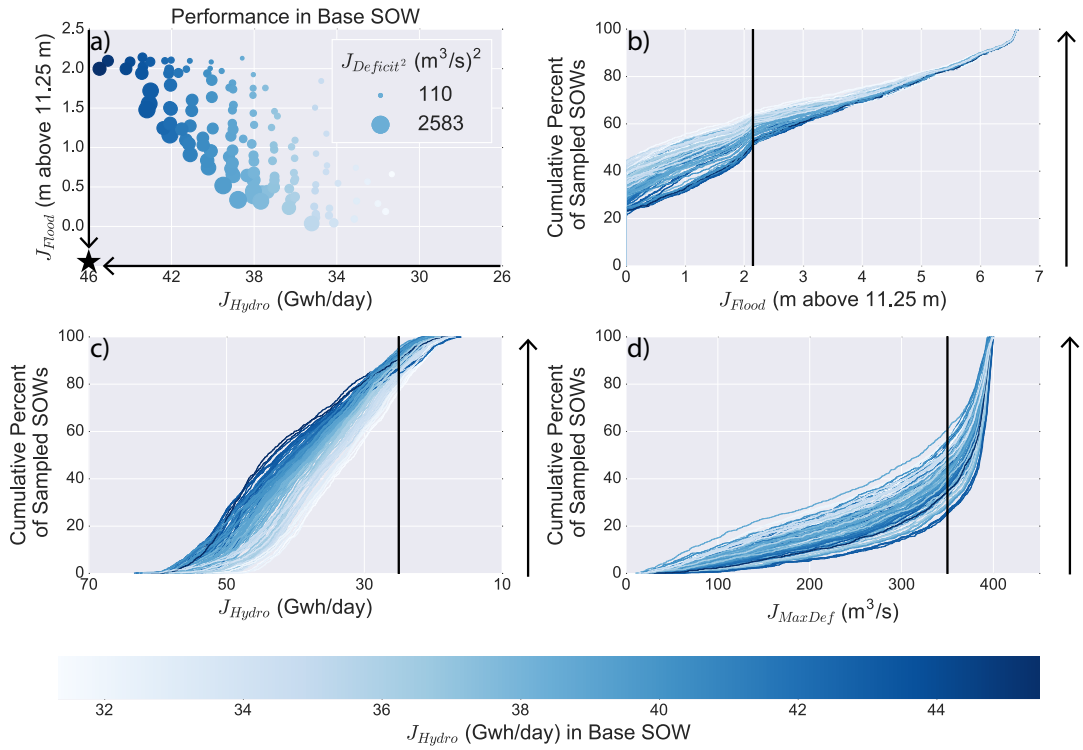

Figure 5.5: Pareto approximate solutions to the Red River optimization problem described in Section 5.3 and their re-evaluated performance across the deeply uncertain SOWs generated as described in Section 5.4.2. Panel a shows a 2-D glyph plot of the tradeoffs in the base SOW, while panels b, c and d show the re-evaluated performance of each solution across the deeply uncertain SOWs on the flood, hydropower and deficit objectives, respectively. All solutions are shaded by their performance on the hydropower objective in the base SOW. Black lines in panels b, c and d indicate minimum performance thresholds on the three objectives and show that the Pareto approximate operating policies achieve the flood threshold in about 50-65% of SOWs, the hydropower threshold in about 80-95%, and the deficit threshold in about 25-60%.

achieves different levels of performance for each objective. Panel b displays performance on the flood objective, panel c on the hydropower objective, and panel d on the deficit objective. In each of these panels, the favorable direction is up, as this means a given solution achieves the level of performance given on

the x-axis in a greater percent of the SOWs than solutions below it. The lines are shaded by the performance of that solution on the hydropower objective in the base SOW, providing some insight into where the solutions in panel a map to in panels b, c and d.

Examining the re-evaluated performance of the Pareto approximate solutions on the flood objective (panel b), the most striking observation is the color gradient from dark blue to light blue in moving from the bottom solutions to the top. This indicates that the lower a solution's hydropower production in the base SOW, the more robust it is on the flood objective. Tracing the solutions from bottom to top along the black line drawn at 2.15 m, it can be seen that the solutions in the approximate Pareto set are able to provide protection to the 100-yr flood in approximately 50-65% of the generated SOWs. As the flood events get larger and more catastrophic, the variability in performance across solutions becomes much smaller, indicating that in really wet worlds the reservoir control operations from the tradeoff solutions in panel a are futile, as the system is driven entirely by hydrology. As the flood events get smaller, though, variability across solutions increases, and operations have a greater effect on system performance.

Moving to panel c, the color gradient generally reverses, with solutions that achieve greater hydropower production in the base SOW continuing to achieve greater production in the alternative SOWs. However, this is only true within a range of about 30-50 Gwh/day, approximately the range of production achieved by the Pareto approximate solutions in the base SOW. As the solutions are subjected to SOWs in which production falls outside this range, the best solutions degrade more and the gradient becomes muddled. This illustrates that for the

hydropower objective, there is a tradeoff between optimality to the base SOW and robustness across SOWs, a classic example of why operators may want to trade optimal performance for less sensitivity to assumptions [Lempert & Collins, 2007]. Alternatively, operators could consider re-optimizing operating policies only if the system appears to be moving toward SOWs in which >50 Gwh/day or <30 Gwh/day are achieved in the worst first percentile year. Since Vietnam is heavily investing in additional sources of electric power [Asian Development Bank, 2016], we assume they may be willing to sacrifice some production in exchange for better performance on other system objectives. As such, we assume 25 Gwh/day would be a satisfactory performance level in the worst first percentile year. Figure 5.5c shows that all solutions meet this performance threshold in about 80-95% of the SOWs, showing that the system's hydropower production is fairly robust to potential changes in hydrologic and socioeconomic conditions.

Finally, the robustness of the Pareto approximate operating policies on the deficit objective is shown in Figure 5.5d. Unlike the flood and hydropower objectives, there is no obvious color gradient across solutions in this figure, indicating a more complex relationship between a solution's hydropower production in the base SOW and its robustness on the deficit objective. However, similar to the flood objective, there is little change in the rank order of solutions along the range of deficits observed in the alternative SOWs. Therefore, there is not a strong tension between a given policy's performance on the deficit objective in the base SOW and its robustness on that objective across the broader ensemble of SOWs. Additionally, demand management has the potential to improve the robustness of all of these solutions on that objective. In the base SOW, the worst-performing solution had a maximum daily deficit in the worst first per-

centile year of 322 m^3/s. Assuming a 10% increase is tolerable, we consider a maximum daily deficit of 350 m^3/s in the worst first percentile year to be acceptable. Examining Figure 5.5d, we see that the solutions in the Pareto approximate set are able to achieve this minimum performance level in about 25-60% of generated SOWs. This low level of robustness and high variability in performance across policies suggests that the deficit may be highly sensitive to changes in hydrologic and socioeconomic conditions under some, but not all, operating policies.

Knowing there is great variability in performance on the deficit objective across policies, and that there is a strong tradeoff between flooding and hydropower under the base SOW, Figure 5.5 suggests that achieving high levels of robust performance across all three objectives simultaneously may pose a significant challenge. Figure 5.6 illustrates just how challenging this is using a parallel axis plot in which each blue line represents a different operating policy, shaded according to the percent of generated SOWs in which it achieves all three performance thresholds (i.e. $J_{Flood} \leq 2.15$ m, $J_{Hydro} \geq 30$ Gwh/day and $J_{MaxDef} \leq 350$ m^3/s), with darker shades indicatings greater robustness. Each line representing a different operating policy crosses each vertical axis at the objective value it achieves in the base SOW, with the favorable direction along each axis being down.

Across the solutions in the Pareto approximate set, all three performance goals can only be met simultaneously in about 10-30% of generated SOWs. While this may sound low, the generated worlds are not equally likely. In fact, some may be extremely unlikely. As such, this metric should not be interpreted as a reliability but as a measure for comparing the relative robustness of alterna-

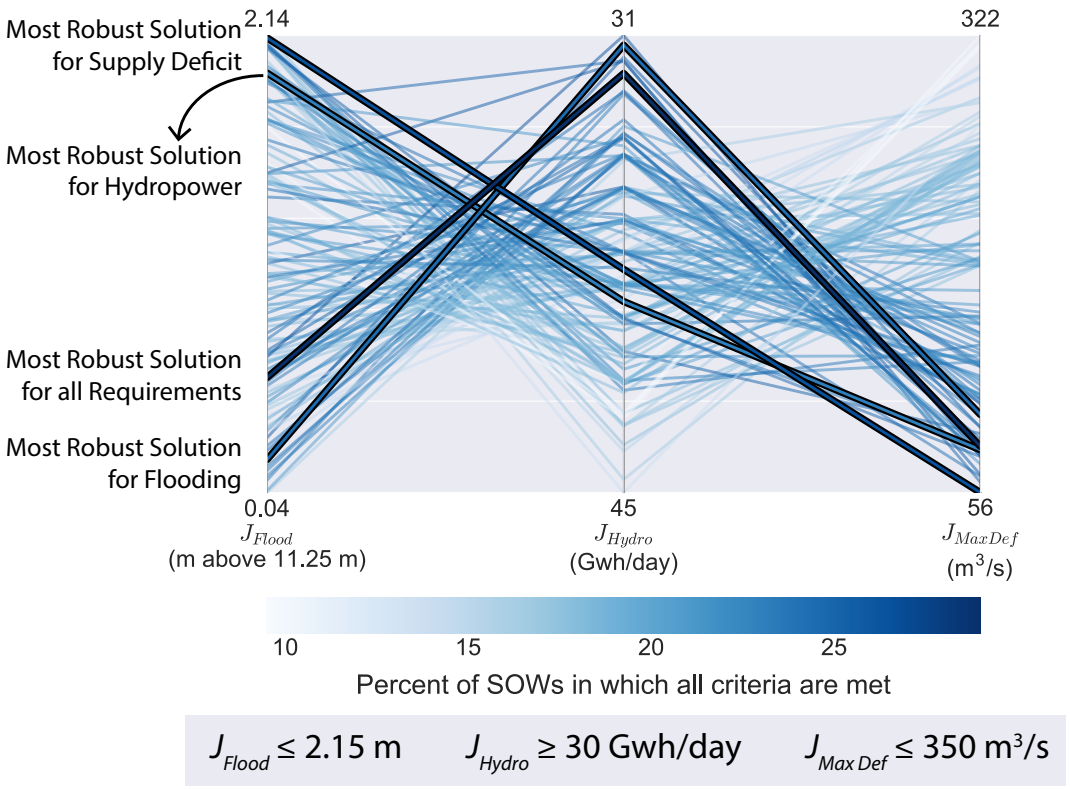

Figure 5.6: Parallel axis plot of the tradeoff set in the base SOW, with each solution shaded by its robustness across all system objectives. Each line represents a different operating policy, crossing the three axes at the objective value it achieves on the corresponding objective in the base SOW. The shading for robustness corresponds to the percent of generated SOWs in which each solution meets all of the minimum performance criteria for the three objectives. The most robust solutions for flooding and the deficit are near optimal on those objectives in the base SOW, while the most robust solution for hydropower production is sub-optimal in the base SOW. The most robust solution across all objectives favors flood protection and deficit minimization, with strong sacrifices in hydropower performance.

tive solutions. Looking first at the most robust solutions with respect to flooding and the deficit, highlighted in Figure 5.6, one can see that they are near-optimal on those objectives in the base SOW. However, the most robust solution for hy-

dropower production is sub-optimal in the base SOW. The reason for this is because the satisficing threshold of 25 Gwh/day in the worst first percentile year is below the 30-50 Gwh/day range in which the near-optimal hydropower solutions are most robust (see Figure 5.5). To achieve satisfactory performance across all objectives, the most robust solution, shaded dark blue, heavily favors flood protection and deficit-minimization over hydropower production in the base SOW. In Section 5.5.2 we investigate which combinations of deeply uncertain factors cause each of these four highlighted solutions to fail to meet the minimum performance criteria for each objective.

5.5.2 Scenario Discovery

We further analyze how the deeply uncertain hydrologic and socioeconomic factors influence the performance of alternative Red River operating policies using logistic regression. This allows us to model the probability that a given control policy is able to satisfy a particular performance criterion as a function of the deeply uncertain factors. While we may not know the probability that a given SOW will occur, we can use logistic regression to predict the probability that a policy will be able to meet satisfactory performance levels in that SOW, should it occur. Stakeholders can then divide the space of deeply uncertain factors into success and failure regions based on the probability with which they would like to satisfy different performance criteria. A detailed description of the logistic regression modeling is provided in Appendix C.

Figure 5.7 illustrates this analysis for the most robust solution with respect to flooding to show where even this policy is unable to meet the flooding perfor-

mance criterion. According to the logistic regression model, the most important factors in determining whether or not this solution can succeed on the flooding objective are the multipliers on the log-space mean, m_μ, the log-space standard deviation, m_σ, and the log-space amplitude of the first harmonic, m_{C_1}. Figures 5.7a and b show where the most robust solution for flooding is able to provide protection to the 100-yr flood as a function of these three factors. These panels represent two dimensional projections of the 1000 SOWs generated in our scenario discovery experiment, with each circle representing a different SOW. The circles are shaded light blue if the policy was able to meet the flooding criterion in that world and dark red if it was not. The x-axis in both panels is the log-space mean multiplier, while the y-axis in panel a is the log-space amplitude of the first harmonic multiplier and the y-axis in panel b is the log-space standard deviation multiplier.

Not surprisingly, wetter worlds ($m_\mu > 1$) increase the probability of failure. However, if the wetter worlds are accompanied by decreases in the log-space amplitude of the first harmonic ($m_{C_1} < 1$) or the annual standard deviation ($m_\sigma < 1$), the most robust flood solution can tolerate greater increases in the log-space mean flow. This is because a decrease in the amplitude of the first harmonic results in drier monsoons and wetter dry seasons. Since flooding is only of concern during the monsoon season, an increase in the mean annual flow can be tolerated if the within-year flow distribution becomes more even, reducing flow peaks during the monsoon. Similarly, a decrease in the annual standard deviation results in less extreme annual highs and lows, so an increase in the mean flow can be tolerated if inter-annual variability decreases.

This interaction between the three multipliers is fairly linear, and approxi-

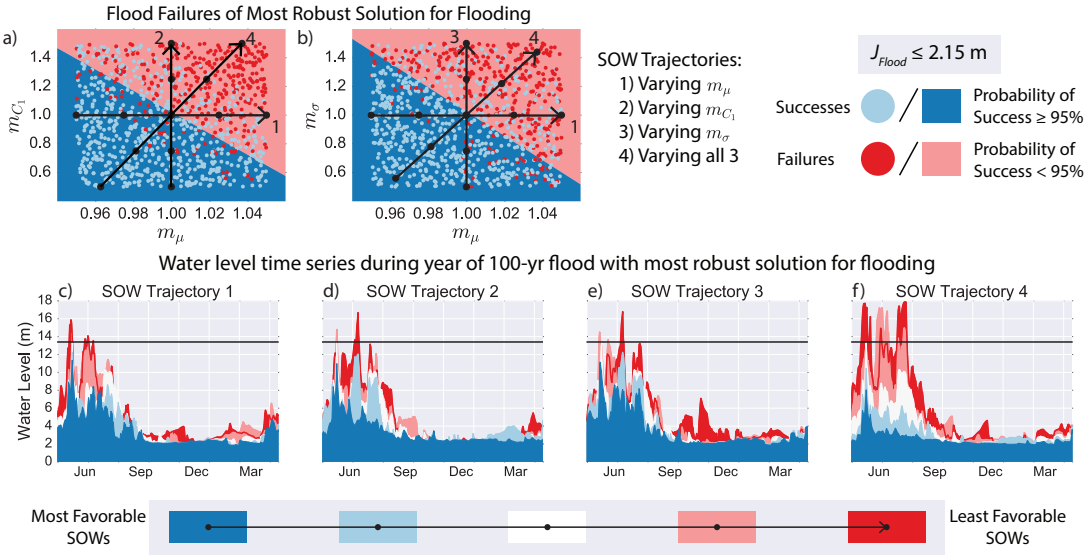

Figure 5.7: Successes and failures of the most robust solution for flooding (panels a and b). The log-space mean multiplier, m_μ, amplitude of the first harmonic multiplier, m_{C_1}, and standard deviation multiplier, m_μ are the most important factors controlling performance. Trajectories along each of these factors individually and jointly in the direction of greatest variability correspond to the flood events illustrated in panels c-f. Panel c shows the effect of the most important factor when moving along the points in Trajectory 1 (varying m_μ), panel d shows the same for Trajectory 2 (varying m_{C_1}), panel e for Trajectory 3 (varying m_σ), and panel f for Trajectory 4 (varying all three). Blue events correspond to the most favorable worlds along the trajectory, while red events represent the least favorable, with a gradient for the worlds sampled in between. The black lines in each panel represent the minimum performance thresholds for the maximum water level at Hanoi in the worst first percentile year.

mate failure boundaries can be defined using logistic regression. Since the effects of overtopping could be catastrophic for Hanoi, we assume a conservative classifier of success and failure regions in which the success region is defined by having at least a 95% chance of providing protection to the 100-yr flood within that SOW. The success region is shaded dark blue under the dots in Figures 5.7a

and b, while the failure region is shaded light red. With this classifier, it can be seen that any increase in one of the three factors from their base value of 1 results in movement to the failure region, indicating high sensitivity of even the most robust solution for flooding to each of these hydrologic factors. This strong sensitivity of flood protection to one's estimate of simple streamflow statistics underscores the potential dangers of optimizing to stationary hydrology [Milly et al., 2008].

To visualize how flood events change as a function of each of these deeply uncertain hydrologic factors, we have plotted time series of flows during the year of the 100-yr flood when operating with the most robust solution for flooding in different SOWs. These events are shown in Figures 5.7c-f. Each of the events illustrated in panels c-f corresponds to a different point along the SOW trajectories shown in panels a and b. The dark blue events illustrate the dynamics resulting from operating in the most favorable SOW along each trajectory, while the dark red events illustrate the dynamics resulting from operating in the least favorable SOW, with a gradient of colors representing the SOWs in between. The events in panel c correspond to the points along SOW Trajectory 1, drawn in panels a and b, in which only the log-space mean multiplier, m_μ, is varied while all other factors are held constant at their base values. Panel d shows how events change along SOW Trajectory 2, drawn in panel a, in which only the log-space amplitude of the first harmonic, m_{C_1}, is varied. Panel e illustrates the same for SOW Trajectory 3, drawn in panel b, in which only the log-space annual standard deviation, m_σ, is varied. Finally, panel f shows how events change along SOW Trajectory 4, drawn in both panels a and b, in which all three of these hydrologic factors vary together perpendicular to the classifying boundary.

The first notable observation in looking at the flood events in Figures 5.7c-f is that even though the reservoir inflows applied in each SOW are simply monotonic transformations of the same set of stationary synthetic streamflows (see Section 5.4.2), the downstream water levels are not, illustrating the complex, non-linear effects of the reservoir control policies. In fact, different years from the stationary synthetic streamflows result in the 100-yr event in different SOWs due to the different ways in which the transformations influence system dynamics. This illustrates the importance of sampling a range of plausible changes in streamflow characteristics beyond just the mean. In particular, the log-space mean, log-space standard deviation, and log-space amplitude of the monsoonal cycle all influence the real-space mean, but change flood events in different ways.

Looking first at the influence of the log-space mean on flood events shown in panel c, when the most robust solution for flooding operates in the base SOW shown in white, it just barely provides protection to the 100-yr event. As the mean decreases/increases (blue/red events), the water levels decrease/increase at all times of the year, with greater differences observed during the monsoon season since the mean multiplier is applied in log-space. This year-round increase in the flows results in near-overtopping routinely occurring in really wet worlds. When only the log-space amplitude of the first harmonic is varied (panel d), there is little difference in flows outside of the monsoon season, and changes during the monsoon season apply mostly to the peaks. As a result, increasingly flashy flows cause overtopping or near-overtopping when only the amplitude of the first harmonic increases. Floods therefore become more severe, but of shorter duration. The effect of the log-space annual standard deviation (panel e) lies somewhere in between: while severe flood events are gen-

erated, those causing near overtopping are not as flashy as those caused solely by increases in the log-space amplitude of the first harmonic, nor as persistent as those caused by increases in the log-space mean. Considering all of these changes together (panel f), it can be seen how the effects of these factors compound each other to result in severe flood events of long duration.

Moving to Figure 5.8 the same analysis is applied to analyze the performance of the most robust solution for hydropower and its sensitivity to the deeply uncertain factors. Once again, hydropower performance failures for this solution were found to be most influenced by the log-space annual mean multiplier, m_μ (x-axis in panels a and b), log-space multiplier on the amplitude of the first harmonic, m_{C_1} (y axis in panel a), and log-space annual standard deviation multiplier, m_σ (y axis in panel b). To classify successes and failures on the hydropower threshold, we choose the 50% probability contour from the logistic regression model. Since alternative energy sources can be substituted for hydropower, stakeholders will likely be less concerned about always being able to satisfy this performance measure. With this classification, the hydropower success region for the most robust solution for hydropower encompasses nearly all SOWs.

Not only is this policy's ability to provide satisfactory hydropower production less sensitive to the deeply uncertain factors than the ability of the most robust solution for flooding to provide flood protection, but its success and failure regions are opposite of those for flooding. For hydropower production, wetter worlds ($m_\mu > 1$) increase production, especially if accompanied by an increase in the amplitude of the annual cycle ($m_{C_1} > 1$), resulting in greater production during the monsoon. This can be seen in both panels a defining success and

failure regions, as well as panels c and d illustrating how hydropower production varies as a function of m_μ and m_{C_1}, respectively. While the amplitude of the annual cycle impacts both hydropower production and flooding, its effect on hydropower is weaker than for flooding because an increased amplitude will also result in drier dry seasons. Consequently, energy production during the dry season will decrease in these worlds, negating some of the monsoonal benefits. This results in a steeper slope dividing the success and failure regions as a function of the mean annual flow and monsoonal amplitude for hydropower than for flooding. While such intra-annual variability would normally decrease hydropower production, because the multipliers are applied in log-space, an increase in the log-space amplitude of the first harmonic increases the real-space mean. Consequently, increased production in the monsoon exceeds the decrease in the dry season.

The interaction between the log-space mean annual flow and its standard deviation is also weaker for hydropower performance than for flooding, as shown in panel b. This is because m_σ has a weaker influence on hydropower production than m_μ, illustrated by the small changes in hydropower production along SOW Trajectory 3 in panel e and the fact that trajectories along all three factors (panel f) look very similar to trajectories solely along the mean and first harmonic (panels c and d). The interaction between m_μ and m_σ is also of opposite sign for hydropower than for flooding. Increased standard deviations result in higher annual highs and lower annual lows, but since the hydropower objective specifically targets the worst first percentile, the lower lows result in more failures. Consequently, an increase in the standard deviation of annual flow is bad for both hydropower production and flooding, whereas an increase in the mean or amplitude of the first harmonic is bad for flooding, but good for hydropower

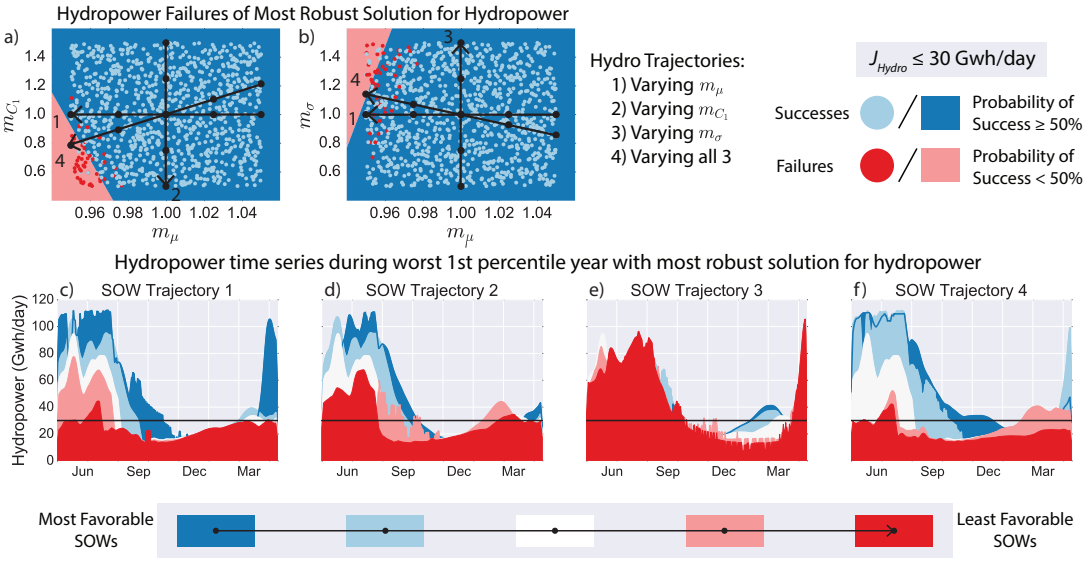

Figure 5.8: Successes and failures of the most robust solution for hydropower (panels a and b). The log-space mean multiplier, m_μ, amplitude of the first harmonic multiplier, m_{C_1}, and standard deviation multiplier, m_μ are the most important factors controlling performance. Trajectories along each of these factors individually and jointly in the direction of greatest variability correspond to the hydropower time series illustrated in panels c-f. Panel c shows the effect of the most important factor when moving along the points in Trajectory 1 (varying m_μ), panel d shows the same for Trajectory 2 (varying m_{C_1}), panel e for Trajectory 3 (varying m_σ), and panel f for Trajectory 4 (varying all three). Blue events correspond to the most favorable worlds along the trajectory, while red events represent the least favorable, with a gradient for the worlds sampled in between. The black lines in each panel represent the minimum performance thresholds for average daily production in the worst first percentile year.

production.

For the deficit objective, Figure 5.9 shows that a different suite of deeply uncertain factors emerge as most important in determining when the most robust solution for the deficit fails to meet the deficit satisficing threshold. The log-

space mean flow is still important, but instead of interacting primarily with the amplitude of the first harmonic and inter-annual standard deviation, it interacts more with the multipliers on agricultural demand, m_{ag} (y-axis in panels a and b), and other demands such as industrial and municipal, m_o (x-axis in panel b). In fact, the greatest interaction is between the agricultural and other demand growth multipliers (panel b).

For the deficit objective, we define failure boundaries using the 75% probability contour from the logistic regression model. Since 70% of the population is employed in agriculture [Nguyen et al., 2002], performing well on the deficit is important for the region's food security, but the consequences of failing to provide sufficient water supply are not as severe as the consequences of flooding Hanoi. The 75% cut-off is therefore chosen to model a moderately risk-averse stakeholder who values confidence in performance on the deficit more than for hydropower, but less than for flooding. Panel a shows that, with this classifier, any decrease in the log-space mean or increase in agricultural demand results in movement to the failure state, as the boundary falls right along the base values of 1.0 on each of these multipliers. Moving to panel b, one can see that greater increases in other demands can be tolerated, as the failure boundary lies at a multiplier value of 2.75 on this factor when there is no growth in agricultural demand.

Figures 5.9c-f show how deficits change as a function of these hydrologic and socioeconomic factors. Across all panels, the deficits follow an approximately biweekly pattern caused by the tidal cycle's influence on the water volume in the irrigation canals. Looking first at the effects of the log-space mean annual flow shown in panel c, one can see that as the mean increases, deficits are only

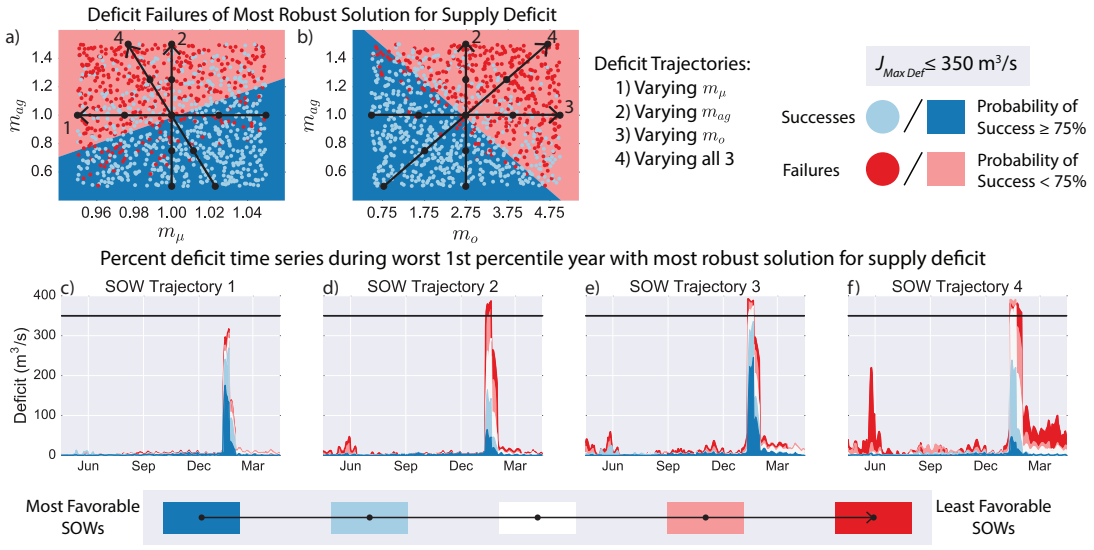

Figure 5.9: Successes and failures of the most robust solutions for the deficit (panels a and b). The log-space mean multiplier, m_μ, agricultural demand multiplier, m_{ag} and other demand multiplier, m_o, are the most important factors controlling performance. Trajectories along each of these factors individually and jointly in the direction of greatest variability correspond to the deficit time series illustrated in panels c-f. Panel c shows the effect of the most important factor when moving along the points in Trajectory 1 (varying m_μ), panel d shows the same for Trajectory 2 (varying m_{ag}), panel e for Trajectory 3 (varying m_o), and panel f for Trajectory 4 (varying all three). Blue events correspond to the most favorable worlds along the trajectory, while red events represent the least favorable, with a gradient for the worlds sampled in between. The black lines in each panel represent the minimum performance thresholds for the maximum daily deficit in the worst first percentile year.

impacted during the dry season, particularly at the time of planting in February (see Figure 5.3). Across all worlds, deficits are low at all other times of the year. However, if the mean annual flow stays constant while agricultural demand increases (panel b), deficits begin to worsen at the end of the dry season and beginning of the monsoon as well. Furthermore, increases in the deficit during

the time of planting become severe, resulting in failure to keep the maximum deficit below 350 m^3/s in the worst first percentile year. The same is true if only other sectoral demands increase (panel c), but the effect is even greater from this sector than from agriculture, as we sample greater potential growth in these sectors. If all of these factors change simultaneously in the direction of greatest variability (panel d), some high deficits begin to emerge during the late dry season and early monsoon season in addition to more prolonged, severe deficits at the time of planting.

5.5.3 Scenario Evaluation

Figures 5.7-5.9 provide insight into which deeply uncertain factors increase the probability that a control policy is able to meet individual performance objectives in different SOWs and why. Decision makers evaluating operating policies for the multi-reservoir Red River system should closely track each of these factors as additional observations become available to determine if they should re-optimize the reservoir operating policies, incentivize demand management strategies, or perhaps build new infrastructure or import additional power or food. Given the proximity of the base hydrologic factors to the flooding boundary, they may even want to re-optimize policies now across a range of plausible deeply uncertain SOWs to improve robustness on the flooding objective. While determining a plausible range of deeply uncertain SOWs is challenging, projected climatic changes from GCMs can at least inform decision makers of which changes may be more likely than others.

To further guide operators on which of these conditions they should be most

vigilant about tracking, we have traced each of the most important hydrologic factors through statistically downscaled climate projections from the CMIP5 ensemble. As illustrated in Figure 5.4, total monthly rainfall is a strong proxy for monthly streamflow. Consequently, we can use the downscaled precipitation projections to estimate changes in the log-space mean annual flow, amplitude of the annual monsoonal cycle, and annual standard deviation. Defining 1976-2005 as climatology within each projection, we re-estimate each of these three factors over 30-year moving windows to calculate time-varying multipliers equivalent to those sampled in our scenario discovery analysis. By observing how these multipliers change in time, we can identify which combinations of hydrologic changes may be more likely to occur than others, and whether or not those combinations fall outside a given policy's success region. Here we define a success region using the most robust solution across all requirements (see Figure 5.6). Failure boundaries on the flooding and hydropower objectives were again determined using logistic regression, with success probabilities of at least 95% and 50% used to define the success region on the flooding and hydropower satisficing criteria, respectively. Since the deficit is controlled primarily by socioeconomic factors and the climate projections can only be used as proxies for hydrologic changes, we do not consider failures on the deficit in defining the success region for this analysis.

The spread of traces across the CMIP5 models is illustrated in Figure 5.10 on top of the success and failure regions of the most robust solution across all requirements. Each CMIP5 model is represented by a different shape and is colored according to the RCP under which the projection was run. Many model/color combinations repeat because an ensemble of initial conditions were run. Three snapshots of these traces are shown: 1980-2009 (panels a and

d), 2025-2054 (panels b and e) and 2070-2099 (panels c and f). The first row of Figure 5.10 shows combinations of the multiplier on the log-space annual mean flow, m_μ, on the x-axis and the multiplier on the log-space amplitude of the first harmonic, m_{C_1}, on the y-axis. The second row does the same but with the multiplier on the log-space annual standard deviation, m_σ, on the y-axis.

While the first time stamp is only four years after the defined climatology from 1976-2005, at which each multiplier is set to 1, after only four years, the model projections begin to spread along the boundary of the success region. Marching forward in time, the spread in model projections significantly increases, many of them to regions well outside the success region. In particular, the projections tend to move to the right of the success region, where flooding is of great concern. Like the flooding events shown in Figure 5.7, this highlights the potential dangers of optimizing operations assuming stationary hydrology and rigidly fixing operational control rules designed under these assumptions.

While the projections do not seem to cluster by RCP, they do cluster by model and there is a strong negative correlation in where the models move, as wetter worlds ($m_\mu > 1$) tend to coincide with less intra-annual variability (decreased m_{C_1}). This correlation structure of the hydrologic factors suggests that when the log-space mean annual flow increases or decreases, most of the changes are driven by dry season flows. Increases in the amplitude of the first harmonic should result in relatively wetter monsoons and drier dry seasons compared to the mean, so if this occurs in conjunction with a decrease in the log-space mean, most of the decreased mean must occur during the dry season. Similarly, decreases in the amplitude of the first harmonic should result in relatively drier monsoons and wetter dry seasons compared to the mean, so if this is accom-

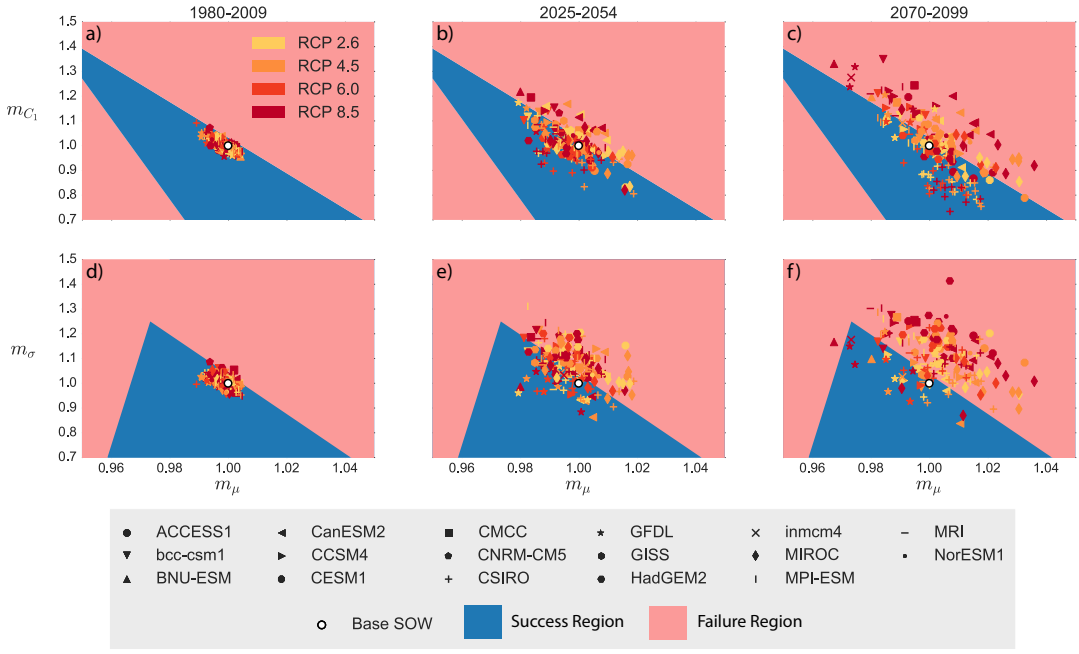

Figure 5.10: Time-varying hydrologic factors tracked through CMIP5 model projections laid atop the success and failure regions of the most robust solution across all requirements shown in Figure 5.6. With respect to the log-space annual mean, m_μ, and log-space amplitude of the first harmonic, m_{C_1} (panels a-c), the projections quickly depart from the base SOW to the upper left and lower right. These regions represent drier, more variable worlds and wetter, less variable worlds, respectively. Of particular concern are the many projections that move to the right of the success region, resulting in failure to provide protection to the 100-yr flood. With respect to the log-space annual mean and standard deviation, m_σ, (panels d-f), there is no noticeable correlation between these factors. However, most of the GCM projections experience increased interannual variability, migrating to the flooding failure region by the end of the century.

panied by an increase in the log-space mean, most of the increased mean must occur during the dry season.

While the dry season tends to control both increases and decreases in the

log-space mean flow, there is no general tendency across models for m_μ or m_{C_1} to increase or decrease. m_σ, however, increases more frequently than it decreases, which is consistent with the theory that global warming will increase climate variability [Trenberth, 2011]. Given that increased variability results in poorer performance on both flooding and hydropower production in the tails (see Figures 5.7-5.8), operators may want to consider re-optimizing policies to more variable worlds. If operators feel comfortable re-optimizing operations across even more deeply uncertain SOWs to improve policy robustness, optimizing over the elliptical cloud of factors spanned by these model projections may be a good place to start.

5.6 Conclusions and Future Work

This study advances scenario discovery analyses for multi-reservoir systems, contributing a new method for jointly exploring the effects of changing monsoonal dynamics and socioeconomic demands on river basin management. We demonstrate this method on the Red River basin in Vietnam to evaluate how food-energy-water tradeoffs in the basin are impacted by different plausible futures. While we found the system to be fairly robust with respect to hydropower production, several hydrologic and socioeconomic factors threatened the system's ability to protect Hanoi from severe flooding and to provide sufficient water supply for agriculture. Specifically, we found the log-space mean annual flow, annual standard deviation, and amplitude of the annual monsoonal cycle to critically impact flood risks to Hanoi. We also found that agricultural and other demand growth could result in severe deficits during the dry season, particularly around the time of planting.

Combining the influences of all of these factors on food-energy-water trade-offs in the basin, it was found that reservoir operating policies optimized to stationary conditions could only satisfy performance requirements across the three sectors within a corridor of hydrologic conditions. Furthermore, a majority of downscaled climate projections evolve to SOWs far from this region, in particular to regions threatened by flooding. Fortunately, knowing which hydrologic factors drive system performance enables the design of more robust and adaptive policies. For example, the most important factors can be used as signposts of change, with factor combinations near the boundaries of the success region serving as triggers for re-operations. Alternatively, the state-based dependence of the operating policies can be exploited to design more adaptive operations by explicitly incorporating information on the most important factors, automatically updating operations as planners observe climatic changes that refine their estimates of these factors through time. Given the close proximity of the failure boundary to current hydrologic conditions, decision makers may even want to re-optimize operations over a broader range of SOWs that appear more plausible based on the climate projections without fear of short-term regrets counteracting long-term robustness.

While the discovery of hydrologic factors most influencing system performance could guide water managers in designing more robust operations for the Red River basin, there are many additional uncertain factors beyond the hydrologic and socioeconomic factors sampled here that could further influence operational success. For example, sedimentation of dams could reduce hydropower production and degrade the downstream ecology, as could sand mining now prevalent in the delta. Sea level rise will almost certainly increase flood risk, while also degrading the quality of freshwater resources in the irrigation canals

through salinization. Furthermore, the log-space rescaling of stationary synthetic flows used to generate alternative hydrologic worlds in this study did not enable an analysis of event-scale changes in flows, such as increased daily variability. While the analysis performed here is an informative first cut examination of how operations should adapt to better meet food-energy-water needs in the Red River basin, future work should explore the impacts of these additional uncertainties in order to further improve system robustness.

CHAPTER 6

CONTRIBUTIONS & FUTURE WORK

6.1 Conclusions & Contributions

Managing coupled human-natural systems requires careful characterization of which system uncertainties drive their dynamics and how human actions interact with the natural system to create feedbacks. This dissertation has advanced diagnostic methods for discovering these drivers and feedbacks using two different socio-ecological systems serving multiple conflicting objectives: a managed lake (Chapter 3), and a multi-reservoir system (Chapters 4 and 5). In both of these systems, exploratory modeling and visual analytics were advanced to discover and understand how different model parameters, objectives or stochastic inputs influence system dynamics and performance. These studies also illustrated how these impacts can differ under alternative human control strategies due to complex interactions between human and natural systems, and their conclusions have important implications for managing several common challenges in socio-ecological systems: tipping points, problem formulation uncertainty and risk characterization.

Chapter 3 contributed an in-depth robustness analysis of the effects of alternative phosphorous (P) pollution control strategies and deep parametric model uncertainty on a town's ability to balance economic benefits and ecological costs to a nearby lake whose water quality exhibits threshold behavior. It was found that state-dependent P control policies designed with direct policy search (DPS) were more robust to deep uncertainties in both ecological and economic parameters than static, temporal control rules often used to design environmental

and climate management plans. In particular, the DPS control strategies were found to be insensitive to uncertainties in the discount parameter used to calculate economic benefits, which was not true using static temporal operations, illustrating how alternative human actions interact differently with the natural system. The reduced sensitivity of DPS control strategies to system uncertainties shows promise for designing adaptive control strategies for other systems with tipping points, such as the Earth system [Anderies et al., 2013; Rockstrom et al., 2009; Lenton et al., 2008].

In Chapter 4, different quantifications of system objectives were found to greatly influence optimal control strategies and how they interact with the natural system. In particular, minimizing expected flood damages, even in the worst case across an ensemble, was found to result in optimized control strategies that failed to reduce the magnitude of catastrophic flood events. Minimizing the 100-yr flood proved much more effective. Not only did the objectives influence the resulting natural system behavior, but they also interacted with each other in unpredictable ways. Worst first percentile objectives were found to result in stable control strategies whose performance in optimization generalized well in validation, but only if all objectives were calculated in this way. If worst first percentile objectives were optimized in concert with expectation objectives, control strategies over-fit to the stochastic inputs they were optimized over, providing less stable representations of system objectives. These insights can inform how multi-objective water systems optimization problems are formulated in the future. However, they also highlight that problems should always be formulated constructively [Tsoukias, 2008], and analysts should investigate how human actions optimized to alternative problem framings influence both the human and natural system.

Finally, Chapter 5 illustrated the importance of employing exploratory modeling strategies over a wide range of deeply uncertain characteristics to discover important drivers of system dynamics and performance. Specifically, Chapter 5 introduced a new stochastic streamflow generation scheme for analyzing the impacts of changing hydrologic characteristics on the performance of multi-reservoir operating policies, with a specific focus on altered monsoonal dynamics. Visual analytics were used to illustrate how extreme flood, hydropower and deficit events changed under alternative operating policies as a function of different combinations of hydrologic and socioeconomic factors, providing insight into how changes in the natural system interact with each other and human actions to influence system dynamics. Interactions between the hydrologic factors highlighted the importance of sampling more than just mean changes, as is commonly done [Moody & Brown, 2013; Brown et al., 2012], and of the benefits of visual analytics to understand how combined natural and human factors influence system dynamics.

The contributions of this dissertation have been or will be disseminated through peer-reviewed journal articles. Chapter 3 was drawn from Quinn et al. [2017], published in *Environmental Modelling & Software*. An abbreviated version of Chapter 4 was recently accepted for publication in *Water Resources Research* [Quinn et al., Accepted]. Chapter 5 is currently in preparation for submission to *Water Resources Research*. This work has also contributed to eight conference presentations: two at the American Geophysical Union Fall Meeting (one poster, one oral), two at the ASCE World Environmental & Water Resources Congress (both oral), one at the International Environmental Modelling and Software Society Conference (oral), two at the Society for Decision Making under Deep Uncertainty Meeting (one poster, one oral), and one at the NSF Sustainability

Research Network Awardees Conference (poster).

Other publications and articles in preparation have contributed to or built off of the work presented in this dissertation. Chapter 4 builds off of Giuliani et al. [In Press] (to which the author contributed), which utilizes the Red River test case to illustrate the computational power of coupling Evolutionary Many-Objective Direct Policy Search (EMODPS) with Multi-Master Borg to reliably solve complex, multi-objective, multi-state control problems that would be intractable with dynamic programming methods. It was through this work that some of the unintended consequences of the Worst Case formulation of the Red River problem explored in Chapter 4 were discovered, motivating the exploration of alternative formulations. Also building off of Giuliani et al. [In Press], Zatarain Salazar et al. [In Review] (to which the author contributed) compare different parallelization schemes of the Multi-Master Borg on a many-objective stochastic control problem in the Susquehanna River Basin. The goal is to explore the tradeoff between coordinated search across co-evolving islands and greater depth of search within the same wall clock using fewer islands. Making the tradeoff evaluation more interesting and complex, they also explore the effects of better representing uncertainty in the stochastic optimization by optimizing over a large ensemble of synthetic streamflows vs. approximating the streamflow distribution with a smaller ensemble, thereby reducing the computational time per function evaluation and allowing for greater depth of search.

6.2 Future Work

As briefly discussed in the conclusions to Chapters 3-5, there are many opportunities to build off of the work in this dissertation to improve our ability to characterize and manage deeply uncertain risks in coupled human and natural systems. In particular, all three studies highlighted the promise of EMODPS in formulating adaptive policies to build more robust systems. Central to this need are effective methods of information selection and policy formulation. Here we discuss how future work in information selection and policy formulation (Section 6.2.1) can be used to influence learning and adaptation (Section 6.2.2) to build not only more robust, but more resilient coupled human and natural systems (Section 6.2.3).

6.2.1 Information Selection and Policy Formulation

Chapter 4 highlighted the unique ability of the EMODPS simulation-optimization approach to design and solve multi-objective management problems with a mixture of mathematically complex objective functions [Giuliani et al., 2016b]. It also illustrated that this can be both a blessing and a curse. In that study, risk-averse objective preferences were better captured by minimizing non-traditional worst first percentile objectives than minimizing more traditional penalty functions in expectation. However, the ability of EMODPS to combine expectation and worst first percentile objectives in a single problem formulation resulted in overfitting of worst first percentile objectives, as the value of those objectives degraded when optimized policies were re-evaluated on a larger set of stochastic streamflows to which they were not optimized. This

was not true when only worst first percentile objectives were optimized, though.

These observations regarding objective functions are true of operating policies as well. While the universal approximators utilized by EMODPS to formulate operating policies are flexible and can easily be conditioned on a number of inputs, the black-box policy form also has its disadvantages. First, the many-to-one mapping of inputs to decisions is difficult to visualize in two dimensions for policies with more than two inputs. Even if the policies can be visualized, care must be taken in interpreting them since only some regions of the multi-dimensional input space are likely to be reached due to correlation among inputs. This difficulty relates to the second challenge of determining which and how many inputs to include in the policies in order to best balance information gain vs. over-specification and redundancy [Galelli et al., 2014], a challenge which also applies to the decision of how many basis functions to include. Furthermore, which inputs are most appropriate to include will also depend on the system objectives and how they're quantified [Tejada-Guibert et al., 1995]. In choosing how to represent time in the policies optimized in Chapters 4 and 5, different representations were found to favor different regions of the Pareto set, missing other regions entirely. Heuristic rules have been developed for different classes of objectives [Lund & Guzman, 1999], but determining how to combine these to balance multiple objectives is not straightforward, and simple rules for doing so should be dominated by policies optimized through search.

To date, efforts related to formulating policies and determining their inputs have been mostly ad-hoc, exploring alternative functional forms heuristically [Giuliani et al., 2016b], selecting inputs iteratively based on how well policies with those inputs predict the optimal releases under perfect information [Giu-

liani et al., 2015], and using rules of thumb to decide how many basis functions to use [Giuliani et al., 2015]. One potentially promising, although challenging, alternative would be to optimize the inputs, number of basis functions and parameters of the policies simultaneously, using metrics from information theory to attempt to balance information gain and overfitting. This would be similar in concept to the optimization approach employed by Kollat et al. [2008] for designing long-term groundwater monitoring networks. In their application, a Bayesian network is built over the decision space to model the dependencies between potential monitoring sites, where the decision whether or not to monitor a particular site depends on whether or not another site is chosen for monitoring. ε-hBOA, a probabilistic model building genetic algorithm (PMBGA; Pelikan [2005]) combining the concepts of ε-dominance and adaptive population sizing with the hierarchical Bayesian Optimization Algorithm (hBOA; Pelikan [2005]) is then used to generate new solutions by sampling from the Bayesian network and then Pareto sorting the new solutions using ε-dominance. Applied in the context of policy formulation for EMODPS, the decision whether or not to include certain inputs or additional basis functions could depend on what other inputs are included. A second optimization stage would have to be added to generate parameter values.

While employing this strategy would be more complicated in the two-stage, mixed discrete-continuous policy formulation decision problem, the idea of explicitly accounting for the covariance structure of inputs would be especially attractive for designing adaptive policies for the Red River basin. As illustrated in Figure 5.10, the hydrologic factors most critical to system performance in the Red River are highly anti-correlated, such that they might provide redundant information. An algorithm that models the dependency across inputs would

recognize this and adjust the probability of including one as an input conditional on whether or not the other is included. In this test case, a human could recognize that dependency, but non-linear dependencies would be harder to notice. Another avenue for future research could be optimizing the functional form of the policy itself, similar to symbolic regression [Koza, 1992], rather than independently optimizing separate forms and comparing them afterward. Of course, both of these potential advances could further degrade policy interpretability, an area in which improvements in visual analytics would be valuable [Thomas & Cook, 2006].

6.2.2 Learning and Adaptation

Closely tied to the problem of input selection are the concepts of learning and adaptation. As discussed in Chapters 3 and 5, EMODPS can facilitate these two concepts by incorporating inputs in its policies that are re-estimated over time through statistical learning, resulting in adaptive policies that evolve as one refines his/her estimates of the inputs. These dynamic policies should improve system robustness relative to static policies. Singh et al. [In Review] (to which the author has contributed) are beginning to apply these ideas to the lake problem described in Chapter 3, investigating the ability of different statistical learning techniques to estimate the location of a tipping point before it is crossed. Currently, observations of lake P concentration are used to inform estimates of the b and q model parameters and their corresponding critical P threshold. However, other indicators may be more appropriate for informing adaptive policies, relating again to the problem of input selection. For example, Scheffer et al. [2009] discuss several early-warning signals for critical transitions

that may be informative indicators, including "critical slowing down" [Wissel, 1984], "flickering" back and forth over an unstable point [Berglund & Gentz, 2002], and increased autocorrelation [Dakos et al., 2008]. The second of these would unfortunately not be a possible signal in irreversible systems, because there is no return after crossing the tipping point. In reversible or hysteretic systems, however, flickering could be an important signal for building resilient systems, discussed below.

6.2.3 Building Resilience in Socio-Ecological Systems

Chapters 3 and 5 of this dissertation focused on designing robust management plans for environmental systems that perform well across a broad range of possible system conditions and external stressors. However both of these studies ignored "resilience," a term which means different things to different people. Holling [1973] distinguishes between engineering and ecological resilience, where the former emphasizes stability (i.e. the ability to recover or return to a stable state [Hashimoto et al., 1982]), while the latter emphasizes persistence and the ability to absorb change. Persistence in this case is distinct from stability, as it does not mean persistence in the same state, but long-term survival. Persistence in this sense may actually *require* instability, as controlled systems exposed to variability better learn to recover and adapt [Carpenter et al., 2015; Holling, 1973].

Both of these concepts of resilience are applicable to work in this dissertation. In systems with irreversible tipping points like the version of the lake problem explored in Chapter 3, engineering resilience is critical. Staying in a stable state

is necessary for persistence, as it is impossible to return after crossing the tipping point. Fortunately, Chapter 3 suggests that EMODPS is a promising method of designing engineering-resilient policies for these systems. The Earth system may also possess irreversible thresholds, as Rockstrom et al. [2009] highlight nine "planetary boundaries" defining a "safe operating space" for humans on Earth, and crossing one or more of these boundaries could result in irreversible climate change. In systems with reversible or hysteretic tipping points, learning about the threshold becomes more important for building resilient systems capable of withstanding shocks. The adaptability of EMODPS could therefore help design ecologically resilient policies for these systems as well.

However, building resilient systems first requires careful characterization of the greatest risks to system stability and persistence. This dissertation has advanced exploratory modeling techniques to characterize such risks and inform the design of improved management strategies for coupled human-natural systems. In particular, this work has demonstrated the value of applying scenario discovery techniques to detect failure states for socio-ecological systems with tipping points and multi-reservoir systems threatened by flooding, as well as to discover unintended consequences of alternative many-objective problem formulations used to design management strategies for coupled human-natural systems. The techniques developed and illustrated in this dissertation can help improve understanding of which uncertainties and external stressors most influence performance in such complex systems. Future work building off this dissertation should explore how this information can be used to design adaptive management strategies and build system resilience.

APPENDIX A

FORMULATION OF OBJECTIVES IN RIVAL FRAMINGS OF RED RIVER CONTROL PROBLEM

This appendix provides a detailed, mathematical description of the objectives in Equation 4.10 under each of the four problem formulations. Recall from Equation 4.8 that the d-th objective, J_d, is calculated by aggregating a daily metric, $g_d(t,i)$, over a T-year simulation (indexed by t) using some operator, Φ, and then filtering the result over an ensemble of N of these simulations (indexed by i) using some statistic, Ψ. For the WC formulation, objectives are calculated across N=50 ensemble members in which simulations are of length T=20 years, while for the other formulations N=1000 ensemble members and T=1 year.

A.1 Hydropower Production

Across all formulations, total daily hydropower production $\eta_{t,i}$ from the four reservoirs in the i-th ensemble member, $g_{Hydro}(t,i)$, is averaged over the simulation length of each ensemble member:

$$\Phi_{Hydro}(i) = \mathbb{E}_{365T}\left[g_{Hydro}(t,i)\right] = \frac{1}{365T}\sum_{t=1}^{365T}\left[\sum_{j=4}^{4}\eta_{t,i}^{j}\right]. \tag{A.1}$$

J_{Hydro}^{WC} is then calculated as the minimum value of $\Phi_{Hydro}(i)$ across the N ensemble members, J_{Hydro}^{WP1} as the 1st percentile, and J_{Hydro}^{EV} as the average:

$$J^{WC}_{Hydro} = \underset{i\in(1,\ldots,N)}{\Psi}\Big[\Phi_{Hydro}(i)\Big] = \min_{i\in(1,\ldots,N)}\Big[\Phi_{Hydro}(i)\Big], \tag{A.2}$$

$$J^{WP1}_{Hydro} = \underset{i\in(1,\ldots,N)}{\Psi}\Big[\Phi_{Hydro}(i)\Big] = \underset{i\in(1,\ldots,N)}{\text{quantile}}\Big\{\Phi_{Hydro}(i), 0.01\Big\}, \text{ and} \tag{A.3}$$

$$J^{EV}_{Hydro} = \underset{i\in(1,\ldots,N)}{\Psi}\Big[\Phi_{Hydro}(i)\Big] = \mathbb{E}_N\Big[\Phi_{Hydro}(i)\Big] = \frac{1}{N}\sum_{i=1}^{N}\Phi_{Hydro}(i). \tag{A.4}$$

A.2 Squared Water Supply Deficit

Across all formulations, the daily squared water supply deficit in the i-th ensemble member, $g_{Deficit^2}(t,i)$, is first averaged over the simulation length of each ensemble member:

$$\Phi_{Deficit^2}(i) = \mathbb{E}_{365T}[g_{Deficit^2}(t,i)] = \frac{1}{365T}\sum_{t=1}^{365T} D^2_{t,i}. \tag{A.5}$$

$J^{WC}_{Deficit^2}$ is then calculated as the maximum value of $\Phi_{Deficit^2}(i)$ across the N ensemble members, $J^{WP1}_{Deficit^2}$ as the 99$^{\text{th}}$ percentile, and $J^{EV}_{Deficit^2}$ as the average:

$$J^{WC}_{Deficit^2} = \underset{i\in(1,\ldots,N)}{\Psi}\Big[\Phi_{Deficit^2}(i)\Big] = \max_{i\in(1,\ldots,N)}\Big[\Phi_{Deficit^2}(i)\Big] \tag{A.6}$$

$$J^{WP1}_{Deficit^2} = \underset{i\in(1,\ldots,N)}{\Psi}\Big[\Phi_{Deficit^2}(i)\Big] = \underset{i\in(1,\ldots,N)}{\text{quantile}}\Big\{\Phi_{Deficit^2}(i), 0.99\Big\}, \text{ and} \tag{A.7}$$

$$J^{EV}_{Deficit^2} = \underset{i\in(1,\ldots,N)}{\Psi}\Big[\Phi_{Deficit^2}(i)\Big] = \mathbb{E}_N\Big[\Phi_{Deficit^2}(i)\Big] = \frac{1}{N}\sum_{i=1}^{N}\Phi_{Deficit^2}(i). \tag{A.8}$$

A.3 Flood Damages and Vulnerability

In the WC formulation, the daily value of the flooding objective in the i-th ensemble member, $g^{WC}_{Flood}(t,i)$, is calculated using the penalty function displayed in Figure 4.3, which approximates damages as a function of the water level at Hanoi, $z^{HN}_{t,i}$. The damage function, $F(z^{HN}_{t,i})$, is a piecewise polynomial described by the following equation:

$$F(z^{HN}_{t,i}) = \begin{cases} 0, & z^{HN}_{t,i} \leq 6.0 \text{ m} \\ \frac{75{,}000}{5.25}(z^{HN}_{t,i} - 6), & 6.0 \text{ m} < z^{HN}_{t,i} \leq 11.25 \text{ m} \\ 1.5 \times 10^6 (z^{HN}_{t,i})^4 - 7.00 \times 10^7 (z^{HN}_{t,i})^3 & \\ +1.22 \times 10^9 (z^{HN}_{t,i})^2 - 9.45 \times 10^9 z^{HN}_{t,i}\,, & z^{HN}_{t,i} \geq 11.25 \text{ m} \\ +2.74 \times 10^{10} & \end{cases} . \quad \text{(A.9)}$$

Within each ensemble member, the daily value of the damage function is averaged over the simulation length:

$$\Phi^{WC}_{Flood}(i) = \mathbb{E}_{365T}\left[g^{WC}_{Flood}(t,i)\right] = \frac{1}{365T}\sum_{t=1}^{365T} F(z^{HN}_{t,i}). \quad \text{(A.10)}$$

J^{WC}_{Flood} is then calculated as the maximum value of $\Phi^{WC}_{Flood}(i)$ across all N ensemble members:

$$J^{WC}_{Flood} = \underset{i\in(1,\ldots,N)}{\Psi}\left[\Phi^{WC}_{Flood}(i)\right] = \max_{i\in(1,\ldots,N)}\left[\Phi^{WC}_{Flood}(i)\right]. \quad \text{(A.11)}$$

In the WP1, EV and EV&SD$_{\text{H}}$ formulations, the flooding objective is framed

as a flood vulnerability objective rather than a flood damage objective. Within each ensemble member i, $g^{WP1}_{Flood}(t,i)$ is defined as the daily water level at Hanoi in excess of 11.25 m, and $\Phi^{WP1}_{Flood}(i)$ is defined as the maximum value of $g^{WP1}_{Flood}(t,i)$ over the simulation length:

$$\Phi^{WP1}_{Flood}(i) = \max_{t\in(1,\ldots,365T)} [g^{WP1}_{Flood}(t,i)] = \max_{t\in(1,\ldots,365T)} [\max(z^{HN}_{t,i} - 11.25\text{ m}, 0)] \tag{A.12}$$

J^{WP1}_{Flood} is then calculated as the 99^{th} percentile of $\Phi^{WP1}_{Flood}(i)$ across the N ensemble members and is constrained to be ≤ 2.15 m:

$$J^{WP1}_{Flood} = \Psi_{i\in(1,\ldots,N)}[\Phi^{WP1}_{Flood}(i)] = \underset{i\in(1,\ldots,N)}{\text{quantile}}\left\{\Phi^{WP1}_{Flood}(i), 0.99\right\} \text{ and} \tag{A.13}$$

$$J^{WP1}_{Flood} \leq 2.15\text{ m}. \tag{A.14}$$

A.4 Flood Resilience/Recovery Time

The WP1, EV and EV&SD_H formulations all include an additional flooding objective to the flood vulnerability objective which represents the inverse of flood resilience. This objective, $J_{Recovery}$, indicates the time to "recover" once the water level at Hanoi exceeds 6 m. First, the within-ensemble average recovery time, $\Phi_{Recovery}(i)$, is calculated as:

$$\Phi_{Recovery}(i) = \frac{\sum_{t=1}^{365T} I_{t,i}}{\sum_{t=1}^{365T} \nu_{t,i}} \tag{A.15}$$

where

$$I_{t,i} = \begin{cases} 0, & z_{t,i}^{HN} \leq 6 \text{ m} \\ 1, & z_{t,i}^{HN} > 6 \text{ m} \end{cases} \text{ and} \tag{A.16}$$

$$\nu_{t,i} = \begin{cases} 1, & t = 1 \text{ and } z_{t,i}^{HN} > 6 \text{ m or} \\ & t > 1,\ z_{t,i}^{HN} > 6 \text{ m and } z_{t-1,i}^{HN} \leq 6 \text{ m} \cdot \\ 0, & \text{otherwise} \end{cases} \tag{A.17}$$

$I_{t,i}$ is an indicator variable signifying if the water level at Hanoi is above 6 m, while $\nu_{t,i}$ is an indicator variable signifying if a 6 m flood event has just begun. In the WP1 formulation, the 99^{th} percentile value of $\Phi_{Recovery}(i)$ across the N ensemble members is minimized, while in the EV and EV&SD_H formulations the average is minimized:

$$J_{Recovery}^{WP1} = \underset{i\in(1,\ldots,N)}{\Psi}\left[\Phi_{Recovery}(i)\right] = \underset{i\in(1,\ldots,N)}{\text{quantile}}\left\{\Phi_{Recovery}(i), 0.99\right\} \text{ and} \tag{A.18}$$

$$J_{Recovery}^{EV} = \underset{i\in(1,\ldots,N)}{\Psi}\left[\Phi_{Recovery}(i)\right] = \mathbb{E}_N\left[\Phi_{Recovery}(i)\right] = \frac{1}{N}\sum_{i=1}^{N}\Phi_{Recovery}(i). \tag{A.19}$$

A.5 Standard Deviation of Annual Hydropower Production

In the EV&SD_H formulation, $J_{Hydro\ Std}^{EV\&SD_H}$ is defined as the standard deviation in average annual hydropower production, $\Phi_{Hydro}(i)$, across the N ensemble members:

$$J^{EV\&SD_H}_{Hydro\ Std} = \underset{i\in(1,\ldots,N)}{\Psi}\left[\Phi_{Hydro}(i)\right] = \mathrm{std}_N\left[\Phi_{Hydro}(i)\right] = \left[\frac{1}{N-1}\sum_{i=1}^{N}\left(\Phi_{Hydro}(i) - J^{EV}_{Hydro}\right)^2\right]^{1/2}. \tag{A.20}$$

This is the only formulation which explicitly includes the inter-annual variability in hydropower production as an objective.

APPENDIX B

SYNTHETIC STREAMFLOW GENERATION AND VERIFICATION

B.1 Synthetic Streamflow Generation

Synthetic streamflows along the five tributaries of the Red River basin are first generated on a monthly time step using the method of Kirsch et al. [2013]. This method is described in Section B.1.1. Monthly streamflows at each site are then disaggregated to daily flows using the method of Nowak et al. [2010], described in Section B.1.2.

B.1.1 Monthly Streamflow Generation

For a given site, we denote the set of historical streamflows as $\mathbf{Q_H} \in \mathbb{R}^{N_H \times T}$ and the set of synthetic streamflows as $\mathbf{Q_S} \in \mathbb{R}^{N_S \times T}$, where N_H and N_S are the number of years in the historical and synthetic records, respectively, and T is the number of time steps per year. Here T=12 for 12 months. We first log-transform the historical monthly flows at each site to yield the matrix $Y_{H_{i,j}} = \ln(Q_{H_{i,j}})$, where i and j are the year and month of the historical record. These flows are then standardized to form the matrix $\mathbf{Z_H} \in \mathbb{R}^{N_H \times T}$ according to equation B.1:

$$Z_{H_{i,j}} = \frac{Y_{H_{i,j}} - \hat{\mu_j}}{\hat{\sigma}_j} \tag{B.1}$$

where $\hat{\mu_j}$ and $\hat{\sigma}_j$ are the sample mean and sample standard deviation of the j-th month's log-transformed flows, respectively. These flows follow a standard normal distribution: $Z_{H_{i,j}} \sim \mathcal{N}(0, 1)$.

For each site, we generate standard normal synthetic flows that reproduce the statistics of $\mathbf{Z_H}$ by first creating a matrix $\mathbf{C} \in \mathbb{R}^{N_S \times T}$ of randomly sampled standard normal flows from $\mathbf{Z_H}$. This is done by formulating a random matrix $\mathbf{M} \in \mathbb{R}^{N_S \times T}$ whose elements are independently sampled integers from $(1, 2, ..., N_H)$. Each element of $\mathbf{C}$ is then assigned the value $C_{i,j} = Z_{H_{i,j},j}$, i.e. the elements in each column of $\mathbf{C}$ are randomly sampled standard normal flows from the same column (month) of $\mathbf{Z_H}$. In order to preserve the historical cross-site correlation, the same matrix $\mathbf{M}$ is used to generate $\mathbf{C}$ for each site.

Because of the random sampling used to populate $\mathbf{C}$, an additional step is needed to generate auto-correlated standard normal synthetic flows, $\mathbf{Z_S}$. Denoting the historical auto-correlation $\mathbf{P_H}$=corr($\mathbf{Z_H}$), where corr($\mathbf{Z_H}$) is the historical correlation between standardized flows in months i and j (columns of $\mathbf{Z_H}$), an upper right triangular matrix, $\mathbf{U}$, can be found using Cholesky decomposition such that $\mathbf{P_H} = \mathbf{U}^\intercal\mathbf{U}$. $\mathbf{Z_S}$ is then generated as $\mathbf{Z_S} = \mathbf{CU}$. Finally, for each site, the auto-correlated synthetic standard normal flows $\mathbf{Z_S}$ are converted back to log-space flows $\mathbf{Y_S}$ according to $Y_{S_{i,j}} = \hat{\mu}_j + Z_{S_{i,j}}\hat{\sigma}_j$. These are then transformed back to real-space flows $\mathbf{Q_S}$ according to $Q_{S_{i,j}} = \exp(Y_{S_{i,j}})$.

While this method reproduces the within-year log-space auto-correlation, it does not preserve year to-year correlation, i.e. concatenating rows of $\mathbf{Q_S}$ to yield a vector of length $N_S \times T$ will yield discontinuities in the auto-correlation from month 12 of one year to month 1 of the next. To resolve this issue, Kirsch et al. [2013] repeat the method described above with a historical matrix $\mathbf{Q'_H} \in \mathbb{R}^{N_{H-1} \times T}$, where each row i of $\mathbf{Q'_H}$ contains historical data from month 7 of year i to month 6 of year i+1, removing the first and last 6 months of streamflows from the historical record. $\mathbf{U}'$ is then generated from $\mathbf{Q'_H}$ in the same way as $\mathbf{U}$ is generated

from $\mathbf{Q_H}$, while $\mathbf{C}'$ is generated from $\mathbf{C}$ in the same was as $\mathbf{Q'_H}$ is generated from $\mathbf{Q_H}$. As before, $\mathbf{Z'_S}$ is then calculated as $\mathbf{Z'_S} = \mathbf{C}'\mathbf{U}'$. Concatenating the last 6 columns of $\mathbf{Z'_S}$ (months 1-6) beginning from row 1 and the last 6 columns of $\mathbf{Z_S}$ (months 7-12) beginning from row 2 yields a set of synthetic standard normal flows that preserve correlation between the last month of the year and the first month of the following year. As before, these are then de-standardized and back-transformed to real space.

B.1.2 Daily Streamflow Generation

After generating monthly flows as described in Section B.1.1, a nearest-neighbor approach described by Nowak et al. [2010] is used to disaggregate these flows to daily values. The first step in this method is to calculate the k nearest neighbors from the historical monthly flows for each synthetically-generated monthly flow. Nearness is determined by the Euclidean distance, d, in real-space flows at the 5 sites (equation B.2):

$$d = \left[\sum_{m=1}^{5} \left((q_S)_m - (q_H)_m \right)^2 \right]^{1/2} \tag{B.2}$$

where $(q_S)_m$ is the real-space synthetic monthly flow generated at site m and $(q_H)_m$ is the real-space historical monthly flow at site m. For each synthetically-generated flow in month j, d is calculated for all historical flows in month j. The k-nearest are then sorted from i=1 for the closest to $i = k$ for the furthest, and probabilistically selected for proportionally scaling flows in disaggregation. We use the Kernel estimator given by Lall & Sharma [1996] to assign the probability p_n of selecting neighbor n (equation B.3):

$$p_n = \frac{\frac{1}{n}}{\sum_{i=1}^{k} \frac{1}{i}} \tag{B.3}$$

Following Lall & Sharma [1996] and Nowak et al. [2010], we use $k = \left\lfloor N_H^{1/2} \right\rceil$. After a neighbor is selected, the final step in disaggregation is to proportionally scale all of the historical daily flows at site m from the selected neighbor so that they sum to the synthetically generated monthly flow at site m. That is, if the first day of the month of the selected historical neighbor represented 5% of that month's historical flow, the first day of the month of the synthetic series would represent 5% of that month's synthetically-generated flow.

When simulating optimized policies over a larger ensemble of streamflows to generate Figures 7 and 8 in the main text, an improvement was made to the synthetic generator. Since the proportional rescaling by month described above results in scaled versions of historical flow patterns occurring on the same days each year, a wider network of neighbors was considered by including monthly totals within a moving window of +/- 7 days of the month when selecting neighbors. That is, rather than only considering historical January flows as neighbors to the synthetic January flows, for example, total flows over 31 consecutive days within the period from the last week of December to the first week of February were considered.

B.2 Verification of Synthetic Streamflow Statistics

As stated in Section B.1, the goal of the streamflow generator is to produce a time series of synthetic streamflows that expand upon those in the historical record

while reproducing their statistics. Log-space historical and synthetic flow duration curves (FDCs) of daily flows at the five sites (Figure B.1) indicate that the first is true, as the synthetic flows generate more extreme high and low flows. The flows also appear unbiased, as this expansion is relatively equal in both directions. Finally, the synthetic FDCs also follow the same shape as the historical FDCs, indicating that they reproduce the within-year distribution of daily flows. FDCs in Figure B.1 were generated from the 1000 years of synthetic streamflows to which policies from the WP1, EV and EV&SD_H formulations were optimized.

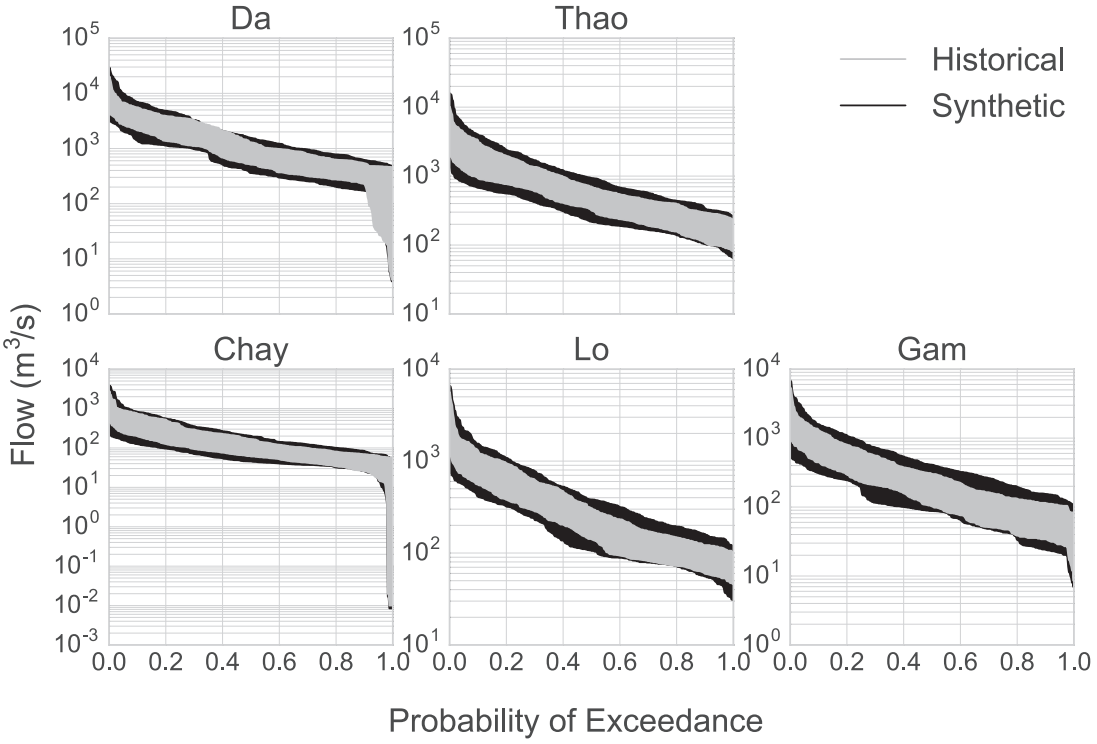

Figure B.1: Flow duration curves of the historical (gray) and synthetic (black) streamflows on each of the five major tributaries of the Red River basin. The synthetic flows substantially increase the range of flows over which the reservoir operating policies are optimized.

To more formally confirm that the synthetic flows are unbiased and follow the same distribution as the historical flows, we test whether or not the synthetic median and variance of real-space monthly flows are statistically different from the historical. The results of these tests are shown in Figure B.2 for the Da River, which provides roughly half of the system flow. This figure was generated from a 100-member ensemble of synthetic series of length 100 years, and a bootstrapped ensemble of historical years of the same size and length. Panel a shows boxplots of the real-space historical and synthetic monthly flows, while panels b and c show boxplots of their means and standard deviations, respectively. Because the real-space flows are not normally distributed, the non-parametric Wilcoxon rank-sum test and Levene's test were used to test whether or not the synthetic monthly medians and variances were statistically different from the historical. The p-values associated with these tests are shown in Figures B.2d and B.2e, respectively. While none of the synthetic variances are statistically different from the historical at a significance level of 0.05, the month 10 median is, and the medians of several other months nearly are. However, panel b indicates that while the medians may be statistically significantly different, the magnitudes of their differences are small and do not represent a consequential bias.

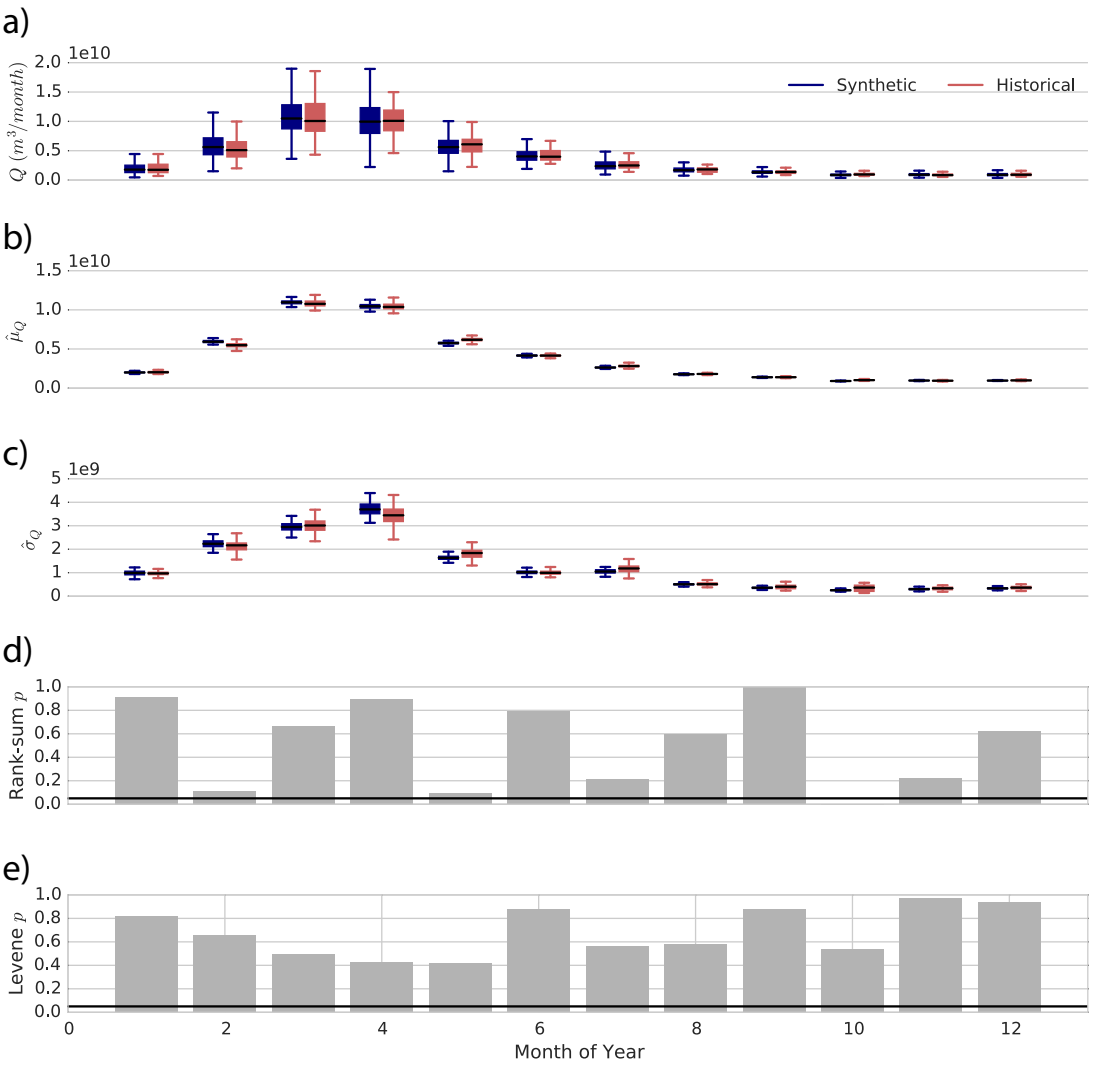

Figure B.2: Boxplots of the historical (pink) and synthetic (blue) total monthly flows (panel a), mean monthly flows (panel b) and standard deviation of monthly flows (panel c) as well as p-values for differences in median (panel d) and variance (panel e) of monthly flows. In the boxplots, a black line is drawn at the median, while the box edges extend to the quartiles and the whiskers to 1.5 times the interquartile range beyond the quartiles. p-values for differences in median were determined by a rank sum test, while those for differences in variance were determined by Levene's test.

In addition to verifying that the synthetic generator reproduces the first two moments of the historical monthly streamflows, we also verify that it reproduces both the historical auto-correlation and cross-site correlation in monthly and daily flows. The results of this analysis, preformed on the same synthetic and bootstrapped historical ensembles as for Figure B.2, are shown in Figure B.3. Panels a and b show the auto-correlation function of historical and synthetic real-space flows in the Da River for up to 12 lags of monthly flows (panel a) and 365 lags of daily flows (panel b). Also shown are 95% confidence intervals on the historical auto-correlations at each lag. The range of auto-correlations generated by the synthetic series expands upon that observed in the historical while remaining within the 95% confidence intervals, suggesting that the historical statistics are well-preserved. Panels c and d show boxplots of the cross-site correlation in monthly (panel c) and daily (panel d) real-space flows for all pairwise combinations of sites. The synthetic generator greatly expands upon the range of cross-site correlations observed in the historical record, both above and below. While Wilcoxon rank sum tests (not shown) for differences in median monthly correlations indicate all pairwise correlations are statistically different between the synthetic and historical series, biases in panel c appear small. Cross-site daily correlation is better preserved due to the scaled re-sampling scheme used to generate daily flows, with statistical differences in mean daily correlations only being statistically different for site pair 5. In summary, Figures B.1-B.3 indicate that the streamflow generator is reasonably reproducing historical statistics, while also expanding on the observed record to allow a more thorough stress test of the Red River's multi-reservoir operating policies.

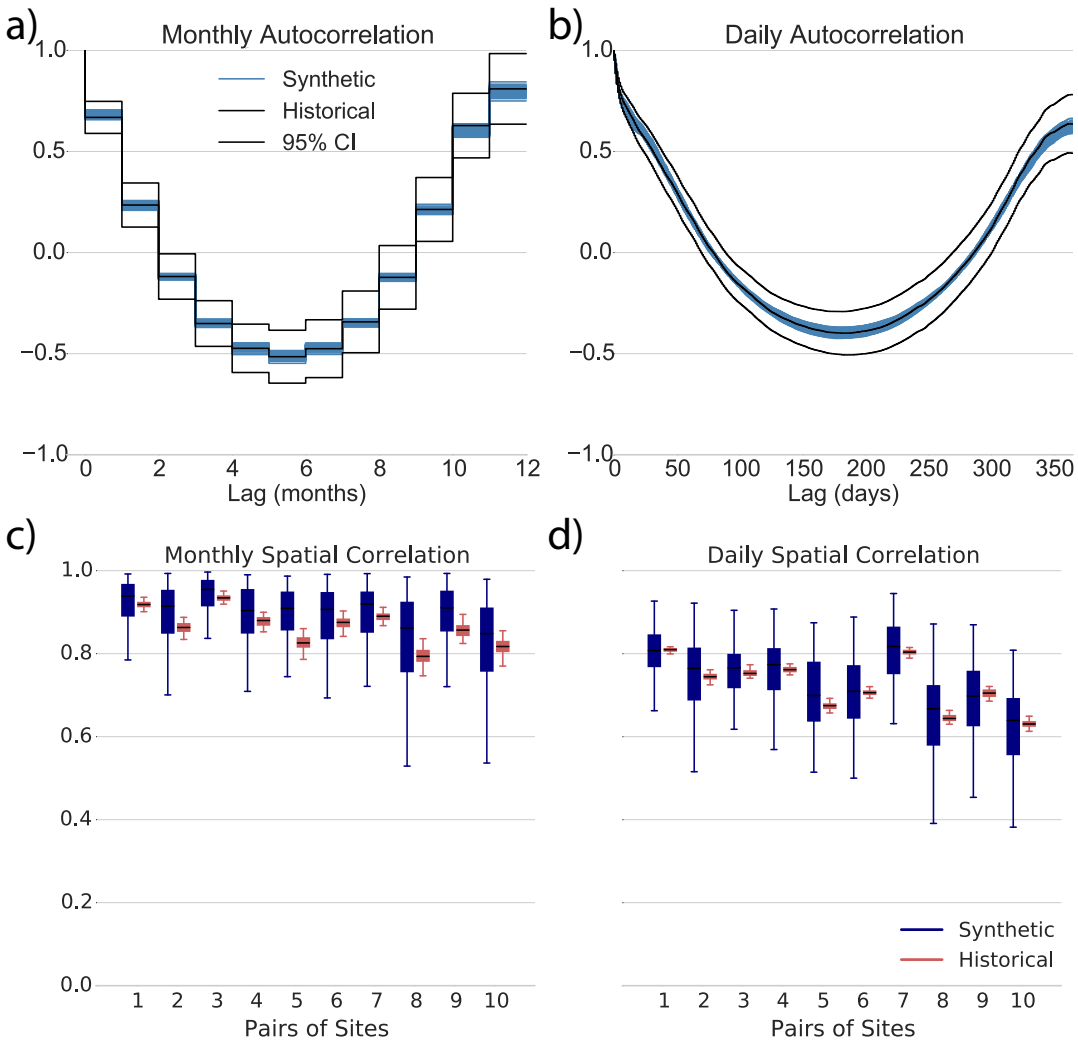

Figure B.3: Historical (black) and synthetic (blue) monthly (panel a) and daily (panel b) auto-correlation functions for streamflow time series, as well as boxplots of pairwise cross-correlations in monthly (panel c) and daily (panel d) historical (pink) and synthetic (blue) streamflows between sites. Black lines in panels a and b show both the mean and 95% confidence interval bounds on the historical auto-correlation. In the boxplots, a black line is drawn at the median, while the box edges extend to the quartiles and the whiskers to 1.5 times the interquartile range beyond the quartiles.

APPENDIX C

LOGISTIC REGRESSION FOR SCENARIO DISCOVERY ANALYSIS

This appendix provides a summary of the binary logistic regression models employed in the scenario discovery analysis described in Section 5.5.2. Binary logistic regression models estimate the probability that an event is classified as a success (1) as opposed to a failure (0) as a function of different covariates. In this study, we build logistic regression models to estimate the probability that a control policy is able to meet a minimum performance threshold on a particular objective in different states of the world (SOWs), where the covariates are deeply uncertain factors describing the SOW. The form of the logistic regression model is given by Equation C.1, where p_i represents the probability that performance in the i-th SOW is classified as a success and $\mathbf{X_i}$ represents a vector of covariates (deeply uncertain factors) describing the i-th SOW:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \mathbf{X_i^T}\beta. \tag{C.1}$$

The coefficients, β, on the covariates are estimated using Maximum Likelihood Estimation. After fitting the logistic model, the hyperplane of deeply uncertain factor combinations defining a probability contour can be determined by setting p to the desired probability level in the above equation. To plot 2-D projections of that hyperplane, the values of the other covariates can be held constant at their base values.

For each logistic regression model, we add covariates to the vector $\mathbf{X}$ sequentially based on which deeply uncertain factors most increase McFadden's pseudo-R^2. McFadden's pseudo-R^2, $R^2_{McFadden}$, is given by equation C.2:

$$R^2_{McFadden} = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})} \tag{C.2}$$

where $\ln \hat{L}(M_{Full})$ is the log-likelihood of the full model and $\ln \hat{L}(M_{Intercept})$ is the log-likelihood of the intercept model, i.e. a model with no covariates beyond the intercept. The intercept model therefore predicts the mean probability of success across all SOWs. $R^2_{McFadden}$ is a measure of improvement of the full model over the intercept model.

Table C.1 lists the values of $R^2_{McFadden}$ associated with different sets of covariates for modeling the probability of success on the flooding performance threshold when operating with the most robust solution for flooding in the *N=1000* generated SOWs. All covariates were first normalized on [0,1] before fitting the models. This was done so that the range of changes sampled across the deeply uncertain factors in the scenario discovery experiment would not impact the slope of the probability contours, and consequently the slope of the perpendicular trajectories shown in Figures 5.7-5.9. The normalization of the covariates does not impact the value of $R^2_{McFadden}$, but does change the value of the estimated coefficients, β. See Section 5.4.2 for a description of the covariates and explanation of the notation.

Based on the results in Table C.1, multipliers on the log-space amplitude of the first harmonic, m_{C_1}, log-space standard deviation, m_σ, and log-space mean, m_μ, were chosen as covariates for the logistic regression model. The value of $R^2_{McFadden}$ for this model is 0.795. The coefficients, β, on each of these predictors as well as their standard errors, $\sigma(\beta)$, z-scores, and p-values are reported in Table C.2. All covariates are significant to at least the 0.001 level. The negative coefficients on each of the predictors indicate that the probability of success on

Table C.1: $R^2_{McFadden}$ values associated with different logistic regression models for predicting the probability of success on the flooding criterion when operating with the most robust solution for flooding. The best singular predictor is shown in bold. Covariates are then added one by one based on which most increase $R^2_{McFadden}$. The best 2-covariate and 3-covariate models are also shown in bold.

Deeply Uncertain Factors Included in Model	$R^2_{McFadden}$
m_μ	0.114
m_σ	0.156
$\boldsymbol{m_{C_1}}$	**0.234**
d_{ϕ_1}	0.006
m_{C_2}	0.014
d_{ϕ_2}	0.005
d_e	0.000
m_{ag}	0.002
m_{aq}	0.000
m_o	0.005
d_D	0.002
$\boldsymbol{m_{C_1}, m_\sigma}$	**0.481**
m_{C_1}, m_μ	0.389
$\boldsymbol{m_{C_1}, m_\sigma, m_\mu}$	**0.796**

the flooding criterion decreases as m_{C_1}, m_σ and m_μ increase. Since all covariates were normalized on [0,1] and are independent, the relative magnitude of the coefficients signifies the relative importance of each predictor over the range of values sampled across the deeply uncertain factors. The decreasing absolute value of the coefficients from m_{C_1} to m_σ to m_μ therefore indicates that m_{C_1} has

Table C.2: Summary statistics of logistic regression model for predicting the probability of success on the flooding criterion when operating with the most robust solution for flooding.

X	β	$\sigma(\beta)$	z	p-value
Intercept	27.680	2.463		
m_{C_1}	-18.557	1.706	-10.874	0.000
m_σ	-16.313	1.512	-10.791	0.000
m_μ	-13.788	1.310	-10.526	0.000

the greatest impact on flooding within the range of sampled changes, then m_σ, then m_μ. The same conclusions are drawn from the individual values of $R^2_{McFadden}$ on these covariates given in Table C.1. These findings confirm the observations made from Figure 5.7.

Table C.3 lists the values of $R^2_{McFadden}$ associated with different sets of covariates for modeling the probability of success on the hydropower performance threshold when operating with the most robust solution for hydropower in the *N=1000* generated SOWs. Covariates were again normalized on [0,1] before fitting all models. Based on these results, multipliers on the log-space standard deviation, m_σ, log-space amplitude of the first harmonic, m_{C_1}, and log-space mean, m_μ were again chosen as covariates for the logistic regression model. The value of $R^2_{McFadden}$ for this model is 0.968, indicating very strong model performance. The coefficients, β, on each of the predictors as well as their standard errors, $\sigma(\beta)$, z-scores, and p-values are reported in Table C.4. All covariates are significant to at least the 0.01 level. The negative coefficient on m_σ indicates that the probability of success on the hydropower criterion decreases as m_σ increases, while the positive coefficients on m_μ and m_{C_1} indicate that the probability of success increases as these factors increase. The decreasing absolute value of the

coefficients from m_μ to m_{C_1} to m_σ indicate that m_μ has the greatest impact on hydropower production within the range of sampled changes on each factor, then m_{C_1}, then m_σ. The same conclusions are drawn from the individual values of $R^2_{McFadden}$ on these covariates given in Table C.3. These findings confirm the observations made from Figure 5.8.

Table C.3: $R^2_{McFadden}$ values associated with different logistic regression models for predicting the probability of success on the hydropower criterion when operating with the most robust solution for hydropower. The best singular predictor is shown in bold. Covariates are then added one by one based on which most increase $R^2_{McFadden}$. The best 2-covariate and 3-covariate models are also shown in bold.

Deeply Uncertain Factors Included in Model	$R^2_{McFadden}$
$\boldsymbol{m_\mu}$	**0.392**
m_σ	0.073
m_{C_1}	0.190
d_{ϕ_1}	0.000
m_{C_2}	0.000
d_{ϕ_2}	0.003
d_e	0.000
m_{ag}	0.001
m_{aq}	0.007
m_o	0.000
d_D	0.000
$\boldsymbol{m_\mu}$, $\boldsymbol{m_{C_1}}$	**0.702**
m_μ, m_σ	0.492
$\boldsymbol{m_\mu}$, $\boldsymbol{m_{C_1}}$, $\boldsymbol{m_\sigma}$	**0.968**

Table C.4: Summary statistics of logistic regression model for predicting the probability of success on the hydropower criterion when operating with the most robust solution for hydropower.

X	β	$\sigma(\beta)$	z	p-value
Intercept	-6.860	4.164		
m_μ	182.870	65.423	2.795	0.005
m_{C_1}	78.284	28.653	2.732	0.006
m_σ	-52.167	18.813	-2.773	0.006

Table C.5 lists the values of $R^2_{McFadden}$ associated with different sets of covariates for modeling the probability of success on the deficit performance threshold when operating with the most robust solution for the deficit in the *N=1000* generated SOWs. Covariates were again normalized on [0,1] before fitting all models. Based on these results, multipliers on the agricultural demand, m_{ag} other demands, m_o, and the log-space mean, m_μ, were chosen as covariates for the logistic regression model. The value of $R^2_{McFadden}$ for this model is 0.544, indicating that it is harder to predict the probability of success on the deficit satisficing criterion than the hydropower and flooding criteria. The coefficients, β, on each of the predictors as well as their standard errors, $\sigma(\beta)$, z-scores and p-values are reported in Table C.6. All the covariates are significant to at least the 0.001 level. The negative coefficients on m_{ag} and m_o indicate that the probability of success on the deficit criterion decreases as agricultural and other demand increase, while the positive coefficient on m_μ indicates that the probability of success increases as the log-space mean increases. The decreasing absolute value of the coefficients from m_{ag} to m_o to m_μ indicate that m_{ag} has the greatest impact on deficits within the range of sampled changes on each factor, then m_o, then m_μ. The same conclusions are drawn from the individual values of $R^2_{McFadden}$ on

these covariates given in Table C.5. These findings confirm the observations made from Figure 5.9.

Table C.5: $R^2_{McFadden}$ values associated with different logistic regression models for predicting the probability of success on the deficit criterion when operating with the most robust solution for the deficit. The best singular predictor is shown in bold. Covariates are then added one by one based on which most increase $R^2_{McFadden}$. The best 2-covariate and 3-covariate models are also shown in bold.

Deeply Uncertain Factors Included in Model	$R^2_{McFadden}$
m_μ	0.052
m_σ	0.001
m_{C_1}	0.000
d_{ϕ_1}	0.010
m_{C_2}	0.002
d_{ϕ_2}	0.000
d_e	0.000
$\boldsymbol{m_{ag}}$	**0.238**
m_{aq}	0.001
m_o	0.130
d_D	0.003
$\boldsymbol{m_{ag}, m_o}$	**0.455**
m_{ag}, m_μ	0.302
$\boldsymbol{m_{ag}, m_o, m_\mu}$	**0.544**

Finally, the same covariates used to model the probability of success on the flooding and hydropower objectives when operating with the most robust solutions for those respective objectives were used to model the probability of

Table C.6: Summary statistics of logistic regression model for predicting the probability of success on the deficit criterion when operating with the most robust solution for the deficit.

X	β	$\sigma(\beta)$	z-score	p-value
Intercept	5.626	0.452		
m_{ag}	-8.683	0.589	-14.744	0.000
m_o	-6.958	0.515	-13.501	0.000
m_μ	3.960	0.407	9.736	0.000

success on the same objectives when operating with the most robust solution across all requirements. These two models were used to define the success regions plotted in Figure 5.10. Summary statistics for the two models are shown in Tables C.7 and C.8. These models had $R^2_{McFadden}$ values of 0.834 and 0.915 for flooding successes and hydropower success, respectively. All covariates were significant to at least the 0.001 level across both models.

Table C.7: Summary statistics of logistic regression model for predicting the probability of success on the flooding criterion when operating with the most robust solution for all requirements.

X	β	$\sigma(\beta)$	z-score	p-value
Intercept	34.440	3.418		
m_{C_1}	-23.477	2.387	-9.837	0.000
m_σ	-21.087	2.136	-9.872	0.000
m_μ	-16.884	1.750	-9.648	0.000

Table C.8: Summary statistics of logistic regression model for predicting the probability of success on the hydropower criterion when operating with the most robust solution for all requirements.

X	β	$\sigma(\beta)$	z-score	p-value
Intercept	-23.607	3.936		
m_μ	69.180	11.450	6.042	0.000
m_{C_1}	42.534	7.025	6.054	0.000
m_σ	-18.439	3.180	-5.799	0.000

BIBLIOGRAPHY

Admiraal, J. F., Wossink, A., de Groot, W. T., & de Snoo, G. R. (2013). More than total economic value: How to combine economic valuation of biodiversity with ecological resilience. *Ecological Economics*, *89*, 115–122.

Anderies, J., Walker, B., & Kinzig, A. (2006). Fifteen weddings and a funeral: case studies and resilience-based management. *Ecology and Society*, *11*.

Anderies, J. M., Carpenter, S. R., Steffen, W., & Rockström, J. (2013). The topology of non-linear global carbon dynamics: From tipping points to planetary boundaries. *Environmental Research Letters*, *8*, 044048.

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *J Polit Econ*, *58*, 328–346.

Asefa, T., Clayton, J., Adams, A., & Anderson, D. (2014). Performance evaluation of a water resources system under varying climatic conditions: Reliability, resilience, vulnerability and beyond. *Journal of Hydrology*, *508*, 53–65.

Asian Development Bank (2016). *Viet Nam: Energy Sector Assessment, Stragey, and Road Map*. Technical Report Asian Development Bank.

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations research*, *41*, 435–449.

Basdekas, L. (2014). Is multiobjective optimization ready for water resources practitioners? utilitys drought policy investigation. *Journal of Water Resources Planning and Management*, *140*, 275–276. URL: http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000415. doi:10.1061/(ASCE)WR.1943-5452.0000415.

Bayazit, M., & Ünal, N. (1990). Effects of hedging on reservoir performance. *Water resources research*, *26*, 713–719.

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

Bellman, R., Glicksberg, I., & Gross, O. (1955). *On the'bang-bang'control problem*. Technical Report DTIC Document.

Ben-Haim, Y. (2004). Uncertainty, probability and information-gaps. *Reliability Engineering & System Safety*, *85*, 249–266.

Berglund, N., & Gentz, B. (2002). Metastability in simple climate models: pathwise analysis of slowly driven langevin equations. *Stochastics and Dynamics*, *2*, 327–356.

Bernardi, D., Than Bui, M., Micotti, M., Dinh, Q., Nguyen, X., Schippa, L., Schmitt, R., Truong, V., Nam Vu, P., & Weber, E. (2014). *Report D5.1: Identifying the Model*. Technical Report Integrated and sustainabale water Management of Red-Thai Binh River System in a changing climate.

Bertsekas, D. P. (1995). *Dynamic programming and optimal control* volume 1. Athena Scientific Belmont, MA.

Beyer, H.-G., & Sendhoff, B. (2007). Robust optimization a comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, *196*, 3190–3218. URL: http://www.sciencedirect.com/science/article/pii/S0045782507001259. doi:10.1016/j.cma.2007.03.003.

Borgomeo, E., Farmer, C. L., & Hall, J. W. (2015). Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments. *Water Resources Research*, *51*, 5382–5405.

Bosomworth, K., Leith, P., Harwood, A., & Wallis, P. J. (2017). Whats the problem in adaptation pathways planning? the potential of a diagnostic problem-structuring approach. *Environmental Science & Policy*, *76*, 23–28.

Bouwer, H. (2000). Integrated water management: emerging issues and challenges. *Agricultural water management*, *45*, 217–228.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Brown, C., Ghile, Y., Laverty, M., & Li, K. (2012). Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. *Water Resources Research*, *48*, W09537.

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, *77*, 34–49.

Burke, E., Kendall, G., Newall, J., Hart, E., Ross, P., & Schulenburg, S. (2003). Hyper-heuristics: An emerging direction in modern search technology. In *Handbook of metaheuristics* (pp. 457–474). Springer.

Buşoniu, L., Ernst, D., De Schutter, B., & Babuška, R. (2011). Cross-entropy optimization of control policies with adaptive basis functions. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *41*, 196–209.

Cancelliere, A., Giuliano, G., Ancarani, A., & Rossi, G. (2002). A neural networks approach for deriving irrigation reservoir operating rules. *Water resources management*, *16*, 71–88.

Carpenter, S. R., Brock, W. A., Folke, C., van Nes, E. H., & Scheffer, M. (2015). Allowing variance may enlarge the safe operating space for exploited ecosystems. *Proceedings of the National Academy of Sciences*, *112*, 14384–14389.

Carpenter, S. R., Ludwig, D., & Brock, W. A. (1999). Management of eutrophication for lakes subject to potentially irreversible change. *Ecological applications*, *9*, 751–771.

Castelletti, A., Pianosi, F., Quach, X., & Soncini-Sessa, R. (2012a). Assessing water reservoirs management and development in northern vietnam. *Hydrol. Earth Syst. Sci*, *16*, 189–199.

Castelletti, A., Pianosi, F., & Soncini-Sessa, R. (2012b). Stochastic and robust control of water resource systems: Concepts, methods and applications. In *System identification, environmental modelling, and control system design* (pp. 383–401). Springer.

Chankong, V., & Haimes, Y. (1983). *Multiobjective Decision Making: Theory and Methodology*. New York: North-Holland.

Cohon, J., & Marks, D. (1975). A review and evaluation of multiobjective programming techniques. *Water Resources Research*, *11*, 208–220.

Coumou, D., & Rahmstorf, S. (2012). A decade of weather extremes. *Nature Climate Change*, *2*, 491–496.

Culley, S., Noble, S., Yates, A., Timbs, M., Westra, S., Maier, H., Giuliani, M., & Castelletti, A. (2016). A bottom-up approach to identifying the maximum op-

erational adaptive capacity of water resource systems to a changing climate. *Water Resources Research*, *52*, 6751–6768.

Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., & Held, H. (2008). Slowing down as an early warning signal for abrupt climate change. *Proceedings of the National Academy of Sciences*, *105*, 14308–14312.

Dalal, S., Han, B., Lempert, R., Jaycocks, A., & Hackbarth, A. (2013). Improving scenario discovery using orthogonal rotations. *Environmental Modelling & Software*, *48*, 49–64. URL: http://www.sciencedirect.com/science/article/pii/S1364815213001345.

De Kort, I. A., & Booij, M. J. (2007). Decision making under uncertainty in a decision support system for the red river. *Environmental Modelling & Software*, *22*, 128–136.

Deb, K., & Agrawal, R. B. (1994). *Simulated binary crossover for continuous search space*. Technical Report IITK/ME/SMD-94027 Indian Institute of Technology Kanpur.

Deb, K., Joshi, D., & Anand, A. (2002). Real-coded evolutionary algorithms with parent-centric re-combination. In *Proceedings of the World Congress on Computational Intelligence* (pp. 61–66).

Deb, K., Mohan, M., & Mishra, S. (2005). Evaluating the epsilon-domination based multiobjective evolutionary algorithm for a quick computation of pareto-optimal solutions. *Evolutionary Computation Journal*, *13*, 501–525.

Dessai, S., Hulme, M., Lempert, R., & Pielke Jr, R. (2009). Climate prediction: a limit to adaptation. *Adapting to climate change: thresholds, values, governance*, (pp. 64–78).

Dinh, N. Q. (2015). *Multi-objective evolutionary algorithm, dynamic and non-dynamic emulators in the design of optimal policies for water resources management*. Ph.D. thesis Politecnico di Milano.

Dittrich, R., Wreford, A., & Moran, D. (2016). A survey of decision-making approaches for climate change adaptation: Are robust methods the way forward? *Ecological Economics*, *122*, 79–89.

Field, C. B., Barros, V. R., Mach, K., & Mastrandrea, M. (2014). Climate change 2014: impacts, adaptation, and vulnerability. *Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change*, .

Filatova, T., Polhill, J. G., & van Ewijk, S. (2016). Regime shifts in coupled socio-environmental systems: review of modelling challenges and approaches. *Environmental modelling & software*, *75*, 333–347.

Fowler, H. J., Kilsby, C., & O'Connell, P. (2003). Modeling the impacts of climatic change and variability on the reliability, resilience, and vulnerability of a water resource system. *Water Resources Research*, *39*.

Friedman, J. H., & Fisher, N. (1999). Bump hunting in high-dimensional data. *Stat. Comput.*, *9*, 123–143.

Galelli, S., & Castelletti, A. (2013). Tree-based iterative input variable selection for hydrological modeling. *Water Resources Research*, *49*, 4295–4310. URL: http://onlinelibrary.wiley.com/doi/10.1002/wrcr.20339/full.

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

http://www.sciencedirect.com/science/article/pii/S1364815214002394. doi:10.1016/j.envsoft.2014.08.015.

Gebretsadik, Y., Fant, C., & Strzepek, K. (2012). *Impact of climate change on irrigation, crops and hydropower in Vietnam*. 2012/79. WIDER Working Paper.

Giuliani, M., Anghileri, D., Castelletti, A., Vu, P. N., & Soncini-Sessa, R. (2016a). Large storage operations under climate change: expanding uncertainties and evolving tradeoffs. *Environmental Research Letters*, *11*, 035009.

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

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., & Reed, P. M. (2016b). Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations. *Journal of Water Resources Planning and Management*, (p. 04015050).

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

Giuliani, M., Quinn, J. D., Herman, J. D., Castelletti, A., & Reed, P. M. (In Press). Scalable multi-objective control for large scale water resources systems under uncertainty. *IEEE Transactions on Control Systems Technology*, . doi:10.1109/TCST.2017.2705162.

Groves, D., Bloom, E., Lempert, R., Fischbach, J., Nevills, J., & Goshi, B. (2014). Developing key indicators for adaptive water planning. *Journal of Water Resources Planning and Management*, *0*, 05014008.

URL: http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000471. doi:10.1061/(ASCE)WR.1943-5452.0000471.

Groves, D. G., Kuhn, K., Fischbach, J., Johnson, D. R., & Syme, J. (2016). *Analysis to Support Louisiana's Flood Risk and Resilience Program and Application to the National Disaster Resilience Competition*. RAND.

Guariso, G., Rinaldi, S., & Soncini-Sessa, R. (1986). The management of lake como: A multiobjective analysis. *Water Resources Research*, *22*, 109–120.

Guo, H., Xu, M., & Hu, Q. (2011). Changes in near-surface wind speed in china: 1969–2005. *International Journal of Climatology*, *31*, 349–358.

Haasnoot, M., Kwakkel, J. H., Walker, W. E., & ter Maat, J. (2013). Dynamic adaptive policy pathways: A method for crafting robust decisions for a deeply uncertain world. *Global Environmental Change*, *23*, 485–498. URL: http://www.sciencedirect.com/science/article/pii/S095937801200146X. doi:10.1016/j.gloenvcha.2012.12.006.

Haasnoot, M., Middelkoop, H., Offermans, A., Van Beek, E., & Van Deursen, W. P. (2012). Exploring pathways for sustainable water management in river deltas in a changing environment. *Climatic Change*, *115*, 795–819.

Hadka, D., Herman, J., Reed, P., & Keller, K. (2015). An open source framework for many-objective robust decision making. *Environmental Modelling & Software*, *74*, 114–129.

Hadka, D., & Reed, P. (2013). Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computation*, *21*, 231–259.

Hadka, D., & Reed, P. (2015). Large-scale parallelization of the borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems. *Environmental Modelling & Software*, *69*, 353–369.

Haimes, Y. Y., & Hall, W. A. (1977). Sensitivity, responsivity, stability, and irreversability as multiple objectives in civil systems. *Advances in Water Resources*, *1*, 71–81.

Hall, J., Grey, D., Garrick, D., Fung, F., Brown, C., Dadson, S., & Sadoff, C. (2014). Coping with the curse of freshwater variability. *Science*, *346*, 429–430.

Hall, J., Lempert, R. J., Keller, K., Hackbarth, A., Mijere, C., & McInerney, D. J. (2012). Robust climate policies under uncertainty: A comparison of robust decision making and info-gap methods. *Risk Analysis*, . doi:10.111▮.1539-6924.2012.01802.x.

Hallegatte, S., Bangalore, M., Bonzanigo, L., Fay, M., Kane, T., Ulf, N., Rozenberg, J., Treguer, D., & Vogt-Schilb, A. (2015). *Shock Waves: Managing the Impacts of Climate Change on Poverty*. World Bank.

Hansson, K., & Ekenberg, L. (2002). Flood mitigation strategies for the red river delta. In *Proceeding of the 2002 Joint CSCE/EWRI of ASCE International Conference on Environmental Engineering, An International Perspective on Environmental Engineering, Niagara Falls, Ont., Canada, July* (pp. 21–24).

Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A. et al. (2013). Observations: atmosphere and surface, .

Hashimoto, T., Stedinger, J. R., & Loucks, D. P. (1982). Reliability, resiliency

and vulnerability criteria for water resource system performance evaluation. *Water Resources Research*, *18*, 14–20.

Herman, J., & Giuliani, M. (In Review). Policy tree optimization for adaptive management of water resources systems. *Environmental Modelling & Software*, .

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

Herman, J. D., Zeff, H. B., Lamontagne, J. R., Reed, P. M., & Characklis, G. W. (2016). Synthetic drought scenario generation to support bottom-up water supply vulnerability assessments. *Journal of Water Resources Planning and Management*, *142*, 04016050.

Herman, J. D., Zeff, H. B., Reed, P. M., & Characklis, G. W. (2014). Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. *Water Resources Research*, *50*, 7692–7713.

Hermans, L. M., Haasnoot, M., ter Maat, J., & Kwakkel, J. H. (2017). Designing monitoring arrangements for collaborative learning about adaptation pathways. *Environmental Science & Policy*, *69*, 29–38.

Hidalgo, H. G., Dettinger, M. D., & Cayan, D. R. (2008). Downscaling with constructed analogues: Daily precipitation and temperature fields over the united states. *California Energy Commission PIER Final Project Report CEC-500-2007-123*, .

Hijioka, Y., Lin, E., Pereira, J. J., Corlett, R., Cui, X., Insarov, G., Surjan, A., Field,

C., Barros, V., Mach, K. et al. (2014). Chapter 24: Asia. *Working Group II contribution to the IPCC Fifth Assessment Report Climate Change*, (pp. 1858–1925).

Hine, D., & Hall, J. W. (2010). Information gap analysis of flood model uncertainties and regional frequency analysis. *Water Resources Research*, *46*. doi:10.1029/2008WR007620.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual review of ecology and systematics*, *4*, 1–23.

Hoppe, R. (2011). *The governance of problems: Puzzling, powering and participation*. Policy Press.

Horan, R. D., Fenichel, E. P., Drury, K. L., & Lodge, D. M. (2011). Managing ecological thresholds in coupled environmental–human systems. *Proceedings of the National Academy of Sciences*, *108*, 7333–7338.

Hsu, P.-c., Li, T., & Wang, B. (2011). Trends in global monsoon area and precipitation over the past 30 years. *Geophysical Research Letters*, *38*.

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: review and synthesis. *Journal of Hydrology*, *319*, 83–95.

Huskova, I., Matrosov, E. S., Harou, J. J., Kasprzyk, J. R., & Lambert, C. (2016). Screening robust water infrastructure investments and their trade-offs under global change: A london example. *Global Environmental Change*, *41*, 216–227.

Jiang, Y., Luo, Y., Zhao, Z., & Tao, S. (2010). Changes in wind speed over china during 1956–2004. *Theoretical and Applied Climatology*, *99*, 421.

Kasprzyk, J. R. (2013). *Many Objective Water Resources Planning and Management*

Given Deep Uncertainties, Population Pressures, and Environmental Change. Ph.D. thesis The Pennsylvania State University.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., & Lempert, R. J. (2013). Many objective robust decision making for complex environmental systems undergoing change. *Environmental Modelling and Software*, *42*, 55–71. doi:10.1016/j.envsoft.2012.12.007.

Kasprzyk, J. R., Reed, P. M., Characklis, G. W., & Kirsch, B. R. (2012). Many-objective de novo water supply portfolio planning under deep uncertainty. *Environmental Modelling and Software*, *34*, 87–104. doi:10.1016/j.envsoft.2011.04.003.

Kasprzyk, J. R., Reed, P. M., & Hadka, D. M. (2015). Battling arrows paradox to discover robust water management alternatives. *Journal of Water Resources Planning and Management*, *142*, 04015053.

Kasprzyk, J. R., Reed, P. M., Kirsch, B. R., & Characklis, G. W. (2009). Managing population and drought risks using many-objective water portfolio planning under uncertainty. *Water Resour. Res.*, *45*, n/a–n/a. doi:10.1029/2009wr008121.

Katz, R. W., & Brown, B. G. (1992). Extreme events in a changing climate: variability is more important than averages. *Climatic change*, *21*, 289–302.

Kawachi, T., & Maeda, S. (2004). Optimal management of waste loading into a river system with nonpoint source pollutants. *Proceedings of the Japan Academy, Series B*, *80*, 392–398.

Keller, K., & McInerney, D. (2008). The dynamics of learning

about a climate threshold. *Climate Dynamics*, *30*, 321–332. URL: http://link.springer.com/article/10.1007/s00382-007-0290-5.

Keller, K., Yohe, G., & Schlesinger, M. (2008). Managing the risks of climate thresholds: uncertainties and information needs. *Climatic Change*, *91*, 5–10. URL: http://www.springerlink.com/index/E550754338771K25.pdf.

Keynes, J. M. (1937). The general theory of employment. *The quarterly journal of economics*, *51*, 209–223.

Kirsch, B. R., Characklis, G. W., & Zeff, H. B. (2013). Evaluating the impact of alternative hydro-climate scenarios on transfer agreements: A practical improvement for generating synthetic streamflows. *Journal of Water Resources Planning and Management*, *139*, 396–406. doi:10.1061/(ASCE)WR.1943-5452.0000287.

Kita, H., Ono, I., & Kobayashi, S. (1999). Multi-parental extension of the unimodal normal distribution crossover for real-coded genetic algorithms. In *Congress on Evolutionary Computation* (pp. 1581–1588).

Kjeldsen, T. R., & Rosbjerg, D. (2004). Choice of reliability, resilience, and vulnerability estimators for risk assessments of water resources systems. *Hydrological Sciences Journal*, *49*, 755–767.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Boston, MA: Houghton Mifflin.

Kollat, J. B., & Reed, P. M. (2007). A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. *Advances in Water Resources*, *30*, 335–353.

Kollat, J. B., Reed, P. M., & Kasprzyk, J. R. (2008). A new epsilon-dominance hierarchical bayesian optimization algorithm for large multiobjective monitoring network design problems. *Advances in Water Resources*, *31*, 828–845. doi:10.1016/j.advwatres.2008.01.017.

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

Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection* volume 1. MIT press.

Kwadijk, J. C. J., Haasnoot, M., Mulder, J. P. M., Hoogvliet, M. M. C., Jeuken, A. B. M., van der Krogt, R. A. A., van Oostrom, N. G. C., Schelfhout, H. A., van Velzen, E. H., van Waveren, H., & de Wit, M. J. M. (2010). Using adaptation tipping points to prepare for climate change and sea level rise: a case study in the netherlands. *Wiley Interdisciplinary Reviews: Climate Change*, *1*, 729–740. URL: `http://dx.doi.org/10.1002/wcc.64`. doi:10.1002/wcc.64.

Kwakkel, J., Walker, W., & Marchau, V. (2010). Adaptive airport strategic planning. *EJTIR*, *10*, 249–273.

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

Kwakkel, J. H., Walker, W. E., & Haasnoot, M. (2016b). Coping with the wickedness of public policy problems: Approaches for decision making under deep uncertainty. *Journal of Water Resources Planning and Management*, (p. 01816001).

Labadie, J. W. (2004). Optimal operation of multireservoir systems: State-of-the-art review. *Journal of Water Resources Planning and Management*, *130*, 93–111.

Lall, U., & Sharma, A. (1996). A nearest neighbor bootstrap for resampling hydrologic time series. *Water Resources Research*, *32*, 679–693.

Laumanns, M., Thiele, L., Deb, K., & Zitzler, E. (2002). Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary computation*, *10*, 263–282.

Le Ngo, L., Madsen, H., & Rosbjerg, D. (2007). Simulation and optimisation modelling approach for operation of the hoa binh reservoir, vietnam. *Journal of Hydrology*, *336*, 269–281.

Lempert, R., Popper, S., & Bankes, S. (2002). Confronting surprise. *Social Science Computer Review*, *20*, 420–440.

Lempert, R. J., Bryant, B. P., & Bankes, S. C. (2008). *Comparing Algorithms for Scenario Discovery*. Technical Report WR-557-NSF RAND.

Lempert, R. J., & Collins, M. (2007). Managing the risk of an uncertain threshold response: Comparison of robust, optimimum, and precautionary approaches. *Risk Analysis*, *27*, 1009–1026.

Lempert, R. J., & Groves, D. G. (2010). Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the American west. *Technological Forecasting and Social Change*, *77*, 960–974.

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A general, analytic method for generating robust strategies and narrative scenarios. *Management Science*, *52*, 514–528.

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2003). *Shaping the next one hundred years: new methods for quantitative, long-term policy analysis*. Santa Monica, CA: RAND.

Lempert, R. J., Popper, S. W., & Bankes, S. C. (2010). Robust decision making: coping with uncertainty. *The Futurist*, *44*, 47.

Lenton, T. M. (2013). Environmental tipping points. *Annual Review of Environment and Resources*, *38*, 1–29.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the earth's climate system. *Proceedings of the national Academy of Sciences*, *105*, 1786–1793.

Li, H., Sheffield, J., & Wood, E. F. (2010). Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change ar4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres*, *115*.

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J. et al. (2007). Complexity of coupled human and natural systems. *science*, *317*, 1513–1516.

Lund, J., & Guzman, J. (1999). Derived operating rules for reservoirs in series or in parallel. *Journal of Water Resources Planning and Management*, *125*, 143–153.

Lund, J. R. (2002). Floodplain planning with risk-based optimization. *Journal of water resources planning and management*, *128*, 202–207.

Maass, A. (1967). Benefit-cost analysis: Its relevance to public investment decisions: Reply. *The Quarterly Journal of Economics*, *81*, 700–702.

Maass, A., Hufschmidt, M. M., Dorfman, R., Thomas, Jr., H. A., Marglin, S. A., & Fair, G. M. (1962). *Design of Water-Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning*. Cambridge: Harvard University Press.

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

Maier, H. R., Lence, B. J., Tolson, B. A., & Foschi, R. O. (2001). First-order reliability method for estimating reliability, vulnerability, and resilience. *Water Resources Research*, *37*, 779–790.

Majone, G., & Quade, E. S. (1980). *Pitfalls of analysis*. Wiley.

Malekmohammadi, B., Kerachian, R., & Zahraie, B. (2009). Developing monthly operating rules for a cascade system of reservoirs: application of bayesian networks. *Environmental Modelling & Software*, *24*, 1420–1432.

Matalas, N. C., & Fiering, M. B. (1977). Water-resource systems planning. *Climate, Climatic Change, and Water Supply. Studies in Geophysics, National Academy of Sciences, Washington, D. C.*, (pp. 99–110).

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Climate change: Stationarity is dead: Whither water management? *Science*, *319*, 573–574. URL: `http://dx.doi.org/10.1126/science.1151915`. doi:10.1126/science.1151915.

Moody, P., & Brown, C. (2012). Modeling stakeholder-defined climate risk on the upper great lakes. *Water Resources Research*, *48*.

Moody, P., & Brown, C. (2013). Robustness indicators for evaluation under climate change: Application to the upper great lakes. *Water Resources Research*, *49*, W20228.

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

Moy, W., Cohon, J. L., & Revelle, C. S. (1986). A programming model for analysis of the reliability, resilience, and vulnerability of a water supply reservoir. *Water Resources Research*, *22*, 489–498.

Nayfeh, A. H., & Balachandran, B. (2008). *Applied nonlinear dynamics: analytical, computational and experimental methods*. John Wiley & Sons.

Nazemi, A., Wheater, H. S., Chun, K. P., & Elshorbagy, A. (2013). A stochastic reconstruction framework for analysis of water resource system vulnerability to climate-induced changes in river flow regime. *Water Resources Research*, *49*, 291–305. URL: `http://onlinelibrary.wiley.com/doi/10.1029/2012WR012755/abstract`. doi:10.1029/2012WR012755.

Needham, J. T., Watkins Jr, D. W., Lund, J. R., & Nanda, S. (2000). Linear programming for flood control in the iowa and des moines rivers. *Journal of Water Resources Planning and Management*, *126*, 118–127.

Neelakantan, T., & Pundarikanthan, N. (2000). Neural network-based simulation-optimization model for reservoir operation. *Journal of water resources planning and management*, *126*, 57–64.

Neumann, J. E., Emanuel, K. A., Ravela, S., Ludwig, L. C., & Verly, C. (2015). Risks of coastal storm surge and the effect of sea level rise in the red river delta, vietnam. *Sustainability*, *7*, 6553–6572.

Nguyen, T. C., Do, N. H., Nguyen, T., & Egashira, K. (2002). Agricultural development in the red river delta, vietnam-water management, land use, and rice production, .

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., & Zechman, E. (2010). State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management*, *136*, 412–432.

Nordhaus, W. D. (2013). *The climate casino: Risk, uncertainty, and economics for a warming world*. Yale University Press.

Nowak, K., Prairie, J., Rajagopalan, B., & Lall, U. (2010). A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow. *Water Resources Research*, *46*. doi:10.1029/2009WR008530.

Oliveira, R., & Loucks, D. P. (1997). Operating rules for multireservoir systems. *Water Resources Research*, *33*, 839–852.

Olsson, L., Opondo, M., Tschakert, P., Agrawal, A., Eriksen, S. E. et al. (2014). Livelihoods and poverty, .

Orlovski, S., Rinaldi, S., & Soncini-Sessa, R. (1984). A min-max approach to reservoir management. *Water Resources Research*, *20*, 1506–1514.

Palmer, R. N., Werick, W. J., MacEwan, A., & Woods, A. W. (1990). Modeling water resources opportunities, challenges, and tradeoffs: The use of shared vision modeling for negotiation and conflict resolution. In *Proceedings of ASCE's 26th Annual Conference on Water Resources Planning and Management*. ASCE.

Pareto, V. (1896). *Cours D'Economie Politique*. Lausanne: Rouge.

Parry, M. L. (2007). *Climate change 2007-impacts, adaptation and vulnerability: Working group II contribution to the fourth assessment report of the IPCC* volume 4. Cambridge University Press.

Pelikan, M. (2005). *Hierarchical bayesian optimization algorithm, toward a new generation of evolutionary algorithms, series: Studies in fuzziness and soft computing*.

Pennell, C., & Reichler, T. (2011). On the effective number of climate models. *Journal of Climate*, *24*, 2358–2367.

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., & Hegewisch, K. C. (2015). Improved bias correction techniques for hydrological simulations of climate change. *Journal of Hydrometeorology*, *16*, 2421–2442.

Pierce, D. W., Cayan, D. R., & Thrasher, B. L. (2014). Statistical downscaling using localized constructed analogs (loca). *Journal of Hydrometeorology*, *15*, 2558–2585.

Poff, N. L., Brown, C. M., Grantham, T. E., Matthews, J. H., Palmer, M. A., Spence, C. M., Wilby, R. L., Haasnoot, M., Mendoza, G. F., Dominique, K. C. et al. (2015). Sustainable water management under future uncertainty with eco-engineering decision scaling. *Nature Climate Change*, .

Pruyt, E., & Islam, T. (2016). On generating and exploring the behavior space of complex models. *System Dynamics Review*, .

Quinn, J. D., Reed, P. M., Giuliani, M., & Castelletti, A. (Accepted). Rival framings: A framework for discovering how problem formulation uncertainties shape risk management trade-offs in water resources systems. *Water Resources Research*, (pp. n/a–n/a). URL: http://dx.doi.org/10.1002/2017WR020524. doi:10.1002/2017WR020524.

Quinn, J. D., Reed, P. M., Giuliani, M., Castelletti, A., Oyler, J., & Nicholas, R. (In Prep). Exploring how changing monsoonal dynamics and human pressures challenge multi-reservoir management of food-energy-water tradeoffs, .

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

Raman, H., & Chandramouli, V. (1996). Deriving a general operating policy for reservoirs using neural network. *Journal of Water Resources Planning and Management*, *122*, 342–347.

Ray, P. A., Watkins Jr, D. W., Vogel, R. M., & Kirshen, P. H. (2013). Performance-based evaluation of an improved robust optimization formulation. *Journal of Water Resources Planning and Management*, *140*, 04014006.

Reed, P. M., & Hadka, D. (2014). Evolving many-objective water management to exploit exascale computing. *Water Resources Research*, (pp. n/a–n/a). URL: http://onlinelibrary.wiley.com/doi/10.1002/2014WR015976/abstract. doi:10.1002/2014WR015976.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present and future. *Advances in Water Resources*, *51*, 438–456.

Reed, P. M., & Kasprzyk, J. R. (2009). Water resources management: The myth, the wicked, and the future. *Journal of Water Resources Planning and Management*, *135*, 411–413.

Revelle, C., Joeres, E., & Kirby, W. (1969). The linear decision rule in reservoir management and design: 1, development of the stochastic model. *Water Resources Research*, *5*, 767–777.

Roach, T., Kapelan, Z., Ledbetter, R., & Ledbetter, M. (2016). Comparison of robust optimization and info-gap methods for water resource management under deep uncertainty. *Journal of Water Resources Planning and Management*, *142*, 04016028.

Rockstrom, J., Steffen, W., Noone, K., Persson, A., Chapin, F., Lambin, E., Lenton, T., Scheffer, M., Folke, C., Schellnhuber, H., Nykvist, B., Wit, C. D., Hughes, T., van der Leeuw, S., Rodhe, H., Sorlin, S., Snyder, P., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R., Fabry, V., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., & Foley, J. (2009). Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and Society*, *14*.

Rosenstein, M. T., & Barto, A. G. (2001). Robot weightlifting by direct policy search. In *International Joint Conference on Artificial Intelligence* (pp. 839–846). Citeseer volume 17.

Roy, B. (1990). Decision-aid and decision-making. *European Journal of Operational Research*, *45*, 324–331.

Roy, B. (2010). Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research*, *200*, 629–638.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, *145*, 280–297.

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

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, *461*, 53–59.

Scheffer, M., Hosper, S., Meijer, M., Moss, B., & Jeppesen, E. (1993). Alternative equilibria in shallow lakes. *Trends in ecology & evolution*, *8*, 275–279.

Singh, R., Quinn, J. D., M., R. P., & Keller, K. (In Review). Skill (or lack thereof) of data-model fusion techniques to provide an early warning signal for an approaching tipping point. *Ecosystems*, .

Singh, R., Reed, P. M., & Keller, K. (2015). Many-objective robust decision making for managing an ecosystem with a deeply uncertain threshold response. *Ecol. Soc*, *20*, 12.

Sivapalan, M., Savenije, H. H., & Blöschl, G. (2012). Socio-hydrology: A new science of people and water. *Hydrological Processes*, *26*, 1270–1276.

Smith, R., Kasprzyk, J., & Dilling, L. (2017). Participatory framework for assessment and improvement of tools (parfait): Increas-

ing the impact and relevance of water management decision support research. *Environmental Modelling & Software*, (pp. –). URL: http://www.sciencedirect.com/science/article/pii/S1364815216308362. doi:https://doi.org/10.1016/j.envsoft.2017.05.004.

Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*, 271–280.

Solomon, S., Plattner, G., Knutti, R., & Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. *Proceedings of the National Academy of Sciences*, *106*, 1704 –1709.

Soncini-Sessa, R., Weber, E., & Castelletti, A. (2007). *Integrated and participatory water resources management-theory* volume 1. Elsevier.

Soncini-Sessa, R., Zuleta, J., & Piccardi, C. (1990). Remarks on the application of a risk-averse approach to the management of el carrizalreservoir. *Advances in Water Resources*, *13*, 76–84.

Starr, C. (1969). Social benefit versus technological risk. *Science*, *165*, 1232–1238.

Starr, M. K. (1962). *Product Design and Decision Theory*. Prentice-Hall, Englewood Cliffs, N.J.

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In D. R. Maidment (Ed.), *Handbook of Hydrology* chapter 18. New York: McGraw-Hill.

Steinschneider, S., McCrary, R., Mearns, L. O., & Brown, C. (2015a). The effects of climate model similarity on probabilistic climate projections and the impli-

cations for local, risk-based adaptation planning. *Geophysical Research Letters*, *42*, 5014–5044.

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

Stirling, A. (2008). opening up? and closing down? power, participation, and pluralism in the social appraisal of technology. *Science, Technology, & Human Values*, *33*, 262–294.

Storn, R., & Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*, 341–359.

Taguchi, G. (1986). *Introduction to quality engineering: designing quality into products and processes*.

Takriti, S., & Ahmed, S. (2004). On robust optimization of two-stage systems. *Mathematical Programming*, *99*, 109–126.

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

Tejada-Guibert, J. A., Johnson, S. A., & Stedinger, J. R. (1995). The value of hydrologic information in stochastic dynamic programming models of a multireservoir system. *Water Resour. Res.*, *31*, 2571–2579. URL: `http://dx.doi.org/10.1029/95wr02172`. doi:10.1029/95wr02172.

Thierens, D., Goldberg, D. E., & Pereira, A. G. (1998). Domino convergence, drift, and the temporal-salience structure of problems. In *The 1998 IEEE International Conference on Evolutionary Computation* (pp. 535–540).

Thomas, J., & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications*, *26*, 10–13.

Thomas, Jr., H. A., & Fiering, M. B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In A. Maass, M. M. Hufschmidt, R. Dorfman, H. A. Thomas, Jr., S. A. Marglin, & G. M. Fair (Eds.), *Design of Water-Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning*. Cambridge: Harvard University Press.

Thornton, P. K., Ericksen, P. J., Herrero, M., & Challinor, A. J. (2014). Climate variability and vulnerability to climate change: a review. *Global change biology*, *20*, 3313–3328.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, *47*, 123–138.

Trindade, B., Reed, P., Herman, J., Zeff, H., & 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.

Tsoukias, A. (2008). From decision theory to decision aiding methodology. *European Journal of Operational Research*, *187*, 138–161.

Tsutsui, S., Yamamura, M., & Higuchi, T. (1999). Multi-parent recombination

with simplex crossover in real coded genetic algorithms. In *Genetic and Evolutionary Computation Conference (GECCO 1999)*.

Vautard, R., Cattiaux, J., Yiou, P., Thépaut, J.-N., & Ciais, P. (2010). Northern hemisphere atmospheric stilling partly attributed to an increase in surface roughness. *Nature Geoscience*, *3*, 756–761.

Vinh Hung, H., Shaw, R., & Kobayashi, M. (2007). Flood risk management for the rua of hanoi: importance of community perception of catastrophic flood risk in disaster risk planning. *Disaster Prevention and Management: An International Journal*, *16*, 245–258.

Vörösmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: vulnerability from climate change and population growth. *science*, *289*, 284.

Wagener, T., Sivapalan, M., Troch, P., McGlynn, B., Harman, C., Gupta, H., Kumar, P., Rao, P., Basu, N., & Wilson, J. (2010). The future of hydrology: An evolving science for a changing world. *Water Resources Research*, *46*, W05301.

Wald, A. (1992). Statistical decision functions. In *Breakthroughs in Statistics* (pp. 342–357). Springer.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. *Integrated assessment*, *4*, 5–17.

Walker, W. E., Rahman, S. A., & Cave, J. (2001). Adaptive policies, policy analysis, and policy-making. *European Journal of Operational Research*, *128*, 282–289. URL:

http://www.sciencedirect.com/science/article/pii/S0377221700000710. doi:10.1016/S0377-2217(00)00071-0.

Wang, B., Liu, J., Kim, H.-J., Webster, P. J., & Yim, S.-Y. (2012). Recent change of the global monsoon precipitation (1979–2008). *Climate Dynamics*, *39*, 1123–1135.

Ward, V. L., Singh, R., Reed, P. M., & Keller, K. (2015). Confronting tipping points: Can multi-objective evolutionary algorithms discover pollution control tradeoffs given environmental thresholds? *Environmental Modelling & Software*, *73*, 27–43.

Watkins, D. W., & McKinney, D. C. (1997). Finding robust solutions to water resources problems. *Journal of Water Resources Planning and Management*, *123*, 49–58. URL: http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9496(1997)123:1(49

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

Werners, S. E., Pfenninger, S., van Slobbe, E., Haasnoot, M., Kwakkel, J. H., & Swart, R. J. (2013). Thresholds, tipping and turning points for sustainability under climate change. *Current opinion in environmental sustainability*, *5*, 334–340.

Whateley, S., Steinschneider, S., & Brown, C. (2014). A climate change range-based method for estimating robustness for water resources supply. *Water Resources Research*, *50*, 8944–8961.

Windsor, J. S. (1973). Optimization model for the operation of flood control systems. *Water Resources Research*, *9*, 1219–1226.

Wissel, C. (1984). A universal law of the characteristic return time near thresholds. *Oecologia*, *65*, 101–107.

World Bank (2016). *High and Dry: Climate Change, Water and the Economy*. World Bank, Washington, DC.

Yakowitz, S. (1982). Dynamic programming applications in water resources. *Water Resour. Res.*, *18*, 673–696. URL: http://dx.doi.org/10.1029/wr018i004p00673. doi:10.1029/wr018i004p00673.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., & Kitoh, A. (2012). Aphrodite: Constructing a long-term daily gridded precipitation dataset for asia based on a dense network of rain gauges. *Bulletin of the American Meteorological Society*, *93*, 1401–1415.

Yeh, W. W.-G. (1985). Reservoir management and operations models: A state-of-the-art review. *Water Resources Research*, *21*, 1797–1818.

Yu, B., Zhu, T., Breisinger, C., Hai, N. M. et al. (2010). Impacts of climate change on agriculture and policy options for adaptation. *International Food Policy Research Institute (IFPRI)*, .

Zatarain Salazar, J., Reed, P., Quinn, J. D., Giuliani, M., & Castelletti, A. (In Review). Balancing exploration, uncertainty and computational demands in many objective reservoir optimization. *Advances in Water Resources*, .

Zatarain Salazar, J., Reed, P. M., Herman, J. D., Giuliani, M., & Castelletti, A. (2016). A diagnostic assessment of evolutionary algorithms for multi-

objective surface water reservoir control. *Advances in Water Resources*, *92*, 172–185.

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

Zeff, H. B., Kasprzyk, J. R., Herman, J. D., Reed, P. M., & Characklis, G. W. (2014). Navigating financial and supply reliability tradeoffs in regional drought management portfolios. *Water Resources Research*, *50*, 4906–4923.

Zeleny, M. (1981). On the squandering of resources and profits via linear programming. *Interfaces*, *11*, 101–107.

Zeleny, M. (1989). Cognitive equilibrium: A new paradigm of decision making? *Human Systems Management*, *8*, 185–188.

Zhou, T., Gong, D., Li, J., & Li, B. (2009). Detecting and understanding the multi-decadal variability of the east asian summer monsoon–recent progress and state of affairs. *Meteorologische Zeitschrift*, *18*, 455–467.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., & Fonseca, V. G. d. (2003). Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Transactions on Evolutionary Computation*, *7*, 117–132. doi:10.1109/tevc.2003.810758.