ADVANCING MODEL DIAGNOSTICS TO SUPPORT HYDROLOGIC PREDICTION AND WATER RESOURCES PLANNING UNDER UNCERTAINTY

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

Jonathan Drew Herman

May 2015

ADVANCING MODEL DIAGNOSTICS TO SUPPORT HYDROLOGIC PREDICTION AND WATER RESOURCES PLANNING UNDER UNCERTAINTY

Jonathan Drew Herman, Ph.D.

Cornell University 2015

Computational models are essential tools for prediction and planning in water resources systems to ensure human water security and environmental health. Water systems models merely approximate the processes by which water moves through natural and built environments; their value depends on assumptions regarding climate, demand, land use, and other uncertain factors that may influence decision making. Numerical techniques to explore the role of these uncertain factors, known as diagnostic methods, can highlight opportunities to improve the accuracy of prediction as well as identify influential uncertainties to inform additional research and policy. This dissertation advances diagnostic methods for water resources models to identify (1) time-varying dominant processes driving modeled hydrologic predictions in flood forecasting, and (2) tradeoffs and vulnerabilities to changing climate and demands in regional urban water supply systems planning for drought. This work proposes diagnostic methods as a key element of *a posteriori* decision support, in which decision alternatives and vulnerable scenarios are identified following computational modeling and data analysis. Consistent with this theme, this work follows a multi-objective approach in which stakeholders can analyze tradeoffs between conflicting objectives as part of an iterative constructive learning process. For a spatially distributed flood forecasting model, results show that

dominant uncertainties vary in space and time, and can inform model-based scientific inference as well as decision making. Similarly, the results of the urban water supply study indicate that sensitivity analysis can suggest cost-effective paths to mitigate vulnerability to deeply uncertain future scenarios, for which likelihoods remain unknown or disputed. The multi-objective approach allows stakeholders to explore tradeoffs in their modeled robustness to inform intra-regional policies such as transfer contracts and shared infrastructure investments. Bridging the areas of hydrology and water systems planning is increasingly valuable, as hydrologic modelers begin to incorporate anthropogenic influences on the water cycle, and water systems planners begin to explore uncertainty in hydrologic process representation. In summary, this work develops diagnostic methods to identify time-varying dominant processes in distributed flood forecasting as well as tradeoffs and vulnerabilities under change in regional urban water supply, ultimately seeking to improve model-based planning for extreme floods and droughts in water resources systems.

BIOGRAPHICAL SKETCH

Jon Herman grew up in the Pocono Mountains of Pennsylvania. He attended Dartmouth College, where he received an A.B. in Engineering Sciences and a B.E. with environmental concentration in 2010. Jon then began graduate work at Penn State University in water resources engineering, earning his M.S. before transitioning to the Cornell Environmental & Water Resources Systems program in 2013. After graduation he will begin a faculty position at the University of California, Davis in the Department of Civil & Environmental Engineering.

To my future wife Jenny, who brightens every day;
and to my family, who taught me the value of hard work
and intellectual humility. *Hochmut muss leide*.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Prof. Patrick Reed, for serving as a constant source of good ideas, technical expertise, and professional mentorship over the years. I especially appreciate the opportunities he has provided for collaborations with researchers across disciplines and around the world, which have allowed me to broaden my horizons and improve my own research. I would also like to thank the members of my Ph.D. committee, Prof. Jery Stedinger and Prof. Todd Walter, for their support and constructive feedback during my time at Cornell. Finally, I would like to thank Prof. Thorsten Wagener for encouraging thoughtful experiments and creative data visualizations over the course of my M.S. degree at Penn State.

I've been fortunate to have met many new friends and colleagues during my time in graduate school. I'm grateful to Prof. Joe Kasprzyk for his friendship and support of my technical and professional endeavors, and to Dr. Josh Kollat for his collaboration and pragmatic approach to research computing. I'd also like to thank Dr. Dave Hadka, the 10x programmer who has made all of us more productive. Thanks to the many fellow graduate students and postdocs who helped create an atmosphere of camaraderie and sanity-preserving support: at Penn State, Dr. Christa Kelleher, Prof. Riddhi Singh, Dr. Keith Sawicz, Matt Woodruff, Dr. Martha Butler, Rachel Urban, Dr. Alisha Fernandez, Emily Bernzott, Adam Wlostowski, and Evan Thomas; at Cornell, Dr. Jon Lamontagne, Dr. Tom Wild, Sue Nee Tan, Calvin Whealton, Jared Smith, Tori Ward, Min Pang, Jazmin Zatarain-Salazar, Allison Truhlar, Julie Quinn, Candice Yu, and Bernardo Trindade.

In addition, I've been doubly fortunate for the chance to work with talented researchers from other universities. I'd like to thank Nate Chaney at Princeton

for his unique ability to debug parallel programs on a deadline. At UNC Chapel Hill, I'd like to thank Prof. Greg Characklis and Harrison Zeff for lending their economic expertise to our work. Finally, at Politecnico Di Milano, I'd like to thank Prof. Andrea Castelletti and Dr. Matteo Giuliani for sharing their clever ideas about policy optimization.

This research was funded by several sources: the National Science Foundation (NSF) under grants EAR-0838357 and OCI-0821527; the National Oceanic and Atmospheric Administration (NOAA) Climate Program Office under award NA110AS2310144; the United States Department of Agriculture (USDA) under WSC Agreement 2014-67003-22076; the Cornell School of Civil & Environmental Engineering Olin Fellowship; and the Cornell University Bernard Meyers Graduate Fellowship. High-performance computing resources were provided by the Texas Advanced Computing Center (TACC). I gratefully acknowledge the support of these programs.

Finally, I'm incredibly lucky to have the love and support of my fiancée Jenny as well as my parents, whose selfless hard work and commitment to my education have allowed me to pursue an academic career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Current Challenges for Model-Based Decision Support in Water Resources Systems

Computational models are essential tools for prediction and planning in water resources systems to ensure human water security and environmental health. Water systems models by definition only approximate the processes by which water moves through natural and built environments; as Rosenblueth and Wiener (1945) noted, "partial models, imperfect as they may be, are the only means developed by science for understanding the universe". Therefore, the value of computational models for decision making depends on assumptions regarding climate, demand, land use, and other uncertain factors that may influence planning and management decisions. Numerical techniques to explore the role of these uncertain factors in a model—i.e., sensitivity analysis, or more generally, model diagnostics—can highlight opportunities to improve the accuracy of prediction as well as identify influential uncertainties to direct additional research and inform policy (Saltelli and Funtowicz, 2013). This dissertation advances diagnostic methods for water resources prediction and planning models to identify (1) time-varying dominant processes driving modeled hydrologic predictions in flood forecasting, and (2) tradeoffs and vulnerabilities to changing climate and demands in regional urban water supply systems planning for drought. In doing so, this work seeks to strengthen model-based decision support in water resources systems in the presence of conflicting objectives and uncertain extreme events.

In general, a simulation model of an environmental system can be thought of as a function mapping decision variables, forcing data, and parameters to one or more performance objectives. A model may range in complexity from a simple empirical relationship, such as the rational formula for estimating peak runoff (Dooge, 1957), to physically-based watershed models with spatially distributed storage elements (Abbott et al., 1986; Beven, 1989). Regardless of complexity, water resources models are typically calibrated to historical observations, yet must maintain fidelity under as-yet-unobserved forcing conditions to offer value to decision makers. To address this problem, model diagnostic methods such as sensitivity analysis can be used to discover the parameters and processes most responsible for changes in the performance objectives, to identify the causes of model deficiencies and ultimately improve process representation. Once stakeholders trust model predictions under forcing conditions outside the range of historical observations, the type and degree of uncertain parameters can be expanded to discover scenarios in which the system may fail (i.e., extreme events such as floods and droughts). Furthermore, water resources systems must consider multiple stakeholders with potentially conflicting objectives (Maass et al., 1962; Cohon and Marks, 1973; Haimes and Hall, 1974). These challenges to model-based decision support will be described in detail in the following subsections for operational flood forecasting and regional portfolio planning to improve robustness to droughts in urban water supply systems.

1.1.1 Hydrologic Process Representation in Flood Forecasting

Watershed models are valuable tools for forecasting flood events, particularly when historical rainfall-runoff relationships are available for calibration. How-

ever, the extent to which a particular model structure can adequately capture the true watershed processes remains a challenge (Gupta et al., 2008; Wagener et al., 2010; Ehret et al., 2013; Gupta and Nearing, 2014; Mendoza et al., 2015). This task is vital for predictions under changing hydrologic conditions, where a modeler must rely on process fidelity rather than an observed input-output relationship (Ewen and Parkin, 1996; Van Werkhoven et al., 2008a; Wagener et al., 2010). In other words, while a model can be made to reproduce historical observations, it must also do so "for the right reasons" to be of value (Kirchner, 2006). The sensitivity of uncertain model parameters can be used to identify dominant processes in a model, serving as a diagnostic tool to trace the causes of undesirable performance and improve the representation of the system. Importantly, modeled dominant processes vary across space (e.g., van Griensven et al., 2006; Tang, 2007; Van Werkhoven et al., 2008a) and time (McCuen, 1973; Reusser et al., 2009; Reusser and Zehe, 2011). Efforts to analyze the spatiotemporal dynamics of parameter sensitivity have been limited by computational and methodological constraints, and have typically attempted to aggregate parameter values across the model grid (e.g., Carpenter et al., 2001; Sieber and Uhlenbrook, 2005; Cuo et al., 2011; Guse et al., 2013) or to aggregate temporally based on hydrologic events (Tang, 2007; Van Werkhoven et al., 2008b; Wagener et al., 2009b). Furthermore, such analyses are rarely performed at a timescale commensurate with the needs of flood forecasting applications (typically a 6-hour resolution or finer). This dissertation contributes a novel approach to time-varying sensitivity analysis of a spatially distributed flood forecasting model, first advancing methods capable of analyzing a high-dimensional parameter space (Chapter 3), then providing conceptual support to explore the dynamic model processes driving the prediction of flood events (Chapter 4).

1.1.2 Urban Water Supply Planning under Change

A need for novel model diagnostic methods also exists in the field of urban water supply planning, where traditional planning methods based on historical drought frequency are challenged by long-term changes in climate (Milly et al., 2008; IPCC, 2013) and demand (Frederick and Major, 1997). While process accuracy is not guaranteed, a larger concern is the inability to predict the climate and demand conditions under which a system will be expected to provide reliable water supply. With sufficient data, it may be possible to account for these uncertainties probabilistically, such as an expected cost-benefit analysis (Schoemaker, 1982; Banzhaf, 2009). However, it is often the case in long-term systems planning that the distributions of uncertain variables are themselves uncertain, a condition referred to as deep uncertainty (Knight, 1921; Langlois and Cosgel, 1993; Lempert, 2002). Deep uncertainty may appear even assuming stationary conditions, albeit to a lesser extent, since the distribution family and parameter values can never be known exactly. Nonstationarity is a long-recognized concept (Matalas and Fiering, 1977) which, while not fully invalidating the historical record (Matalas, 2012), has nevertheless motivated the discussion of deep uncertainty in the literature and the development of new decision analysis frameworks to address it.

Deep uncertainty does not reflect a state of complete ignorance, since it is possible to generate plausible states of the world and potentially rank them in order of likelihood following additional analysis (Kwakkel and Pruyt, 2013). However, we would prefer to do so *a posteriori*, after identifying the most vulnerable states of the world for which likelihoods should be derived. This "scenario-neutral" (Prudhomme et al., 2010) or "bottom-up" approach (Weaver et al., 2013;

Nazemi and Wheater, 2014) has been advocated by several recent frameworks for decision making under uncertainty, including Many-Objective Robust Decision Making (MORDM) (Kasprzyk et al., 2013), which is extended in this work. Bottom-up frameworks stand in contrast to a "top-down" approach, in which analysts seek to predict future climate and demand conditions prior to assessing their impact on the system. This dissertation reviews the field of bottom-up decision analysis frameworks under deep uncertainty, and advances the methodology for multi-stakeholder urban water supply systems (Chapters 5 and 6).

1.1.3 Navigating Tradeoffs in Multi-Objective, Multi-Stakeholder Systems

An additional outstanding challenge in water resources systems is the need to balance conflicting objectives across multiple stakeholders. Multiple objectives may be aggregated *a priori* with weighted preferences (Reuss, 2003; Banzhaf, 2009), an approach which allows simple solution techniques but requires decision makers to accurately assign preference weighting prior to reviewing alternatives (Bond et al., 2008, 2010), which cannot be guaranteed. An alternative approach is to seek the full set of tradeoff solutions which cannot improve one objective without degrading another, also known as the Pareto set (Pareto, 1896). Stakeholder preferences can then be discussed *a posteriori* with knowledge of which candidate solutions are approximately Pareto-optimal. This *a posteriori* approach to decision support, reviewed by Tsoukias (2008), allows constructive learning with stakeholder feedback (Roy, 1971, 1990), in which problem formulations may compete as multiple working hypotheses (Chamberlin, 1890).

Modern search tools such as multi-objective evolutionary algorithms (MOEAs) allow the approximation of the Pareto set in a single optimization run. This dissertation employs MOEAs to discover promising portfolios of drought mitigation strategies, balancing objectives such as volumetric reliability and cost across multiple water utilities in a region (Chapters 5 and 6).

In summary, this work addresses the outstanding challenges of hydrologic process fidelity, water supply vulnerability to deeply uncertain change, and multi-objective decision support under uncertainty through the common thread of model diagnostics. The following chapters will develop diagnostic methods to explore the representation of dominant hydrologic processes and identify vulnerabilities to nonstationary model assumptions, ultimately seeking to improve model-based planning for extreme floods and droughts in water resources systems.

1.2 Scope & Organization

1.2.1 Chapter 2: Methodological Components

Chapter 2 will provide technical background on the computational methods extended in this dissertation. These include: methods for sensitivity analysis to identify input factors, such as Sobol variance decomposition and Morris screening analysis; concepts and methods for multi-objective optimization using evolutionary algorithms; and concepts dealing with decision support under deep uncertainty. The rationale for employing these methods is to identify the most influential uncertainties in a system along with candidate solutions to balance

tradeoffs between conflicting objectives.

1.2.2 Chapter 3: Efficient Sensitivity Analysis for Spatially Distributed Flood Forecasting

Chapter 3 seeks a computationally efficient approach to identifying dominant processes in a spatially distributed flood forecasting model. Despite the increased use of spatially distributed models in hydrology, there remains a need for diagnostic methods capable of identifying dominant processes in complex, highly parameterized models without aggregating parameter values. Sobol sensitivity analysis (Section 2.1.1) has proven to be a valuable tool for diagnostic analyses of hydrologic models; however, for many spatially distributed models, the Sobol method requires a prohibitive number of model evaluations to reliably decompose output variance across the full set of parameters. We investigate the potential of the method of Morris (Section 2.1.2), a screening-based sensitivity approach, to provide results sufficiently similar to those of the Sobol method at a greatly reduced computational expense. The methods are benchmarked on the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) model (Koren et al., 2004) over a six-month period in the Blue River Watershed, Oklahoma, USA. Results of this comparison indicate that the method of Morris correctly identifies sensitive and insensitive parameters with 300 times fewer model evaluations than the Sobol method. Method of Morris proves to be a promising diagnostic approach for global sensitivity analysis of highly parameterized, spatially distributed hydrologic models. Chapter 3 is drawn from Herman et al. (2013a), published in *Hydrology & Earth System Sci-*

ences.

1.2.3 Chapter 4: High-Resolution Time-Varying Sensitivity Analysis for Spatially Distributed Watershed Models

Chapter 4 builds on the results of the previous chapter to perform a time-varying sensitivity analysis of a spatially distributed flood forecasting model, seeking to align diagnostic methods with the models' operational and scientific goals (i.e., dynamic responses to spatially variable precipitation at sub-daily timescales). Chapter 4 contributes a novel approach for computing and visualizing time-varying global sensitivity indices for spatially distributed model parameters. The high-resolution model diagnostics developed in this chapter employ the method of Morris to identify evolving patterns in dominant model processes at sub-daily timescales over a six-month period. The method is demonstrated on the HL-RDHM model in the Blue River watershed. Three hydrologic events are selected from the six-month period to investigate the patterns in spatiotemporal sensitivities that emerge as a function of forcing patterns as well as wet-to-dry transitions. Events with similar magnitudes and durations exhibit significantly different performance controls in space and time, indicating that the diagnostic inferences drawn from representative events will be biased by the *a priori* selection of those events. By contrast, this chapter demonstrates high-resolution time-varying sensitivity analysis, requiring no assumptions regarding representative events and allowing modelers to identify transitions between sets of dominant parameters or processes *a posteriori*. The proposed approach details the dynamics of parameter sensitivity in nearly continuous time, provid-

ing decision-relevant diagnostic insights into the underlying model processes driving predictions. Furthermore, the approach offers the potential to identify transition points between dominant parameters and processes in the absence of observations, such as under nonstationarity. Chapter 3 is drawn from Herman et al. (2013b), published in *Hydrology & Earth System Sciences*.

1.2.4 Chapter 5: Robustness-Based Portfolio Planning for Regional Drought Management under Deep Uncertainty

Chapter 5 advances diagnostic methods for water supply planning models, moving from the traditional approach of seeking optimal solutions in the projected future toward seeking robust solutions capable of acceptable performance across a range of possible futures. While optimality is a foundational mathematical concept in water resources planning and management, "optimal" solutions may be vulnerable to failure if deeply uncertain future conditions deviate from those assumed during optimization. These vulnerabilities may produce asymmetric impacts across a region, making it vital to evaluate the robustness of management strategies as well as their impacts for regional stakeholders. Chapter 5 contributes a multi-stakeholder many-objective robust decision making (MORDM) framework that blends many-objective search and uncertainty analysis tools to discover tradeoffs between water supply alternatives and their robustness to deep uncertainties (e.g., population pressures, climate change, financial risks, etc.). The proposed framework is demonstrated for four interconnected water utilities representing major stakeholders in the "Research Triangle" region of North Carolina, USA. The utilities supply well over one mil-

lion customers and have the ability to collectively manage drought via transfer agreements and shared infrastructure. Results show that water portfolios for this region that compose optimal tradeoffs (i.e., Pareto-approximate solutions) under projected future conditions may suffer degraded performance with only modest changes in deeply uncertain hydrologic and economic factors. We then use the Patient Rule Induction Method (PRIM, Section 2.1.3) to identify which uncertain factors drive the individual and collective vulnerabilities for the four cooperating utilities. The framework identifies stakeholder dependencies and robustness tradeoffs associated with cooperative regional planning, which illuminate the tensions between individual versus regional water supply goals. Cooperative demand management was found to be the key factor controlling the robustness of regional water supply planning, dominating other hydroclimatic and economic uncertainties through the 2025 planning horizon. Results suggest that a modest reduction in the projected rate of demand growth (from approximately 3% per year to 2.4%) will substantially improve the utilities' robustness to future uncertainty and reduce the potential for regional tensions. The proposed multi-stakeholder MORDM framework offers model-based insights into the risks and challenges posed by rising water demands and hydrological uncertainties, providing a planning template for regions now forced to confront rapidly evolving water scarcity risks. Chapter 5 is drawn from Herman et al. (2014), published in *Water Resources Research*.

1.2.5 Chapter 6: How Should Robustness be Defined for Water Systems Planning under Change?

While Chapter 5 develops a methodology for discovering robust solutions under deep uncertainty, Chapter 6 reviews other emerging methods in the field of bottom-up decision analysis frameworks and compares the implications of their recommendations. Bottom-up approaches identify vulnerable scenarios prior to assigning likelihoods; examples include Robust Decision Making (RDM), Decision Scaling, Info-Gap, and Many-Objective Robust Decision Making (MORDM). Chapter 6 proposes a taxonomy of robustness frameworks to compare and contrast these approaches, based on their methods of (1) alternative generation, (2) sampling of states of the world, (3) quantification of robustness measures, and (4) sensitivity analysis to identify important uncertainties. Building from the proposed taxonomy, we use a regional urban water supply case study in the Research Triangle region of North Carolina to illustrate the decision-relevant consequences that emerge from each of these choices. Results indicate that the methodological choices in the taxonomy lead to the selection of substantially different planning alternatives, underscoring the importance of an informed definition of robustness. Moreover, the results show that some commonly employed methodological choices and definitions of robustness can have undesired consequences when ranking decision alternatives. For the demonstrated test case, recommendations for overcoming these issues include: (1) that decision alternatives should be searched rather than prespecified; (2) dominant uncertainties should be discovered via sensitivity analysis rather than assumed; and (3) that a carefully elicited multivariate satisficing measure of robustness allows stakeholders to achieve their problem-specific performance requirements.

This work emphasizes the importance of an informed problem formulation for systems facing challenging performance tradeoffs, and provides a common vocabulary to link the robustness frameworks widely used in the field of water systems planning. Chapter 6 is drawn from Herman et al. (2015), forthcoming in *Journal of Water Resources Planning & Management*.

1.2.6 Chapter 7: Contributions & Future Work

Finally, Chapter 7 summarizes the findings of this dissertation and its contributions to model-based decision support in water resources systems under uncertainty. Recommendations for future work are provided, including the potential for new software tools to facilitate stakeholder interaction, real-time information and data assimilation to reduce uncertainty and improve model predictions, and the design of policies based on decision triggers to improve adaptability to future change. This work contributes to ongoing computational and conceptual improvements in water systems modeling to overcome the challenges to decision making posed by conflicting objectives and uncertain extreme events.

1.2.7 Author Contributions for Published Work

Chapters 3 and 4: Jonathan Herman conceived the studies and led the modeling, data analysis, and writing. Joshua Kollat led the parallel computing formulation and assisted with the data analysis and writing. Patrick Reed and Thorsten Wagener supervised the experiments and contributed to the data analysis and writing.

Chapters 5 and 6: Jonathan Herman conceived the studies and led the computational experiments, data analysis, and writing. Harrison Zeff led the design of financial instruments and contributed to the modeling, data analysis and writing. Patrick Reed and Greg Characklis supervised the experiments and contributed to the data analysis and writing.

CHAPTER 2

METHODOLOGICAL COMPONENTS

2.1 Sensitivity Analysis

2.1.1 Sobol sensitivity analysis

Sobol sensitivity analysis (Sobol, 1993; Sobol, 2001; Saltelli, 2002) is a global, variance-based method that attributes the total variance of output Y to individual inputs and their interactions, following an ANOVA-like decomposition (Efron and Stein, 1981; Iooss and Lemaître, 2014):

$$\mathrm{Var}(Y) = \sum_i D_i(Y) + \sum_{i<j} D_{ij}(Y) + \sum_{i<j<k} D_{ijk}(Y) + D_{12\ldots n}(Y) \tag{2.1}$$

where D_i is the first-order variance contribution of the i^{th} parameter, D_{ij} is the second-order contribution of the interaction between parameters i and j, D_{ijk} represents the third-order interaction between parameters i, j, and k, and $D_{12\ldots n}$ contains all interactions higher than third-order, up to n total inputs. The first-order and total-order sensitivity indices are defined as follows:

$$S_i = \frac{\mathrm{Var}[\mathrm{E}(Y|X_i)]}{\mathrm{Var}(Y)} = \frac{D_i}{\mathrm{Var}(Y)} \tag{2.2}$$

$$S_{T_i} = \frac{\mathrm{Var}[\mathrm{E}(Y|\mathbf{X}_{\sim i})]}{\mathrm{Var}(Y)} = \frac{D_{\sim i}}{\mathrm{Var}(Y)} \tag{2.3}$$

The first-order index represents the expected reduction in variance if input i were fixed, not accounting for interactions with other inputs. The total-order index represents the reduction in variance ($D_{\sim i}$) that would occur if all inputs except i were fixed. The total-order index incorporates estimates of all interactions with other parameters. The difference between a parameter's first and

total order indices represents the effects of its interactions with other parameters.

Estimators for S_i and S_{T_i} are reviewed by Saltelli et al. (2010); here, we follow the Monte Carlo scheme proposed by Sobol (2001) and Saltelli (2002); Saltelli et al. (2008). First, a global sample of the parameter space is taken using a quasi-random Sobol sequence (Sobol, 2001) in place of standard uniform sampling to improve convergence. For a sample size n and number of parameters p, two $(n \times p)$ matrices A and B are sampled. The samples from matrix A are evaluated in the model, and the result is used to estimate the mean and variance of the output Y:

$$\hat{\mu_Y} = \frac{1}{n}\sum_{s=1}^{n} f(\theta_s), \quad \hat{\sigma_Y^2} = \frac{1}{n}\sum_{s=1}^{n} f^2(\theta_s) - \hat{\mu_Y}^2. \tag{2.4}$$

Here, f is the function representing the model and θ_s represents the parameter set associated with sample s. For adaptations to equation 2.4 which aim to improve the convergence of Sobol indices, the reader is referred to (Saltelli, 2002) and (Nossent and Bauwens, 2012b).

The estimated variance contributions $\hat{D}_i$ and $\hat{D_{\sim i}}$ are calculated according to Sobol (2001) and Saltelli et al. (2008). The second sample matrix B is used to "cross-sample" each parameter value to compute the variance contributions as follows:

$$\hat{D}_i = \frac{1}{n}\sum_{s=1}^{n} f\left(\theta_s^A\right) f\left(\theta_{\sim is}^B, \theta_{is}^A\right) - \hat{\mu_Y}^2 \tag{2.5}$$

$$\hat{D_{\sim i}} = \frac{1}{n}\sum_{s=1}^{n} f\left(\theta_s^A\right) f\left(\theta_{\sim is}^A, \theta_{is}^B\right) - \hat{\mu_Y}^2. \tag{2.6}$$

In Equations 2.5 and 2.6, the parameter sets θ_i are superscripted to indicate which parameters are sampled from which set. The sample set is denoted by the superscript A or B; the parameters taken from that set are denoted either by i (the i-th parameter) or $\sim i$ (all parameters except i). This results in a total of

$n(p+2)$ model evaluations that must be performed, representing the computational cost of the method. A suitable sample size n is often not known *a priori*; bootstrap confidence intervals are developed to assess the convergence of the sensitivity indices (Efron, 1979). The estimators described above require a specific sampling technique and cannot be applied to given (arbitrary) input-output data. Following Plischke et al. (2013), it is also possible to estimate first-order Sobol indices from given data, noting that the calculation of total-order indices from given data remains an open question.

2.1.2 Method of Morris

The method of Morris (1991) derives measures of global sensitivity from a set of local derivatives, or elementary effects, sampled on a grid throughout the parameter space. It is based on one-at-a-time (OAT) methods, in which each parameter x_i is perturbed along a grid of size Δ_i to create a trajectory through the parameter space. For a given model with p parameters, one trajectory will contain a sequence of p such perturbations. Each trajectory yields one estimate of the elementary effect for each parameter (i.e., the ratio of the change in model output to the change in that parameter). Equation (2.7) shows the calculation of a single elementary effect for the i-th parameter.

$$EE_i = \frac{f(x_1, \ldots, x_i + \Delta_i, \ldots, x_p) - f(\mathbf{x})}{\Delta_i} \tag{2.7}$$

where $f(\mathbf{x})$ represents the prior point in the trajectory. In alternative formulations, both the numerator and denominator are normalized by the values of the function and parameter x_i, respectively, at the prior point $\mathbf{x}$ (van Griensven et al., 2006). Using the single trajectory shown in Eq. (2.7), one can calculate the elementary effects of each parameter with only $p+1$ model evaluations. However,

by using only a single trajectory, this OAT method is highly dependent on the location of the initial point x in the parameter space and does not account for interactions between parameters. For this reason, the method of Morris (1991) performs the OAT method over N trajectories through the parameter space. This study employs the sampling approach originally proposed by Morris (1991), in which trajectories through the parameter space are generated by randomly perturbing one factor at a time. Recent advances in this area by (Campolongo et al., 2007, 2011) and (Ruano et al., 2012) provide trajectories which maximize coverage of the parameter space, ensuring that the sampled elementary effects yield accurate estimates of global sensitivity. These improvements suggest promising directions for future investigation. Once trajectories are sampled, the resulting set of elementary effects is then averaged to give μ, which serves as an estimate of total-order effects. Similarly, the standard deviation of the set of elementary effects σ describes the variability throughout the parameter space and thus the extent to which parameter interactions are present. This study uses the improvement of (Campolongo et al., 2007) in which an estimate of total-order sensitivity of the i-th parameter, μ_i^*, is computed from the mean of the absolute values of the elementary effects over the set of N trajectories as shown in Eq. (2.8).

$$\mu_i^* = \frac{1}{N} \sum_{j=1}^{N} |EE_i^j| \tag{2.8}$$

2.1.3 Patient Rule Induction Method

The Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999) is a data mining technique to identify ranges of uncertain input factors that are likely to cause a particular outcome. In sensitivity analysis parlance, PRIM may be

thought of as a "factor mapping" approach (Saltelli et al., 2008) given its focus on finding sensitive ranges in the input space. The outcome of interest may be based on a single value (e.g., reliability below 99%), the intersection of multiple outcomes (reliability below 99% and cost above $30 million), or the union of multiple outcomes (reliability below 99% or cost above $30 million).

Consider a sample of N states of the world comprised of k uncertain factors $x_i, i \in (1, \ldots, k)$ and a simulation model $y = f(s, \mathbf{x})$ that computes an output of interest y for a particular strategy s. We additionally define a vector $\mathbf{Y}^*$ containing m policy-relevant thresholds that will be met in a set of "cases of interest" (Bryant and Lempert, 2010):

$$\mathcal{I}_{all} = \{\mathbf{x} \mid f(s, \mathbf{x}) \geq Y_1^*\} \cap \ldots \cap \{\mathbf{x} \mid f(s, \mathbf{x}) \geq Y_m^*\} \tag{2.9}$$

Equation 2.9 refers to the cases of interest $\mathcal{I}_{all}$ for which all thresholds $\mathbf{Y}^*$ are satisfied. Note that the thresholds can, in general, be expressed as any combination of greater-than or less-than requirements. It is also possible to define cases of interest for which any of the conditions $\mathbf{Y}^*$ are satisfied:

$$\mathcal{I}_{any} = \{\mathbf{x} \mid f(s, \mathbf{x}) \geq Y_1^*\} \cup \ldots \cup \{\mathbf{x} \mid f(s, \mathbf{x}) \geq Y_m^*\} \tag{2.10}$$

Following Bryant and Lempert (2010), PRIM seeks to discover a set of box constraints $B = \{\mathbf{x} \mid a \leq x_j \leq b, j \in L\}$, where $L \subseteq \{1, \ldots, k\}$, to best describe the occurrence of cases of interest $\mathcal{I}$. Uncertain factors $x_{j \notin L}$ are not constrained by B.

Defining cases of interest allows PRIM to consider a binary simulation output: $y_i' = 1$ if $\mathbf{x}_i \in \mathcal{I}$ and $y_i' = 0$ otherwise. A good candidate box B must serve several purposes. First, it must contain a high proportion of cases of interest, i.e., it should maximize *coverage*:

$$C = \sum_{\mathbf{x}_i \in B} y_i' \Big/ \sum_{\mathbf{x}_i \in \mathcal{I}} 1 \tag{2.11}$$

Additionally, the states of the world $\mathbf{x}_i$ captured by the box must be primarily cases of interest, i.e., it should maximize *density*:

$$\mathcal{D} = \sum_{\mathbf{x}_i \in B} y_i' \Big/ \sum_{\mathbf{x}_i \in B} 1 \tag{2.12}$$

Coverage and density are conflicting metrics. For example, wide constraints will cover nearly all of the cases of interest, but will also contain many other cases, leading to a poor density. Similarly, a set of narrow constraints could contain exclusively cases of interest, but would fail to capture many of them, leading to a poor coverage metric.

PRIM works by "peeling" away layers of the uncertainty space by constraining one dimension at a time, based on the constraint that will maximize the mean value of y' for $\mathbf{x} \in B_1$. Once the candidate box B_1 is defined, the points $\{(\mathbf{x}, y') \mid \mathbf{x} \in B_1\}$ are removed from the dataset and PRIM continues to define a new box B_2 using the same approach. Friedman and Fisher (1999) refer to the set of candidate boxes as the "peeling trajectory". The coverage and density are calculated for each box, and the conflict between these two metrics is addressed by the user interactively selecting a candidate box B from the coverage-density tradeoff. The result is a set of box constraints on the uncertain factors determined to be influential; factors with little or no influence are ignored. PRIM has been adopted as a key component of the Robust Decision Making framework (Lempert, 2002; Lempert et al., 2006) given its ability to identify multivariate vulnerabilities even when the model mapping inputs to outcomes is nonlinear or otherwise challenging. In this work, we employ an open-source implementation of PRIM written in the R language, which is distributed with the Scenario Discovery Toolkit (Bryant, 2009).

2.2 Multi-Objective Evolutionary Optimization

Multi-objective Evolutionary Algorithms (MOEAs) belong to a broader class of population-based metaheuristics which employ structured randomness to explore the search space and locate near-optimal solutions (Maier et al., 2014). They draw inspiration from biological evolution, including the concepts of mutation, crossover, and selection based on "fitness" (i.e., performance in one or more objective values) (Nicklow et al., 2010). These concepts are implemented using one or more operators, such as polynomial mutation or simulated binary crossover, to create potentially promising new solutions from existing ones. The algorithms are, in general, designed to solve multi-objective (vector) optimization problems of the following type, with D objectives, M inequality constraints, and N equality constraints:

$$\begin{aligned} \min_{\mathbf{x}\in\Omega} \quad & F(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots f_D(\mathbf{x})] \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \forall i \in (1, .., M) \\ & h_j(\mathbf{x}) = 0, \forall j \in (1, .., N) \end{aligned}$$

where $f_i(\mathbf{x})$ are the objective functions and $\mathbf{x}$ are the decision variables. Importantly, problems of this type do not result in a single optimal solution, but rather a set of solutions that are identified as "Pareto-optimal" or "non-dominated", as will be described in Section 2.2.1. Initial efforts to solve problems of this type relied on repeated solving of weighted linear programming problems by iteratively adjusting the weights (Cohon and Marks, 1973, 1975), which required the problems to be linearly separable (i.e., where the objective function is a weighted sum of decision variables) and convex. Because classical optimization relies on the gradient of the function, it becomes untenable when faced with unknown analytical derivatives, nonsmooth functions, and/or a high-dimensional

search space, forcing the analyst to simplify the problem to accommodate the solution method rather than vice versa. Particularly in water resources systems, it cannot be assumed that an optimization problem will be linear, convex, or continuous since the field deals with stochastic systems, discrete decisions, and complex relationships between decisions and outcomes (Reed et al., 2013). For this reason, evolutionary optimization has become increasingly popular in the water resources field since the initial applications developed by, for example, Goldberg (1989), Dougherty and Marryott (1991), and McKinney and Lin (1994). Potential drawbacks include the corresponding increase in computational power required for evolutionary algorithms to succeed relative to classical approaches, and the guarantee only of an approximately optimal solution rather than an exact one. However, evolutionary methods advocate solving the exact problem approximately rather than solving the approximate problem exactly (Maier et al., 2014). Nicklow et al. (2010) and Maier et al. (2014) provide reviews of single and multi-objective evolutionary algorithms in water resources. Coello Coello (2007) provides an extensive review of MOEA applications across many scientific and engineering fields. Finally, Reed et al. (2013) provide a comprehensive diagnostic assessment of the efficiency and reliability of modern MOEAs.

2.2.1 Pareto Dominance and ε-Dominance

MOEAs aim to find the Pareto-optimal set of solutions. For a D-objective minimization problem, an objective vector $F(\mathbf{x})$ dominates another vector $F(\mathbf{y})$, denoted $F(\mathbf{x}) \succ F(\mathbf{y})$, if and only if $f_i(\mathbf{x}) \leq f_i(\mathbf{y})\ \forall i \in (1, \ldots, D)$ and there exists a $j \in (1, \ldots, D)$ for which $f_j(\mathbf{x}) < f_j(\mathbf{y})$ (Reed et al., 2013). In other words, $F(\mathbf{x})$

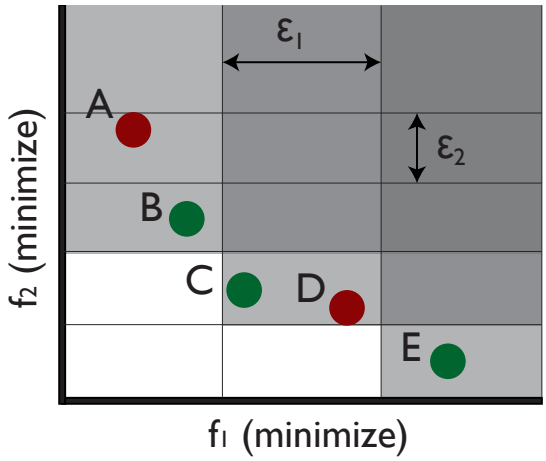

Figure 2.1: A two-objective minimization problem illustrating the concept of epsilon dominance, adapted from Kasprzyk (2013). The epsilon resolution for each objective is indicated. Only one solution is allowed per grid box, and other solutions are eliminated via intra-box sorting (e.g., solution C dominates solution D). Inter-box sorting is performed with respect to entire grid boxes (e.g., solution B box-dominates solution A). In this example, the non-dominated members of the set are B, C, and E (green), while solutions A and D (red) are dominated.

performs at least as well as $F(\mathbf{y})$ in all objectives, so the decision vector $\mathbf{y}$ can be removed from consideration regardless of the decision maker's preference. The Pareto-optimal set of objective vectors contains the solutions which are not dominated by any other solution. In practice, it is not possible to enumerate all possible solutions in order to identify those that are Pareto-optimal. Instead, for engineering applications it is common to use MOEAs to identify the best-known approximation to the Pareto set (a set of vectors termed "Pareto-approximate") which are not dominated by any solution found during search.

Desirable properties of MOEAs include convergence to the true Pareto-optimal set and diversity across the set. The process of ε-dominance (Laumanns et al., 2002) guarantees both properties provided sufficient search time, and also provides a theoretical upper bound for the size of the Pareto-approximate set, a

valuable contribution from a decision making standpoint. ε-dominance requires a resolution (or tolerance) $\varepsilon_i > 0$ to be defined for all objectives $i \in (1, \ldots, D)$, which creates a grid in the objective space as shown in Figure 2.1. An objective vector $F(\mathbf{x})$ ε-dominates another vector $F(\mathbf{y})$, denoted $F(\mathbf{x}) \succ_\varepsilon F(\mathbf{y})$, if and only if $(1 - \varepsilon_i) f_i(\mathbf{x}) \leq f_i(\mathbf{y})$ $\forall i \in (1, \ldots, D)$ (Laumanns et al., 2002). The set of ε-nondominated vectors can be approximated numerically by allowing only one solution to survive in each grid box, then performing a Pareto sort on the grid boxes themselves (see Figure 2.1). The concept of ε-dominance limits the size of the Pareto-approximate set, improving the rate of convergence and also the interpretability of the results for decision making—most decision makers will be able to specify a significant precision for each objective within which small variations are considered inconsequential.

2.2.2 Borg MOEA

In practice, the crossover and mutation search operators used in MOEAs (introduced in Section 2.2) must be parameterized by the user. These may include the probability of crossover, the magnitude of allowable mutations, etc., where a suitable value is unknown *a priori*. For this reason, recent algorithm development has focused on adaptive operators which can alter the parameter values or the operators used during the course of the search (Vrugt and Robinson, 2007; Vrugt et al., 2009). The Borg MOEA (Hadka and Reed, 2013) extends this concept by combining multiple operators to search the solution space, selecting operators adaptively based on their demonstrated probability of improvement during the optimization. The search operators encompassed by Borg include: simulated binary crossover (SBX; Deb and Agrawal, 1994), dif-

ferential evolution (DE; Storn and Price, 1997), parent-centric crossover (PCX; Deb et al., 2002), unimodal normal distribution crossover (UNDX; Kita et al., 1999), simplex crossover (SPX; Tsutsui et al., 1999), polynomial mutation (PM), and uniform mutation (UM) applied to each of D objectives with probability $1/D$ (Hadka and Reed, 2013; Reed et al., 2013). Example distributions of the new solutions produced by these operators are shown in Figure 2.2 (Hadka and Reed, 2013). Borg also assimilates several recent advances in the evolutionary algorithms literature, including epsilon-dominance archiving (Laumanns et al., 2002), adaptive population sizing (Kollat and Reed, 2007a), and a steady-state algorithm structure (Deb et al., 2005), which refers to the replacement of only one solution in the population per iteration (in contrast to generational algorithms, which replace the entire population). Borg has shown superior performance on challenging nonlinear, non-convex, multimodal problems (Hadka and Reed, 2012; Hadka et al., 2012; Reed et al., 2013). In this work, the Borg MOEA is used to identify multi-objective tradeoffs in water resources problems.

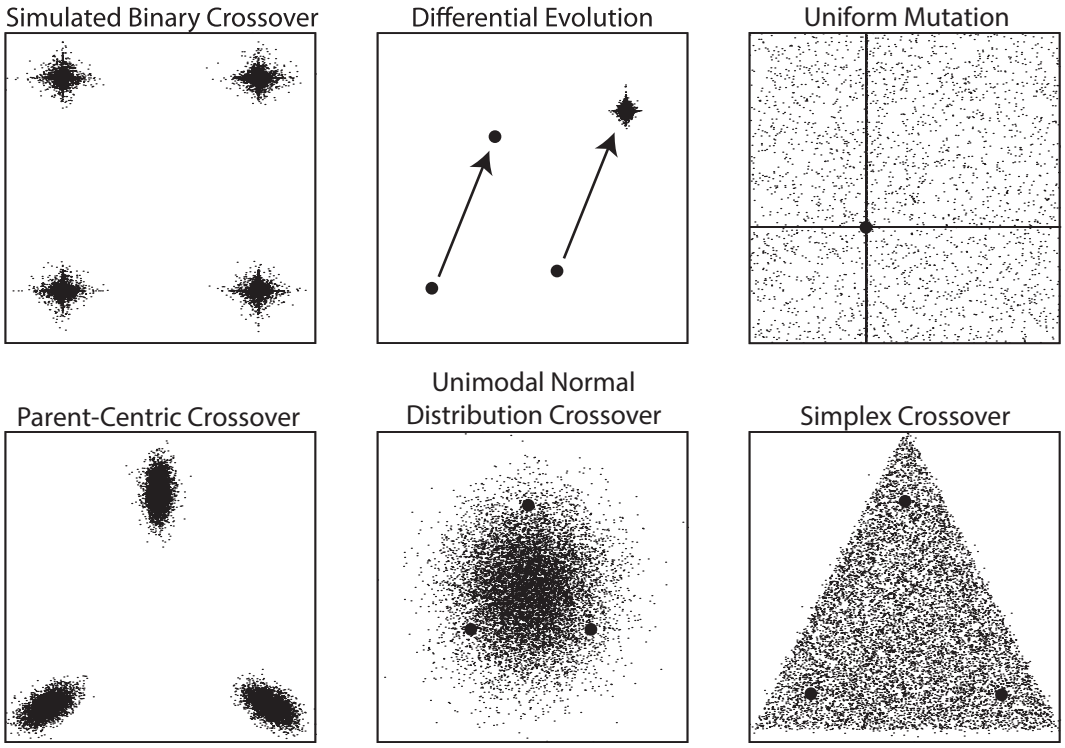

Figure 2.2: Example distributions of new candidate solutions produced by the recombination operators of the Borg algorithm, in a hypothetical two-dimensional space of decision variables. The initial "parent" solutions are indicated by larger circles. From Hadka and Reed (2013).

CHAPTER 3

EFFICIENT SENSITIVITY ANALYSIS METHODOLOGY FOR SPATIALLY DISTRIBUTED FLOOD FORECASTING

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

Herman, J.D., Kollat, J.B., Reed, P.M., and Wagener, T. (2013). Technical Note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. Hydrol. Earth Syst. Sci., 17(7), 2893-2903.

This work was partially supported by the U.S. National Science Foundation under grants EAR-0838357 and OCI-0821527. The opinions, findings, and conclusions are solely those of the authors.

3.1 Abstract

The increase in spatially distributed hydrologic modeling warrants a corresponding increase in diagnostic methods capable of analyzing complex models with large numbers of parameters. Sobol sensitivity analysis has proven to be a valuable tool for diagnostic analyses of hydrologic models. However, for many spatially distributed models, the Sobol method requires a prohibitive number of model evaluations to reliably decompose output variance across the full set of parameters. We investigate the potential of the method of Morris, a screening-based sensitivity approach, to provide results sufficiently similar to those of the Sobol method at a greatly reduced computational expense. The methods are benchmarked on the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) model over a six-month period in the Blue River Watershed, Oklahoma, USA. The Sobol method required over six million model eval-

uations to ensure reliable sensitivity indices, corresponding to more than 30 000 computing hours and roughly 180 gigabytes of storage space. We find that the method of Morris correctly identifies sensitive and insensitive parameters with 300 times fewer model evaluations, requiring only 100 computing hours and 1 gigabyte of storage space. Method of Morris proves to be a promising diagnostic approach for global sensitivity analysis of highly parameterized, spatially distributed hydrologic models.

3.2 Introduction: Diagnostics for Spatially Distributed Watershed Models

Distributed hydrologic models aim to improve simulations of watershed behavior by allowing forcing data and model parameters to vary across a spatial grid. Recent advances in hydrologic data collection and computing power have increased the appeal of distributed models while also allowing further increases in complexity (Smith et al., 2004, 2012). This added complexity is not without cost; a typical distributed model usually contains thousands more parameters than a lumped model, causing a commensurate leap in computational requirements as well as challenges in diagnosing model behavior (van Griensven et al., 2006; Gupta et al., 2008). Calibration of such highly parameterized models remains difficult, not only due to the computation involved, but also because of their highly interactive parameter spaces and nonlinear, multimodal objective spaces (Gupta et al., 1998; Carpenter et al., 2001). To address these challenges, this study explores diagnostic methods capable of characterizing the complex relationships between distributed model parameters and objectives efficiently

and accurately.

Sensitivity analysis has long been used to derive diagnostic insight from hydrologic models by identifying the input factors controlling model performance (Hornberger and Spear, 1981; Franchini et al., 1996; Freer et al., 1996; Wagener et al., 2001; Muleta and Nicklow, 2005; Sieber and Uhlenbrook, 2005; Bastidas et al., 2006; Demaria et al., 2007; Cloke et al., 2008; Van Werkhoven et al., 2008a, 2009; Wagener et al., 2009a; Reusser et al., 2011; Reusser and Zehe, 2011; Herman et al., 2013c). The most common applications of sensitivity analysis include factor fixing, in which insensitive inputs are assigned fixed values to simplify further analysis; factor prioritization, in which the most sensitive inputs are identified; and factor mapping, which identifies the regions of the input space in which a particular input is most sensitive (Saltelli et al., 2008). A number of input factors can be explored in sensitivity analysis, including forcing variables, but in diagnostic applications it is common to analyze model parameters directly. In this study, we aim to analyze the ranking of sensitive model parameters (*i.e.*, both those that are sensitive and insensitive) as well as to compare their quantitative measures of sensitivity.

Sensitivity methods can be broadly divided into local methods and global methods. Local methods provide measures of importance around a single point in the parameter space. Global methods aim to reflect the importance of a parameter throughout the full multivariate space of a model. Relatively few studies have performed global sensitivity analysis for spatially distributed models due to the severe computational demands posed by sampling their high-dimensional parameter spaces. Distributed sensitivity studies in hydrology and land surface modeling have often addressed this problem by aggregating pa-

rameter values across the model grid or subgrids (e.g., Carpenter et al., 2001; Hall et al., 2005; Sieber and Uhlenbrook, 2005; Zaehle et al., 2005; Alton et al., 2006). Fewer still are studies which have performed sensitivity analysis on a full set of spatially distributed parameters (e.g., Muleta and Nicklow, 2005; van Griensven et al., 2006; Tang et al., 2007a; Van Werkhoven et al., 2008b). These studies clearly show the benefits of performing global sensitivity analysis on a distributed model without sacrificing resolution in the parameter space. This study hypothesizes that the need for such sacrifices (i.e., to reduce computational demands) can be reduced with a careful choice of the sensitivity analysis method.

This study compares the efficiency and effectiveness of two state-of-the-art global sensitivity analysis methods, Sobol sensitivity analysis (Sobol, 2001; Saltelli, 2002) and the method of Morris (1991). Sobol sensitivity analysis is a variance-based method that attributes variance in the model output to individual parameters and their interactions. In a comparison of several widely-used sensitivity methods, the Sobol method was found to provide the most accurate and robust sensitivity indices, particularly in nonlinear models with strong parameter interactions (Tang et al., 2007b; Yang, 2011). However, the number of model evaluations required by the Sobol method increases rapidly with the number of parameters, making its efficiency questionable in the distributed case. The method of Morris (1991) measures global sensitivity using a set of local derivative approximations (elementary effects) taken at points sampled throughout the parameter space. The method of Morris can estimate parameter interactions by considering both the mean and variance of the elementary effects. Full descriptions of the Sobol and Morris method are given in Sections 2.1.1 and 2.1.2, respectively.

We implement the two sensitivity analysis methods for the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) (Koren et al., 2004; Reed et al., 2004; Smith et al., 2004; Moreda et al., 2006), developed by the United States National Weather Service (NWS). The model is used to simulate the Blue River Watershed, Oklahoma, USA, over a six-month period using hourly timesteps and forcing data. Sensitivity results from the Sobol and Morris methods are compared spatially and statistically to determine the extent to which the method of Morris provides computational savings while maintaining sensitivity indices sufficiently similar to those of the Sobol method. In turn, we investigate whether the method of Morris is a promising candidate to overcome the challenges to diagnostic analysis posed by the high-dimensional parameter spaces of distributed hydrologic models.

3.3 Model and study area

3.3.1 HL-RDHM model

The HL-RDHM, developed by the United States NWS, is a modeling framework for building lumped, semi-distributed, and fully distributed hydrologic models (Koren et al., 2004; Reed et al., 2004; Smith et al., 2004; Moreda et al., 2006). The model is structured using a 4 km $\times$ 4 km grid resolution derived from the Hydrologic Rainfall Analysis Project (HRAP), which corresponds to the NEXRAD precipitation products developed by the US NWS. The water balance in each grid cell is modeled with the Sacramento Soil Moisture Accounting (SAC-SMA) model (Burnash and Singh, 1995). Figure 3.1c shows the water balance com-

ponents of the SAC-SMA model in each grid cell, including impervious area parameters (*PCTIM* and *ADIMP*), the upper and lower storage zones (*UZ*– and *LZ*–), and the percolation functions connecting the upper and lower zones (Z_{Perc}, R_{Exp}, and P_{Free}).. Routing between grid cells is modeled with a kinematic wave approximation to the St. Venant equations. This study performs sensitivity analysis on 14 parameters of the SAC-SMA model in each cell of the HRAP grid as shown in Fig. 3.1c. Since the model contains 78 grid cells, a total of $78 \times 14 = 1092$ parameters are required to perform sensitivity analysis without spatial aggregation. The sampling ranges for these parameters are derived from prior work (Van Werkhoven et al., 2008b) and in consultation with the National Weather Service. Note that the choice of plausible sampling ranges is critical to ensure representative model performance (Sobol, 2001; Nossent and Bauwens, 2012a).

3.3.2 Study area: Blue River, Oklahoma

The computational experiments in this study were performed for the Blue River Basin in southern Oklahoma, USA, one of the basins included in the Distributed Model Inter-comparison Project Phase 2 (DMIP2) (Smith et al., 2012). Figure 3.1a shows the location of the Blue River. The watershed is represented by 78 HRAP grid cells, as shown in Fig. 3.1b, resulting in a total basin area of 1248 km^2. The model was forced using hourly NEXRAD precipitation data over the 6 month period from 16 November 2000 to 15 May 2001, preceded by a 3 week warmup period. Figure 3.2 shows the hourly precipitation and streamflow data for the Blue River during the selected simulation period. As Fig. 3.2 indicates, the Blue River remains at low flow during much of the simulation period, punctuated by a series of large rainfall events.

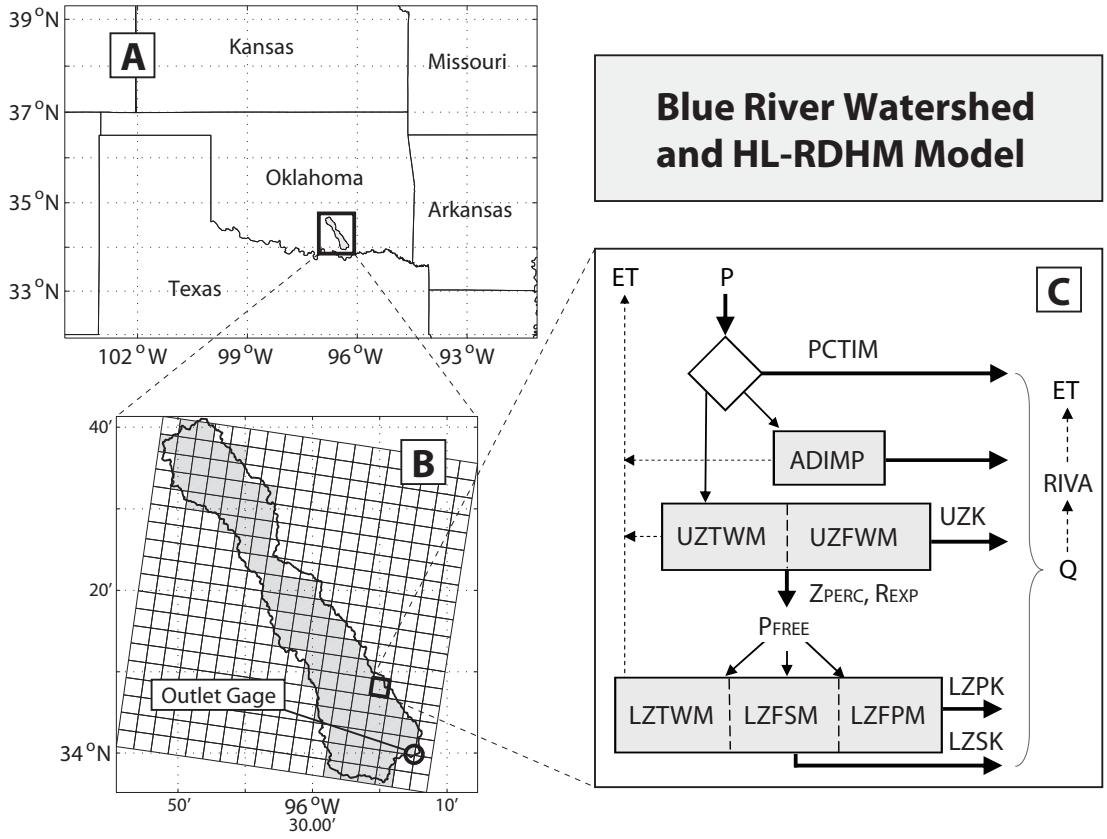

Figure 3.1: **(A)** Location of the Blue River Basin in southern Oklahoma, USA. **(B)** The 78 HRAP grid cells of the Blue River Basin (shaded). **(C)** The Sacramento Soil Moisture Accounting (SAC-SMA) model, which simulates the water balance in each grid cell.

3.4 Computational Experiment

The sensitivity analyses were performed on the 14 SAC-SMA model parameters as indicated in Fig. 3.1. The lower and upper bounds for each parameter are based on the a priori gridded parameter values derived by the NWS (Koren et al., 2004) and extended for sensitivity analysis by (Van Werkhoven et al., 2008b). These parameter ranges are included in Table 3.1. Parameter values for each grid cell were sampled separately from uniform distributions. Rather than

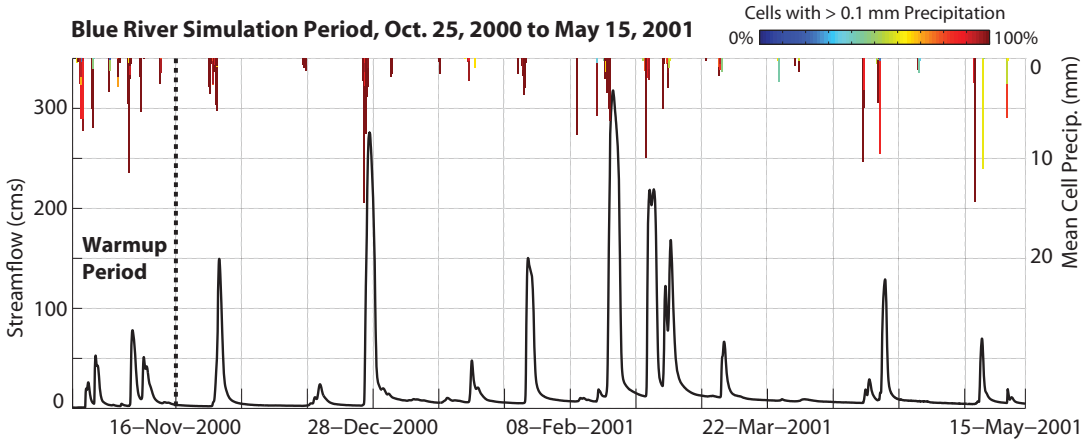

Figure 3.2: The hourly hydrograph of the 6 month simulation period for the Blue River Basin, with a 3 week warmup period. The precipitation amounts are based on the mean value across the 78 HRAP grid cells in the basin. The colors of the precipitation bars indicate the fraction of grid cells receiving more than 0.1 mm precipitation, representing the spatial distribution of each hourly rainfall value.

measure the sensitivity of the output streamflow directly, we measure the sensitivity of the root mean squared error (RMSE) metric, calculated using the known hourly streamflow values over the 6 month simulation period. This ensures that our sensitivity indices are grounded relative to the observed streamflow and describe the controls on model performance.

The sample sizes and corresponding number of model evaluations required for both the Sobol and Morris methods are shown in Table 3.2. For the Sobol method, sample sizes of $N = 1000$ and $N = 6000$ were used, resulting in just over 1 million and 6 million model evaluations, respectively. The latter value represents the limit of computational feasibility for this model at an hourly timestep, to derive maximally accurate baseline values of the sensitivity indices. We chose the two sample sizes to verify convergence of the Sobol indices. Confidence in-

Table 3.1: HL-RDHM parameters and their uniform sampling ranges for sensitivity analysis.

Parameter	Description	Units	Lower Bound	Upper Bound
LZFPM	Lower zone primary maximum storage	mm	8.8	58.8
LZFSM	Lower zone supplemental maximum storage	mm	19.2	193.2
LZPK	Lower zone primary withdrawal rate	day^{-1}	0.0408	0.264
LZSK	Lower zone supplemental withdrawal rate	day^{-1}	0.00168	0.0175
LZTWM	Lower zone tension water maximum storage	mm	61.6	249.6
PCTIM	Permanent impervious area	%	0.0	5.0
PFREE	Percolation to lower zone	%	16.0	55.0
REXP	Percolation equation exponent	–	1.69	3.47
UZFWM	Upper zone free water maximum storage	mm	8.8	64.8
UZK	Upper zone free water withdrawal rate	day^{-1}	0.19	0.76
UZTWM	Upper zone tension water maximum storage	mm	19.2	78
ZPERC	Maximum percolation rate under dry conditions	mm/day	27.2	140.4
ADIMP	Saturated impervious area	%	0.0	20.0
RIVA	Riparian vegetation area	%	0.0	20.0

Table 3.2: Sample sizes and number of model runs performed for each of the sensitivity analysis methods.

Method	Sample size	Model evaluations
Sobol'	1000	1 094 000
	6000	6 564 000
Morris	20	21 840
	40	43 680
	60	65 520
	80	87 360
	100	109 200

tervals for the sensitivity indices derived from the bootstrap method (Archer et al., 1997) were monitored to ensure convergence of the Sobol method at the $N = 6000$ level. We considered convergence acceptable if the 95% confidence interval represented less than 10% of the sensitivity index value for the most sensitive parameters. For the method of Morris, sample sizes ranging from $N = 20$ to $N = 100$ were chosen to determine if the approach can provide suitable results with orders of magnitude fewer model evaluations. The sensitivity analyses were performed using the CyberSTAR high-performance cluster at Penn State University, which contains a combination quad-core AMD Shanghai processors (2.7 Ghz) and Intel Nehalem processors (2.66 Ghz). Approximately 50 000 computing hours were required to complete the experiment.

3.5 Results and discussion

The results of the sensitivity analyses can be addressed through the lens of two primary questions: (1) what is the sample size required for the Sobol method to return reliable sensitivity indices; and (2) how suitable are the indices returned by the method of Morris relative to the baseline created by the Sobol method.

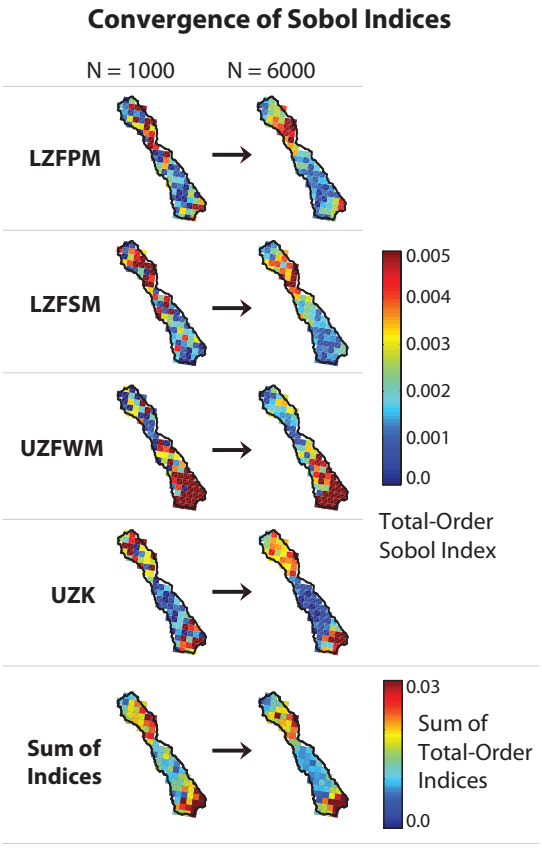

Figure 3.3: Maps of the total-order Sobol sensitivity indices for the four most sensitive parameters as well as the total sum in each grid cell. The maps are shown for the $N = 1000$ and $N = 6000$ sample sizes. The lower sample size shows a coarse identification of sensitive and insensitive cells. The $N = 6000$ sample size shows smoother spatial patterns of sensitivity indices, suggesting that this level of sampling is required for reliable Sobol indices.

3.5.1 Convergence of Sobol indices

Figure 3.3 shows the spatial maps of total-order Sobol sensitivity indices for the sample sizes $N = 1000$ and $N = 6000$. The four most sensitive parameters of the SAC-SMA model are shown, as well as the cell-level sum of sensitivity indices. The total-order indices vary over a small range since the output variance must be distributed across the full set of distributed parameters, 1092 in total.

Figure 3.3 shows several interesting spatial patterns of sensitivity. First, the most sensitive parameters are primarily upper and lower storage zone maxima. The lower-zone storage maxima, LZFPM and LZFSM, are most sensitive in the headwater portion of the basin, while the upper-zone storage maximum UZFWM is most sensitive toward the outlet of the basin. The resulting summation of sensitivity indices shows a division of the most active cells, with one group in the headwaters and another near the outlet.

From Fig. 3.2, it is clear that most precipitation events during the simulation period are distributed across nearly all grid cells in the watershed. This suggests that much of the spatial variability of sensitivity in Fig. 3.3 is due to processes within the model itself rather than forcing patterns. The RMSE metric is most sensitive to errors in peak flows, so the sensitivity indices in Fig. 3.3 can be interpreted in the context of the several high-flow events shown in the hydrograph in Fig. 3.2. Toward the outlet of the basin, the primary runoff-generating mechanisms in the model are overflow from the upper-zone free water storage (UZFWM) and flow out of the upper zone (controlled by UZK). The fact that the lower zone rate constants, LZPK and LZSK, are not sensitive indicates that they act on a slower timescale and thus do not affect RMSE. In the headwaters, the lower zone storage maxima (LZFPM and LZFSM) and the rate constant UZK

are most sensitive, likely because these parameters must not allow too much direct runoff from the headwater region to prevent the model from overshooting the observed flow peaks and causing poor RMSE performance. While the temporal distribution of forcing can affect the sensitivity indices shown in Fig. 3.3, the spatial distribution can be restricted to the processes occurring within the model.

Also visible in Fig. 3.3 is the difference in Sobol sensitivity indices as a function of sample size. At a sample size of $N = 1000$, the most sensitive cells are identified, but it is clear that cells with intermediate sensitivity values largely remain unidentified. For example, it is common to see sensitive cells (red) adjacent to insensitive cells. Intuitively, we should expect to see a smoother spatial gradient of sensitivity in which the most sensitive cells are adjacent to intermediate-sensitivity cells, which in turn are adjacent to low-sensitivity cells. This is achieved to a larger extent with a sample size of $N = 6000$. Here, the sensitivity indices vary more smoothly in space, indicating that the $N = 6000$ case provides a baseline for total-order sensitivity indices. The bootstrap confidence intervals confirm convergence for the $N = 6000$ sample size. The $N = 1000$ case would not be sufficient to capture the full range of sensitivity, a fact which underscores the high computational requirements of the Sobol method. It is worth noting that the slow convergence of the Sobol method for this model is related to the large number of parameters over which variance must be decomposed, leading to small sensitivity values and a correspondingly narrow range of acceptable confidence bounds (Nossent et al., 2011).

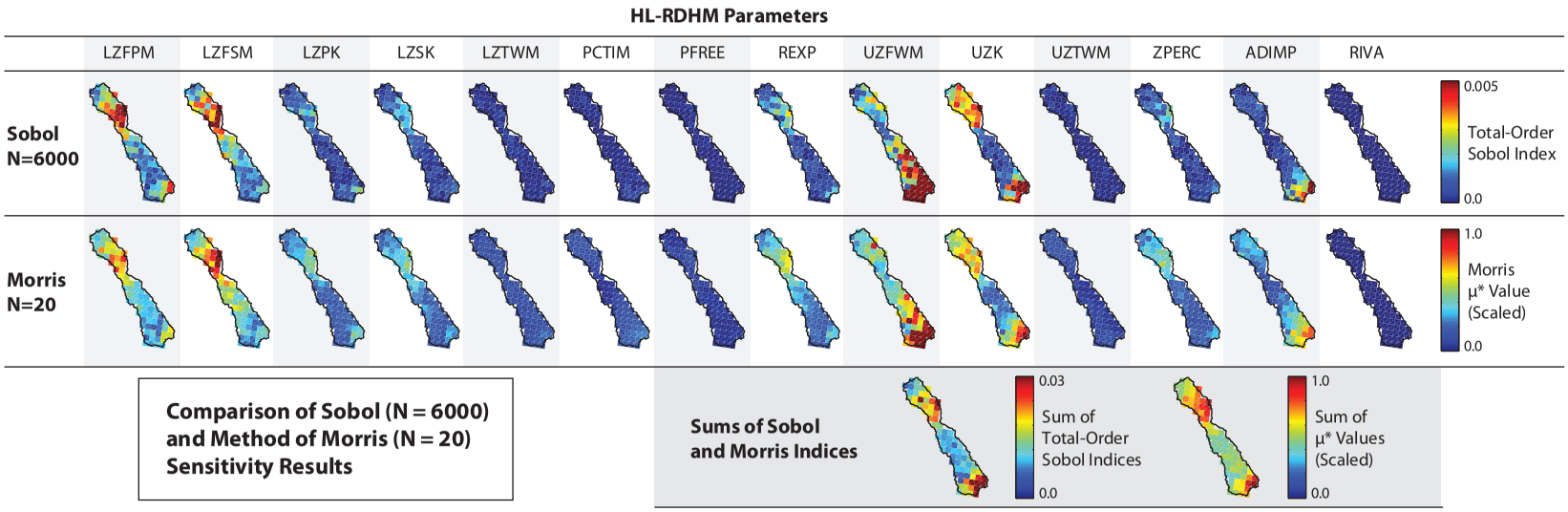

Figure 3.4: Total-order sensitivity indices calculated by the Sobol method, with sample size $N = 6000$, and the method of Morris, with $N = 20$. The Morris indices are normalized to the range $[0, 1]$ since they do not offer a quantitative interpretation of percent variance. The method of Morris is able to correctly identify sensitive and insensitive parameters, as well as their spatial patterns, with far fewer model evaluations. The sums of sensitivity indices represent the addition of parameter indices within each grid cell.

3.5.2 Comparison of Sobol and Morris indices

The Sobol sensitivity indices from the $N = 6000$ case form a set of target values against which the method of Morris will be compared. Figure 3.4 compares this target to the lowest-sample Morris experiment, $N = 20$, for all 14 of the SAC-SMA parameters and the sums of parameter indices for each of the 78 grid cells. The Sobol indices offer a quantitative interpretation as a fraction of total variance, but the Morris indices do not; the latter are normalized to avoid this misinterpretation.

Figure 3.4 shows that the total-order indices calculated by the method of Morris with only $N = 20$ samples successfully capture the spatial patterns of the Sobol indices with $N = 6000$ samples. The Morris indices are able to isolate the most sensitive parameters, along with their correct locations in the watershed: LZFPM, LZFSM, and UZK in the headwaters, and UZFWM, UZK, and ADIMP near the outlet. It also correctly identifies the parameters that are insensitive over the simulation period: LZTWM, PCTIM, PFREE, UZTWM, and RIVA. The sums of indices are comparable between the Sobol and Morris methods, as well, with sensitive areas near the headwaters and outlet, and intermediate sums of sensitivity in the rest of the basin. In general, the Morris indices follow smooth spatial patterns, which aligns with intuition regarding sensitive regions of the watershed. From the sensitivity maps in Fig. 3.4, The method of Morris with a sample size of $N = 20$ is able to correctly identify sensitive and insensitive parameters, as well as their spatial patterns, at greatly reduced computational expense relative to the Sobol method.

The Morris sensitivity indices can also be compared statistically to the Sobol indices for the $N = 6000$ case to ensure sufficient similarity. Figure 3.5 compares

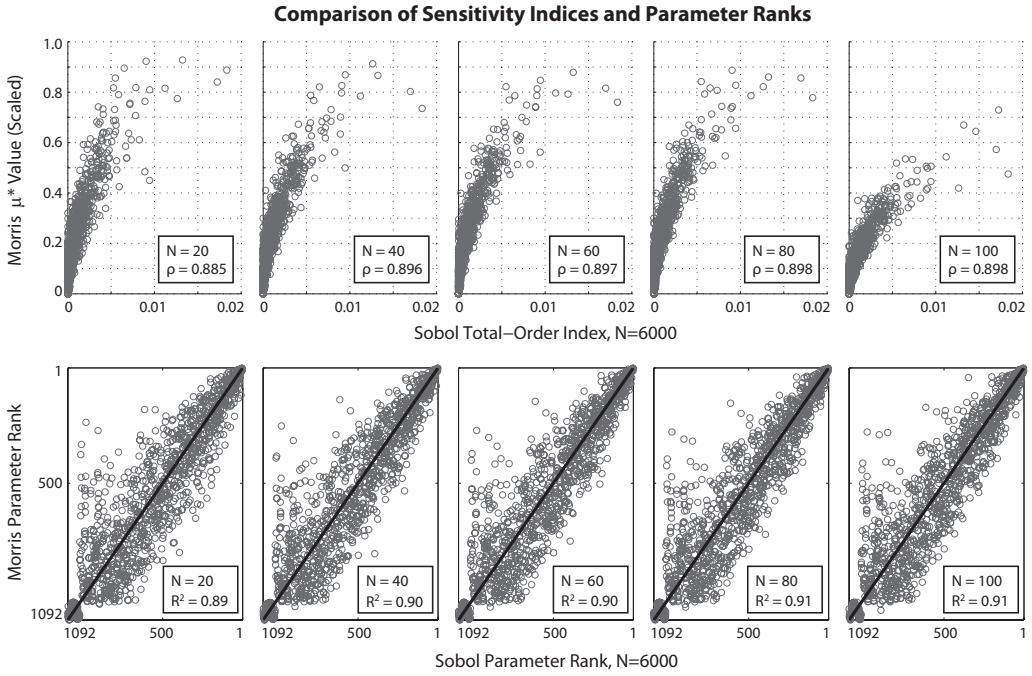

Figure 3.5: Statistical comparison of sensitivity indices and sensitivity ranks (1–1092) between the Sobol method ($N = 6000$) and the method of Morris with sample sizes from $N = 20$ to $N = 100$. The sensitivity indices are compared using a nonlinear Spearman correlation coefficient (ρ), while the rankings are compared with a linear correlation coefficient (R^2). Each plot contains all 14 parameters from each grid cell, for a total of 1092 points.

the sensitivity indices for each method, as well as the sensitivity ranks (1–1092), for all of the Morris sample sizes from $N = 20$ to $N = 100$. The sensitivity indices are compared using a nonlinear Spearman correlation coefficient, because a one-to-one correspondence between Sobol and Morris indices is not necessary. The rankings are compared with a linear correlation coefficient, because these ideally will exhibit a one-to-one correspondence.

The top row of Fig. 3.5 shows that the Morris μ^* values for all sample sizes are well-correlated with the Sobol indices with a sample size of $N = 6000$. Im-

portantly, there appears to be little benefit in running the method of Morris for sample sizes greater than $N = 20$, since the correlation remains similar for higher sample sizes. The relationship between Morris μ^* values and Sobol indices is approximately linear for low-sensitivity parameters. However, the relationship becomes nonlinear for high-sensitivity parameters, where the Morris μ^* values appear to flatten out. This suggests that the Sobol method can better distinguish between the most sensitive parameters, whereas the method of Morris cannot. Still, the method of Morris successfully distinguishes sensitive from insensitive parameters, and a sample size of $N = 20$ is clearly sufficient to achieve this.

The bottom row of Fig. 3.5 shows that the sensitivity rankings given by the method of Morris are well-correlated with those given by the Sobol method with $N = 6000$. Again, a sample size of $N = 20$ for the method of Morris appears sufficient to achieve a good correlation, and little is gained by increasing the sample size further. Of particular interest are the clusters of highly correlated parameters ranked near the most and least sensitive (ranks 1 and 1092, respectively). This indicates that the method of Morris can isolate the most and least sensitive parameters with high reliability, reinforcing its utility as a screening method. The outliers in the bottom row of Fig. 3.5 reinforce the difficulty for the method of Morris to distinguish between sensitive parameters; it correctly identifies them as sensitive, but struggles to rank them quantitatively. The largest outliers occur in the upper-left of each plot, where the method of Morris attributes erroneously high rankings to certain parameters. These outliers correspond to parameters of average rank, whose low (but non-zero) sensitivity values are extremely difficult to differentiate from one another. Thus, these points highlight the limitations of the method of Morris for this model, but they do not detract from the success of the approach in correctly classifying the parameters with the

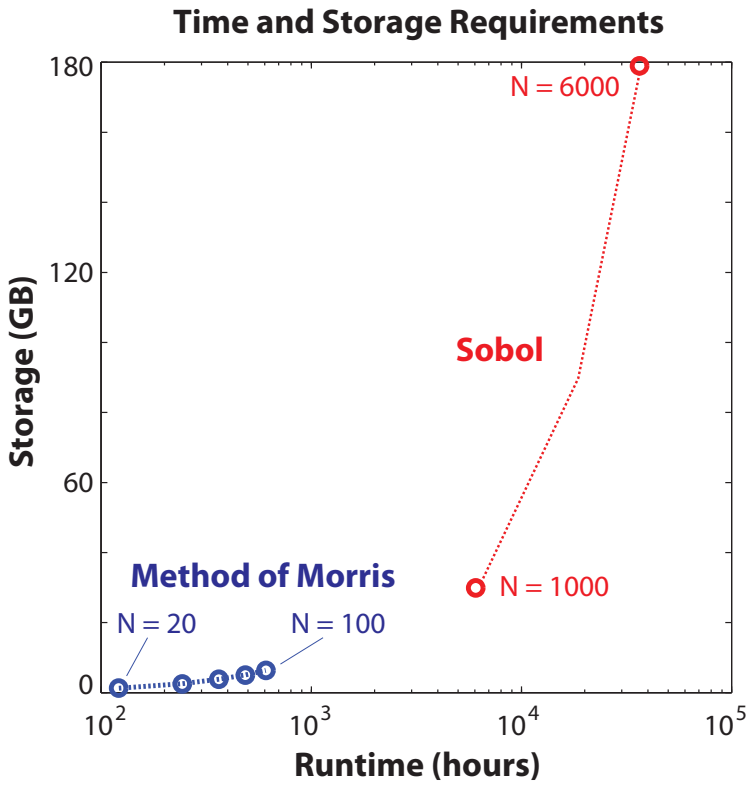

Figure 3.6: Computation time (hours) and storage (gigabytes) required for each experiment. The method of Morris with $N = 20$ represents a factor of 300 computational savings compared to the Sobol method with $N = 6000$.

highest and lowest sensitivity values.

Given that both the spatial and statistical comparisons between the Sobol and Morris sensitivity indices indicate the success of the method of Morris, it is worth exploring the amount of computation saved to achieve a highly similar set of sensitivity results. Figure 3.6 shows the location of each experiment in the space defined by the computation time and storage required. The largest Sobol experiment, with $N = 6000$, required over 6 million model evaluations, leading to more than 30 000 hours of computation time and approximately 180 gigabytes of storage space to store the model output. By contrast, the small-

est Morris experiment, with $N = 20$, required roughly 100 hours of computation and 1 gigabytes of storage space. This represents a factor of 300 savings in both the runtime and storage dimensions relative to the Sobol method. As shown in Figs. 3.4 and 3.5, the sensitivity indices calculated by this lowest-sample Morris experiment are spatially and statistically comparable to those calculated by the highest-sample Sobol experiment, indicating that the method of Morris provides significant computational savings without significant degradation of solution quality.

3.6 Conclusions

The method of Morris is able to correctly identify sensitive and insensitive parameters for a highly parameterized, spatially distributed watershed model with 300 times fewer model evaluations than the Sobol method. Even for this complex model, the efficient factorial sampling scheme of the method of Morris is sufficient to isolate the controls on model performance without any prior assumptions on the form of the model output. The Sobol method confers several advantages, including the first order sensitivity indices, and a large ensemble of model evaluations to be used in an uncertainty analysis or likelihood-based optimization framework, which the method of Morris does not provide. However, for many distributed modeling applications, the Sobol method requires a prohibitive number of model evaluations. In light of these results, the method of Morris proves to be a promising way forward for efficent global sensitivity analysis of distributed models. It also holds promise as a screening technique, identifying parameters to remove prior to more complex analyses such as the Sobol method or model calibration. Future work will include an investigation

of time-varying sensitivity to determine the extent to which spatial sensitivity patterns change during wet and dry periods. The increasing use of spatially distributed hydrologic models requires that diagnostics such as these sensitivity analysis methods be evaluated not only in terms of their statistical effectiveness but also by their efficiency, to ensure that hydrologic modelers can obtain maximally reliable diagnostic insights at a reasonable computational cost.

CHAPTER 4

HIGH-RESOLUTION TIME-VARYING SENSITIVITY ANALYSIS FOR SPATIALLY DISTRIBUTED WATERSHED MODELS

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

Herman, J.D., Kollat, J.B., Reed, P.M., and Wagener, T. (2013). From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models. Hydrol. Earth Syst. Sci., 17(12), 5109-5125.

This work was partially supported by the U.S. National Science Foundation under grants EAR-0838357 and OCI-0821527. The opinions, findings, and conclusions are solely those of the authors.

4.1 Abstract

Distributed watershed models are now widely used in practice to simulate runoff responses at high spatial and temporal resolutions. Counter to this purpose, diagnostic analyses of distributed models currently aggregate performance measures in space and/or time and are thus disconnected from the models' operational and scientific goals. To address this disconnect, this study contributes a novel approach for computing and visualizing time-varying global sensitivity indices for spatially distributed model parameters. The high-resolution model diagnostics employ the method of Morris to identify evolving patterns in dominant model processes at sub-daily timescales over a six-month period. The method is demonstrated on the United States National Weather Service's Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) in the Blue River watershed, Oklahoma, USA. Three hydrologic events

are selected from within the six-month period to investigate the patterns in spatiotemporal sensitivities that emerge as a function of forcing patterns as well as wet-to-dry transitions. Events with similar magnitudes and durations exhibit significantly different performance controls in space and time, indicating that the diagnostic inferences drawn from representative events will be heavily biased by the *a priori* selection of those events. By contrast, this study demonstrates high-resolution time-varying sensitivity analysis, requiring no assumptions regarding representative events and allowing modelers to identify transitions between sets of dominant parameters or processes *a posteriori*. The proposed approach details the dynamics of parameter sensitivity in nearly continuous time, providing critical diagnostic insights into the underlying model processes driving predictions. Furthermore, the approach offers the potential to identify transition points between dominant parameters and processes in the absence of observations, such as under nonstationarity.

4.2 Introduction: Parameter uncertainty in time and space

Distributed rainfall-runoff models allow model parameters and forcing data to vary on a spatial grid, aiming to better represent the spatial variability of watershed processes at the cost of increasing model complexity. This added complexity poses several key challenges, most notably: (1) the difficulty of identifying appropriate parameter sets in a highly interactive, nonlinear, multimodal objective space (Gupta et al., 1998; Carpenter et al., 2001), and (2) the related difficulty of tracing the causes of desirable or undesirable model performance (*i.e.*, diagnosing model behavior) (van Griensven et al., 2006; Gupta et al., 2008; Reusser et al., 2009). Considering the widespread operational use of distributed water-

shed models, there remains a need for diagnostic methods capable of studying such models at their full spatial and temporal complexity by avoiding aggregation in either dimension to the extent permitted by computational constraints.

Sensitivity analysis is a foundational diagnostic approach in the hydrologic modeling literature, (e.g., Hornberger and Spear, 1981; Franchini et al., 1996; Freer et al., 1996; Wagener et al., 2001; Muleta and Nicklow, 2005; Sieber and Uhlenbrook, 2005; Bastidas et al., 2006; Demaria et al., 2007; Cloke et al., 2008; Van Werkhoven et al., 2008a, 2009; Wagener et al., 2009a; Reusser et al., 2011; Reusser and Zehe, 2011; Herman et al., 2013a,c)). However, very few studies have performed global sensitivity analysis for spatially distributed watershed models due to the computational demands posed by the high dimension of their parameter spaces. Sensitivity analyses of distributed hydrologic and land surface models have frequently addressed this problem by aggregating parameter values across the model grid or subgrids (e.g., Carpenter et al., 2001; Hall et al., 2005; Sieber and Uhlenbrook, 2005; Zaehle et al., 2005; Alton et al., 2006; Cuo et al., 2011; Guse et al., 2013). Few studies have performed global sensitivity analysis on a full set of spatially distributed parameters. The studies that do exist have been limited to event-scale analyses, which reported highly complex spatial sensitivities arising from the interplay between forcing heterogeneity, proximity to observations, and the timescale of model performance metrics explored (e.g., Muleta and Nicklow, 2005; van Griensven et al., 2006; Tang et al., 2007a; Van Werkhoven et al., 2008b; Yatheendradas et al., 2008). Although these studies suggest the potential for time-varying spatial sensitivity analyses, computational demands have limited their exploration of this issue.

More recent studies have explored time-varying sensitivities at predefined

intervals throughout the model simulation, revealing the dynamics of model controls under changing conditions (Wagener et al., 2003; Van Werkhoven et al., 2008a; Reusser and Zehe, 2011; Reusser et al., 2011; Herman et al., 2013c). This approach has largely been limited to lumped models, with the exception of (Reusser et al., 2011) which analyzed a spatially explicit model. The studies that have focused on event-scale spatial sensitivities (Tang et al., 2007a; Van Werkhoven et al., 2008b; Wagener et al., 2009b; Herman et al., 2013a) have proposed using observations to identify representative events for a watershed, a valid concept as long as such representative events exist. However, if the dynamics of a watershed cannot be accurately restricted to one of several event classifications, the *a priori* selection of representative events introduces diagnostic biases that fail to account for the full range of process variability. In this work, we aim to extend the event scale approach to explore the dynamic controls of a distributed watershed model at a finely resolved sub-daily timestep, as well as to advance methods capable of computing and visualizing the results of this analysis.

This study proposes high-resolution time-varying sensitivity analysis for a spatially distributed rainfall-runoff model, avoiding the biases introduced by representative event selection by identifying key transitions between dominant parameters and processes *a posteriori*.These parameters dominate the performance of the model at a particular time, distinct from the true dominant watershed processes independent of our modeling efforts. Our high-resolution global sensitivity analysis employs the method of Morris (1991), which has recently been shown to attain accurate spatially distributed sensitivities at substantially lower computational expense than Sobol variance decomposition over a temporally aggregated six-month time period (Herman et al., 2013a). The high-

resolution sensitivity analysis is applied to the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) (Koren et al., 2004; Reed et al., 2004; Smith et al., 2004; Moreda et al., 2006), developed by the United States National Weather Service (NWS). The model test case focuses on the Blue River Basin, Oklahoma, USA, over a six-month period using hourly timesteps and spatially gridded forcing data. The sensitivity of model performance metrics is computed for the full period, the event scale, and a high-resolution moving window with a 3-hour timestep to demonstrate the benefit of investigating the full dynamics of spatially distributed model controls. This approach represents a novel, computationally efficient contribution to identify the dynamics of dominant model drivers under changing hydrologic conditions for highly parameterized distributed watershed models.

4.3 Model & Study Area

The HL-RDHM is a distributed rainfall-runoff model with surface-connected grid cells, where the water balance in each grid cell is modeled with the Sacramento Soil Moisture Accounting (SAC-SMA) model. A full description of the HL-RDHM is given in Section 3.3.1. This study performs sensitivity analysis on the 14 parameters defined in Table 3.1 for the SAC-SMA model. These 14 parameters are allowed to vary independently within each of the 78 cells of the HRAP grid shown in Fig. 3.1c, yielding 1092 parameters in total for the diagnostic analysis.

Herman et al. (2013c) showed that time-varying parameter sensitivity can be linked to the underlying mechanisms of a model. Here, studying the formula-

tion of the SAC-SMA model allows the development of hypotheses regarding the expected parameter sensitivities, and how these might change in space and time. At each timestep, evaporation first occurs from the additional impervious store, both upper zone stores, and the lower zone tension store. In all cases, evaporation is proportional to the saturation level of the storage element. Next, direct runoff occurs from the impervious area, specified by *PCTIM*, and the additional impervious area due to saturation, specified by *ADIMP*. Precipitation not assigned to direct runoff enters the upper zone free water store. Gravity drainage occurs from the upper and lower zones according to the rate constants *UZK*, *LZPK*, and *LZSK*, and is linearly proportional to the amount of water in each respective store. Finally, runoff is also generated when the storage capacity of the upper zone (*UZFWM*) is exceeded. The same process occurs when all of the lower zone storage capacities are exceeded (*LZTWM*, *LZFSM*, and *LZFPM*), but otherwise excess from any of the lower zones will spill into another.

After the runoff generation mechanisms have occurred, each timestep of the model concludes with a redistribution of water between stores according to their saturation levels. First, any deficiencies in the upper and lower tension stores are filled by the free water in their respective zones. Next, percolation occurs from the upper zone free water store to the lower zone based on the saturation level of the lower zone. It is important to note that the lower zone controls percolation in the SAC-SMA model, unlike many other water balance models where percolation is equivalent to spillover from the upper zone. The amount of percolation varies with the parameters Z_{Perc}, the maximum percolation rate under dry conditions, and R_{Exp}, the unitless exponent of the percolation equation (Koren et al., 2004). Finally, the parameter P_{Free} determines the fraction of percolation that enters the primary and secondary free water stores in the lower

zone.

From this description of model mechanisms, we can hypothesize which parameters might be most sensitive in space and time. During and immediately after precipitation events, the parameters associated with quick responses should be most sensitive. This includes the impervious area parameters and the upper and lower zone storage maxima, which can cause direct runoff via overflow. We might expect these sensitive parameters to be spatially concentrated near the outlet of the watershed, since only this area will have sufficient time to contribute to streamflow while the event is occurring. Between precipitation events, the primary streamflow generation mechanism will be drainage from the storage zones, controlled by the rate constants *UZK*, *LZPK*, and *LZSK*; we would expect these to be most sensitive in the time following an event, and with a broader spatial distribution to reflect their slower response. As found in prior work (Herman et al., 2013c), the percolation parameters are unlikely to be highly sensitive at any time, for two reasons. First, the amount of percolation is controlled by the moisture deficiency in the lower zone, so the parameter *LZTWM* (for example) has more influence on the magnitude of percolation than do the percolation parameters themselves. Second, the percolation parameters do not contribute directly to streamflow, so their signature may be obscured by intermediate processes. In general, we expect the lower zone parameters to exhibit higher sensitivity over the course of the simulation than upper zone parameters, because the lower zone deficiencies are filled first during the redistribution routine. It is important to note that the spatiotemporal parameter sensitivities will depend on the metric chosen. For example, the sensitivity of the root mean squared error metric on a short timescale will emphasize transitions between quick-response processes, while a water balance error metric on

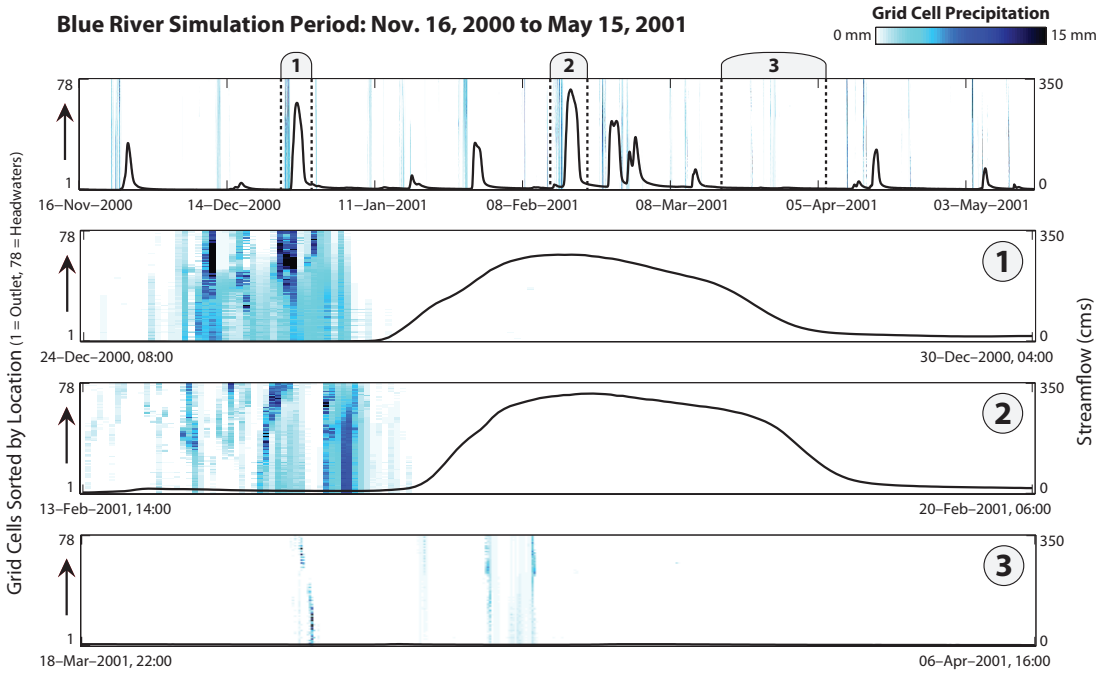

Figure 4.1: Hourly streamflow and precipitation values during the 6 month simulation period for the Blue River Basin. The *y*-axis on the left side of each plot represents abstracted spatial information, where the 78 grid cells of the basin model are sorted from the outlet cell (1) to the cell furthest from the outlet (78). The *y*-axis on the right side of each plot shows the magnitude of streamflow. The color corresponds to the amount of precipitation at each hourly timestep. Time periods **(1)**, **(2)**, and **(3)** are highlighted for further analysis, with **(1)** and **(2)** representing large events with different spatial distributions of precipitation, and **(3)** representing a low-flow period. Note that hours with high precipitation are more visible in time periods **(1)**-**(3)** than in the 6 month simulation period due to the reduced width of hourly intervals when plotting over the full period.

a longer timescale will capture the integrated effects of interacting states and fluxes.

This study focuses on the Blue River Basin in southern Oklahoma, USA, building on its inclusion in the Distributed Model Inter-comparison Project

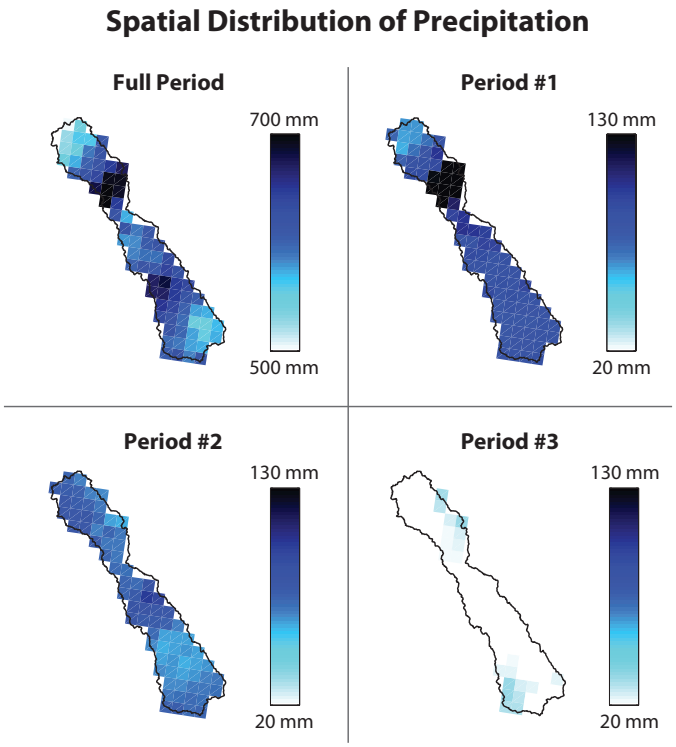

Figure 4.2: Maps of precipitation in the Blue River Basin over the full six-month simulation period and the three sub-periods defined in Figure 4.1. Over the full period, precipitation is roughly even across the watershed. The event during Period 1 is focused in the headwaters, while the event during Period 2 is more evenly spread. Finally, Period 3 represents a dry period with little streamflow.

Phase 2 (DMIP2) (Smith et al., 2012). Figure 3.1a shows the location of the Blue River in Oklahoma. The watershed is represented by 78 HRAP grid cells, as shown in Figure 3.1b, resulting in a total basin area of 1248 km^2. The model was forced using hourly NEXRAD precipitation data over the 6 month period from 16 November 2000 to 15 May 2001, preceded by a 3 week warmup period. Figure 4.1 shows the hourly precipitation and streamflow data for the Blue River during the selected simulation period. The selected period reflects a significant wet-to-dry transition during the 6-month period, increasing the efficacy of the model warmup. The time period selection was also influenced by the availabil-

ity of hourly NEXRAD data. The vertical axis of Fig. 4.1 contains the 78 HRAP grid cells of the watershed, arranged according to distance from the outlet cell. As Fig. 4.1 indicates, the Blue River basin experiences a series of large rainfall events early in the period before entering a long dry period in the late spring.

We begin by computing parameter sensitivity over the full simulation period. Then, in order to explore the potential consequences of event scale diagnostics, we select *a priori* three sub-periods to represent watershed dynamics. These are highlighted in Fig. 4.1 for further analysis: (1) a large rainfall event with the highest intensity precipitation focused in the headwaters; (2) a large rainfall event with similar cumulative precipitation but uniform intensity throughout the basin, and (3) a prolonged dry period with low flow. Figure 4.2 shows the spatial distribution of forcing for each of the three selected sub-periods. We utilize these three sub-periods to explore the relationship between parameter sensitivities over the full period and those derived for shorter events. We then advance this comparison by computing spatially distributed parameter sensitivities at a high-resolution moving window with a 3-hour timestep. In summary, the experiment consists of sensitivity analysis at three temporal resolutions: the full 6-month period, three representative sub-periods, and the high-resolution moving window. We seek to understand the similarities and differences in dominant model behavior at each of these temporal resolutions. In the absence of process-level watershed data, our diagnostic analysis focuses on the transitions between dominant modeled processes under changing hydrologic conditions.

4.4 Computational Experiment

The method of Morris (described in Section 2.1.2) was performed on the 14 SAC-SMA model parameters in each grid cell of the HL-RDHM model as indicated in Fig. 3.1. The uniform sampling bounds for each parameter given in Table 3.1 are based on the *a priori* gridded parameter values derived by the NWS (Koren et al., 2004) and extended for the event-scale sensitivity analysis performed by (Van Werkhoven et al., 2008b). Parameter values for each grid cell were sampled separately, resulting in a total of $78 \times 14 = 1092$ total sampled parameters. Rather than measure the sensitivity of the output streamflow directly, we measure the sensitivity of model performance metrics, calculated using the known hourly streamflow values over the 6-month simluation period. This ensures that our sensitivity indices properly incorporate measures of model accuracy, an approach strongly supported by recent literature (van Griensven et al., 2006; Demaria et al., 2007; Cloke et al., 2008; Pappenberger et al., 2008; Van Werkhoven et al., 2008a; Reusser et al., 2011; Rosolem et al., 2012). We compute sensitivity indices at the event scale using the root mean squared error (RMSE) and runoff coefficient error (ROCE) metrics. The RMSE metric represents the sum of squared residuals over a particular time window:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{s,i} - Q_{o,i})^2} \tag{4.1}$$

where Q_s and Q_o are the simulated and observed flows, respectively. The ROCE metric represents the error in the water balance, calculated as a percentage bias:

$$\text{ROCE} = \frac{\left|\sum_{i=1}^{n} Q_{s,i} - \sum_{i=1}^{n} Q_{o,i}\right|}{\sum_{i=1}^{n} Q_{o,i}} \tag{4.2}$$

The RMSE metric focuses on quick responses, while the ROCE metric

highlights the long-term bias of the water balance calculated by the model (Van Werkhoven et al., 2008a). These two metrics combine to provide a comprehensive understanding of model response at the event scale. The high-resolution sensitivity analysis is performed using a 24-hour moving window with a 3-hour timestep; only the sensitivity of the RMSE metric is computed here, since a water balance metric would be inappropriate for such a short timescale. With this high-resolution moving window, the sensitivity indices of all 1092 parameters are calculated at a total of 1457 intervals over the course of the 6-month simulation period.

We calculate sensitivity indices using a sample size of $N = 20$, corresponding to $21,860$ model evaluations. This represents a significant computational savings compared to a typical global sensitivity analysis method. Herman et al. (2013a) showed that the method of Morris using $N = 20$ for the full simulation period of this study was capable of providing sensitivity results comparable to the Sobol method using $N = 6,000$, which required over 6.5 million model evaluations. The high-resolution sensitivity analysis investigated here is only computationally tractable due to the demonstrated efficiency of the method of Morris (Herman et al., 2013a).

The sensitivity analyses were performed using the NSF CyberSTAR high-performance cluster at Penn State University, which contains a combination of quad-core AMD Shanghai processors (2.7 Ghz) and Intel Nehalem processors (2.66 Ghz). An open-source implementation of the method of Morris was used from the R Sensitivity Package (Pujol et al., 2013), which includes the methodological improvement of Campolongo et al. (2007). Approximately 100 computing hours were required for the model evaluations at the $N = 20$ sample size,

with an additional 100 hours needed to compute the sensitivity indices for each of the nearly 1500 sub-intervals.

4.5 Results and Discussion

Sensitivity results are presented in order of increasing temporal resolution. We begin with the full period and event scale sensitivity indices (Figs. 4.3-4.4) before proceeding to the high-resolution results (Figs. 4.5-4.7). These results can be interpreted in the context of the precipitation patterns shown in Figs. 4.1-4.2. This sequence of results is designed to explore the potential shortcomings of the aggregated approaches along with the additional insights provided by the high-resolution approach.

4.5.1 Full Period and Event-Scale Sensitivity Analysis

The sensitivity indices for the root mean squared error (RMSE) metric are shown in Figure 4.3 for the full simulation period and the three selected events. The μ^* values from the method of Morris are normalized to the range $[0, 1]$ to facilitate comparison across experiments, from an initial range of $[0, 0.08]$. For the full six-month period, the spatial distribution of parameter sensitivity appears bimodal: a concentrated high-sensitivity area occurs in the headwaters, particularly for the lower zone storage maxima *LZFPM* and *LZFSM*, and a second concentration occurs near the outlet of the watershed, particularly for the upper zone parameters *UZFWM* and *UZK*. Considering the forcing patterns shown in Fig. 4.2, the RMSE for the full period is likely dominated by several large

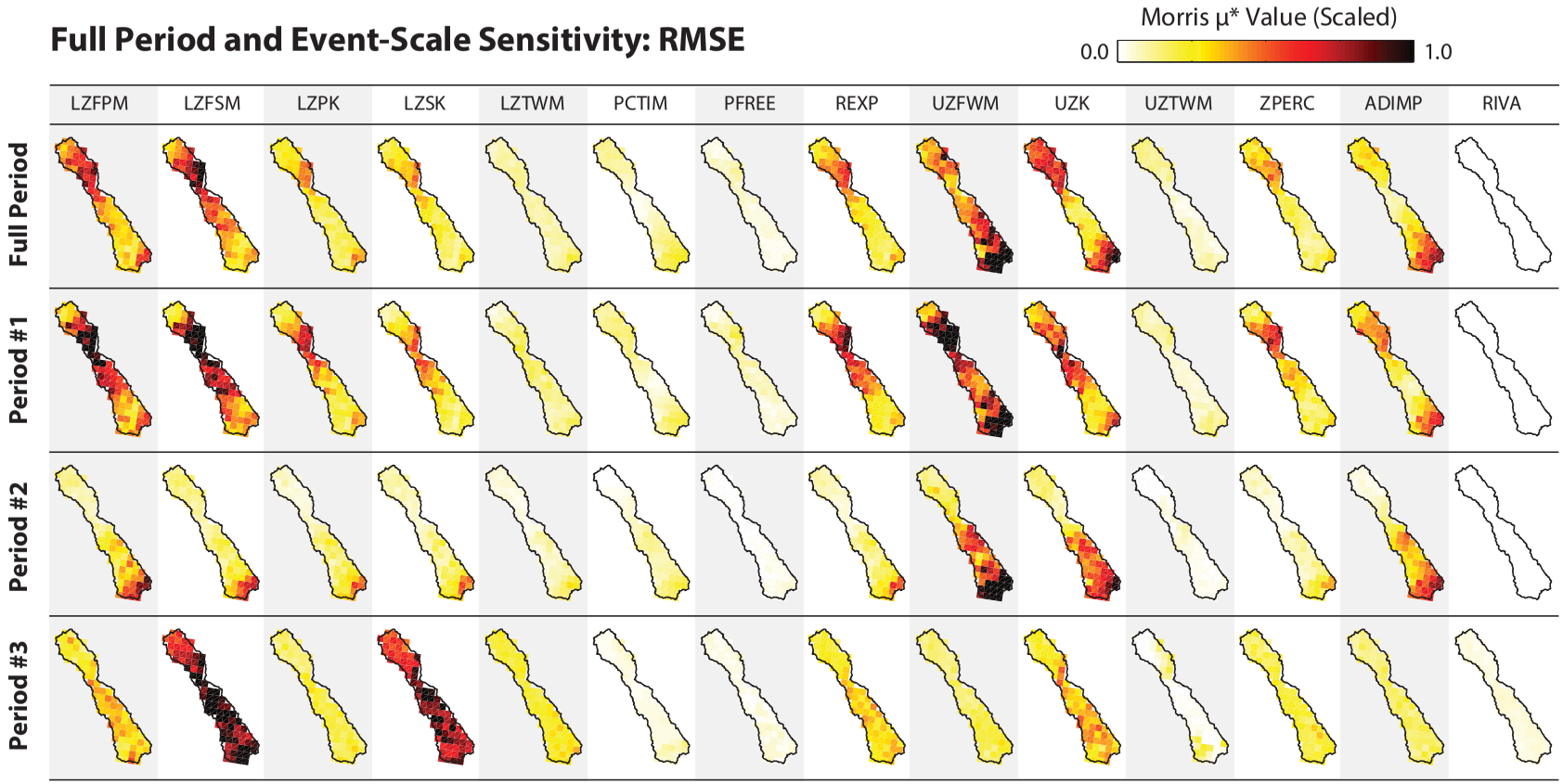

Figure 4.3: Full period and event-scale sensitivities of the root mean squared error (RMSE) metric over the full six-month simulation and three selected sub-periods. The μ^* values from the method of Morris are scaled to the range $[0, 1]$ from an initial range of $[0, 0.08]$. The RMSE metric focuses on the model's ability to reproduce observed streamflow peaks. The event-scale sensitivity indices differ significantly from those in the aggregated full period depending on the magnitude of the event and the spatial distribution of precipitation.

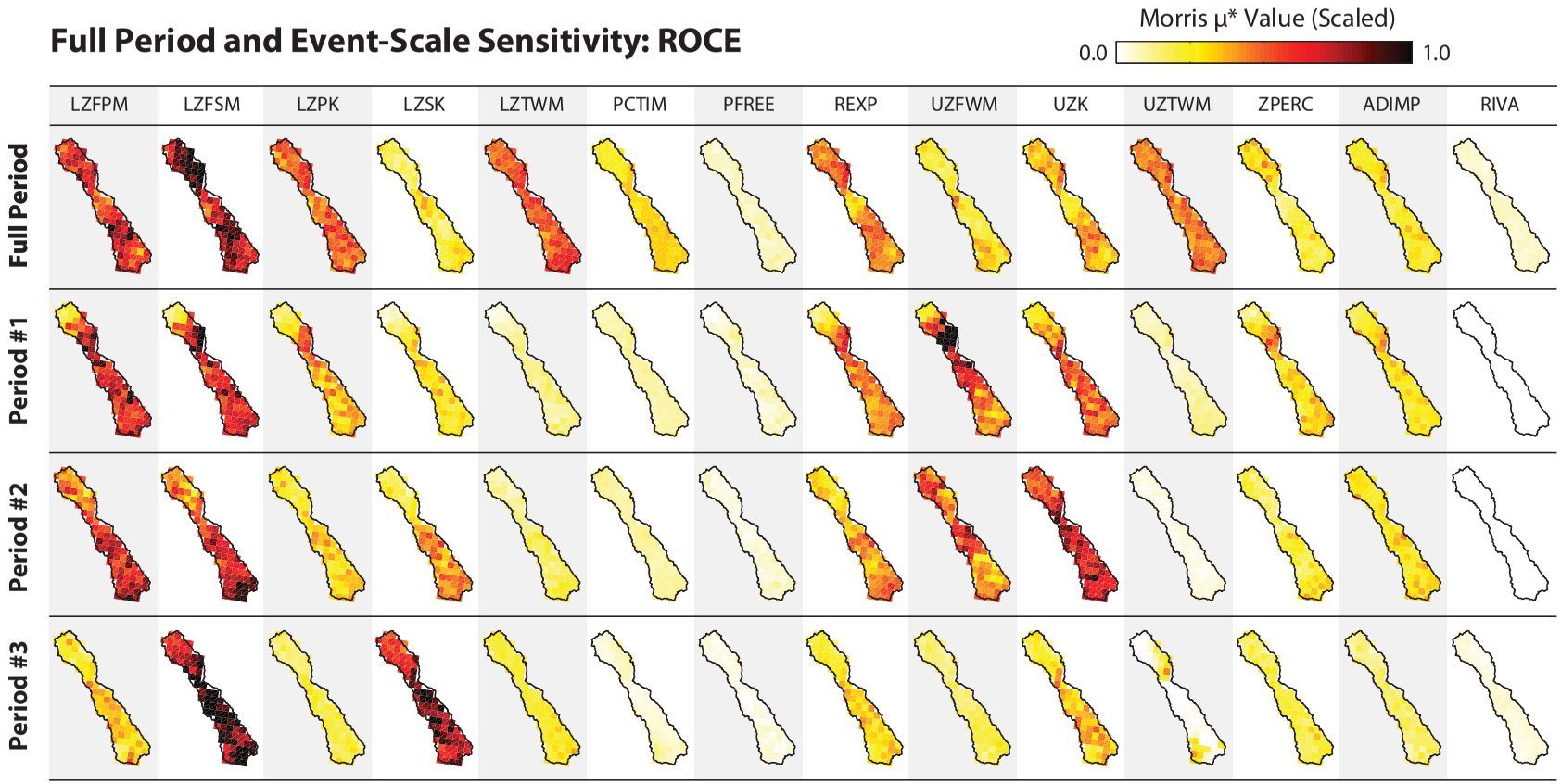

Figure 4.4: Full period and event-scale sensitivities of the runoff coefficient error (ROCE) metric over the full six-month simulation and three selected sub-periods. The μ^* values from the method of Morris are scaled to the range $[0, 1]$ from an initial range of $[0, 0.08]$. The ROCE metric focuses on the model's ability to reproduce the observed long-term water balance. The spatial controls on the water balance are more evenly spread across the watershed compared to the RMSE metric in Figure 4.3.

events, some of which are concentrated in the headwaters of the basin such as the event during Period 1. This explains the bimodal division of RMSE sensitivity between the headwaters (due to concentrated forcing during large events) and the outlet cells (due to gauge proximity). Note that both the upper zone storage maximum, *UZFWM*, and its associated drainage coefficient, *UZK*, are sensitive during this period, whereas only the storage maxima in the lower zone are sensitive. This difference indicates that flow from the upper zone is generated from a combination of storage overflow and gravity drainage, while flow from the lower zone is primarily generated by storage overflow alone. This result highlights the importance of understanding which flow generation mechanisms dominate model performance during peak events over the course of the simulation.

Period 1 exhibits a strong concentration of parameter sensitivity in the headwater cells of the basin. As illustrated in Fig. 4.2, the large precipitation event during this period occurs primarily in the headwaters, so this result is expected. Even the upper zone parameters are most sensitive in the headwaters during Period 1 (with the exception of *UZFWM*), despite the typical dependence of upper zone sensitivity on gauge proximity. This result contrasts with Period 2, which exhibits very little sensitivity in the headwaters. In Period 2, the majority of high-sensitivity cells appear near the watershed outlet, even for the lower zone parameters. The contrast between Periods 1 and 2 reveals the effect of the spatial distribution of forcing on parameter sensitivity. The headwater cells of the model are only activated when precipitation is concentrated in this region, and flow is generated by exceeding storage maxima in both the upper and lower zones. Conversely, when precipitation is distributed across the basin, model performance is dominated only by the cells near the outlet gauge where

flow is generated in the upper zone by a combination of storage exceedance and gravity drainage. These differences in the responses of Periods 1 and 2 are potentially complicated by internal model states, such as antecedent moisture conditions, which could alter the response signatures. However, as shown in the supplemental material, zooming in to the time-varying sensitivity of Periods 1 and 2 clearly reveals the differences in the spatial distribution of precipitation between the two events. The sensitivity responses to each event begin almost immediately following the precipitation, and thus their differences may be traced primarily to the precipitation distribution over the watershed. With no information regarding the "true" watershed processes, it is worth noting that a significant portion of the model remains insensitive during both large events. The differences between sensitivity patterns in Periods 1 and 2 would be very difficult to predict in advance, and thus underscore the need for diagnostic methods that do not depend on spatial aggregation.

Finally, the dry Period 3 exhibits very different sensitivity patterns from any of the other periods. Here, sensitivity is effectively concentrated in the lower zone secondary storage element, with maximum *LZFSM* and drainage coefficient *LZSK*. The lower zone secondary storage is very likely to be the last element containing water during dry periods, as it has the slowest drainage constant as shown in Table 3.1. Therefore, this element controls model performance after the other storage zones have drained or evaporated. The sensitivities of parameters *LZFSM* and *LZSK* are distributed across the entire watershed, unlike during Periods 1 and 2 where the most sensitive parameters only occur in concentrated areas. This suggests that dry periods may provide valuable identifiability information for cells which are otherwise inactive, particularly for these slow-draining storage elements in the lower zone of the model.

At the event scale, it is valuable to assess the sensitivity of multiple diagnostic measures to obtain a more thorough understanding of controls on model performance. In general, model error can be decomposed into bias and variance (Gupta et al., 2009). The RMSE metric, with its dependence on quick runoff response, is most closely related to variance; we also investigate the runoff coefficient error (ROCE), a water balance metric related to model bias (Van Werkhoven et al., 2008a). The event-scale sensitivity indices for the ROCE metric are shown in Figure 4.4. Compared to the RMSE metric shown in Figure 4.3, the sensitivity of ROCE is spread across a larger number of parameters, and more evenly distributed across the spatial extent of the watershed. Whereas RMSE is controlled by a few cells depending on their proximity to precipitation and/or the outlet gauge, the water balance error depends on the soil moisture calculations in all cells. In general, many of the same lower and upper zone parameters dominate the ROCE and RMSE performance metrics: the storage maxima *LZFPM*, *LZFSM*, and *UZFWM*, and the drainage constants *LZPK* and *UZK*. Similar to RMSE, the ROCE metric depends on flow generation via storage exceedance as well as gravity drainage processes in the model. Compared to Fig. 4.3, the differences between Periods 1 and 2 are far less pronounced in Fig. 4.4, indicating that the spatial distribution of precipitation does not affect the water balance error as much as it does peak flows. Period 3 shows the most similarity to the corresponding RMSE result, as its ROCE metric is still controlled primarily by the secondary storage parameters *LZFSM* and *LZSK*. The apparent independence of the ROCE metric to forcing and gauging locations suggests that this measure of performance succeeds in extracting information from a larger spatial area of the model, potentially providing benefits for identifiability. However, the ROCE metric alone will not account for the timing of

flow peaks, and is therefore best applied in conjunction with a timing-based metric such as RMSE.

The event-scale sensitivity results shown in Figs. 4.3 and 4.4 provide useful diagnostic insight for the full simulation period and selected sub-periods. These findings align with previous work: spatially concentrated precipitation will cause parameter sensitivity to appear in a similar pattern as the precipitation, whereas uniformly distributed precipitation will cause sensitivity in cells near the outlet (Tang et al., 2007a; Van Werkhoven et al., 2008b). However, the event-scale analysis also contains several weaknesses. First, the results are highly dependent on the choice of events to study, as illustrated by the differences in controls across the selected periods. It would be prohibitively difficult to design or select representative events which fully capture the range of model responses. Instead, it is beneficial to analyze the emergent model responses in nearly-continuous time, and select sub-periods of interest *a posteriori*. Second, the event-scale results do not indicate when these parameters become sensitive relative to changing hydrologic conditions. Consequently, the sensitivity indices shown in Figs. 4.3 and 4.4 for the full simulation period are strongly influenced by only a few large events, with their dynamics obscured by aggregation. It has been noted in previous work that the value of streamflow observations for identifying distributed model parameters may be limited by the location and intensity of forcing, particularly if the period of analysis is defined to include only a single rainfall event (Van Werkhoven et al., 2008b). We hypothesize that allowing distributed parameter sensitivity to vary in nearly-continuous time will extract more value from streamflow observations by highlighting parameter activation across a much broader range of hydrologic conditions.

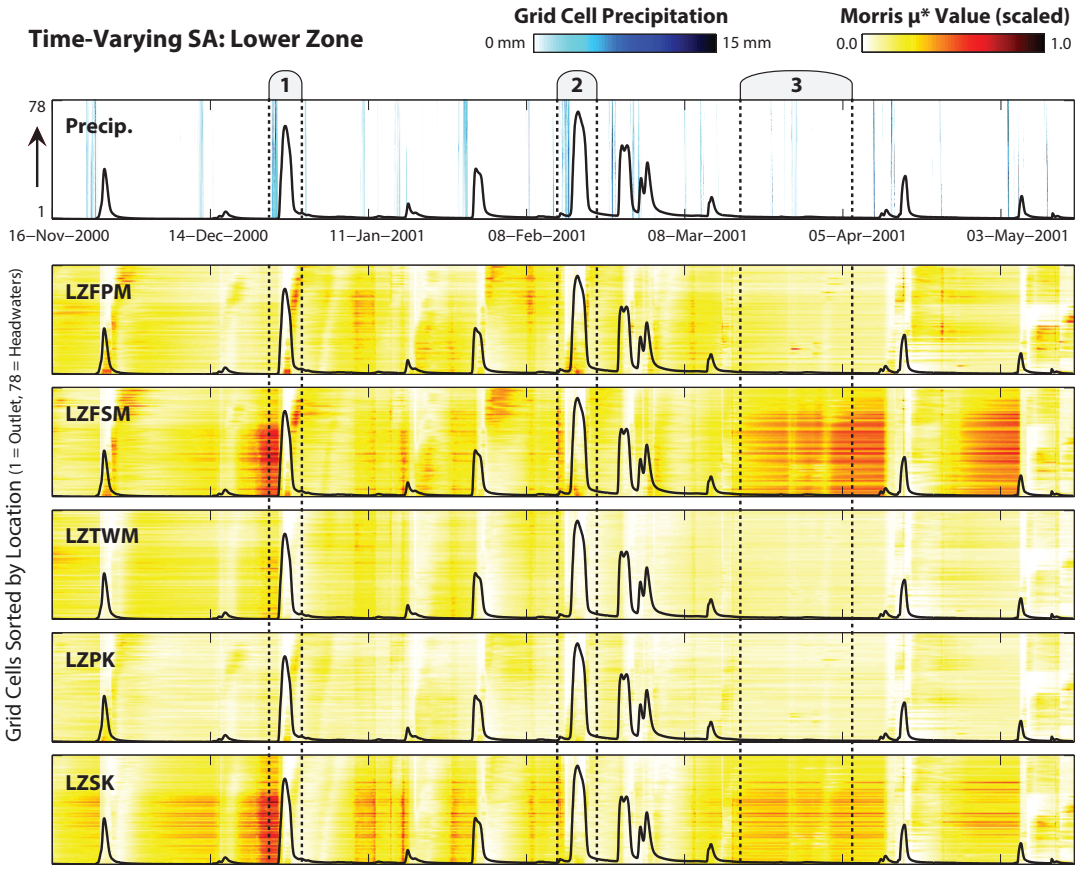

Figure 4.5: Time-varying sensitivity of the RMSE metric for the five lower zone parameters of the HL-RDHM model. The indices are calculated for a 24-hour moving window with a 3-hour timestep. The y-axis arranges the 78 grid cells based on their distance from the watershed outlet, from the outlet ($y = 1$) to the furthest headwater cell ($y = 78$). The μ^* values from the method of Morris are scaled to the range $[0, 1]$ from an initial range of $[0, 0.2]$. The lower zone parameters maintain a consistent, moderate level of sensitivity throughout the simulation. Exceptions occur during large events, when the lower zone parameters are mostly insensitive. The secondary storage parameters, *LZFSM* and *LZSK*, are particularly sensitive during low-flow periods.

4.5.2 High-Resolution Distributed Sensitivity

This study aims to elucidate the time-varying nature of these distributed parameter sensitivities by performing global sensitivity analysis using a high-resolution moving window with a sub-daily timestep. Parameter sensitivity is calculated on a 24-hour moving window with a 3-hour timestep, resulting in 1457 intervals over the six-month simulation period.

The choice of window size contains two competing considerations: it must be large enough to smooth out any noise in the performance metric calculations, yet also small enough to capture dominant model processes that occur on fast timescales. The moving window size and timestep used in this study reflect a balance between these two issues. Since the model runs on an hourly timestep, a 24-hour window size (with significant overlap due to the 3-hour timestep) will smooth noise in the calculation of the RMSE metric. Additionally, since most large events during the simulation period are approximately 48-72 hours in length, the 24-hour window is also sufficiently small to capture quick responses, which is critical from a flood forecasting standpoint.

Thus, each parameter has a time series of sensitivity indices for each grid cell (*i.e.*, the results summarize time-evolving sensitivity maps across all spatial grid cells in the model). These sensitivity indices are shown in Figures 4.5, 4.6, and 4.7, corresponding to the lower zone parameters, upper zone parameters, and remaining parameters, respectively. Each of the three plots contains the same hourly precipitation data in the top panel, as well as the same hourly hydrograph data superimposed on each subplot. The sensitivity indices are aligned at the center of each moving window interval. The μ^* values from the method of Morris are normalized to the range $[0, 1]$ to facilitate comparison across ex-

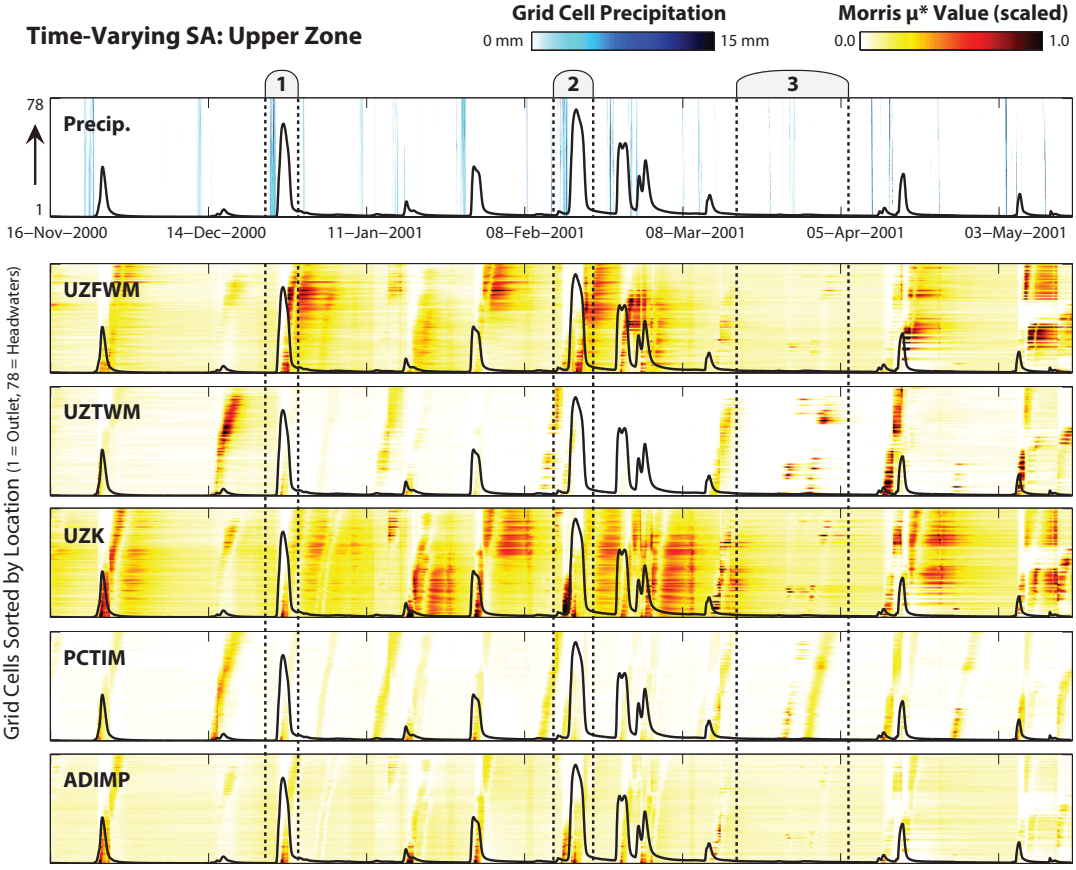

Figure 4.6: Time-varying sensitivity of the RMSE metric for the five upper zone parameters of the HL-RDHM model. The indices are calculated for a 24-hour moving window with a 3-hour timestep. The y-axis arranges the 78 grid cells based on their distance from the watershed outlet, from the outlet ($y = 1$) to the furthest headwater cell ($y = 78$). The μ^* values from the method of Morris are scaled to the range $[0, 1]$ from an initial range of $[0, 0.2]$. The parameters controlling the upper zone free water element, *UZFWM* and *UZK*, are highly sensitive during large events. The high sensitivity of these parameters typically begins near the watershed outlet during the rising limb of the hydrograph, and transitions toward the headwater cells during the falling limb.

periments, from an initial range of $[0, 0.2]$. While these figures are designed for journal format, animations of time-varying sensitivity indices are available as a multimedia supplement.

In Figures 4.5, 4.6, and 4.7, the two spatial dimensions of the watershed are compressed into the y-axis, where the 78 grid cells are arranged according to their distance from the watershed outlet. The bottom of each subplot ($y = 1$) represents the outlet cell, while the top of each subplot ($y = 78$) represents the headwater cell furthest from the outlet. This configuration allows us to visualize both space and time on the same axes without drawing a large number of maps of the watershed. This plotting approach is particularly effective for the Blue River basin, which has a long, narrow shape.

Figure 4.5 shows that the lower zone parameters maintain a moderate level of sensitivity throughout the simulation. The influence of all lower zone parameters clearly recedes during large events, except for the parameters *LZFPM* and *LZFSM* in the cells nearest to the watershed outlet. This indicates that the only contribution to large streamflow events from the lower zone occurs due to exceeding the storage maxima, not due to gravity drainage. The drainage processes occur on slower timescales and would not contribute significantly to peak flows. The effect of slow drainage processes is clear from the high sensitivity of the secondary storage parameters, *LZFSM* and *LZSK* during low-flow periods, after other storage elements have been emptied. These insights largely align with those found at the event scale in Figures 4.3 and 4.4. However, when comparing sensitivity indices across temporal resolutions, it is important to note that parameter sensitivity is measured relative to other parameters. Thus, a larger time window may cause some parameters to appear insensitive at cer-

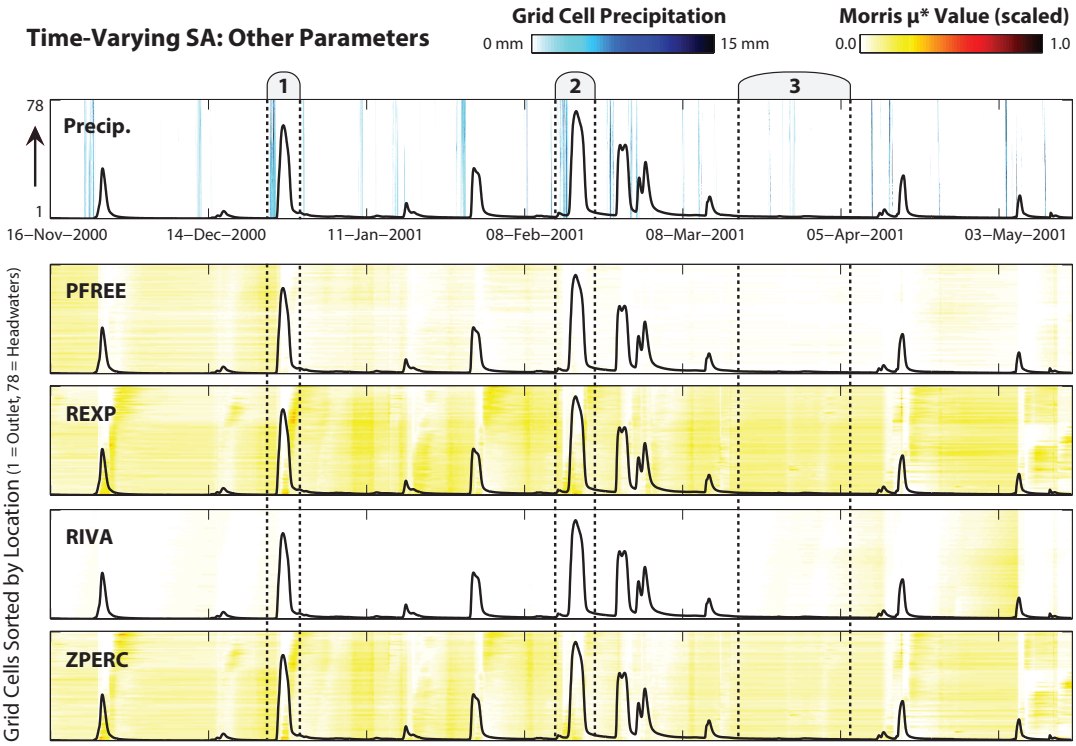

Figure 4.7: Time-varying sensitivity of the RMSE metric for the four remaining parameters of the HL-RDHM model. The indices are calculated for a 24-hour moving window with a 3-hour timestep. The y-axis arranges the 78 grid cells based on their distance from the watershed outlet, from the outlet ($y = 1$) to the furthest headwater cell ($y = 78$). The μ^* values from the method of Morris are scaled to the range $[0, 1]$ from an initial range of $[0, 0.2]$. These parameters influence model performance significantly less than the lower zone parameters (Figure 4.5) or the upper zone parameters (Figure 4.6), and never appear highly sensitive during the course of the simulation.

tain locations due to the dominance of others. This phenomenon is visible, for example, for the *LZSK* parameter near the outlet of the watershed, where it is sensitive in Figure 4.5 but not in Figure 4.3.Compared to the event scale analysis, the high temporal resolution in Figure 4.5 has the advantage of clarifying the timing of parameter activation. For example, it has been noted in prior studies that the lower zone parameters frequently control the performance of the SAC-SMA model (Van Werkhoven et al., 2008a; Herman et al., 2013c), from which it might be concluded that the lower zone contributes significantly to flow peaks. However, Figure 4.5 indicates that the lower zone contributes to performance primarily during non-peak periods, which, when aggregated, may yield higher levels of sensitivity depending on the period studied.

By contrast, the upper zone parameters are clearly activated during and after streamflow events, as shown in Figure 4.6. In particular, the upper zone free water parameters (*UZFWM* and drainage coefficient *UZK*) become dominant during large events. The propagation of sensitivity upward through the watershed is clear for these parameters, starting at the outlet cells during the rising limb of each event and moving toward the headwaters during the falling limb. As expected, there is a lag between the time at which the event begins and the time at which the headwater cells begin to affect the model performance due to routing. Similarly, the timing of activation for the outlet cells depends on the event; compare the parameters *UZFWM* and *UZK* during Period 1, where the outlet cells are activated midway through the event, to Period 2, where *UZK* is activated immediately during the rising limb of the hydrograph. The lag in parameter sensitivity during Period 1 is likely due to the headwater–focused precipitation event, while in Period 2 the precipitation occurs closer to the outlet. Interestingly, Figure 4.6 shows that the additional impervious area param-

eter, *ADIMP*, is only sensitive for cells near the outlet during events, but this signal does not propagate to the headwaters. The impervious area only affects model performance in cells close to the outlet, since these directly control the quick response during the rising limb of each event. As Figure 4.6 indicates, the upper zone parameters typically do not control model performance during low-flow periods and small events. There are two interesting exceptions to this, however. First, the drainage rate *UZK* remains sensitive for several hours after each event as the upper zone drains its storage. That is, gravity drainage from the upper zone typically occurs slowly enough such that it continues to release water well into the low-flow periods. Second, the parameters *UZTWM* (upper zone tension storage) and *PCTIM* (percent impervious area) are most sensitive following rainfall events which do not lead to large streamflow events. In other words, these parameters are most important when model performance requires the absence of a response. If impervious area is too high (causing high direct runoff), or tension storage capacity too low (causing runoff via overflow), the model may overestimate streamflow and create significant errors in the RMSE metric. This phenomenon is not visible for the events analyzed in Figure 4.3 because it would be difficult to predict at the time of *a priori* event selection. When a large streamflow response is required, the parameters *UZFWM* and *UZK* dominate instead, as these become the primary mechanism by which large events are generated. An additional difference between Figures 4.3 and 4.6 is that, while the *UZK* parameter is sensitive throughout the watershed at the event scale (e.g., during Period 1), it only appears sensitive near the outlet at the high-resolution timescale. This could be due to the dominance of other parameters in the upper region of the watershed at the high-resolution timescale, or simply a difference in the sensitivity of the upper zone drainage response at different timescales.

Finally, the *UZFWM* parameter during Period 2 is only sensitive near the watershed outlet when considered at the event scale, but the high-resolution results in Figure 4.6 show that the sensitivity of *UZFWM* propagates to the headwaters during the falling limb of the event. Since the RMSE metric focuses on peak flows, the falling limb does not play a significant role in the calculation of aggregated sensitivity for each period, even though these are clearly visible at the high-resolution timescale.

Finally, Figure 4.7 shows the time-varying sensitivity indices for the four remaining parameters in the model. These parameters rarely dominate the upper and lower zone parameters shown in Figures 4.5 and 4.6. The percolation parameters Z_{Perc} and R_{Exp} maintain a low, but non-zero, level of sensitivity throughout the simulation, which decreases to zero during streamflow peaks. The percolation parameter P_{Free} follows a similar trend, but becomes inactive during the dry Period 3. The riparian vegetation area, *RIVA*, is only activated in the late spring season when evapotranspiration becomes more pronounced, as expected. Comparing to the event scale results in Figure 4.3, these parameters are generally much less sensitive at the high-resolution timescale. For example, the R_{Exp} parameter modifies the rate of percolation from the upper to lower zone but does not cause runoff directly, which makes it less likely to influence model performance during the 24-hour moving window than over the course of an aggregated event. The parameters shown in Figure 4.7 play a small role in model performance, as evidenced by their moderate but non-zero sensitivities through time. However, the sensitivities of the upper and lower zone parameters are typically much larger (see Figures 4.5 and 4.6) and thus these remaining parameters rarely control model performance. This result suggests a potential identifiability problem for these less-sensitive parameters, as they are rarely ac-

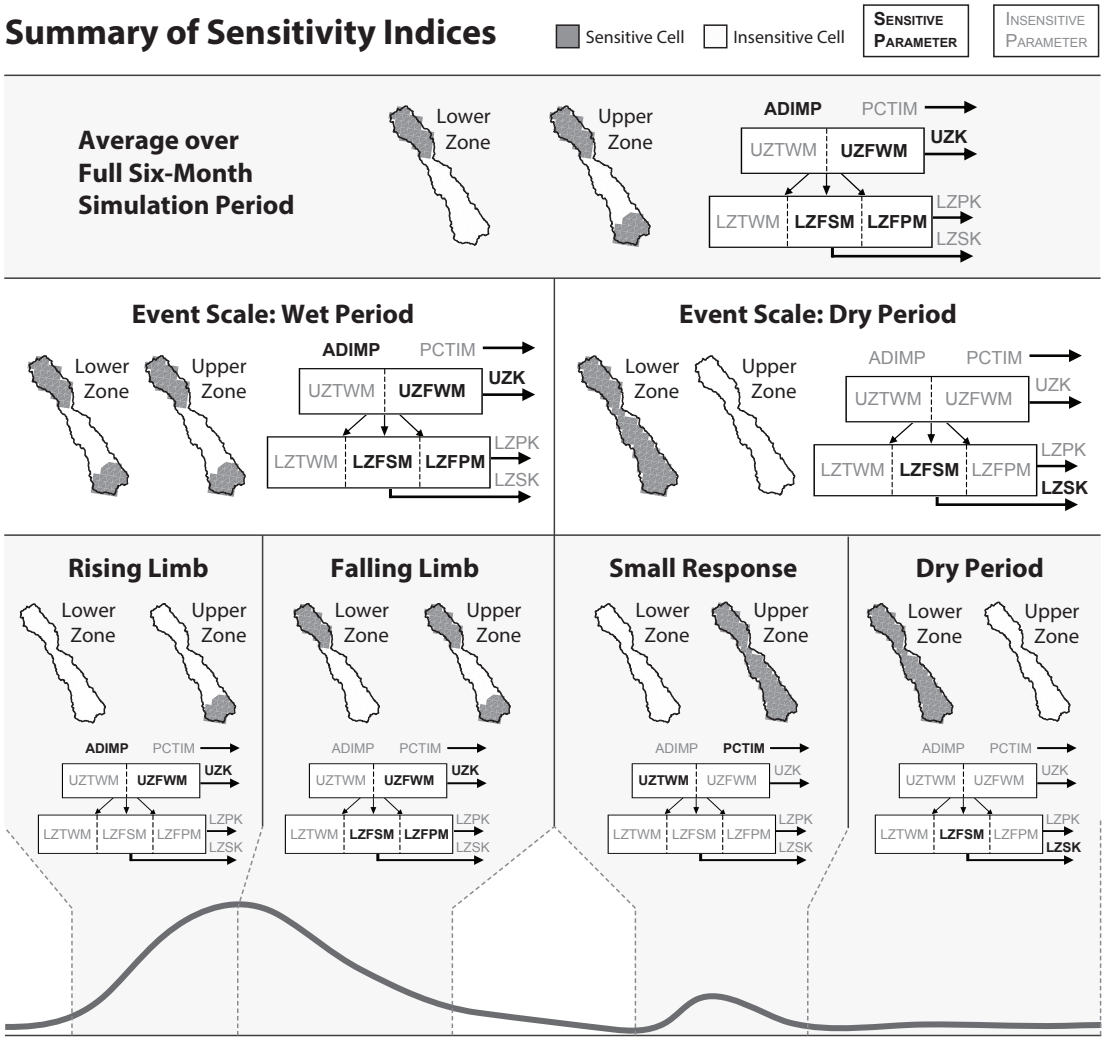

Figure 4.8: Qualitative summary of sensitivity indices for the RMSE metric at increasing temporal resolutions, from the aggregated simulation period (top), to the event scale (middle), and the high-resolution moving window (bottom). Each classification includes spatial maps showing where upper zone and lower zone parameters are sensitive in the watershed, along with a model diagram highlighting the most sensitive parameters during the defined periods. Results shown for the high-resolution sensitivity analysis are a qualitative summary of insights gleaned from Figures 4.5–4.7 and thus do not reflect the dependence of time-varying sensitivity on the spatial distribution of precipitation.

tivated in any of the model grid cells.

4.5.3 Discussion

The event-scale sensitivity maps shown in Figures 4.3 and 4.4 represent a traditional approach to diagnostic analysis of a spatially distributed watershed model, on the relatively few occasions that such analyses have been performed. The event-scale approach shows the spatial distribution of sensitivity for selected intervals, which may change dramatically from one event to another, as the results show for Periods 1, 2, and 3. The choice of representative events for a diagnostic analysis can therefore strongly bias the outcome, particularly considering the complex dependencies between parameter sensitivity, the spatial distribution of forcing of an event, and the proximity of a given cell to the outlet gauge. These results suggest the severe difficulty of selecting *a priori* a set of precipitation events which capture the full range of potential model responses, the approach suggested in prior studies (Tang et al., 2007a; Van Werkhoven et al., 2008b). For example, this difficulty is demonstrated by our inability to foresee the phenomenon in which parameters *UZTWM* and *PCTIM* are most sensitive when modeled responses require attenuation to match observations. Furthermore, the temporal aggregation involved in the event-scale analysis obscures the underlying dynamics of parameter sensitivity. When full temporal resolution is allowed, parameter sensitivity indices show clear patterns of activation before, during, and after streamflow events, as shown in Figures 4.5, 4.6, and 4.7. The high-resolution approach improves on the event-scale analysis by isolating the time and location at which individual parameters and cells are activated, allowing a larger fraction of the model to contribute to its performance mea-

sures throughout the simulation and thus making better use of the information content contained in streamflow observations. Within each event, parameter sensitivity clearly represents a dynamic rather than static quantity, and should be analyzed accordingly as computational costs permit.

A qualitative summary of sensitivity indices at increasing temporal resolution is shown in Fig. 4.8. At the highest resolution (i.e., the moving window), the spatiotemporal sequence of influential parameters in Fig. 4.8 reflects the transitions between sets of dominant parameters and processes. In general, the highlighted sensitive parameters and cells are those with a scaled value of $\mu^* > 0.5$. For the high-resolution timescale, the four cases shown (rising limb, falling limb, small response, and dry period) are intended to reflect general insights from Figures 4.3-4.7. These cases do not necessarily reflect all timesteps in the simulation period, but rather the most interesting classifications that were observed in the results. However, since the Blue River watershed is driven primarily by infrequent large events without significant temperature or elevation effects, it is possible to broadly separate the hydrograph into these four classifications, as shown in the supplemental material.

As Figure 4.8 shows, the dominant controls for the full aggregated period are a combination of lower zone parameters in the headwaters of the basin, and upper zone parameters near both the headwaters and outlet. The full period sensitivities are clearly influenced by the wet periods at the event scale, which exhibit the same responses, indicating that the aggregate period is biased toward these large events (a result consistent with the focus of the RMSE metric). By contrast, dry periods at the event scale exhibit very different sensitivity patterns, centered around slow drainage from the lower zone supplemental store.

The summarized high-resolution sensitivity results in the bottom row of Fig. 4.8 provide a more detailed understanding of model behavior than the full period or the event scale. In general, the parameters that appear most sensitive at the event scale are also the most active for the high-resolution moving window. These primarily include the upper zone parameters *UZFWM* and *UZK* and the lower zone parameters *LZFPM* and *LZPK*. This finding aligns with our initial hypotheses, since gravity drainage and overflow from exceeding storage maxima represent two of the primary runoff generation mechanisms in the model. The most sensitive cells during the rising and falling limbs of large events represent a decomposition of the event scale sensitivity during wet period, which may be particularly valuable depending on the part of the hydrograph being analyzed. As anticipated, the upper zone and impervious area parameters dominate model performance during and immediately following large events, since these create the quick response required to reproduce observed streamflow. The high-resolution dry period exhibits largely the same sensitivities as the event scale, which would be expected considering the lack of dynamic behavior during these dry periods. Finally, the small response reflects the common scenario in which quick runoff must be avoided to achieve good performance, a behavior which remains invisible at the event scale unless a small response event is explicitly chosen for analysis *a priori*.

The high-resolution results in the bottom panel of Fig. 4.8 can also be interpreted to identify transitions between dominant parameters and processes in the model. During the rising limb of streamflow events, the dominant processes in the model are typically direct runoff from impervious area, and overflow/drainage from the upper zone free water store. As might be expected, these processes are most dominant near the outlet of the watershed, reflecting

the need for a quick response to match the observed hydrograph. During the falling limb, the model transitions to a dominant process comprising slower drainage responses from the upper and lower zone. These processes are dominant in the headwaters as well in addition to the cells near the outlet, since the longer time lag allows cells further from the outlet to contribute to streamflow. During small responses, the dominant process consists of direct runoff from impervious area and overflow from upper zone tension water, both of which must be properly attenuated to avoid overshooting the observed peak. Finally, during dry periods, a dominant process consisting of slow release from the lower zone often dominates model performance. These types of insights regarding transitions between modeled processes are not attainable from *a priori* selection of events assumed to be broadly representative. The coarser event scale sensitivities are typically obscured, and are not necessarily consistent even for seemingly similar events (as highlighted in Figures 4.3 and 4.4).

It should be emphasized that even though Fig. 4.8 represents a qualitative aggregation of the high-resolution sensitivity patterns, this aggregation is drawn *a posteriori* from the full range of dynamic parameter activation characterized using the three-hour moving window. The value of the high-resolution approach, as shown in Figs. 4.5-4.7, is its ability to isolate parameter activation in space and time while avoiding the potential biases introduced by *a priori* event selection and aggregation. The high-resolution analysis removes these biases by reducing the size of the interval window such that peak flows do not accumulate undue influence relative to the rest of the examined interval. In the high-resolution results, the dynamic transitions between upper and lower zone sensitivity become clear: the lower zone maintains a fairly constant level of control over model performance throughout the simulation, while the upper zone

dominates during large events. The upper zone storage elements are the first to receive precipitation during large events and therefore exert the most control over the timing and magnitude of the quick response. The lower zone elements release water more slowly and are most responsible for model performance in the absence of large events. In prior analysis of the SAC-SMA model (a spatially lumped version of HL-RDHM), it was found that the lower zone parameters almost exclusively control the RMSE metric at the monthly timescale (Herman et al., 2013c). By zooming in to a 3-hour timestep, the high-resolution method identifies the importance of the upper zone parameters for properly reproducing quick responses. This approach is able to identify sensitive cells which were not visible at the event scale as well as the timing of their activation, making it a valuable addition to traditional diagnostic approaches.

The high-resolution sensitivity approach presented here requires several important considerations. First, as with any sensitivity analysis, the results strongly depend on the choice of performance metric. Figures 4.3 and 4.4 show the substantial differences in sensitivity indices that occur when changing from the RMSE metric to the ROCE metric. We focus on the RMSE metric in this study because its emphasis on large events is consistent with our goal of understanding model behavior in the context of quick-response flood forecasting. Interestingly, as the window size of the analysis decreases, different performance metrics begin to behave similarly (i.e., in the limit as window size approaches zero, most metrics reduce to a percent error at a single point). This leads to the second important consideration, the choice of window size. Modeled processes which dominate performance at one timescale may be invisible at another, so it is crucial to choose a window size commensurate with the purpose of the analysis. Our moving window size of 24 hours (with a 3-hour timestep) reflects the

need to capture dominant processes on a fast timescale while also containing a sufficient number of timesteps to smooth out any noise in the performance metric. Finally, the visualization approach presented in Figures 4.5-4.7 (in which the spatial dimensions are compressed into a single distance measure on the y-axis) is readily applicable to the narrow Blue River watershed, but may be difficult to extend to other watersheds. This is particularly true if significant spatial heterogeneity exists in land cover or soil storage properties. In such instances, it may be preferable to represent the y-axis as, for example, the soil storage capacity of each grid cell, or whichever characteristic is expected to govern grid cell sensitivity. For this case study, the primary characteristic of interest is simply the distance from the watershed outlet, but this may not hold true for all applications.

This high-resolution approach is intended to complement, rather than replace, the insights derived from an event-scale analysis. In this case, neither the event-scale results nor the high-resolution results could have been predicted from the other. The event-scale analysis provides performance controls for selected intervals of interest; the high-resolution analysis contributes clear details of parameter dynamics across the simulation period without focusing on any particular interval. The need for complementary approaches at different temporal resolutions is further highlighted by the fact that a given parameter may only be sensitive at a certain timescale, considering these sensitivity measures describe the effect of a parameter relative to the others, and these relative effects are extremely likely to change depending on the scope of the analysis.

4.6 Conclusions

High-resolution sensitivity analysis explores the full spatial and temporal variability of distributed watershed model controls, highlighting the importance of avoiding confounding aggregation to the extent permitted by computational constraints. The complexity of spatially distributed models typically causes a significant fraction of parameters to be inactive at any particular time, a phenomenon clearly shown in the event-scale results of this study. This sparsity of activation can lead to needless complexity and inappropriate modification of inactive parameters. However, it also presents a valuable opportunity to overcome the complexity of distributed parameter identification by restricting search to only those parameters which are active at a specific time and location. It also suggests an opportunity to identify locations and timing for optimal data collection to improve the modeled representation of hydrogic processes, particularly under nonstationary conditions in which dominant watershed processes fall outside observed ranges. For example, the results of this study indicate that large streamflow events in the model are controlled primarily by upper zone fluxes quite close to the watershed outlet; collecting flux data in only this area during a large event could provide justification to falsify the internal processes of the model, and to improve them by calibrating against the new observations. As demonstrated in this study, spatial variability can easily be visualized as a time series and provides valuable information for analyzing model behavior. In light of these opportunities, it is imperative for diagnostic analyses of distributed models to explore parameter activation at the spatial and temporal scales for which the model was designed. This study represents a novel step in this direction by visualizing spatially explicit, time-varying watershed model

sensitivity. As computational power continues to increase, such methods improve the potential for efficiently isolating distributed model behaviors at high spatial and temporal resolutions, an area which remains largely unexplored relative to similar analyses of simpler lumped models.

CHAPTER 5

ROBUSTNESS-BASED PORTFOLIO PLANNING FOR REGIONAL DROUGHT MANAGEMENT UNDER DEEP UNCERTAINTY

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

Herman, J.D., Zeff, H.B., Reed, P.M., and Characklis, G.W. (2014). Beyond Optimality: Multi-stakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. Water Resources Research, 50(10), 7692-7713.

Sections 5.4.2, 5.4.3, and 5.5.2 have been augmented with material from the supplement of the journal article, available from the publisher's website.

This work was supported by the Sectoral Applications Research Program (SARP) of the United States National Oceanic and Atmospheric Administration (NOAA) Climate Program Office under grant NA110AR4310144. The opinions, findings, and conclusions are solely those of the authors.

5.1 Abstract

While optimality is a foundational mathematical concept in water resources planning and management, "optimal" solutions may be vulnerable to failure if deeply uncertain future conditions deviate from those assumed during optimization. These vulnerabilities may produce asymmetric impacts across a region, making it vital to evaluate the robustness of management strategies as well as their impacts for regional stakeholders. In this study, we contribute a multi-stakeholder many-objective robust decision making (MORDM) framework that blends many-objective search and uncertainty analysis tools to dis-

cover key tradeoffs between water supply alternatives and their robustness to deep uncertainties (e.g., population pressures, climate change, financial risks, etc.). The proposed framework is demonstrated for four interconnected water utilities representing major stakeholders in the "Research Triangle" region of North Carolina, USA. The utilities supply well over one million customers and have the ability to collectively manage drought via transfer agreements and shared infrastructure. We show that water portfolios for this region that compose optimal tradeoffs (i.e., Pareto-approximate solutions) under projected future conditions may suffer significantly degraded performance with only modest changes in deeply uncertain hydrologic and economic factors. We then use the Patient Rule Induction Method (PRIM) to identify which uncertain factors drive the individual and collective vulnerabilities for the four cooperating utilities. Our framework identifies key stakeholder dependencies and robustness tradeoffs associated with cooperative regional planning, which are critical to understanding the tensions between individual versus regional water supply goals. Cooperative demand management was found to be the key factor controlling the robustness of regional water supply planning, dominating other hydroclimatic and economic uncertainties through the 2025 planning horizon. Results suggest that a modest reduction in the projected rate of demand growth (from approximately 3% per year to 2.4%) will substantially improve the utilities' robustness to future uncertainty and reduce the potential for regional tensions. The proposed multi-stakeholder MORDM framework offers critical insights into the risks and challenges posed by rising water demands and hydrological uncertainties, providing a planning template for regions now forced to confront rapidly evolving water scarcity risks.

5.2 Introduction: Regional water supply planning under uncertainty

Cooperative regional water portfolio planning, in which adaptive management strategies are coordinated across multiple water utilities, is a core component of the "soft path" approach for utilizing existing infrastructure more efficiently (Gleick, 2002, 2003). Such portfolios may combine conservation measures (e.g. Renwick and Green, 2000), water transfers (Lund and Israel, 1995; Wilchfort and Lund, 1997; Hadjigeorgalis, 2008), and financial instruments (Brown and Carriquiry, 2007; Zeff and Characklis, 2013; Zeff et al., 2014) to diversify the management of scarcity risks in a flexible manner. Regional water portfolios provide an innovative approach to offset future demand and climate risks, yet their potential vulnerabilities to *deep uncertainties* must be recognized. Deep uncertainty acknowledges that decision makers may not be able to enumerate all sources of uncertainty in a system nor their associated probabilities (Langlois and Cosgel, 1993; Lempert, 2002; Lempert et al., 2003; Polasky et al., 2011; Kasprzyk et al., 2013). Also referred to as Knightian uncertainty (Knight, 1921), a core concern is providing appropriate risk management actions despite the uncertainties in correctly specifying all probability distributions (Friedman, 1976). Deep uncertainty is especially prevalent in complex economic and environmental systems, where rapidly evolving systems may cause management policies to produce severe unintended consequences on stakeholders. These traits have long been recognized in water supply planning, making it a classic example of a "wicked" problem (Rittel and Webber, 1973; Liebman, 1976; Reed and Kasprzyk, 2009). Effectively, the problem of deep uncertainty means that complex water resources systems cannot be designed and managed based on narrow definitions of opti-

mality tied to a single most probable future scenario.

Robust Decision Making (RDM) addresses the problem of deep uncertainty by shifting the focus of policy design from seeking a single optimal solution to seeking a robust solution which provides satisfactory performance across many plausible states of the world (Lempert, 2002; Bankes et al., 2001; Bankes, 2002). By assuming that future states of the world are deeply uncertain, RDM does not attempt to assign probabilities to them, nor choose which is most likely. Instead, the ensemble of states of the world is viewed as a set of "computational experiments" (Bankes et al., 2001), used to identify ranges of deeply uncertain variables for which a particular policy or system design performs poorly (Groves and Lempert, 2007; Lempert and Groves, 2010). The RDM approach eliminates the need to identify a small set of future scenarios *a priori*, as is often the case in the expected value maximization approaches in classical decision theory (Clemen, 1996; Lempert and Collins, 2007). Instead, the emphasis is placed on identifying threshold values for the deeply uncertain variables that will most strongly affect the performance of a policy design, leaving decision makers to discuss possible likelihoods *a posteriori* with the knowledge of which scenarios matter most (Bryant and Lempert, 2010). Decision makers can then select a robust strategy which performs well across a range of plausible states of the world.

In recent work, the RDM approach was extended for many objectives (MORDM) by Kasprzyk et al. (2013). A key contribution of MORDM is the use of global optimization with multiobjective evolutionary algorithms (MOEAs) to discover tradeoffs across a diverse set of planning alternatives. The complexity of water resources systems often means that a randomly chosen solution will

be infeasible or severely suboptimal; the use of MOEAs to discover alternatives overcomes this weakness by intelligently searching for high-quality solutions prior to assessing their robustness (Kasprzyk et al., 2013; Reed et al., 2013). Additionally, MORDM employs interactive visual analytics (Thomas and Cook, 2006; Keim et al., 2006; Kollat and Reed, 2007b; Thomas and Kielman, 2009; Woodruff et al., 2013; Reed and Kollat, 2013) to facilitate the iterative feedback loop between problem formulation, scenario discovery, and the selection of robust solutions. Similar to the scenario discovery process in the RDM framework (Groves and Lempert, 2007), MORDM employs a statistical rule induction algorithm to identify the threshold values of deeply uncertain variables beyond which system performance falls below user-defined constraints in one or more objectives. This represents an important advancement given the long-recognized need to help decision makers understand the critical tradeoffs, dependencies, and vulnerabilities in their management decisions (Maass et al., 1962; Haimes and Hall, 1977; Brill, 1979). MORDM seeks to discover tradeoff solutions whose expected performance is optimal for our best available projection of the future, and whose performance degrades minimally with errors in our assumptions for deeply uncertain factors (i.e., robust satisficing behavior).

Several recent studies in the water resources literature have emphasized the importance of evaluating candidate strategies according to their robustness against future uncertainty. Hine and Hall (2010) perform information gap analysis to assess the sensitivity of flood management decisions to model uncertainty, determining the cost of achieving robustness to floods of different magnitudes. Hall et al. (2012) compare the RDM and Info-Gap (Ben-Haim, 2004) methods for identifying robust climate policies. Matrosov et al. (2013b) perform a similar methodological comparison for water supply infrastructure expansion in

the Thames Basin, UK. Both the RDM and Info-Gap methods focus on identifying the amount of uncertainty that can be tolerated while ensuring a specified level of performance; while RDM begins with a sample of the entire uncertainty space, Info-Gap proceeds outward from an initial "best-estimate" until it identifies thresholds causing poor performance (Matrosov et al., 2013b). Thus, in sensitivity analysis parlance, Info-Gap may be thought of as a local analysis centered at a particular reference policy or design, while RDM is a global analysis over plausible ranges of uncertainty. Finally, Brown et al. (2012) have proposed Decision Scaling, where future climate information is considered when estimating the subjective probabilities of impacts most relevant to water resources decision makers. Whereas traditional assessment of climate risks begins with a "top-down" downscaling of Atmospheric-Ocean General Circulation Model (AOGCM) projections and their associated uncertainties, Decision Scaling advocates a "bottom-up" analysis in which hydrologic and socioeconomic systems are first assessed to identify key climate thresholds that would trigger decision-relevant risks. Then, AOGCM projections can be employed in the final step of the analysis to assess the likelihood of these thresholds being crossed (Brown et al., 2012). For water resources planning in general, mapping climate change scenarios to probabilistic likelihoods remains a severely challenging technical question that has not been resolved (see Leung et al. (2013)). Decision Scaling assumes that discrete management alternatives have been specified *a priori* and, like traditional RDM and Info-Gap, provides a sense of robustness relative to decision-makers' key thresholds. Decision Scaling has successfully demonstrated the importance of robustness indicators for climate-related uncertainty (Moody and Brown, 2013), and has served as an example of bridging decision making and climate modeling (Weaver et al., 2013). MORDM distinguishes it-

self from the above methods by broadening decision-makers' abilities to discover new planning alternatives, understand their tradeoffs, and, in this study, quantify multi-stakeholder robustness dependencies.

This study advances the MORDM framework to evaluate the robustness of coordinated regional water supply portfolios for four water utilities in the "Research Triangle" region of North Carolina, USA, which contains the cities of Chapel Hill/Carrboro, Cary, Raleigh, and Durham. Facing increasing hydrologic uncertainty and rapid population growth, this region of nearly two million residents exemplifies the growing concern for water scarcity in the eastern USA. In prior work, Zeff et al. (2014) perform multi-objective optimization to explore tradeoffs between reliability and financial objectives in this region, concluding that a combination of demand management, transfers, and financial instruments are needed for the region to meet the cooperating utilities' requirements for supply reliability and financial stability. These financial instruments are intended to reduce the risk associated with the revenue and cost fluctuations that occur as a result of demand management during supply shortages; they include the option for self-insurance via annual payments into a contingency fund, as well as the option for third-party insurance contracts.

This study aids the four Research Triangle utilities in evaluating and implementing regional water supply portfolios combining demand management, transfers, and financial instruments as identified by Zeff et al. (2014). This study exposes those water supply portfolios to a wide range of hydroclimatic and economic uncertainties to determine vulnerabilities for the overall region as well as the individual utilities. After identifying these vulnerabilities, we identify key drivers and inter-utility dependencies that maximally impact the robustness

of the region's water supply through 2025. We propose MORDM for multiple stakeholders as a subset of Many-Objective Visual Analytics (MOVA), a decision aiding framework designed to facilitate iterative learning feedbacks between problem formulation and stakeholder preferences. MOVA and MORDM are described in full in Section 5.3. Figure 5.1 formally illustrates the four components of our proposed *a posteriori* multi-stakeholder strategy for evaluating and selecting regional water portfolios: (1) definition of stakeholders' multivariate performance thresholds (or requirements), (2) identification of deeply uncertain factors which may affect performance, (3) evaluation of the vulnerabilities, tradeoffs, and inter-stakeholder dependencies in planning alternatives, and (4) evaluative selection of robust solutions that balance individual and collective stakeholder interests. Importantly, the performance criteria used to define robustness in our study are developed in consultation with decision makers from the four water utilities in the Research Triangle region.

This study represents the first extension of the MORDM framework to a regional system with multiple interacting stakeholders. We employ performance requirements elicited directly from decision makers within the utilities to define the robustness of their shared regional water resources. Furthermore, it advances the MORDM approach to explore the improvements in robustness made possible by managing key uncertainties in the system. This is particularly valuable in a regional system, where the actions of each utility will affect the future robustness of the others. As water demand continues to grow in heavily populated urban centers and drought-induced supply shocks remain a threat, robust cooperative agreements between water utilities offer a promising path for managing water supply risks more efficiently.

5.3 Methodological Framework

5.3.1 Many-Objective Visual Analytics

Decision-making for "wicked" water resources problems, in which the problem formulations themselves are uncertain, requires iterative learning feedbacks across competing problem framing hypotheses and emerging stakeholder preferences as they discover what is possible in the systems being managed. Many-Objective Visual Analytics (MOVA) (Woodruff et al., 2013) provides a decision theoretic foundation to achieve this, in which problem formulation, many-objective optimization, design selection, and interactive visualization combine to provide learning feedbacks between modelers and stakeholders. MOVA formalizes the concept of constructive decision aiding, in which problem framing is performed interactively with stakeholder feedback (Roy, 1971; Tsoukias, 2008). In this study, we focus on two core components of the MOVA framework in particular: negotiated design selection and interactive visualization (Figure 5.1). These processes are critical to balancing optimality in our best available projections of the future state of the world versus robustness to deep uncertainties (Hitch, 1960; Clímaco, 2004), particularly for systems with multiple stakeholders. Figure 5.1 formally illustrates the four components of our proposed *a posteriori* multi-stakeholder strategy for evaluating and selecting regional water portfolios: (1) definition of stakeholders' multivariate performance thresholds (or requirements), (2) identification of deeply uncertain factors which may affect performance, (3) evaluation of the vulnerabilities, tradeoffs, and inter-stakeholder dependencies in planning alternatives, and (4) evaluative selection of robust solutions that balance individual and collective stakeholder interests.

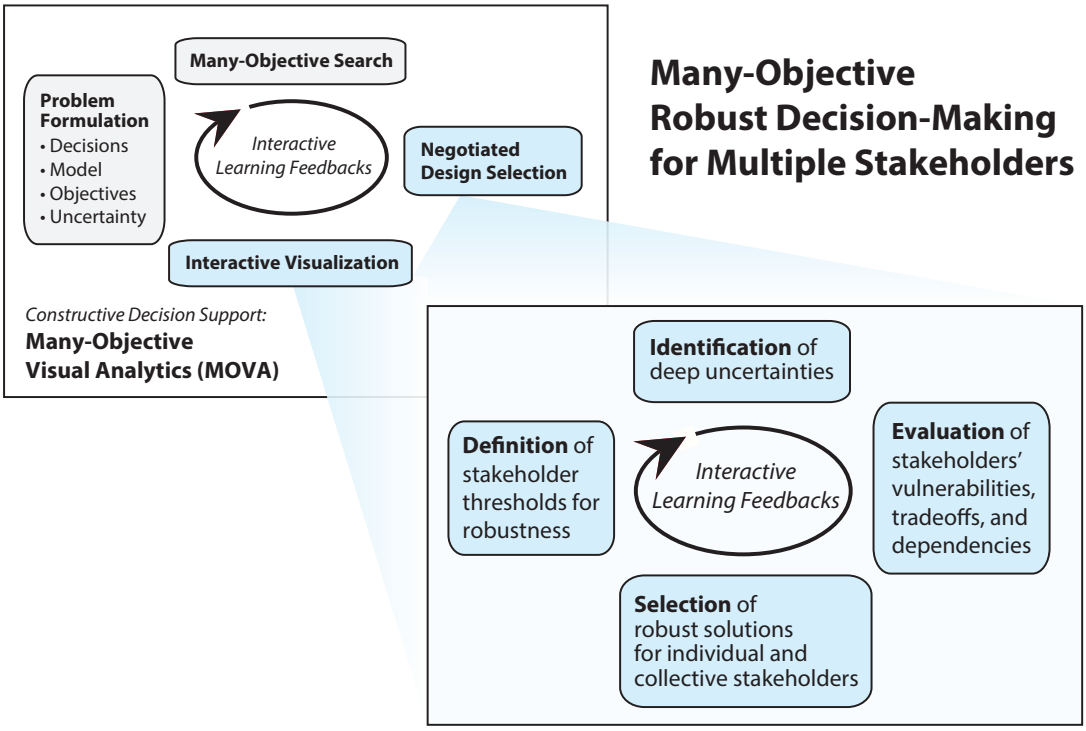

Figure 5.1: Many-Objective Robust Decision Making (MORDM) viewed as sub-processes of the broader Many-Objective Visual Analytics (MOVA) constructive decision support framework. MORDM focuses on two stages of the MOVA framework in particular: negotiated design selection and interactive visualization. The proposed approach consists of four interacting components, to identify a balance between individual and collective robustness for cooperating stakeholders.

Figure 5.1 illustrates that these four components are facilitated by interactive visualization and learning feedbacks.

In this work, we evaluate robustness as a posterior analysis step using a set of solutions discovered with multi-objective optimization in the projected future state of the world, following the MORDM approach (Kasprzyk et al., 2013). This must be distinguished from robust multi-objective optimization approaches (Deb and Gupta, 2006), such as (1) using the expected value of perfor-

mance across plausible scenarios, and (2) using the original objective function, subject to constraints on its deviation from the baseline. In robust optimization, the sampled uncertain values are understood to be based on some form of likelihood or probability distribution, since they are evaluated either in expectation or, conservatively, by guarding against extreme cases. This is suitable for well-characterized uncertainty, but perhaps not deep uncertainty: when the probability distributions themselves are uncertain, actions constrained to perform well across all states of the world may be overly conservative, leading to high opportunity costs or regrets. Assuming high fidelity model projections under well-characterized uncertainties, Pareto approximate solutions provide an optimal balance for competing performance objectives. MORDM performs a posterior sensitivity analysis on the Pareto set to identify those solutions whose performance degrades minimally if the future deviates from the assumptions used during optimization (i.e., they are Pareto satisficing).

5.3.2 Elicited Stakeholder Thresholds

In order to determine the acceptability of a solution's performance in alternative future states of the world, stakeholders must provide a set of minimum acceptable performance requirements. An advantage of the MORDM framework is that these requirements can be elicited while interactively showing stakeholders their attainable performance tradeoffs (i.e., *a posteriori* decision aiding). These requirements would typically be given relative to problem objectives (Figure 5.1) but may also include requirements for decision variables, internal system states, etc. In this work, we focus on requirements associated with specific objective values. The requirement thresholds may be expressed either as the mini-

mum acceptable level of performance, or a value of some variable below which failure occurs; the two need not necessarily intersect. A candidate solution is then given a binary evaluation with respect to these requirements. Furthermore, since MORDM allows performance thresholds to be defined on multiple variables, decision makers can specify whether the set of requirements should be evaluated as a union (where any of the criteria are met) or an intersection (where all criteria must be met).

5.3.3 Identification of Deep Uncertainties

The MORDM framework (Kasprzyk et al., 2013) builds on RDM (Lempert, 2002) to identify potential vulnerabilities in a system. In the prior RDM literature, a problem formulation is abstracted using "XLRM" notation: the deep uncertainties ("X"), levers ("L"), relationship ("R"), and measures ("M") (Lempert et al., 2003, 2006). Here, we substitute conventional optimization terminology: "decisions, model, and objectives" in lieu of "levers, relationship, and measures", respectively. The deep uncertainties ("X") represent exogenous factors with assumed (or estimated) likelihoods in the best available projections of the future state of the world, but which may vary significantly if projections are incorrect (e.g., population demands, climate, price elasticities, etc.). Thus, the key concern is that solutions found to be optimal in the expected state of the world may fail if the future state of the world deviates from modeled projections (i.e., how wrong can we be before it matters?). Following Bryant and Lempert (2010), we construct alternative future states of the world by sampling these deeply uncertain factors over plausible ranges. Each Pareto-approximate point from the many-objective search based on the expected state of the world is re-evaluated

in these alternative states of the world to evaluate the changes in performace that result from incorrect assumptions or unforeseen changes in the system. Thus, MORDM may be thought of as a type of global sensitivity analysis, in which we seek the deeply uncertain factors most responsible for changes in multiobjective performance. In this study, the projected future state of the world is captured by a baseline Monte Carlo simulation (i.e., our best available projections of expected performance). Deep uncertainties are explored using a Latin hypercube sampling (LHS) of the exogenous factors that had to be specified in our baseline state of the world. Each member of the LHS sample specifies the information needed for new Monte Carlo simulations of alternative future states of the world. More details on this step of the MORDM framework can be referenced in Kasprzyk et al. (2013).

5.3.4 Evaluation of Vulnerabilities, Dependencies, and Trade-offs

The *vulnerabilities* in a system are the ranges of deeply uncertain factors observed to cause performance degradation in excess of the elicited requirements. Following Lempert et al. (2008), we apply the Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999) to identify ranges of uncertain factors which are likely to cause poor performance (see Section 2.1.3). Once the system vulnerabilities have been identified, we assess which of these, if any, might be mitigated. Extending this concept, special consideration must be taken in the case of a multi-stakeholder system. Given a method to quantify the robustness of a solution (which will be discussed in the following subsection), we

would like to select solutions which provide future robustness for all stakeholders to the extent possible. However, in complex regional water supply systems, dependencies between stakeholders are often poorly understood, limiting decision makers' understanding of the benefits and key drivers associated with cooperative planning. PRIM provides a mechanism to help identify decision paths and critical deep uncertainties that highlight potential competitive tensions. Thus, the tradeoffs and dependencies between the stakeholders must be explored, acknowledging that robust solutions for one stakeholder may degrade the robustness of others. A core goal of our multi-stakeholder MORDM framework is to provide solutions that are maximally robust and acceptable for a region's major stakeholders.

5.3.5 Selection of Robust Solutions

Building on prior RDM applications (Lempert, 2002; Bankes et al., 2001; Bankes, 2002), the alternative future states of the world are constructed from deeply uncertain factors for which probability distributions are uncertain or unknown. Under this restriction, an appropriate measure of robustness is simply "satisficing over a wide range of futures" (Lempert and Collins, 2007). Hall et al. (2012) describe satisficing as "performing reasonably well compared to the alternatives over a wide range of plausible futures". This approach links to the satisficing criterion of Simon (1959) and to Starr's domain criterion (Starr, 1962; Schneller and Sphicas, 1983), which aims to minimize the volume of the scenario space in which a solution fails (or conversely, to maximize the volume in which a solution performs well). Following this logic, we formally define robustness as *the fraction of sampled states of the world in which a solution satisfies all performance re-*

quirements. This is similar to the robustness indicators proposed by Moody and Brown (2013), except (1) the performance criteria can be defined for many objectives, and (2) the scenarios are summed over an ensemble of discrete evaluations rather than integrated over a continuous space, since the simulation model in this case cannot be stated as a continuous function. Importantly, this satisficing criterion is not intended to reflect a probability or likelihood, only the fraction of sampled states of the world for which one or more requirements are met. Because the performance requirements can be defined on multiple objectives, including cost, this definition offers flexibility to avoid overly conservative solutions commonly associated with robustness defined in terms of reliability alone.

5.4 Model & Study Area

5.4.1 Research Triangle, North Carolina

The Research Triangle region is located in the state of North Carolina in the southeast U.S. (Figure 5.2a). The region's increased water stress reflects a broader trend across the eastern USA, where rapidly growing urban demand combined with the threat of climate-induced droughts have necessitated a transition in water planning to carefully address apparent water scarcity (Frederick and Schwarz, 1999; Lane et al., 1999; Seager et al., 2009; Griffin et al., 2013). The region's nearly two million residents are served by four major water utilities: Raleigh, Durham, Cary, and the Orange Water and Sewer Authority (OWASA, serving the towns of Chapel Hill and Carrboro), operating a total of nine reservoirs as shown in Table 5.1. Durham and OWASA operate their own reservoirs,

Table 5.1: Reservoirs and total storage capacities for the four Research Triangle Water utilities (adapted from Zeff et al. (2014)). Allocation refers to the total fraction of the capacity that has been allocated to the utility; Raleigh's 42.4% allocation to Falls Lake reflects 100% of the municipal supply allocation, while Cary's 12.7% of Jordan Lake reflects 39% of the municipal supply allocation.

Utility	Reservoirs	Total Capacity (BG)	Allocation
Durham	Little River Lake Michie	6.4	100%
OWASA	Cane Creek Stone Quarry University Lake	3.0	100%
Raleigh	Falls Lake Lake Wheeler Lake Benson	34.7	42.4%
Cary	Jordan Lake	45.8	12.7%

while Raleigh and Cary receive municipal supply allocations from reservoirs operated by the U.S. Army Corps of Engineers (USACE). These are Falls Lake and Jordan Lake, respectively. The total supply capacity of each water utility is shown in Figure 5.2b, with projected demand (to 2025) shown in Figure 5.2c. This is the minimum period over which the utilities are expected to remain without any supply expansion, given the increased difficulty of developing new infrastructure in the region (Postel, 2000; Scudder, 2005). Currently, a significant portion of the Jordan Lake municipal supply is unallocated, making it a valuable resource to help meet growing demands in the region (for details regarding current allocations, please see (Palmer and Characklis, 2009) and (Caldwell and Characklis, 2014). Cary has sole access to the water supply in Jordan Lake, but

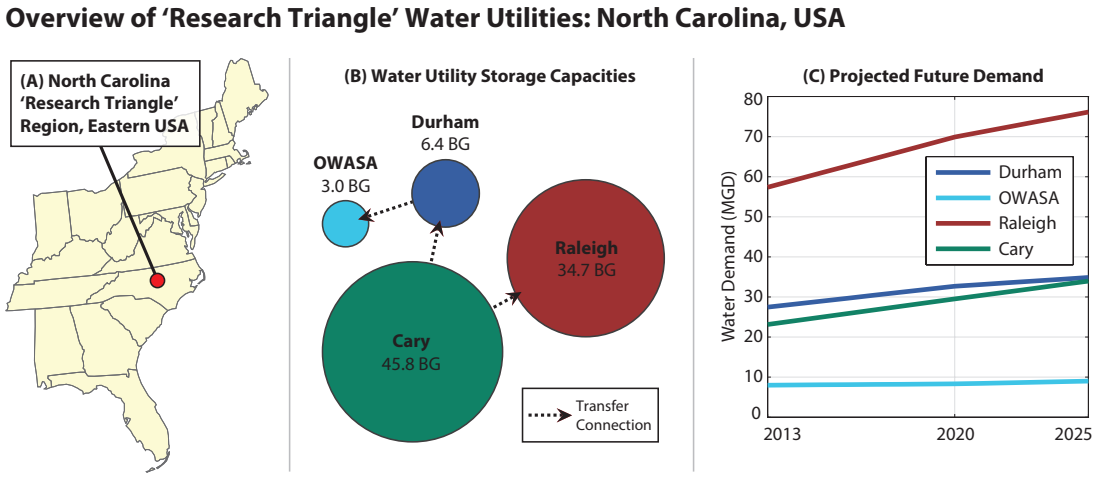

Figure 5.2: Overview of water supplies in the Research Triangle region, North Carolina, USA. (A) Location of the Research Triangle region in the southeast United States. (B) Four water utilities in the region, sized according to their respective supply capacities (BG). Transfer connections are noted with arrows. (C) Projected demand, 2013-2025 (MGD) for each of the four utilities.

as Figure 5.2 indicates, Cary does not experience nearly the largest demand. Thus, Cary has the opportunity to transfer surplus (treated) water from Jordan Lake to the other utilities in times of scarcity. Future allocations from Jordan Lake will represent an important policy decision, for which this type of regional portfolio analysis can propose potential solutions. For a detailed description of water supply in the Research Triangle region, please refer to (Zeff et al., 2014).

5.4.2 Simulation

The Research Triangle water supply model, introduced by (Zeff et al., 2014), simulates the water balance for each reservoir in the region on a weekly timestep over the period 2013-2025. The water balance for reservoir r is defined as fol-

lows:

$$\frac{\Delta S_r}{\Delta t} = I_{n_r} + I_{t_r} - D_r - E_r - C_r - W_r \tag{5.1}$$

where Δt = 1 week, I_{n_r} are the natural inflows, I_{t_r} are the inflows from transfers, D_r are the consumptive demands, E_r is evaporative flux from the reservoir, C_r is the required environmental flows, and W_r represents spillage:

$$W_r = \begin{cases} S_r - S_{max}, & S_r > S_{max} \\ 0, & S_r \leq S_{max} \end{cases} \tag{5.2}$$

Natural inflows I_{n_r} and evaporation E_r are generated synthetically based on 80 years of historical data. Each model evaluation is performed over 1,000 realizations of these synthetic time series to compute expected (baseline) performance accounting for hydrologic uncertainty within historical bounds. The synthetic series are generated using a modified Fractional Gaussian Noise method, which aims to reproduce the standard moments and seasonal variation in autocorrelation that define the dynamics of extreme events (Kirsch et al., 2013). Each term of the water balance is described in detail in the Technical Supplement to this paper.

Weekly demand for each utility D_u is computed by combining projected annual growth (provided by each water utility) with seasonal trends (with multipliers derived from historical data). The seasonal multipliers reflect, for example, increased demand in the summer months due to municipal outdoor use. Mandatory environmental flows C_r exist to protect the areas downstream of the reservoirs. These are implemented in the water balance model as required releases from the reservoirs, prior to any supply considerations.

In the model, the utilities have several options to manage water supply and financial risk. They can implement water use restrictions on consumers, which

preserves supply but causes revenue loss. They can also request water transfers from Jordan Lake, via Cary, in times of shortage. Restrictions and transfers are formulated as decision triggers on the perceived risk-of-failure for each utility. The risk-of-failure, ROF $\in [0, 1]$, is defined as the expected probability that a utility's reservoir storage will fall below 20% of total capacity at any point over the following 12 months. Thus, the metric depends on the current level of reservoir storage as well as the time of year. The ROF calculations have been provided by the utilities. In the event that transfer requests are triggered by multiple utilities simultaneously, the available treatment and conveyance capacity is divided proportional to the utilities' risk-of-failure levels as described by (Caldwell and Characklis, 2014). Thus, a utility facing a higher risk-of-failure will receive a larger proportion of the available transfer capacity.

The financial variability caused by these restrictions and transfers can be mitigated with two forms of insurance: either a self-insured contingency fund, or a third-party contract. Each utility's revenue is modeled based on monthly billing data from the period 2008-2011, where consumption patterns are divided by customer class (residential, commercial, etc.) (Zeff and Characklis, 2013). The relative usage in each tier is a function of seasonal water use patterns, so utility revenues are determined from a weighted average across all pricing tiers. Revenue losses from demand management are calculated as the difference between the revenues from restricted and unrestricted water use. As specified by the utilities, transfer water is currently assumed to have a constant price of $3,000 (USD) per million gallons, the same price charged by Cary to their lowest tier of residential consumers. The amount set aside for both types of insurance is defined as a percentage of annual volumetric revenue (AVR), which keeps the cost of insurance relative to the size of each utility.

5.4.3 Problem Formulations

(Zeff et al., 2014) optimize the Research Triangle water supply model for four problem formulations (sets of decision variables, objectives, and constraints), seeking to balance the tradeoffs between the utilities' reliability and financial objectives. In the most complex problem formulation, the utilities seek to optimize a vector of five objectives $F(\mathbf{x})$ using decision variables $\mathbf{x}$:

$$F(\mathbf{x}) = (f_{\text{REL}}, f_{\text{RF}}, f_{\text{JLA}}, f_{\text{EC}}, f_{\text{WCC}}) \tag{5.3}$$

$$\mathbf{x}_u = (T_{r_u}, T_{t_u}, JLA_u, SI_u, TPI_u),\ u = (1, 2, 3, 4) \tag{5.4}$$

where f_{REL} is the utility storage reliability (maximize), f_{RF} is the frequency of water use restrictions (minimize), f_{JLA} is the total allocation to Jordan Lake (minimize), f_{EC} is the expected cost of drought management (minimize), and f_{WCC} is the "worst-case" cost, i.e., the cost in the highest 1% of Monte Carlo realizations (minimize). The decision variables $\mathbf{x}_u$ for utility u are: T_r, the risk-of-failure (ROF) value that triggers restrictions; T_t, the ROF value to trigger transfer purchases; *JLA*, the Jordan Lake allocation (% of available capacity); *SI*, the annual contribution to self-insurance (contingency fund, % of annual revenue); and *TPI*, the percentage of annual revenue with which to purchase third-party insurance. Each term in the problem formulation is described in detail in the Technical Supplement to this paper.

The use of multiple problem formulations represents a constructive approach to problem framing (Roy, 1971; Tsoukias, 2008). We adapt the same four problem formulations to explore the sources of vulnerability potentially impacting the utilities. The problem formulations provide increasing levels of complexity to explore how regional cooperation and financial decisions impact

the utilities' objectives. The objectives are calculated as a min-max formulation such that the worst-performing utility in the region is optimized, ensuring that the other utilities perform at least as well. The purpose of this formulation is to push the search toward regions of better performance for all utilities simultaneously; while this optimization focuses on utilities facing more severe demand pressures, the results (to be described in Section 5.6) generally indicate that all utilities outperform the status quo with this strategy. The objectives are averaged over the 1,000 Monte Carlo simulations to account for potential variability in reservoir inflows and evaporation. Note that these Monte Carlo simulations represent the baseline projected future state of the world, since they are intended to reproduce the standard moments and autocorrelation of historical hydrology. This ensemble should not be confused with the sampling of deeply uncertain climate and economic variables that comprises MORDM, since these are intended to represent alternative future states of the world to explore deeply uncertain assumptions in our baseline model.

Tables 5.2 and 5.3 summarize the decisions and objectives for each problem formulation. In Formulation I, the utilities may only implement usage restrictions in the event of supply shortfall. In this formulation, they seek to maximize reliability, minimize restriction frequency, and minimize the average cost of mitigation. Formulation I captures the current drought management practices employed by the four water utilities. Formulation II adds the option to transfer water from Cary to the other utilities, with the additional objective of minimizing the total fraction of Jordan Lake allocated for this purpose. The allocation of Jordan Lake is minimized to reflect the utilities' concern for environmental quality, since this is the primary supply in the region with significant surplus capacity. Formulation III enables the utilities to implement a contingency fund,

Table 5.2: Decision variables for each water utility across the four problem formulations. Thus, the model contains a maximum of 20 decision variables in the most complex problem formulation.

Decision Variable	Lower Bound	Upper Bound	Formulation	Description
Restriction Threshold (Risk-of-failure)	0%	100%	I, II, III, IV	Triggers water use restrictions
Transfer Threshold (Risk-of-failure)	0%	100%	II, III, IV	Triggers transfer requests
Jordan Lake Allocation	(Current value)	100%	II, III, IV	Fraction of municipal supply
Self-Insurance Payment (AVR%)	0%	10%	III, IV	Annual fund contribution
Third-Party Insurance (discrete)	1	5	IV	Restriction stage to trigger payments

Table 5.3: Regional performance objectives across the four problem formulations. Objectives are calculated using a min-max formulation, where the worst-performing utility in the region is optimized. This guarantees that the other utilities will perform at least as well. Epsilon refers to the precision used to distinguish between the performance of different solutions during optimization.

Objective	Description	Min/Max	Formulation	Epsilon
Reliability	Probability that reservoir remains above 20% capacity in the worst-performing year (2013-2025)	Max	I, II, III, IV	0.1%
Restriction Frequency	Expected % of years (2013-2025) with at least one week of water use restrictions	Min	I, II, III, IV	5%
Average Cost	Expected revenue shortfalls, transfer costs, and mitigation expenses of a water supply portfolio	Min	I, II, III, IV	0.25%
Jordan Lake Allocation	Total Jordan Lake allocation granted to all four utilities	Min	II, III, IV	2%
Worst-Case Cost	Revenue shortfalls, transfer costs, and mitigation expenses in the worst 1% of simulations	Min	III, IV	0.5%

where a chosen fraction of annual revenue is saved to mitigate future revenue losses due to drought-related conservation. Finally, Formulation IV adds the option of third-party insurance against drought-induced revenue losses. Both Formulations III and IV contain the additional objective of minimizing worst-case costs (the worst-performing 1% of Monte Carlo simulations for a given water portfolio alternative). The sequence of four formulations was carefully designed to provide a constructive sequence of results to distinguish the effects of demand management using risk-of-failure thresholds, inter-utility transfers, and alternative cost mitigation strategies (see (Zeff et al., 2014) for additional details).

5.5 Computational Experiment

5.5.1 Multiobjective Optimization

To discover the tradeoffs in Formulations I-IV for managing the Research Triangle water supply system, (Zeff et al., 2014) employ the Borg multiobjective evolutionary algorithm (MOEA) (Hadka and Reed, 2013), which is described in detail in Section 2.2.2. The optimization was performed using 50 random seeds with 1 million function evaluations (NFE) per seed to ensure the best possible approximation to the Pareto-optimal set.

5.5.2 Identification of Deep Uncertainties

A key component of the MORDM approach is the identification of the deeply uncertain variables used to construct the ensemble of alternative future states of the world. For the Research Triangle baseline water supply model, we identified thirteen deeply uncertain factors and assigned sampling ranges to each, as summarized in Table 5.4. These hydroclimatic and economic variables were selected to reflect critical assumptions in the baseline Monte Carlo model used in the many-objective optimization. Note that these factors could vary significantly with unforeseen changes in the system (i.e., if current projections are accurate, but the system changes) or errors in the projections themselves; the RDM framework does not distinguish between the two. The variable ranges in Table 5.4 represent plausible rather than probable values. They provide a mechanism for understanding "how wrong" our baseline model assumptions can be before significant vulnerabilities to deep uncertainties occur. In order to perform an initial screening of the robustness of candidate solutions, we sample scenarios from the uncertain factors in Table 5.4 using an uncorrelated Latin Hypercube sample.

The uncertain factors listed in Table 5.4 can be described mathematically as follows. The uncertain factors are assigned to d_i, $i = (1, 2, \ldots, 13)$, respectively. In general, the uncertainties modify several terms in the water balance:

$$\frac{\Delta S_r}{\Delta t} = \tilde{I}_{n_r} + I_{t_r} - \tilde{D}_r - \tilde{E}_r - C_r - \tilde{W}_r \tag{5.5}$$

where the spillage depends on the modified capacities $\tilde{S}_{max}$:

$$\tilde{W}_r = \begin{cases} S_r - \tilde{S}_{max}, & S_r > \tilde{S}_{max} \\ 0, & S_r \leq \tilde{S}_{max} \end{cases} \tag{5.6}$$

Table 5.4: Uncertainty factors and sampling ranges for Many-Objective Robust Decision Making

Category	Name	Current	Lower	Upper
Climate	Inflows Multiplier	1.0	0.8	1.2
	Evaporation Multiplier	1.0	0.8	1.2
Demand	Consumer Reductions Multiplier	1.0	0.8	1.2
	Consumer Reductions Lag (weeks)	0	0	4
	Mean Peaking Factor	1.0	0.5	2.0
	Demand Growth Multiplier	1.0	0.5	2.0
	Standard Deviation of Demand Variations	1.0	0.5	2.0
Capacity	Falls Lake Municipal Supply Allocation	1.0	0.8	1.2
	Jordan Lake Municipal Supply Allocation	1.0	0.8	1.2
	Cary Treatment Capacity Multiplier	1.0	1.0	2.0
	Transfer Connection Capacity Multiplier	1.0	1.0	2.0
Costs	Transfer Cost ($/MG)	3000	2500	5000
	Insurance Premium Multiplier	1.2	1.1	1.5

The following subsections will describe how the uncertain factors in Table 5.4 modify the individual terms in Equation 5.5.

Climate & Hydrology

The scaling factors d_1 and d_2 on the reservoir inflows and evaporation, respectively, are used as a simple proxy for near-term changes in climate through 2025. They are simply multiplied by the baseline inflows and evaporation each week for each reservoir; the same multipliers are used for all reservoirs in the region.

$$\tilde{I}_{n_r} = d_1 * I_{n_r} \quad \forall r \tag{5.7}$$

$$\tilde{E}_r = d_2 * E_r \quad \forall r \tag{5.8}$$

Demand

The magnitude of the restrictions depends on the consumer reductions factor, $CRF_{s,t} \in [0, 1]$ which in turn depends on the restriction stage s and season t. In reality, the reduction in water use under restrictions is uncertain—it may be higher or lower than the utilities' estimates. Thus we modify the consumer reductions factor such that $\widetilde{CRF}_{s,t} = d_3 * CRF_{s,t}$. Similarly, the modified demand for each utility is calculated from the unrestricted demand as $\tilde{D}_u = \widetilde{CRF}_{s,t} * UD_u$. In the baseline model, we assume that when a utility stops water use restrictions, the consumers instantly resume their unrestricted water use. This is a best-case assumption for the utilities, as they will resume their expected revenue stream as quickly as possible. A potential concern is "demand hardening", where consumers adjust their water use to comply with restrictions and are slow to resume their normal use patterns when restrictions are stopped. For this reason, we explore the impact of the consumer reductions lag, $d_4 \in [0, 4]$, which specifies the number of weeks after stopping restrictions before consumers resume their normal water use. In the baseline model, $d_4 = 0$. The baseline model developed by (Zeff et al., 2014) employs mean unit demand factors $\bar{\delta}_{u,w}$ for utility u in week w, which are sample averages of historical data. The mean unit demand factors average to one over the year, i.e., $\frac{1}{52} \sum_{w=1}^{52} \bar{\delta}_{u,w} = 1$. These factors incorporate seasonal fluctuations in demand such that the weekly demand for each utility is:

$$D_{u_w} = D_{u_y}(\epsilon s_{\delta_w} + \bar{\delta}_{u,w}) \tag{5.9}$$

where s_{δ_w} is the sample standard deviation of the historical unit demand in week w, and ϵ is a random number drawn from the joint inflow-demand distribution. Since seasonal demand fluctuations are uncertain, we explore the impact of the "mean peaking factor" d_5, which modifies the unit demand value such that:

$$\tilde{\bar{\delta}}_{u_w} = (\bar{\delta}_{u,w} - 1)d_5 + 1 \tag{5.10}$$

A higher value of d_5 increases the emphasis on seasonal demand peaks; since demand in the region is typically higher during irrigation season, this has the effect of biasing the distribution of demand toward summer months. Importantly, it does not change the total annual demand D_{u_y}, only its weekly distribution throughout the year. Similarly, the standard deviation of demand variations is scaled by the factor d_7 to explore the impact of wider fluctuations:

$$\tilde{s}_{\delta_w} = s_{\delta_w} d_7 \tag{5.11}$$

The demand growth multiplier d_6 is used to scale the slope of demand growth. Each utility u has provided projections D_{u_y} of demand in year y through 2025, assuming linear growth. We modify the slope of this projection as follows:

$$\tilde{D}_{u_y} = D_{u_0} + (D_{u_y} - D_{u_0})d_6 \tag{5.12}$$

where D_{u_0} is the projected demand in year 0 of the simulation.

Capacity

Infrastructure uncertainties facing a utility may include: loss of storage capacity due to infrastructure failure, changing infrastructure for *other* utilities (since the region is connected via transfer agreements), and environmental legislation reducing the available fraction of municipal supply from existing sources, among

others. These uncertainties are captured in the factors d_8, d_9, d_{10}, and d_{11}. The baseline municipal supply of Falls Lake, 14.7 BG, is scaled by the factor d_8. Similarly, the factor d_9 is used to scale the municipal supply of Jordan Lake, which in the baseline model is 14.9 BG under current conditions with the majority unallocated. The Cary treatment capacity multiplier (d_{10}) is multiplied by the current treatment capacity (56 MGD) to reflect a potential increase or decrease in treatment capacity. Finally, the transfer connection capacity multiplier (d_{11}) is multiplied by all three transfer connections in the region.

Costs

In the baseline model developed by (Zeff et al., 2014), water transfers have a fixed cost of \$3,000/MG. This is the price currently charged by Cary to their lowest tier of residential consumer. The price of transfers may fluctuate in the future, and can be considered a significant uncertainty for Durham and Raleigh, which plan to rely on Cary for transfers during shortfalls. In this work, the cost of transfers is sampled over the range \$2,500/MG–\$5,000/MG, represented by the uncertainty factor d_{12}. Finally, the insurance premium paid by the utilities for third-party insurance to protect against the cost of drought mitigation is considered uncertain. (Zeff et al., 2014) considered an insurance premium of 20%; here, we sample the uncertain factor d_{13} over the range 10%–50% to explore its impact on the financial stability of the utilities.

It is useful to visualize how the deep uncertainty ensemble modifies the statistical distributions of key factors relative to our baseline Monte Carlo simulation. Figure 5.3 shows an example of these impacts for Raleigh. The top panel contains the range of flow duration curves describing the inflows to Falls Lake.

The black line represents the historical flow duration curve. The yellow area represents the range of inflows generated by the baseline Monte Carlo simulation performed by (Zeff et al., 2014). Finally, the blue area shows the range of flow duration curves constructed by sampling the inflows multiplier shown in Table 5.4. The latter provides a dramatically larger range of inflows, with increased frequency of extreme values. The bottom panel of Figure 5.3 contains a similar plot for demand growth. The current baseline demand projection for Raleigh is shown in black, while the projections associated with each sampled value of the demand growth multiplier (Table 5.4) are colored according to their value of that multiplier. Again, the sampled range of potential demand growth is quite large; simulations performed over this range will identify the point above which performance degrades significantly. The remaining deeply uncertain factors in Table 5.4 produced similar effects to those illustrated in Figure 5.3.

5.5.3 Experiment Design

The deeply uncertain factors defined in Table 5.4 were used to construct a 13-dimensional Latin Hypercube Sample of 10,000 alternative future states of the world. This is not to be confused with the Monte Carlo ensembles of streamflow, evapotranspiration, and demand endogenous to each model evaluation, which represents relatively well-characterized uncertainty compared to the deeply uncertain factors in Table 5.4. Each of the Pareto-approximate water portfolios found by (Zeff et al., 2014) was evaluated in each of these 10,000 deeply uncertain states of the world. The individual performance objectives for each water utility were recorded for each evaluation. An open-source implementation

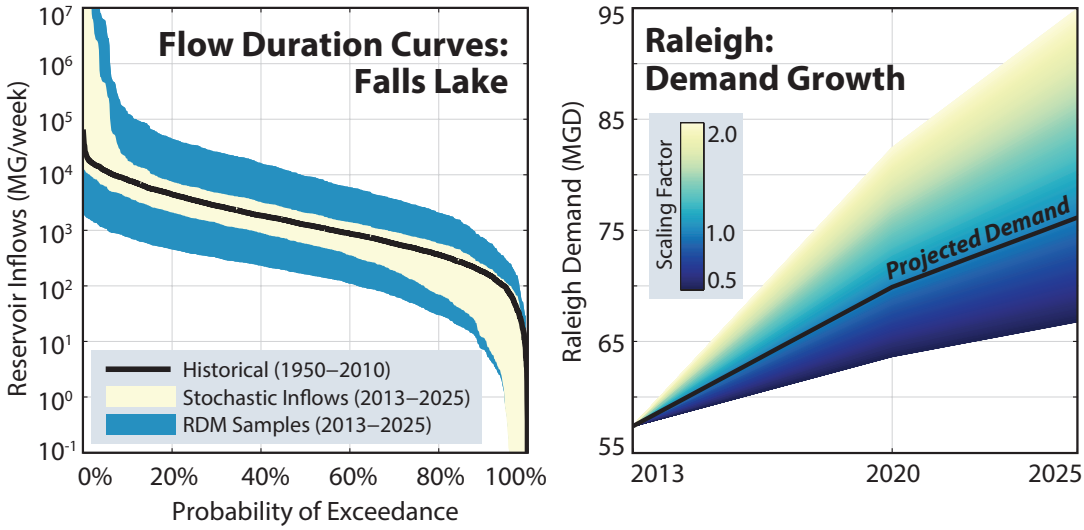

Figure 5.3: Example impacts of sampling deeply uncertain inflow and demand factors. The top panel shows the ranges of flow duration curves for inflows into Falls Lake, where the baseline Monte Carlo samples used by (Zeff et al., 2014) are represented by the yellow range, and the extended ranges for MORDM samples are shown in blue. The bottom panel shows the sampled range of future demand growth for Raleigh, colored according to the demand growth multiplier given in Table 5.4.

of PRIM (Bryant, 2009) was then used for scenario discovery to identify the thresholds of deeply uncertain factors responsible for critical system failures. The model evaluations were performed in parallel using 400 processors of the CyberSTAR high-performance cluster at Penn State University, which contains a combination of quad-core AMD Shanghai processors (2.7 GHz) and Intel Nehalem processors (2.66 GHz).

5.6 Results

5.6.1 Multiobjective Optimization

The Pareto-approximate solutions for each of the four problem formulations found by (Zeff et al., 2014) are displayed on parallel axes in Figure 5.4. The x-axis shows the four regional objectives in the min-max formulation. The Jordan Lake allocation objective is omitted from this plot because it is also a decision variable. The y-axis of each subplot indicates the value of the objective, where ideal performance would be represented by a horizontal line intersecting the bottom of the Reliability, Restriction Frequency, Average Cost, and Worst-Case Cost axes. Diagonal lines across the axes represent strong conflicts between the respective objectives. Figure 5.4 indicates that the more complex formulations, which include financial instruments, provide a more diverse set of options for the decision maker, particularly regarding the ability to reduce worst-case costs. These Pareto-approximate solutions serve as a starting point for our robustness analysis in this study. We aim to facilitate the utilities' selection of robust portfolio instruments (i.e., problem formulations) as well as their specific implementations (i.e., specific solutions).

In a practical decision context, the full range of Pareto-approximate solutions may not be of interest to the stakeholders. For example, several of the solutions in Figure 5.4 contain reliability values near 92%, far lower than the water utilities would tolerate. We can restrict the solutions shown by filtering (or "brushing") the solutions according to a set of constraints. The Research Triangle water utilities have expressed the following set of performance criteria: reliability greater than 99%, restriction frequency less than one in five years

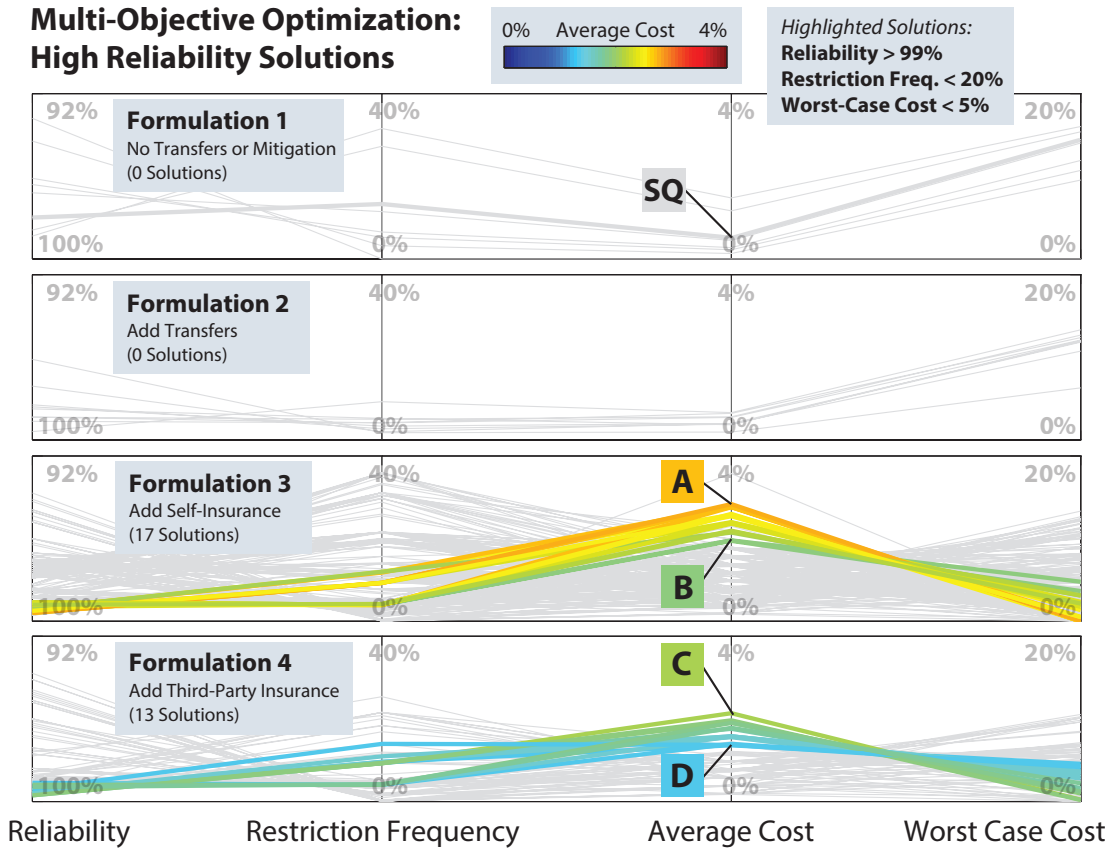

Figure 5.4: Pareto-approximate solutions for each problem formulation as optimized by (Zeff et al., 2014), shown on a parallel coordinate plot. The x-axis shows the four regional objectives in the min-max formulation. The y-axis shows the value of the objective, with the ideal point along the bottom. These solutions represent the non-dominated strategies for the region in the projected future state of the world, and thus provide a valuable starting point for the robustness analysis. Solutions in color satisfy the performance criteria specified by the water utilities: Reliability > 99%, Restriction Frequency < 20%, and Worst-Case Costs < 5%. Solutions failing to satisfy any of these criteria are shown in gray. Four solutions (A–D) are highlighted for further analysis, along with a representative status quo solution (SQ) for the system without transfers or cost mitigation. Note that the simpler problem formulations (which do not include financial instruments) are unable to produce any solutions satisfying the performance criteria.

(20%), and worst-case costs below 5% of annual volumetric revenue. The solutions satisfying these conditions are shown in color Figure 5.4, while those failing to meet the conditions are shown in gray in the background. Figure 5.4 suggests several salient points. First, only Formulations III and IV (which contain financial instruments) are capable of satisfying the utilities' expressed preferences for the region as a whole. This underscores the importance of the financial instruments introduced by (Zeff and Characklis, 2013) and extended by (Zeff et al., 2014). Second, the high-reliability solutions typically have high average cost, since this forms a tradeoff against the other objectives. We select four high-reliability solutions (A–D) in Figure 5.4 to be studied in more depth in our robustness analysis. These solutions represent regional water portfolios that the utilities would likely choose when only considering baseline Monte Carlo evaluations (i.e., the projected future state of the world).

Here, we investigate whether these solutions are robust to uncertainty in addition to being optimal. A core contribution of this work is the MORDM framework's ability to falsify competing formulation hypotheses (Popper, 1963) while also providing a formal mechanism for selecting one or more appropriate solutions from the Pareto approximate sets for further consideration. The *a posteriori* multiobjective decision support literature dating back to (Cohon and Marks, 1973) has assumed that solutions should be selected based solely on their relative tradeoffs, but has failed to consider how multi-stakeholder robustness varies throughout the Pareto-approximate set as addressed here.

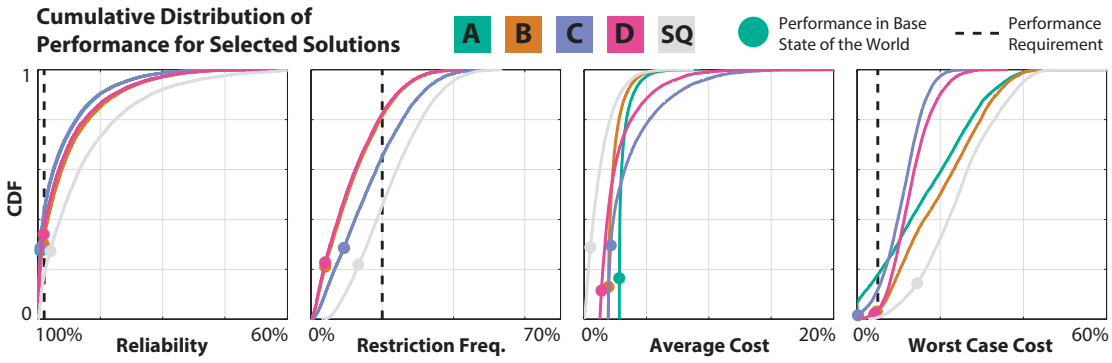

Figure 5.5: Distribution of solution performance across the 10,000 alternative states of the world sampled in the MORDM framework. Five solutions are shown: (A–D), representing high-reliability solutions found in Figure 5.4, along with the status quo solution (SQ) representing no transfers or mitigation. The performance in the expected state of the world is shown with a dot, and the thresholds specified by the water utilities are shown as dashed lines. Across the sampled states of the world, performance degrades dramatically, with a large fraction of the simulations falling well below the performance criteria expressed by the utilities. This indicates that even the solutions which are Pareto-approximate in the projected future remain vulnerable to deep uncertainties.

5.6.2 Solution Degradation under Uncertainty

Solutions (A–D), selected in Figure 5.4, represent an advanced approach to multi-objective decision making, in which the set of Pareto-approximate solutions is filtered *a posteriori* subject to stakeholder preferences. However, because these solutions were optimized according to the baseline Monte Carlo simulation, they may be biased to a particular set of model assumptions. The first task of our robustness analysis is simply to explore the changes in performance that occur as a result of sampling the sources of deep uncertainty given in Table 5.4 to create alternative future states of the world. Figure 5.5 shows the cumulative distributions of four performance objectives across the 10,000 states

of the world. The expected performance (i.e., the Pareto-approximate performance obtained in the baseline state of the world) is shown with a dot for each objective. Five solutions are shown: A–D, which represent high-reliability solutions, and the status quo solution (SQ), representing the current system with no transfers or financial mitigation. The performance thresholds elicited from the water utilities are shown with dashed lines. Clearly, the alternative states of the world sampled here create significant vulnerabilities for the regional water supply system in all objectives. The high-reliability solutions A–D outperform the status quo solution in all objectives (except average cost) in almost all states of the world, yet their performance in many states of the world would not be acceptable to the water utilities. Reliability falls well below 80% for many of the scenarios, with worst-case costs exceeding 20% of annual volumetric revenue. Figure 5.5 clearly indicates that even the solutions found via advanced multi-objective optimization methods—which are Pareto-approximate in the baseline Monte Carlo simulation—may degrade significantly when exposed to sources of deep uncertainty.

5.6.3 Robustness for Individual Stakeholders

The robustness of each solution can be determined using the approach described in Section 5.3.5 (i.e., the fraction of states of the world in which each solution meets the utilities' expressed performance criteria). Although (Zeff et al., 2014) performed the initial multi-objective optimization using a min-max approach for the region, we can further explore the robustness of each solution for each utility to better understand the multi-stakeholder tradeoffs. In this study, we are primarily interested in the Pareto-optimal solutions from Formulations

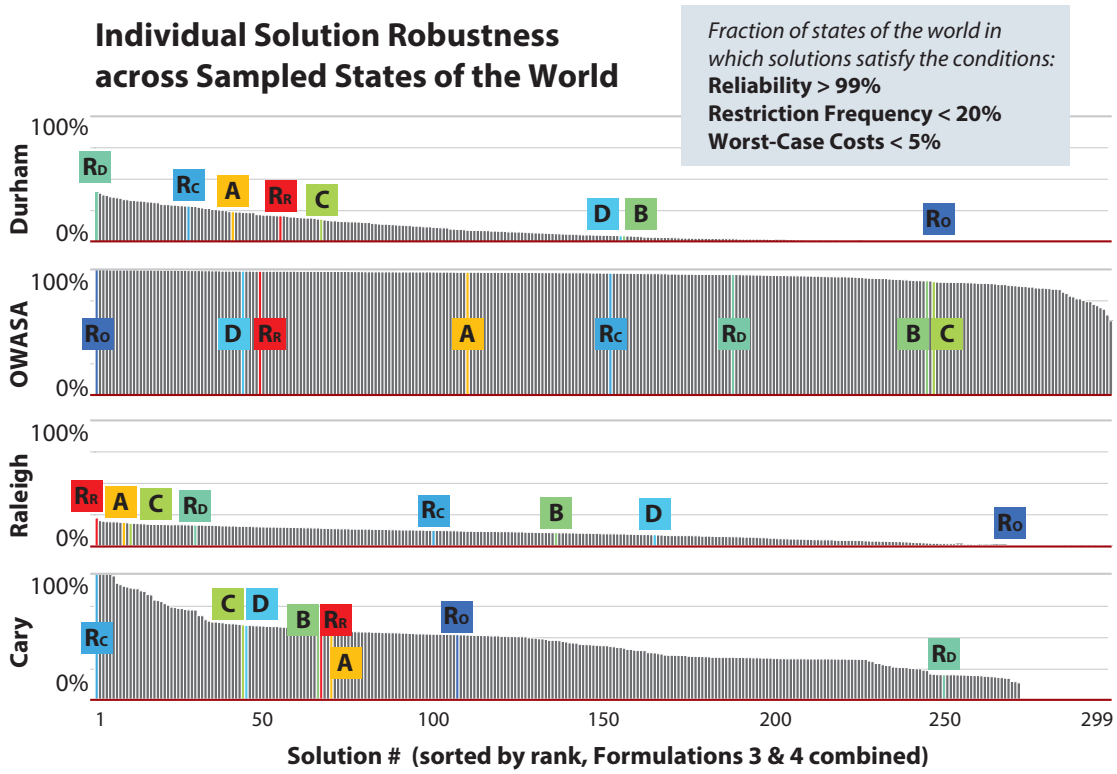

Figure 5.6: Robustness of all Pareto-approximate solutions from Formulations 3 and 4 (299 solutions in total), for each of the four water utilities. Robustness, in this case, is defined as the fraction of sampled states of the world in which a solution meets the utilities' expressed performance criteria, which should not be interpreted as a probability or likelihood. Solutions of interest are highlighted: A–D, from Figures 5.4 and 5.5, and the most robust solutions for each utility (R_D, R_O, R_R, and R_C). The maximum achievable robustness varies significantly between utilities despite their geographic proximity. Furthermore, the robustness of each utility trades off against the others, creating challenges for regional portfolio planning.

3 and 4 (299 solutions in total), since these offer the potential for water transfer and financial mitigation. In addition, they are the only formulations capable of meeting the performance criteria in the baseline Monte Carlo simulation (see Figure 5.4).

Figure 5.6 shows the robustness of these solutions for each of the four water utilities. Robustness values range from 0%, where the solution fails to meet the performance criteria in all alternative future states of the world, to 100%, where the solution satisfies the criteria in all alternative future states of the world. The solutions are ranked in descending order of robustness for each utility. Several solutions of interest are highlighted. First, the solutions A–D identified in Figures 5.4 and 5.5 are shown. While solutions A–D provide near-optimal performance in the baseline Monte Carlo simulation, they are clearly not the most robust to uncertainty, and none of them provides a high degree of robustness for all utilities simultaneously. Second, using the results of this plot, we identify the most robust solution for each utility, denoted with a subscript (R_D for Durham, R_O for OWASA, etc.). The highlighted solutions are colored according to their average cost, following Figure 5.4, where blue represents low average cost and red indicates high average cost.

Perhaps the most striking feature of Figure 5.6 is the large difference in achievable robustness for each utility, if the conditions in the region deviate from those assumed in the baseline Monte Carlo simulation. For example, OWASA demonstrates very high robustness across the sampled states of the world, nearing 100% for many of the solutions. Durham achieves a maximum of 40% robustness for its best solution, R_D. Raleigh only achieves roughly 20% robustness for its best solution, R_R. Finally, Cary shows the potential to achieve 100% robustness for several solutions, but falls in the range 40-60% for the majority of solutions. It is clear that despite their geographic proximity, the four utilities face very different challenges going forward. Particularly for the large urban areas of Raleigh and Durham, significant vulnerabilities may occur. This finding underscores the need to consider the robustness of candidate water

portfolios in addition to their optimality in the baseline Monte Carlo simulation.

Figure 5.6 also shows significant inter-utility robustness tradeoffs. For example, the most robust solution for Durham (R_D) is one of Cary's least robust solutions, because Cary must transfer large amounts of water for Durham to attain this performance. As the supplier of water transfers to the region, Cary has the potential to suffer significant reductions in robustness if the larger cities (i.e., Durham or Raleigh) encounter significant shortages. Meanwhile, Raleigh's most robust solution (R_R) provides reasonable, but not optimal, performance for the other utilities. This raises interesting questions, such as whose robustness should be considered foremost when the utilities agree to act as a regional cooperative, and what are the consequent risks and costs for this priority. Would utilities accept a small reduction in robustness in order to achieve a large improvement for another utility? These issues, critical to the regional planning context, are exemplified in Figure 5.6.

5.6.4 Identification of Key Uncertainties

A first step for improving the solution robustness values shown in Figure 5.6 is to identify the key sources of vulnerability and attempt to mitigate them. Following the MORDM framework, we use the Patient Rule Induction Method (PRIM) (Friedman and Fisher, 1999) for this purpose. PRIM identifies the uncertain variables most responsible for causing performance failures in one or more criteria, along with the ranges of those variables most likely to cause failure. Figure 5.7 summarizes the results from PRIM for selected solutions of inter-

est. The two deeply uncertain variables most responsible for causing failure in one or more objectives are the demand growth scaling factor, and the Falls Lake municipal supply scaling factor. Of these, the former is most important for all tested solutions, while the latter only appears for solutions C, D, and R_O. The vulnerable ranges for the demand growth multiplier indicate that essentially any regional demand growth in excess of the current projections will render the regional performance goals unachievable. Notable for their absence here are the climate uncertainties, the inflow and evaporation multipliers. The results from PRIM indicate that demand growth is the most critical uncertainty in this system, and the one which must be addressed first if the robustness of the regional portfolio and the individual utilities is to be improved.

5.6.5 Cooperative Demand Reduction

Given that demand growth is the key uncertainty facing water utilities in the Research Triangle region, we analyze the impact of restricting demand growth on the robustness of each utility. The original robustness values shown in Figure 5.6 reflected exposure to demand growth multipliers up to a maximum value of 2.0 (double the current projection). Here, we filter the tested states of the world according to decreasing maximum values of demand growth: 1.5, 1.1, 1.0, 0.9, and 0.8. For example, the latter of these limits examines only the deeply uncertain states of the world with demand growth multipliers less than or equal to 80% of the current projection (i.e., a reduction in demand growth from approximately 3% to 2.4%). The rate of demand growth dominates the other hydroclimatic and economic uncertainties over the 2025 planning horizon. A reduction in growth rate from 3% to 2.4% is relatively modest and should prove feasible

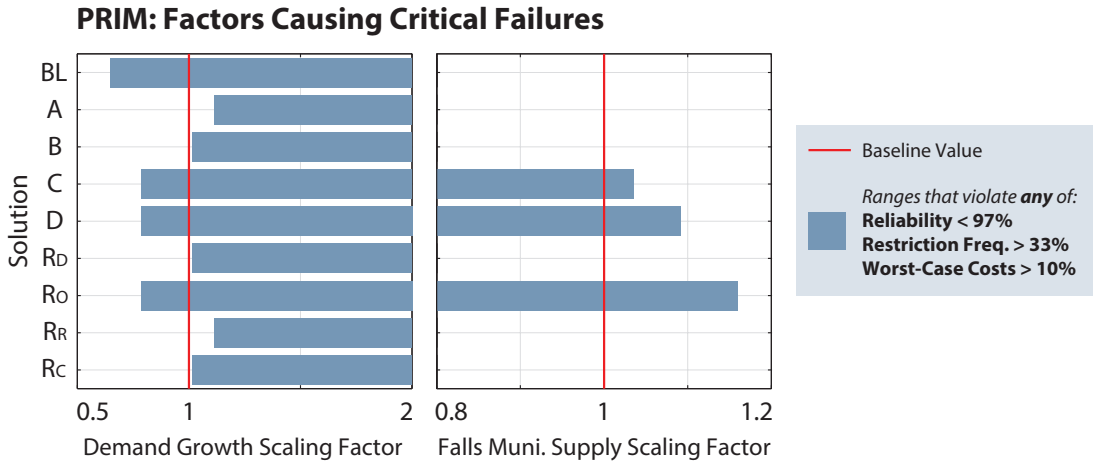

Figure 5.7: Results from the Patient Rule Induction Method (PRIM), performed for a set of selected solutions. The two key uncertainties identified for the regional system are the demand growth multiplier (left) and the Falls Lake municipal supply multiplier (right). The other uncertainties listed in Table 5.4 were not identified as having a significant influence on performance failure, including the climate uncertainties (inflow and evaporation). The vulnerable ranges for each uncertainty are indicated by the shaded bars. Importantly, for all tested solutions, demand growth in excess of the current projection (i.e., greater than 1.0) is likely to cause vulnerability.

for the utilities to achieve collectively. The potential benefits of implementing these limits are shown in Figure 5.8. Note here that we do not evaluate the ability of the water utilities to achieve these limits; we simply state the potential gains in robustness made possible under each restriction level. However, the utilities believe that such a reduction in projected demand growth is achievable given construction of new homes and improved water use efficiency. In each subplot of Figure 5.8, the solutions are ranked from left to right along the x-axis in descending order of their original robustness values (i.e., with demand growth multipliers up to 2.0) for each utility.

Figure 5.8 indicates substantial improvements in robustness for all utilities,

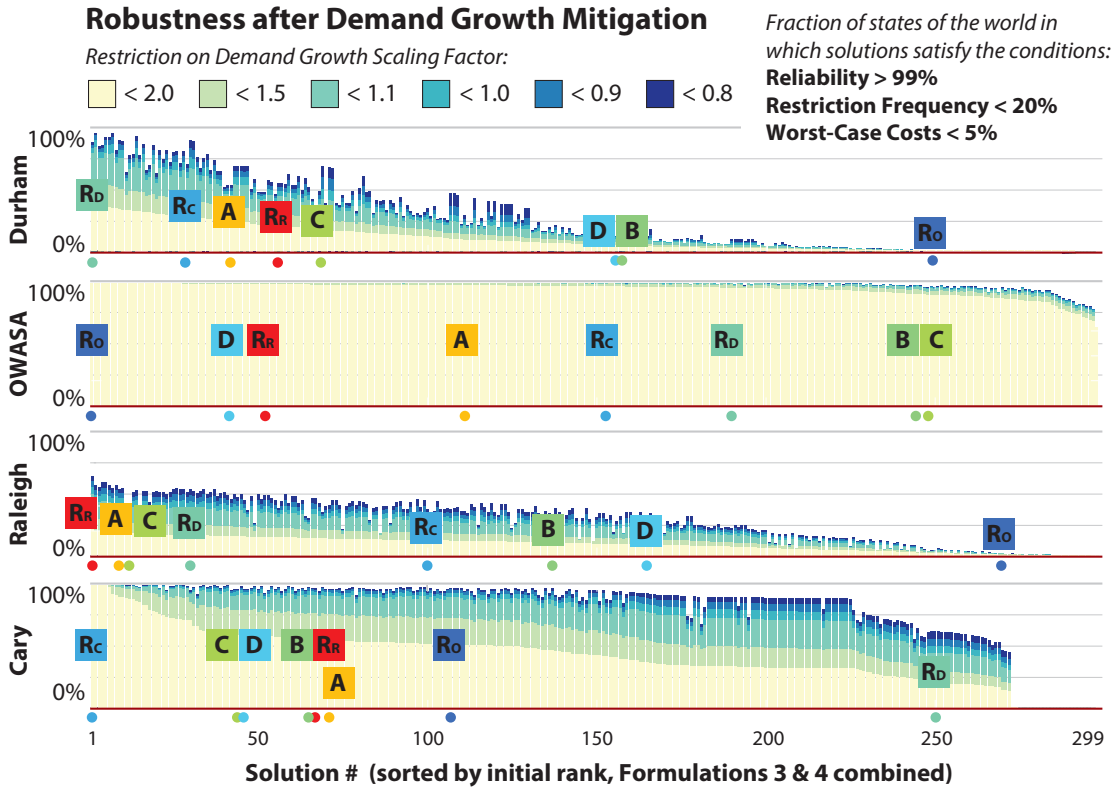

Figure 5.8: Solution robustness for each utility at different levels of restriction on the demand growth scaling factor. These demand growth values could represent cooperative actions taken by the utilities to mitigate the vulnerabilities associated with uncertainty in demand growth. Comparing to Figure 5.6, restricting demand growth causes clear benefits for all utilities. Under this strategy, Durham has the opportunity to achieve robustness values greater than 90%, and Raleigh can exceed 60%. Again, these robustness values reflect the fraction of sampled states of the world in which criteria are met, and should not be interpreted as a probability or likelihood. Cary shows the most consistent improvement, as solutions which demonstrated poor robustness in Figure 5.6 are now greater than 90%.

particularly if demand growth were to be restricted below the currently projected rate (a scaling factor of 1.0). The yellow bars in Figure 5.8 reflect the robustness under a maximum demand growth multiplier of 2.0, which is the original setting shown in Figure 5.6. The ordering of solutions on the x-axis reflects

this original setting, and the reductions in demand growth may improve the solutions by different amounts based on their unequal dependence on particular decision instruments. Durham shows an improvement in maximum achievable robustness from roughly 60% to over 90%, if demand growth were to be confined below 80-90% of the currently projected rate. Raleigh improves from a maximum of only 20% to roughly 60%. Cary shows consistent improvement, with the majority of solutions exceeding 90% robustness. OWASA ostensibly benefits least from these restrictions, but only because their robustness in the absence of demand growth mitigation already measures close to 100% for the majority of solutions. Importantly, the robustness improvements shown in Figure 5.8 reflect cooperative demand reduction. The demand growth multipliers are applied to all utilities equally, and thus mitigation strategies on the part of any utility could potentially create benefits for all others depending on the transfer agreements contained in a given portfolio. There is currently no incentive system in place to encourage utilities for whom robustness is easily attained to reduce demand growth to benefit others. However, our results indicate that such incentives could be designed to improve the robustness of regional water supply.

5.6.6 Robustness Tradeoffs between Utilities

The improved robustness values shown in Figure 5.8 provide a basis for selecting recommended regional water portfolios in conjunction with demand growth mitigation efforts. However, it is clear from Figure 5.8 that the robustness of each utility trades off against the others—that is, the selection of a solution may be complicated by competitive effects between the utilities. There-

fore, these tradeoffs should be explored directly when choosing a solution, as shown in Figure 5.9. Recall in Figure 5.4 we selected solutions A–D according to their performance in the baseline Monte Carlo simulation, which represents traditional *a posteriori* multiobjective decision support that ignores inter-utility robustness tradeoffs. Alternatively, Figure 5.9 illustrates the selection of regional water portfolios based on their demonstrated multi-stakeholder robustness across alternative states of the world.

Figure 5.9 illustrates the robustness tradeoffs between water utilities using parallel axes (5.9a) and glyph views (5.9b-c). Instead of performance objectives, the four metrics under consideration are the robustness values for each of the utilities. The candidate portfolios are plotted six times, once for each demand growth restriction level, with colors corresponding to each level. For example, the tradeoffs between utility robustness values are visible in the parallel axis plot in Figure 5.9a. Each line represents a candidate portfolio, and tradeoffs between utilities are indicated by crossing lines. A sharp tradeoff exists between Cary and both Raleigh and Durham, particularly when the demand growth is not limited (up to 2.0). This is a sensible result since Cary provides water transfers to both Durham and Raleigh, and will be stressed if regional demand growth continues unabated with transfer agreements in place. As seen in Figure 5.8, OWASA performs well regardless of the demand growth rate. Durham and Raleigh show a moderate tradeoff, although not as sharp as might be expected given their large demands. The Durham and Raleigh robustness tradeoffs become more pronounced for cooperative regional demand reduction scenarios with growth rate multipliers below 1.0.

Figures 5.9b and 5.9c provide an alternative glyph view of the significant

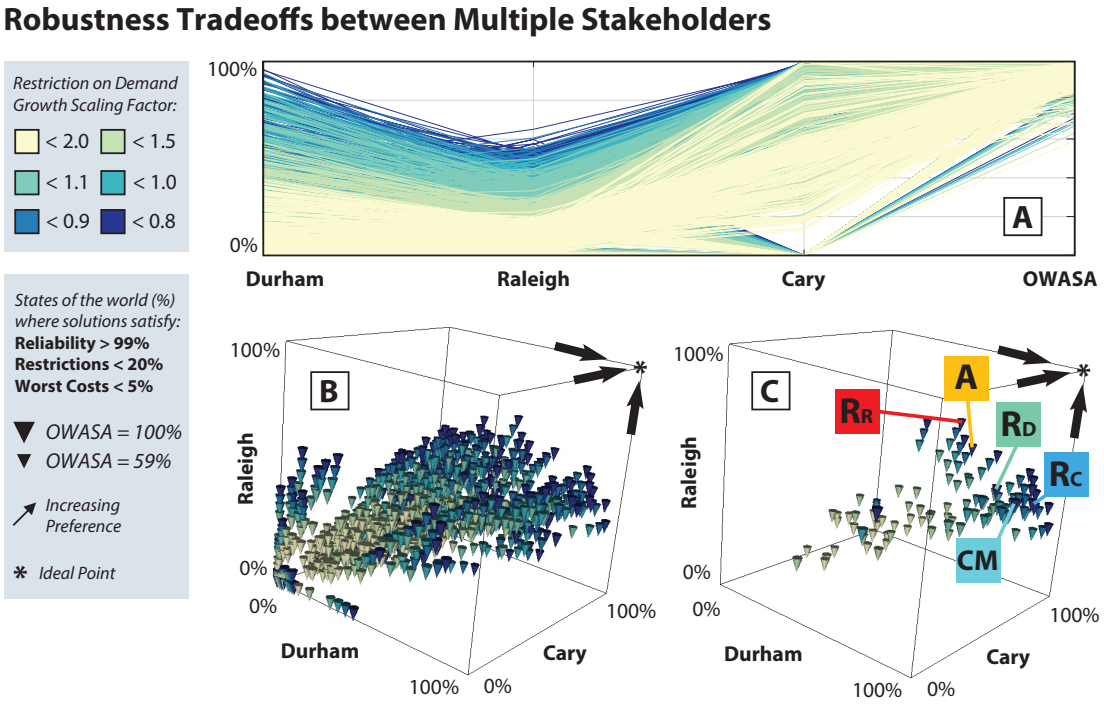

Figure 5.9: Tradeoffs between the robustness values for the four water utilities under different levels of demand growth mitigation, shown as (A) a parallel axis plot, (B) a glyph plot containing all solutions, and (C) a glyph plot containing only the non-dominated solutions in each subset. Clear improvements in robustness are visible as demand growth is restricted to progressively lower values. Solutions of interest are highlighted in panel (C), along with a compromise solution (CM) to balance robustness between the utilities. The solutions highlighted in panel (C) occur in conjunction with demand growth mitigation below 80% of the current projection (blue), as these solutions dominate all other subsets.

gains in robustness made possible by limiting the rate of demand growth. Figure 5.9b contains all solutions, while Figure 5.9c contains only the non-dominated solutions in each subset. "Non-dominated" in this case refers to the robustness space, rather than the objective space (i.e., the solutions that cannot improve the robustness of one utility without decreasing that of another). Decreasing the maximum demand growth multiplier from a value of 2.0 to a max-

Table 5.5: Robustness values for selected solutions, with the demand growth multiplier restricted to values below 0.8. CM represents the compromise solution indicated in Figure 5.9.

Solution	Durham	OWASA	Raleigh	Cary
A	53%	100%	55%	97%
R_D	88%	100%	53%	62%
R_R	46%	100%	65%	97%
R_C	81%	100%	38%	100%
CM	86%	100%	49%	96%

imum of 0.8 causes the solutions to traverse nearly the entire robustness space, toward the ideal point (i.e., 100% robustness for all utilities). In Figure 5.9c, selected solutions which survived the non-domination sort are highlighted, along with a compromise solution (CM). The robustness values for these selected solutions are shown in Table 5.5.

Another implicit tradeoff exists in Figure 5.9, which is the relationship between the demand growth limitation and the multi-stakeholder benefit achieved. Figure 5.9 does not consider, for example, the expense or the feasibility of enforcing the limits on the rate of demand growth. However, any limitation on the upper bound of demand growth will offer improvements in robustness. If the utilities are able to achieve a reduction below 80% of the current projection, the benefits are clear: Table 5.5 shows that the compromise solution (CM) offers robustness values above 85% for all utilities except Raleigh (49%). All of these values are greatly improved compared to the baseline robustness values shown in Figure 5.6. Although the individual utilities may achieve slightly higher robustness values with other portfolios, the compromise solution attempts to balance robustness equitably across the region, given the im-

portance of regional cooperation for transfer and financial agreements.

5.6.7 Robust Portfolio Selection

The compromise solution (CM) selected in Figure 5.9 reflects an acceptable balance of robustness across the four water utilities, in combination with efforts to limit the rate of regional demand growth. It is useful to explore this solution in the decision space to determine the steps necessary for achieving it. The decision profile corresponding to the compromise solution is shown in Figure 5.10, in relation to the decisions for all other solutions from the original Pareto-approximate set (Figure 5.4). Figure 5.10 indicates, for each utility, the extent to which each available portfolio option is used. These include restrictions (R), transfers (T), Jordan Lake allocation (JLA), self-insurance (SI), and third-party insurance (TPI). Decisions for which all Pareto-approximate solutions (represented by gray lines) cluster near a particular value may be inferred to be influential, since other values are unlikely to yield Pareto optimality. For example, Durham and Raleigh show highly active transfer strategies for all of the selected solutions, meaning that they will initiate transfers at the first sign of supply risk. Likewise, Jordan Lake allocations tend to be minimized for all utilities, since this was one of the original optimization objectives. The exceptions are Raleigh, which consistently shows a low but non-zero allocation, and Cary, for which some solutions require a large allocation to offset the impact of revenue losses caused by restrictions. OWASA shows few patterns in its decision profiles, indicating that regional performance and robustness are largely insensitive to its decisions (i.e., OWASA performs well in the majority of simulations regardless of its decisions).

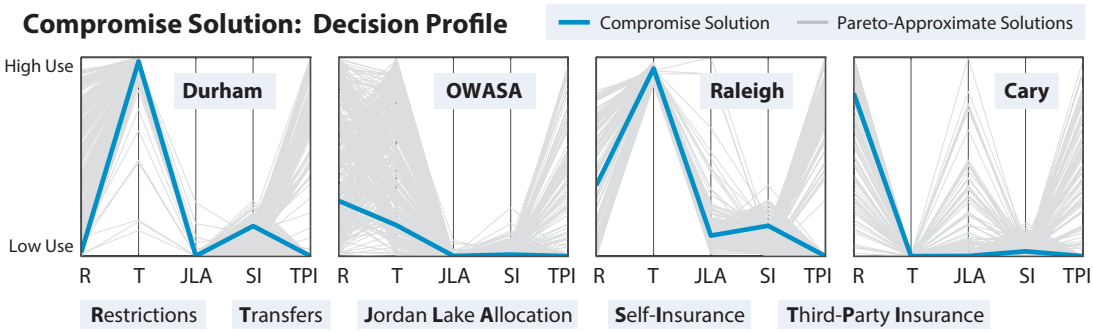

Figure 5.10: Decision profiles for the compromise solution, in relation to all other solutions from the original Pareto-approximate set. The parallel axis plots indicate the extent to which each decision is used in the portfolio: restrictions (R), transfers (T), Jordan Lake allocation (JLA), self-insurance (SI), and third-party insurance (TPI). The top row of plots contains the high-reliability solutions found in the projected future state of the world. The bottom row of plots contains solutions found using robustness values for the utilities.

The compromise solution, shown in blue in Figure 5.10, contains high transfer requirements for Raleigh and Durham. Durham contains a low restriction requirement, while OWASA and Raleigh use restrictions moderately, thus satisfying the performance criterion of implementing restrictions less frequently than once every five years. Cary utilizes restrictions frequently in the compromise solution, for two reasons: first, it cannot request transfers from others, since it controls the Jordan Lake water supply; and second, by placing restrictions on its own consumers, Cary allows a larger portion of Jordan Lake to be used for transfers to the rest of the region. The limits on water transfers from Cary are primarily due to treatment plant capacity, which provides an opportunity for future work to explore the potential benefits of a jointly financed treatment expansion relative to the cost of developing new reservoir supply. Importantly, the issue of allocations to Jordan Lake represents a political challenge to distribute

restriction requirements acceptably between Cary and the broader region. With sole access to the Jordan Lake supply, Cary is likely to be stressed by population growth in the region even though its own demand growth is projected to be moderate. These issues highlight a key role for robustness analysis—to identify key sources of vulnerability, even absent a precise prediction of future conditions. Robustness is achievable in this region, but the transfer agreements between utilities must be designed carefully to avoid undue stress on any individual utility.

Synthesizing the results shown in Figures 5.9 and 5.10, a recommended regional portfolio can be formed. First, robustness for the region is unlikely to be attained without measures in place to limit the rate of demand growth at or below the current projection. Regional robustness largely depends on Durham and Raleigh having access to water transfers, which includes active transfer triggers for both cities, as well as a small allocation to Jordan Lake for Raleigh. Simulations suggest that providing this transfer volume may stress Cary's treatment plant capacity, underscoring the need to carefully design transfer agreements. OWASA demonstrates robust performance with a moderate balance of conservation and financial measures. Finally, self-insurance and third party insurance allow the utilities to mitigate worst-case costs, as shown by (Zeff et al., 2014), without which robust performance would not be possible.

5.7 Conclusions

This study extends the MORDM framework to a regional system with multiple cooperating water utilities, the "Research Triangle" region of North Car-

olina, USA. We compare the robustness of Pareto-approximate water portfolios in combination with strategies to manage the rate of future demand growth, identified as the key vulnerability facing the region. Results indicate that the rate of demand growth represents the key uncertainty for this regional system through the 2025 planning horizon, dominating other hydroclimatic and economic uncertainties. Additionally, actions taken by any utility to achieve a modest reduction in the projected rate of demand growth (from approximately 3% per year to 2.4%) will augment the robustness of the regional water supply as a whole. Clear multi-stakeholder tradeoffs emerge for attaining robustness across the utilities, and a compromise solution is highlighted to balance performance in the projected future state of the world and robustness to sources of future uncertainty. Importantly, these insights are obscured by a traditional multiobjective optimization under projected future conditions, highlighting the importance of evaluating the candidate solutions across alternative plausible future states of the world as espoused by the MORDM framework.

The dependence of robustness on regional demand growth presents several important questions for future investigation. First, the utilities would need to explore strategies to control the rate of demand growth in the region, while population grows at different rates in each of the four municipalities. Negotiating the responsibility for controlling demand growth while accounting for variable populations presents an important challenge. Second, the definition of robustness as presented in this study hinges on the risk-averse preferences expressed by the utilities; the optimal strategies to achieve robustness may be altered under different risk attitudes. Third, the results of this study indicate that even with Pareto-optimal portfolios and demand control measures, Raleigh in particular may remain vulnerable to future sources of uncertainty using "soft-

path" management approaches. Future research will explore the potential for infrastructure improvements to provide additional protection against uncertain climate and demand scenarios. Finally, there is the overarching question of the *value of robustness*: how should utilities assign worth to protecting against these sampled scenarios, for which the probabilities remain unknown? Measures to control demand growth will incur new costs as well as potential revenue losses, which must be compared to the evaluated benefit of the robustness achieved by doing so. Questions such as these continue to challenge the field of water management under uncertainty, and provide promising avenues for future work.

By designing regional cooperative water portfolios, water utilities have the potential to manage existing supply capacity more efficiently. This study supports regional cooperative portfolio planning by advancing methods to identify future vulnerabilities to water scarcity and the strategies capable of properly managing them. The increased efficiency of regional cooperation is particularly needed in urban areas in the eastern U.S., where continually growing demand and hydrologic uncertainties are compounded by the undesirability (or infeasibility) of traditional infrastructure-based solutions to risk. These factors in concert suggest a transition to a paradigm of water scarcity in the eastern U.S., which must be managed with a thorough exploration of available strategies and their associated vulnerabilities across a range of uncertain futures.

CHAPTER 6

HOW SHOULD ROBUSTNESS BE DEFINED FOR WATER SYSTEMS PLANNING UNDER CHANGE?

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

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

Portions of this work were funded under grants from the Sectoral Applications Research Program (SARP) of the United States National Oceanic and Atmospheric Administration (NOAA) Climate Program Office (Award no. NA110AS2310144), the National Science Foundation through the Network for Sustainable Climate Risk Management (SCRiM) (NSF cooperative agreement GEO-1240507), as well as the National Institute of Food and Agriculture, U.S. Department of Agriculture (WSC Agreement No. 2014-67003-22076). The views expressed in this work represent those of the authors and do not necessarily reflect the views or policies of NOAA, NSF, or the USDA.

6.1 Abstract

Water systems planners have long recognized the need for robust solutions capable of withstanding deviations from the conditions for which they were designed. Robustness analyses have shifted from expected utility to exploratory "bottom-up" approaches, which identify vulnerable scenarios prior to assigning likelihoods; examples include Robust Decision Making (RDM), Decision Scaling, Info-Gap, and Many-Objective Robust Decision Making (MORDM). We propose a taxonomy of robustness frameworks to compare and contrast these

approaches, based on their methods of (1) alternative generation, (2) sampling of states of the world, (3) quantification of robustness measures, and (4) sensitivity analysis to identify important uncertainties. Building from the proposed taxonomy, we use a regional urban water supply case study in the Research Triangle region of North Carolina to illustrate the decision-relevant consequences that emerge from each of these choices. Results indicate that the methodological choices in the taxonomy lead to the selection of substantially different planning alternatives, underscoring the importance of an informed definition of robustness. Moreover, the results show that some commonly employed methodological choices and definitions of robustness can have undesired consequences when ranking decision alternatives. For the demonstrated test case, recommendations for overcoming these issues include: (1) that decision alternatives should be searched rather than prespecified; (2) dominant uncertainties should be discovered via sensitivity analysis rather than assumed; and (3) that a carefully elicited multivariate satisficing measure of robustness allows stakeholders to achieve their problem-specific performance requirements. This work emphasizes the importance of an informed problem formulation for systems facing challenging performance tradeoffs, and provides a common vocabulary to link the robustness frameworks widely used in the field of water systems planning.

6.2 Introduction: *a posteriori* decision support

Decision makers in water resources systems aim to achieve multiple performance objectives in the projected future while remaining robust to deviations from these projections. Extensive studies have demonstrated the willingness of decision makers to sacrifice expected performance in order to improve ro-

bustness to uncertainty (Hitch, 1960; Maass et al., 1962; Bonder, 1979; Schneller and Sphicas, 1983; Walker et al., 2001; Lempert, 2002; Clímaco, 2004; Walker et al., 2013; DiFrancesco and Tullos, 2014), signaling a departure from traditional decision theory requiring *a priori* aggregation of costs and benefits. The challenge of simultaneously navigating these goals in complex systems has led to the rise of *a posteriori* decision support, reviewed by Tsoukias (2008), in which the identification of decision alternatives and vulnerable states is preceded by data gathering, numerical modeling, and optimization. This process represents constructive learning with stakeholder feedback (Roy, 1971, 1990), in which problem formulations compete as multiple working hypotheses (Chamberlin, 1890). Given a set of alternatives shown to be near-optimal in the best available projections of the projected future state of the world, it remains a challenge to perform *a posteriori* alternative selection according to a defensible robustness criterion; several competing criteria have been proposed, and their consequences for decision making warrant comparison.

The term *a posteriori* decision support is drawn from the multi-objective optimization literature, referring to the generation of decision alternatives via computational search before imposing stakeholder preference on the problem (Cohon and Marks, 1973, 1975), also known as "Generate-First-Choose-Later" (GFCL) in the systems engineering literature (e.g., Hwang et al., 1979; Crossley et al., 1999; Balling and Richard, 2000; Messac and Mattson, 2002; Reynoso-Meza et al., 2014). Historically, this approach has stood in contrast to *a priori* weighted preference aggregation (Reuss, 2003; Banzhaf, 2009), which allows simpler solution techniques but requires decision makers to accurately assign preference weighting prior to reviewing alternatives (Bond et al., 2008, 2010). The salient implication for robustness analysis is that a single decision alternative produced

by weighted aggregation, while promising in the projected future, may fail under deviations from this projection. A multi-objective, *a posteriori* approach provides a flexible set of alternatives that may be evaluated according to their robustness.

In a related context, the concept of *a posteriori* decision support can be extended to the selection of scenarios, or states of the world, for decision making under uncertainty. In a traditional scenario analysis, uncertain states of the world are prespecified and assigned likelihoods based on model results or stakeholder expertise (e.g., climate change scenarios guiding the Intergovernmental Panel on Climate Change (Moss et al., 2010; IPCC, 2013)). Decision makers may then seek the alternative with the optimal expected performance across states of the world (e.g., Schoemaker, 1982) or the alternative which minimizes the probability of an undesirable outcome (e.g., Hashimoto et al., 1982). In the former case, expected value often fails to adequately capture stakeholder preferences under uncertainty. Maass et al. (1962) note that an alternative with higher expected value "is not invariably to be preferred", citing insurance as a common example of sacrificing expected cost in favor of reducing vulnerability to uncertainty. But the larger issue with *a priori* specification of probabilities on uncertain states of the world is that the distributions themselves are uncertain, a condition known as "deep" or "Knightian" uncertainty (Knight, 1921; Langlois and Cosgel, 1993; Lempert, 2002; Lempert et al., 2003; Polasky et al., 2011; Kasprzyk et al., 2013; Dessai et al., 2013). The challenges posed by deep uncertainty have long confounded water systems planning, making it a classic example of a "wicked" problem (Rittel and Webber, 1973; Liebman, 1976; Reed and Kasprzyk, 2009). Deep uncertainty does not reflect a state of complete ignorance, since it is possible to generate plausible states of the world and poten-

tially rank them in order of likelihood following additional analysis (Kwakkel and Pruyt, 2013). However, we would prefer to do so *a posteriori*, after identifying the most vulnerable states of the world for which likelihoods should be derived. This "scenario-neutral" (Prudhomme et al., 2010) or "bottom-up" approach (Weaver et al., 2013; Nazemi and Wheater, 2014) has been advocated by several recent frameworks for decision making under uncertainty, including Decision Scaling (Brown et al., 2012), Robust Decision Making (RDM) (Lempert, 2002; Lempert et al., 2003), Information-Gap (Info-Gap) (Ben-Haim, 2004), and Many-Objective Robust Decision Making (MORDM) (Kasprzyk et al., 2013). The components of these frameworks will be compared in the following section, along with the extent to which they align with the concepts of *a posteriori* decision support.

In this paper, we explore the idea of selecting alternatives from a Pareto-approximate set according to their robustness, focusing on the definitional challenges that arise in the process. Using model simulations from recent work in multi-objective urban water supply portfolio planning (Zeff et al., 2014; Herman et al., 2014), we illustrate the decision-relevant consequences that emerge from each of these choices. This may be seen as a generalization of the MORDM approach (Kasprzyk et al., 2013), in which alternatives are first generated by multi-objective search prior to their evaluation in deeply uncertain states of the world. Several robustness frameworks instead focus on discrete alternatives prespecified by stakeholders, the potential consequences of which will be discussed. In this work, we caution against *a priori* assumptions throughout the analysis: the prespecification of alternatives, preference weighting, the most important (sensitive) factors, or performance thresholds may produce unintended consequences for decision making. By contrast, this work will demonstrate the

potential for *a posteriori* robustness analysis of decision alternatives to inform and achieve stakeholders' performance requirements.

6.3 A taxonomy of robustness frameworks

We propose a taxonomy of robustness frameworks to compare existing approaches, based on their methods of (1) identifying *alternatives*, (2) sampling *states of the world*, (3) quantification of *robustness measures*, and (4) identification of key uncertainties (or *robustness controls*) using sensitivity analysis. Figure 6.1 outlines the methodological choices that can be made within each of these categories, which will be detailed in the following subsections. Table 6.1 describes current robustness frameworks according to these classifications. It should be noted that the frameworks share many similarities and potentially offer interchangeable components. Table 6.1 simply represents a summary of published work to date under each heading; it is hoped that this comparison allows future studies to draw key ideas more freely from these innovative frameworks.

6.3.1 Alternatives

Planning analyses begin with the identification of decision alternatives to be considered. In the simplest case, a set of discrete alternatives may be prespecified by the decision maker. This reflects a high degree of knowledge about system performance under uncertainty, and is commonly used in the Decision Scaling, Info-Gap, and RDM frameworks (e.g., Hine and Hall, 2010; Hall et al., 2012; Moody and Brown, 2013; Matrosov et al., 2013b; Lempert et al., 2006; Lempert

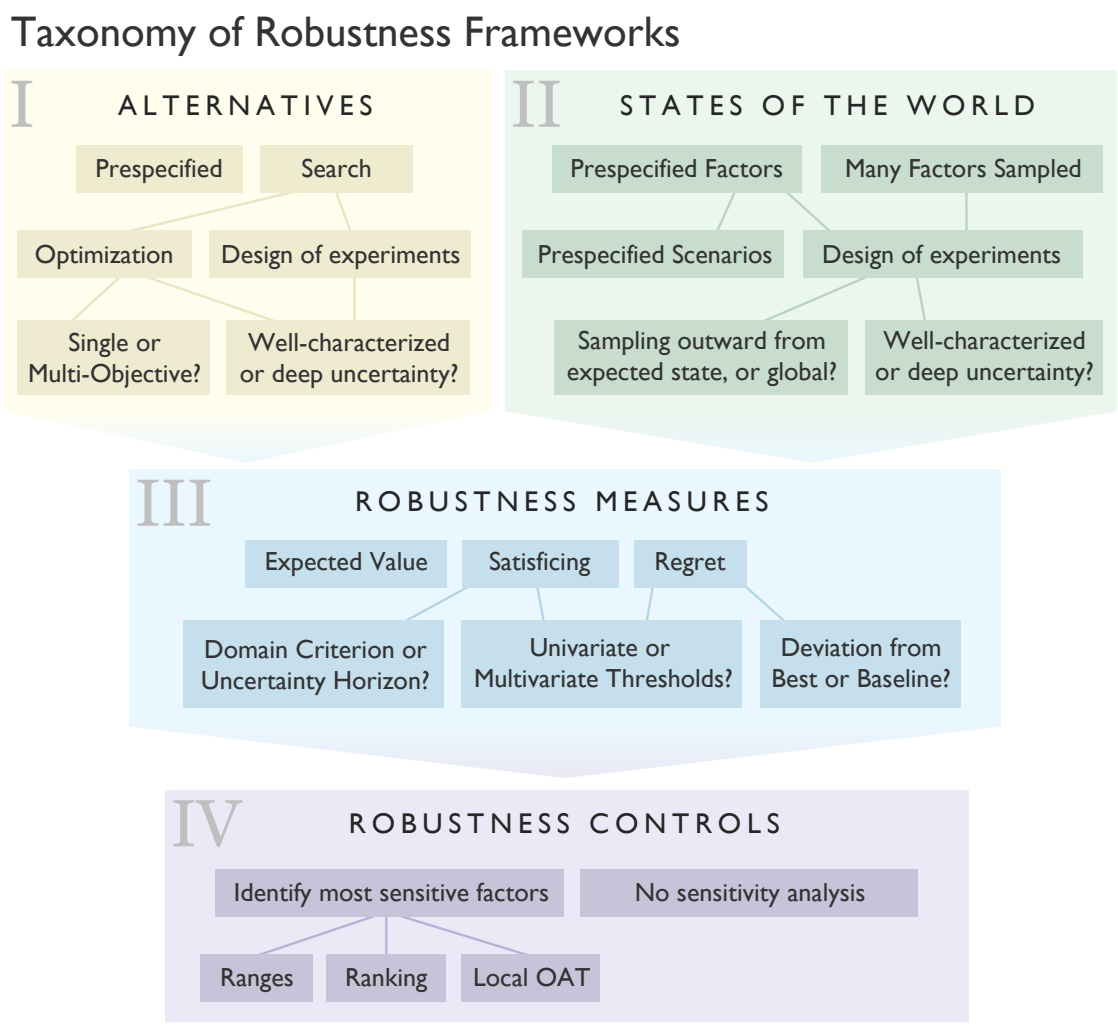

Figure 6.1: Taxonomy of robustness frameworks, based on (1) the selection of decision alternatives, (2) the selection and sampling of the uncertain factors which make up plausible future states of the world, (3) the quantification of robustness measures by which to rank candidate alternatives, and (4) the identification of uncertain factors controlling robustness via sensitivity analysis.

Table 6.1: Classification of current robustness frameworks according to the taxonomy provided in Figure 6.1

Framework	Select Citations	Alternatives	States of the World	Measures	Controls
Decision Scaling	(Brown, 2010; Brown et al., 2012; Brown and Wilby, 2012; Moody and Brown, 2013; Li et al., 2014; Turner et al., 2014; Ghile et al., 2014)	Prespecified	Prespecified or many factors; Global sampling under deep uncertainty	Satisficing (domain criterion)	Climate Response Function (visual), ANOVA ranking
Info-gap	(Hipel and Ben-Haim, 1999; Ben-Haim, 2004; Hine and Hall, 2010; Hall et al., 2012; Ben-Haim, 2012; Matrosov et al., 2013b; Korteling et al., 2013)	Prespecified or enumerative	Many factors sampled, outward from expected state under deep uncertainty	Satisficing (uncertainty horizon)	No sensitivity analysis
Robust Decision Making (RDM)	(Lempert et al., 2003; Lempert and Collins, 2007; Groves and Lempert, 2007; Bryant and Lempert, 2010; Lempert and Groves, 2010; Tingstad et al., 2014)	Prespecified or Design of Experiments	Many factors sampled; Global sampling under deep uncertainty	Satisficing (domain criterion) or Regret (deviation from best)	Ranges (PRIM)
Many-Objective Robust Decision Making (MORDM)	(Kasprzyk et al., 2013; Herman et al., 2014)	Multi-objective search under well-characterized uncertainty	Many factors sampled; Global sampling under deep uncertainty	Satisficing (domain criterion) or Regret (deviation from baseline)	Ranges (PRIM)
Robust Optimization (RO)	(Watkins and McKinney, 1997; Deb and Gupta, 2006; Perelman et al., 2013; Hamarat et al., 2014)	Search; well-characterized or deep uncertainty	Prespecified factors; Typically sampled under well-characterized uncertainty	Number of feasible scenarios, or variance in performance	No sensitivity analysis

and Collins, 2007; Bryant and Lempert, 2010; Tingstad et al., 2014), though several Decision Scaling studies have focused on characterizing the risks of an existing system rather than choosing between planning alternatives (Brown, 2010; Li et al., 2014; Turner et al., 2014; Ghile et al., 2014), as have several RDM studies (e.g. Groves et al., 2014). Conversely, alternatives may arise as the result of computational search. This may take the form of a Latin Hypercube sample or an enumeration of the decision space (Korteling et al., 2013; Matrosov et al., 2013a), or a full multi-objective optimization, as proposed by MORDM (Kasprzyk et al., 2013). Alternatives to be optimized may include binary infrastructure decisions (Matrosov et al., 2013b), continuous variables such as decision triggers or transfer allocations (Zeff et al., 2014), functions mapping observed states to actions (Giuliani et al., 2014), or a combination thereof. The use of optimization to develop alternatives allows decision makers to augment their expertise with computational simulation to compare the estimated future performance of candidate solutions.

When optimization is employed, solutions may be optimized over a range of scenarios defined by well-characterized uncertainty (i.e., where probability distributions can be approximated with reasonable confidence), or over deeply uncertain scenarios (i.e., optimizing for robustness to deep uncertainty directly). The former approach is advocated by MORDM (Kasprzyk et al., 2013) and Paton et al. (2014), while the latter approach is taken by Robust Optimization (RO). Watkins and McKinney (1997), following Mulvey et al. (1995), develop a framework for RO consisting of robust solutions which remain nearly optimal and satisfy stakeholder requirements across scenarios. Later multi-objective RO methods proposed by Deb and Gupta (2006) suggest optimizing with multivariate constraints on the deviation of a solution from its expectation. Similarly,

McInerney et al. (2012) optimize a weighted sum of the expected and worst-case outcomes, while Hamarat et al. (2014) propose an approach for optimizing robustness according to a "signal-to-noise" ratio, which reflects the ratio of the mean performance divided by its standard deviation. However, the applicability of these approaches under deep uncertainty remains unclear. In a thorough review of RO methods, Beyer and Sendhoff (2007) distinguish probabilistic uncertainty, focused on likelihood, from "possibilistic" or epistemic uncertainty, focused on plausibility and arising from lack of knowledge. Beyer and Sendhoff (2007) note that RO methods for the latter type of uncertainty remain relatively unexplored in the literature, and that the measures of expectation and dispersion commonly seen in RO require the analyst to treat the uncertainty as probabilistic, even when this may not be appropriate. For RO under deep uncertainty, constraint violations may be penalized unduly, as these are expected in extreme future states of the world, and the signal-to-noise ratio implicitly assumes some familiarity with the distribution of future scenarios. Robust multi-objective optimization under deep uncertainty remains an open question; the MORDM framework evaluates robustness *a posteriori* to ensure approximate Pareto optimality in the best projection of the future state of the world under well-characterized uncertainty.

6.3.2 States of the World

Whether alternatives are prespecified or optimized, robustness frameworks typically proceed by evaluating their performance across a set of uncertain states of the world. The exception is RO, which essentially combines the "alternatives" and "states of the world" categories in Figure 6.1 into a single optimization. In

the remaining frameworks, the states of the world expand the uncertain factors beyond the ranges considered when determining the alternatives. The uncertain factors may be considered well-characterized if a probability distribution can be identified, but are typically treated as deep uncertainties. This designation acknowledges that any probability distributions assumed when determining alternatives (whether explicitly or implicitly) are themselves uncertain.

One major aim of robustness analyses is to isolate the deeply uncertain factors responsible for system vulnerabilities. If decision makers have reason to believe *a priori* that a certain subset of uncertain factors will be most relevant, these may be prespecified. Alternatively, many uncertain factors may be sampled, with the understanding that some may later be found to be noninfluential. For example, several Decision Scaling studies limit the risk analysis to climate factors only (Brown et al., 2012; Moody and Brown, 2013), while recent work extends Decision Scaling to include hydroeconomic factors (Ghile et al., 2014; Lownsbery, 2014), noting that demand pressures may pose equal or greater risk than climate factors depending on the planning horizon (Frederick and Major, 1997; Lins and Stakhiv, 1998; Vorosmarty, 2000).The Info-Gap, RDM, and MORDM frameworks all typically assume that many factors will be sampled rather than prespecifying only a few.

Once the uncertain factors are identified, specific combinations of them can be sampled to create states of the world. Again, these may be prespecified in the case of traditional scenario analysis, (e.g., Liu et al., 2008; Mahmoud et al., 2009) or sampled, in the case of all robustness frameworks considered here. This sampling is typically performed over noninformative priors, reflecting the exploratory nature of deep uncertainty analyses; however, a well-characterized

probability distribution can be used if known. The manner of sampling differentiates the robustness frameworks considered here. Decision Scaling, RDM, and MORDM typically perform an exploratory "global" sample of uncertain factors over plausible ranges (Lempert et al., 2006; Groves and Lempert, 2007; Bryant and Lempert, 2010). By contrast, Info-Gap samples radially outward from the projected future state in the uncertainty space until it encounters a state that causes failure (Hipel and Ben-Haim, 1999; Ben-Haim, 2004; Matrosov et al., 2013b), with the option to specify a covariance between related factors (Hall et al., 2012). The radius of uncertainty considered is defined by a parameter α, which expands the range of all uncertain factors simultaneously.

6.3.3 Robustness Measures

Given a set of decision alternatives evaluated in a set of states of the world, it remains to quantify their robustness to facilitate the decision making process. The choice of a robustness measure is not straightforward; Schneller and Sphicas (1983) refer to this as "the metaproblem of how to decide how to decide". Figure 6.1 defines the overarching categories of *regret* and *satisficing* measures, adopted from (Lempert and Collins, 2007). Both measures can be employed in either a univariate or multivariate context. Broadly, "regret" quantifies the cost (not necessarily monetary) of choosing incorrectly. It can be defined for a single solution as the deviation from its baseline (expected) performance, for example as employed by Kasprzyk et al. (2013). In this case, the measure of regret focuses on incorrect assumptions regarding the future state of the world. Alternatively, regret may be defined as the difference between a solution's performance and that of the best solution in the prevailing state of the world. Here, the measure

of regret focuses on the incorrect choice of decision alternative, as proposed by Savage (1951).

"Satisficing" refers to the tendency of decision makers to seek outcomes that meet one or more requirements but may not achieve optimal performance (Simon, 1959). One option for a satisficing-based robustness measure is the domain criterion, a variant of which was first introduced by Starr (1962) and further analyzed by Schneller and Sphicas (1983). The domain criterion quantifies the volume of the uncertain factor space in which a solution meets the decision makers' performance requirements. In practice, this is done by calculating the fraction of sampled states of the world in which a solution satisfies one or more performance thresholds, as done by Herman et al. (2014). The domain criterion is also similar to the threshold-based loss function proposed by Brown et al. (2012) and extended by Moody and Brown (2013). Conversely, the Info-Gap framework quantifies robustness according to the "uncertainty horizon", that is, the value of the outward sampling parameter α that can be withstood before the system fails to meet the performance threshold (Hall et al., 2012; Korteling et al., 2013). This approach assumes a continuous region of acceptable performance centered at the projected future state of the world, which must be known or estimated. Finally, these satisficing measures also relate to visualization techniques such as "Robust regions visualization" (Lempert, 2002; Bankes et al., 2001; Bankes, 2002), or the "climate sectors" decision rules demonstrated by Brown et al. (2012), which show the regions of the uncertainty space in which a particular solution is preferred over competing alternatives.

6.3.4 Robustness Controls

After quantifying the robustness of candidate alternatives, a last crucial step of decision support under uncertainty is to isolate the uncertain factors most responsible for failure to satisfy stakeholder requirements (i.e., "robustness controls"). This can be viewed as a consequence-oriented sensitivity analysis. The Info-Gap framework does not provide sensitivity information at the factor level, since all information about the robustness of a solution is contained in the aggregate parameter α. The remaining frameworks largely aim to identify sensitive ranges of uncertain factors likely to cause failure. In the sensitivity analysis literature, this is known as "factor mapping" (Saltelli et al., 2008). For this purpose, the RDM and MORDM frameworks advocate the Patient Rule Induction Method (Friedman and Fisher, 1999), a data mining technique that identifies ranges (or more generally, hypercubes) in the uncertainty space where failures are likely to occur (Groves and Lempert, 2007; Lempert et al., 2008; Bryant and Lempert, 2010). Decision Scaling also aims to identify sensitive ranges of uncertain climate factors using the Climate Response Function (e.g. Brown et al., 2012; Turner et al., 2014). The purpose of a sensitivity analysis in this context is to help decision makers target specific vulnerable factors for adaptation, an issue addressed in the "Dynamic Adaptive Pathways" approach developed by Haasnoot et al. (2013).

One rather unexplored area in the current robustness frameworks is factor prioritization, or the ranking of all uncertain factors in order of their sensitivity (Saltelli et al., 2008). In its simplest form, this can be done using local partial derivatives of each factor independently, known as "one-at-a-time" or OAT (see Dessai and Hulme, 2007). More thorough and informative are global variance

decomposition techniques such as Sobol sensitivity analysis (Sobol, 1993; Sobol, 2001; Saltelli, 2002). Significant potential exists to incorporate such methods into robustness analyses in water systems planning, particularly those which require the ranking of many uncertain factors. Lownsbery (2014) apportions total output uncertainty to its constituent factors using an ANOVA decomposition over time for a water resources system, similar to the partitioning of prediction uncertainty developed in the climate literature (Hawkins and Sutton, 2009, 2011). Sensitivity analysis can provide decision-relevant information on par with the recommendation itself, often leading to a modification of alternatives to account for the controlling uncertainties, as in Herman et al. (2014). Recent studies have called for sensitivity analysis to become a required tool to improve the transparency of model assumptions in policy contexts (Borgonovo and Peccati, 2008; Saltelli and Funtowicz, 2013; Saltelli et al., 2013), an apt recommendation for water systems planning problems.

6.4 Methods

We illustrate the decision-relevant consequences that emerge from the analysis choices in the taxonomy of Figure 6.1 using model results from recent work in multi-objective urban water supply portfolio planning. The portfolio planning approach to balance supply-side and demand-side strategies has been developed extensively in conjunction with multi-objective optimization and iterative problem formulation (Kasprzyk et al., 2009, 2012). In prior work, Zeff et al. (2014) develop an urban water supply model for the "Research Triangle" region of North Carolina and formulate a multi-objective portfolio optimization problem. Herman et al. (2014) then analyze the robustness of the resulting set

of near-optimal (non-dominated) solutions using an extension of the MORDM approach. Key details of these methods are summarized below; the reader is referred to the references above for full details.

6.4.1 North Carolina Research Triangle

The Research Triangle region is located in the state of North Carolina in the southeast U.S., where nearly two million residents are served by four major water utilities: Raleigh, Durham, Cary, and the Orange Water and Sewer Authority (OWASA, serving the towns of Chapel Hill and Carrboro). Here we develop a simulation of the Research Triangle water supply, with synthetic reservoir inflows and weekly consumptive demand data provided by each utility. We seek to optimize the utilities' water supply portfolios over the period 2013–2025, with the following decision instruments: restrictions, transfers, self-insurance (contingency fund), and third-party insurance. The Research Triangle water supply model, data, and optimization problem are described in detail in Section 5.4.

6.4.2 Experiments: Alternatives and States of the World

Following the taxonomy provided by Figure 6.1, we begin by comparing the robustness of decision alternatives found during the multi-objective search with that of a prespecified solution. In this case, the prespecified solution reflects the utilities' status quo prior to the analysis performed by Zeff et al. (2014). The prespecified solution allows utilities to impose water use restrictions, but they may not use water transfers or any financial instruments. We compare this solution

to the Pareto set of solutions resulting from the multi-objective optimization by considering the cumulative distributions of objective values across states of the world; a solution which provides improved performance across many states of the world will, in general, be considered more robust. This approach explores the benefits gained by optimizing the portfolio instruments under consideration, rather than limiting the analysis to a decision alternative prespecified by the utilities.

The second major methodological choice highlighted in Figure 6.1 involves states of the world comprising either prespecified factors, representing *a priori* assumptions of importance, or many sampled factors, with no such assumptions. To represent prespecified factors, we show the success or failure of candidate alternatives in a two-dimensional projection of the uncertainty space focusing solely on the inflows and evaporation, serving as a proxy for a climate change analysis in which precipitation and temperature might be assumed to control solution performance. We contrast this with a similar two-dimensional projection of multipliers on inflow and demand growth, the latter of which was found to dominate the performance of decision alternatives by Herman et al. (2014) through the 2025 planning horizon. This comparison highlights the importance of avoiding *a priori* assumptions regarding the most sensitive factors, since the results of the analysis may confound expectations.

6.4.3 Comparison of Robustness Measures

Using the approximately Pareto-optimal set of decision alternatives identified by Zeff et al. (2014), we quantify and rank their robustness according to four

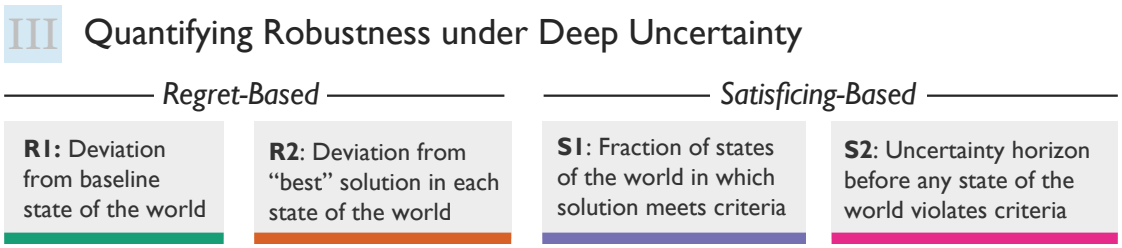

Figure 6.2: Four measures for quantifying robustness. The categories of regret and satisficing categories are adapted from Lempert and Collins (2007). Measure R2 is a variant of Savage regret (Savage, 1951), while S1 links to Starr's domain criterion (Starr, 1962; Schneller and Sphicas, 1983). The uncertainty horizon, S2, is drawn from the Info-Gap method (Ben-Haim, 2004).

different measures as shown in Figure 6.2. The measures include two variants of "regret" and two variants of "satisficing", the categories described by Lempert and Collins (2007). For brevity, the two regret-based measures are named R1 and R2, while the satisficing-based measures are named S1 and S2. First, measure R1 calculates regret as a solution's deviation from its performance in the baseline (expected) state of the world. This is an individual measure that does not depend on other solutions in the set; it was demonstrated, for example, by (Kasprzyk et al., 2013). Here, we define R1 for each decision alternative as the 90^{th} percentile deviation of the worst-case objective:

$$\text{R1} = \max_i \{D_{i,90} : \text{P}[\, D_i \leq D_{i,90} \,] = 0.90\} \tag{6.1}$$

$$D_{i,j} = \frac{|F(\mathbf{x})_{i,j} - F(\mathbf{x})_i^*|}{F(\mathbf{x})_i^*} \tag{6.2}$$

where $F(\mathbf{x})_i^*$ represents the value of objective i in the baseline state of the world, and $F(\mathbf{x})_{i,j}$ represents the value of objective i calculated in state of the world j.

Similarly, measure R2 calculates regret as the deviation from the "best" so-

lution in each state of the world:

$$\text{R2} = \max_i \{D_{i,90} : \text{P}[\ D_i \leq D_{i,90}\ ] = 0.90\} \tag{6.3}$$

$$D_{i,j} = \frac{|F(\mathbf{x})_{i,j} - \min_s F(\mathbf{x}_s)_{i,j}|}{F(\mathbf{x})_{i,j}} \tag{6.4}$$

The "best" value of each objective i is taken across solutions s. Clearly, this measure depends on the other solutions in the set, as it incorporates all of them. The deviation $D_{i,j}$ in Equation 6.4 is normalized by the objective value itself rather than the "best" value, since the latter often approaches zero. The 90^{th} percentile value for the regret-based measures is intended to reflect the tail end of poor performance while reducing susceptibility to outliers.

Measure S1 is defined as the fraction of N states of the world in which a solution meets stakeholders' performance requirements in one or more objectives.

$$\text{S1} = \frac{1}{N} \sum_{j=1}^{N} \Lambda_{s,j} \tag{6.5}$$

where $\Lambda_{s,j} = 1$ if solution s meets requirements in state of the world j and $\Lambda_{s,j} = 0$ otherwise. An important component of this measure is the elicitation of requirements from stakeholders. In this case study, the utilities indicated the following requirements: reliability above 99%, restriction frequency less than 20% (once every 5 years), and worst-case cost less than 5% of annual volumetric revenue (Herman et al., 2014). Controlling the worst-case cost is critical for utilities operating under the "cost-recovery" budget model, where stable revenues are required to recoup large infrastructure investments (Hughes and Leurig, 2013). The use of multivariate requirements has been advocated in the MORDM framework (Kasprzyk et al., 2013) and elsewhere (e.g., Gerst et al., 2013).

Finally, measure S2 represents the uncertainty horizon that can be withstood prior to system failure. Following the Info-Gap approach, the uncertainty horizon (or "robustness") can be defined as (Hall et al., 2012):

$$S2 = \hat{\alpha} = \max\{\alpha : \min_{j \in U(\alpha)} F(\mathbf{x})_j \geq r^*\} \tag{6.6}$$

$\hat{\alpha}$ is the maximum level of uncertainty, measured outward from a baseline point, that can be tolerated without performance falling below r^*. The inner minimization term is applied over all states of the world j that are sampled in the radius α, meaning that the calculation of $\hat{\alpha}$ does not allow any violation of the performance requirement. The Info-Gap framework defines a univariate requirement r^*; while this could be extended to include multivariate thresholds, applications of this method in the literature have not done so. In this study, we use the utilities' reliability requirement to compute S2 (i.e., a univariate value of $r^* = 0.99$.) Because many of the deeply uncertain factors in this study are scaling factors, the baseline point is simply defined where these are equal to one. However, in general it may not be straightforward to define this baseline point.

The robustness calculations in this study are based on an ensemble evaluation of 10,000 alternative states of the world, defined by 13 dimensions of deep uncertainty including climatic and economic factors. Each of 299 Pareto-approximate solutions was evaluated across this ensemble of states of the world to determine the impact of uncertain factors on performance. For details of the experimental setup, please refer to Herman et al. (2014).

6.4.4 Comparison of Robustness Controls

As an illustration of different sensitivity analysis approaches to identify important uncertain factors in the system, we compare the Patient Rule Induction Method, PRIM (Friedman and Fisher, 1999) with Sobol sensitivity analysis (Sobol, 1993; Sobol, 2001; Saltelli, 2002). PRIM is a factor mapping approach aiming to identify sensitive ranges of uncertain factors that are likely to cause a particular outcome, and is described in detail in Section 2.1.3. Sobol sensitivity analysis (Sobol, 1993; Sobol, 2001; Saltelli, 2002) is a global, variance-based method that attributes the total variance of output Y to individual inputs and their interactions, described in detail in Section 2.1.1. The two approaches are complementary: PRIM provides the vulnerable ranges of uncertain factors, while Sobol provides a full ranking of sensitive factors.

6.5 Results

6.5.1 Alternatives and States of the World

Figure 6.3 shows the cumulative distributions of performance in four objectives for the prespecified alternative and those found via multi-objective optimization by Zeff et al. (2014). This relates to the first block of the taxonomy in Figure 6.1, the selection of alternatives. The performance values shown are for Durham, one of the larger water utilities in the region. The prespecified solution, which does not use water transfers or financial instruments, is shown in red. This represents the strategy that the utilities currently follow in the ab-

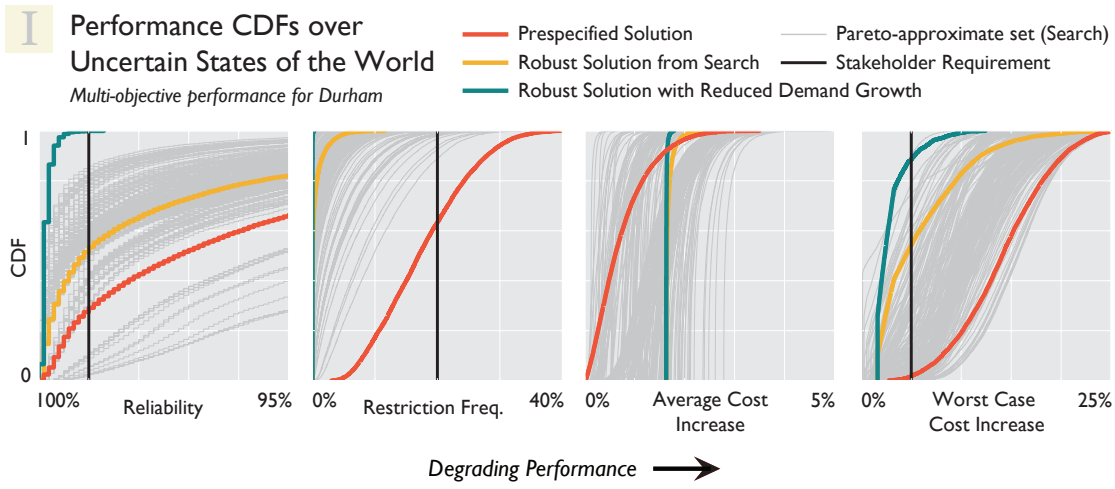

Figure 6.3: Cumulative distributions for Durham's four objectives using the prespecified alternative (red) and those found via multi-objective optimization (gray) by Zeff et al. (2014). An example of a robust solution is shown in yellow, identified by Herman et al. (2014) using a satisficing measure, and further improved (green) by controlling the rate of demand growth, the most influential uncertainty. The Pareto set resulting from optimization provides options for stakeholders to select a robust decision alternative.

sence of an optimized portfolio. An example of a robust solution is shown in yellow, identified by Herman et al. (2014) using a satisficing measure. Finally, the distributions drawn in green represent the same robust solution after performing sensitivity analysis and mitigating the most important uncertainty, the rate of demand growth. The utilities' performance requirements are drawn as vertical black lines. Figure 6.3 highlights several salient points. First, multi-objective optimization greatly improves on the performance of the prespecified solution, not only in the projected future under well-characterized uncertainty, but also across the distribution of plausible states of the world under deep uncertainty. The resulting Pareto set, shown in gray in Figure 6.3, provides options for stakeholders to improve the robustness of the decision alternative. Second,

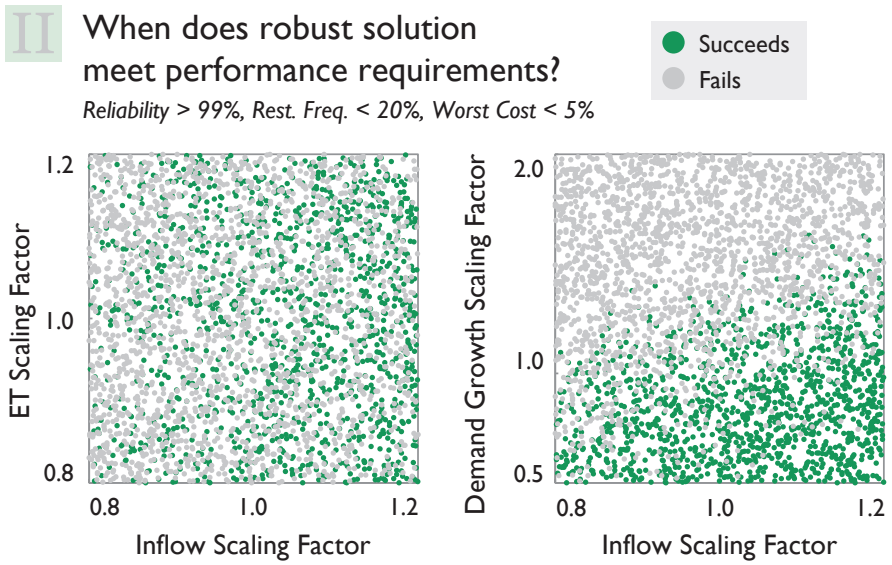

Figure 6.4: States of the world in which the robust solution from Figure 6.3 fails to meet the utilities' performance requirements. Green points indicate success, and gray points indicate failure. Plotting only climate factors (left) does not indicate sensitivity, whereas plotting the sensitive factors identified by Herman et al. (2014) shows a clear separation between states of the world in which the solution succeeds and fails.

robustness can be further improved by augmenting the optimized decision alternatives with actions taken to reduce the impact of key uncertainties, identified with sensitivity analysis. Limiting this analysis to prespecified decision alternatives or prespecified uncertain factors would not provide stakeholders with the same degree of flexibility as the *a posteriori* approach shown here.

To compare the decision consequences of selecting states of the world, Figure 6.4 assesses where, in the space of uncertain factors, the robust solution from Figure 6.3 fails to meet the utilities' performance requirements. It contains a pair of two-dimensional projections of the 13-dimensional uncertainty space. Green points indicate states of the world where the solution successfully meets the requirements; light gray points indicate states of the world where the solution

fails. We would like to identify the uncertainties that are responsible for these differences. On the left, the uncertainties considered are the scaling factors on reservoir inflows and evapotranspiration. This represents the *a priori* assumption that climate factors—for example, precipitation and temperature—can be prespecified as the only factors of interest. As the left panel of Figure 6.4 suggests, these uncertainties do not reveal much information about the successes and failures of this solution. By contrast, the right panel of Figure 6.4 shows the same states of the world in the two-dimensional projection defined by the scaling factors on demand growth and inflows, the first of which was identified by Herman et al. (2014) as the most influential uncertainty in the system. This result suggests that, through the 2025 planning horizon, the utilities may have significantly more control over system vulnerabilities than previously assumed. This plot shows a clear separation between states of the world in which the solution succeeds and fails, reinforcing the sensitivity of the demand growth factor and the importance of withholding assumptions about sensitive factors when sampling states of the world. In this case study, sampling many uncertain factors allow the utilities to understand their vulnerabilities without a potentially limiting focus on climate factors.

6.5.2 Comparison of Robustness Measures

Using the four robustness measures defined in Figure 6.2, we compute the rankings of the Pareto-approximate solutions identified by Zeff et al. (2014). Figure 6.5 shows the "best" solution according to each measure (i.e., a ranking of 1), along with where these fall when solutions are ranked according to the other measures. The status quo solution, or the prespecified solution highlighted in

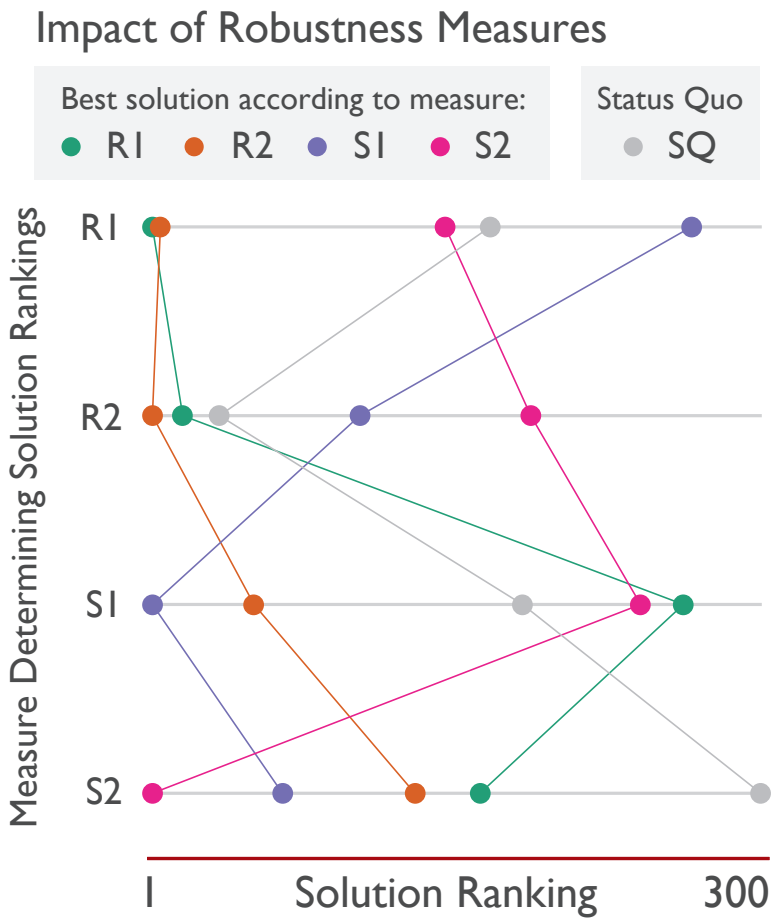

Figure 6.5: Ranking the "best" solution according to each measure, along with where these solutions fall when ranked according to other measures. The status quo (prespecified) solution is shown in gray. The choice of robustness measure will significantly affect the ranking of decision alternatives and thus the recommendation from an analysis.

Figure 6.3, is shown in gray. Figure 6.5 indicates that the choice of robustness measure will significantly affect the choice of decision alternatives. For example, the best solution according to measures R1 (green) and R2 (orange) appear very similar to one another, suggesting that the two regret-based metrics provide similar rankings of solutions. However, these perform poorly when solutions are ranked according to the satisficing measures S1 and S2. Further-

more, these latter two measures tend to offer conflicting rankings, indicated by crossing lines. The metric S1 (satisficing according to absolute criteria) has been widely used in the literature, and this conflicts sharply with the other metrics, particularly the percent deviation measure R1. In summary, Figure 6.5 shows that the four robustness measures provide very different rankings of solutions, underscoring the importance of choosing a measure appropriate to a particular study.

The robust solutions according to each measure can be mapped back into the original objective space (i.e., in the best available projection of the future state of the world under well-characterized uncertainty) to understand the decision consequences of the choice of robustness measure. Figure 6.6 shows where these robust solutions fall in the objective space, using a 4-dimensional scatter plot (top) and a parallel axis plot (bottom). The full Pareto set is shown in gray, while the robust solutions are highlighted in colors consistent with Figure 6.5. The directions of preference are indicated with arrows on the scatter plot, and are directed toward the bottom of each axis in the parallel axis plot.

Figure 6.6 indicates that the four different robustness measures may lead decision makers to very different regions of the objective space, with some of these differences caused by mathematical artifacts rather than system properties. For example, measure R1 (percent deviation relative to the baseline scenario) selects a solution with generally poor performance in the baseline scenario in order to minimize its deviation in other states of the world. Measure R2 (deviation from the best solution in each state of the world) selects a solution with mediocre performance in all objectives, seeking to avoid severe underperformance in any single objective. The satisficing measure S1 places utilities' performance require-

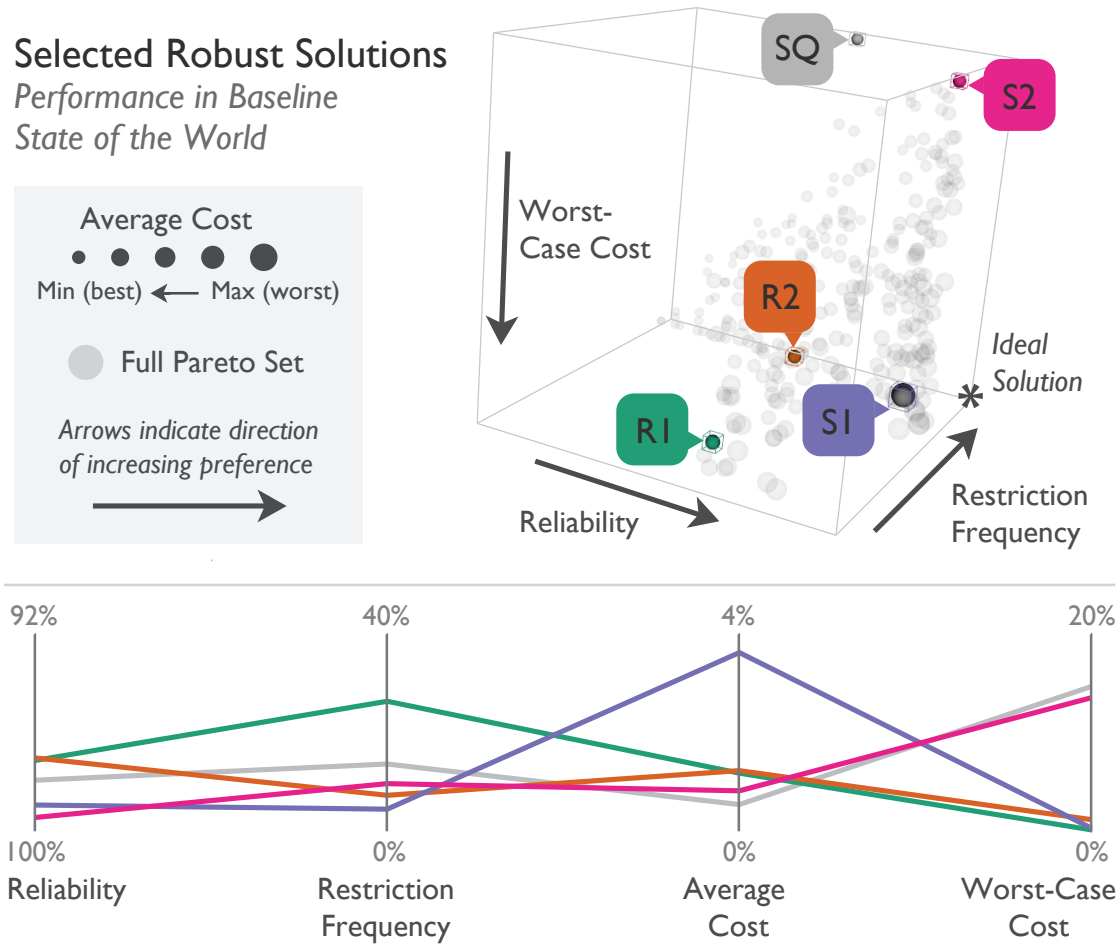

Figure 6.6: The most robust solutions according to each measure, mapped back into the original objective space in the projected future state of the world. The objective space is plotted using a 4-dimensional scatter plot (top) and a parallel axis plot (bottom). The full Pareto set is shown in gray, while the robust solutions are highlighted in colors consistent with Figure 6.5. The directions of preference are indicated with arrows on the scatter plot, and are toward the bottom of each axis in the parallel axis plot. The four different robustness measures may lead decision makers to very different regions of the objective space.

ments on all objectives except average cost; as a result, the solution selected by S1 requires a high average cost in order to achieve the requirements, which reflects the preferences elicited from the utilities. Finally, measure S2 (the uncertainty horizon) selects for high reliability, its single requirement, leaving it vulnerable to poor performance in the other objectives. In general, S1 is the only measure that chooses a solution in line with the utilities' preferences, showing the value of a multivariate satisficing approach. The resulting average cost of 3-4% is not onerous for the utilities, as it can be planned for in order to reduce the worst-case cost in excess of this average (Herman et al., 2014). Figure 6.6 suggests that the choice of robustness measure may produce unintended consequences for recommendations if it does not consider stakeholders' preferences. Importantly, these insights are only possible with a Pareto-approximate set of solutions for comparison, which gives decision makers a clear understanding of the tradeoffs and consequences of each near-optimal alternative in the projected future.

6.5.3 Univariate and Multivariate Satisficing Thresholds

Out of these choices shown in Figure 6.6, only the satisficing measure S1 provides solutions with acceptable performance for the utilities in the projected future state of the world. It is useful to further explore the decision consequences of the individual constraints that compose the multivariate measure. Figure 6.7 shows a parallel axis plot of the solutions that would be chosen if the satisficing measure were only composed of individual requirements. The combined requirement (matching solution S1 from Figure 6.6) is shown in purple. The consequence of satisficing based on a univariate requirement is, of course, that

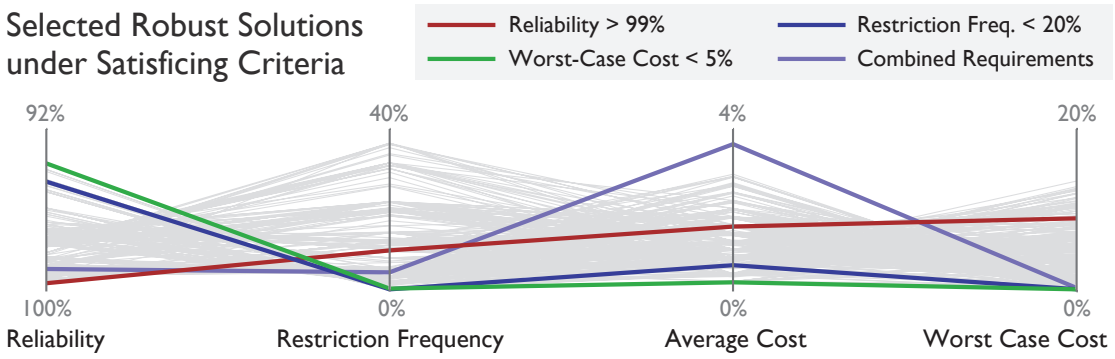

Figure 6.7: Individual compared to multivariate performance requirements for the satisficing measure. Colors indicate which solutions would be chosen if the satisficing measure were only composed of individual requirements. The combined requirement (matching solution S1 from Figure 6.6) is shown in purple. Satisficing based on a single objective will sacrifice performance in the others.

other objectives will be ignored. Despite the simplicity of this fact, it is not widely done; many robustness analyses in the literature could benefit from a multivariate performance requirement to more accurately capture the interests of stakeholders in the system. Satisficing based solely on reliability (red solution in Figure 6.7) yields poor performance in restriction frequency and worst-case cost, and using either restriction frequency or worst-case cost requirements (blue and green, respectively) sacrifices reliability. Given stakeholders' difficulty in correctly specifying all performance requirements *a priori* (Bond et al., 2008), it is important for a robustness measure to maintain the flexibility to adjust multivariate preferences after exploring the alternatives. Moreover, these results emphasize the importance of exploring multiple competing robustness measures to better understand their potential impacts.

6.5.4 Comparison of Robustness Controls

Figure 6.4 demonstrated a qualitative, visual approach to identifying the uncertain factors responsible for system vulnerabilities. However, robustness analyses are strengthened by formal, quantitative sensitivity analyses. Figure 6.8 illustrates the results from a factor mapping sensitivity analysis (PRIM) compared to a factor prioritization (Sobol). PRIM identifies the ranges of one or more sensitive factors likely to cause a particular outcome, seeking the minimum set of rules needed to capture the maximum number of instances of that outcome. For this problem, the method identifies the demand growth scaling factor as the only sensitive uncertainty, and suggests that a value greater than one (i.e., if demand grows faster than projected) will most likely cause system failures. It does not provide information about the sensitivity of the other uncertain factors, other than designating them as noninfluential. By contrast, the Sobol indices shown in the bottom panel of Figure 6.8 provide a full ranking of sensitive factors, since it analyzes the variance contribution of each factor. The impact of the uncertain factors (x-axis) on each of the objectives (y-axis) is quantified with the first-order sensitivity index, where white is insensitive and dark blue is sensitive. An additional objective named "Failures" has been added, which is a binary variable indicating whether any of the failure conditions defined for PRIM have been violated. Sobol identifies the most sensitive factor as the rate of demand growth, which agrees with the PRIM result. Other influential factors include the inflows multiplier, the peaking factor (controlling seasonality of demand) and the uncertainty on the fraction of certain reservoirs allocated for municipal supply. This ranking information complements the ranges identified by PRIM and would be a useful addition to the robustness analyses performed in the water systems field.

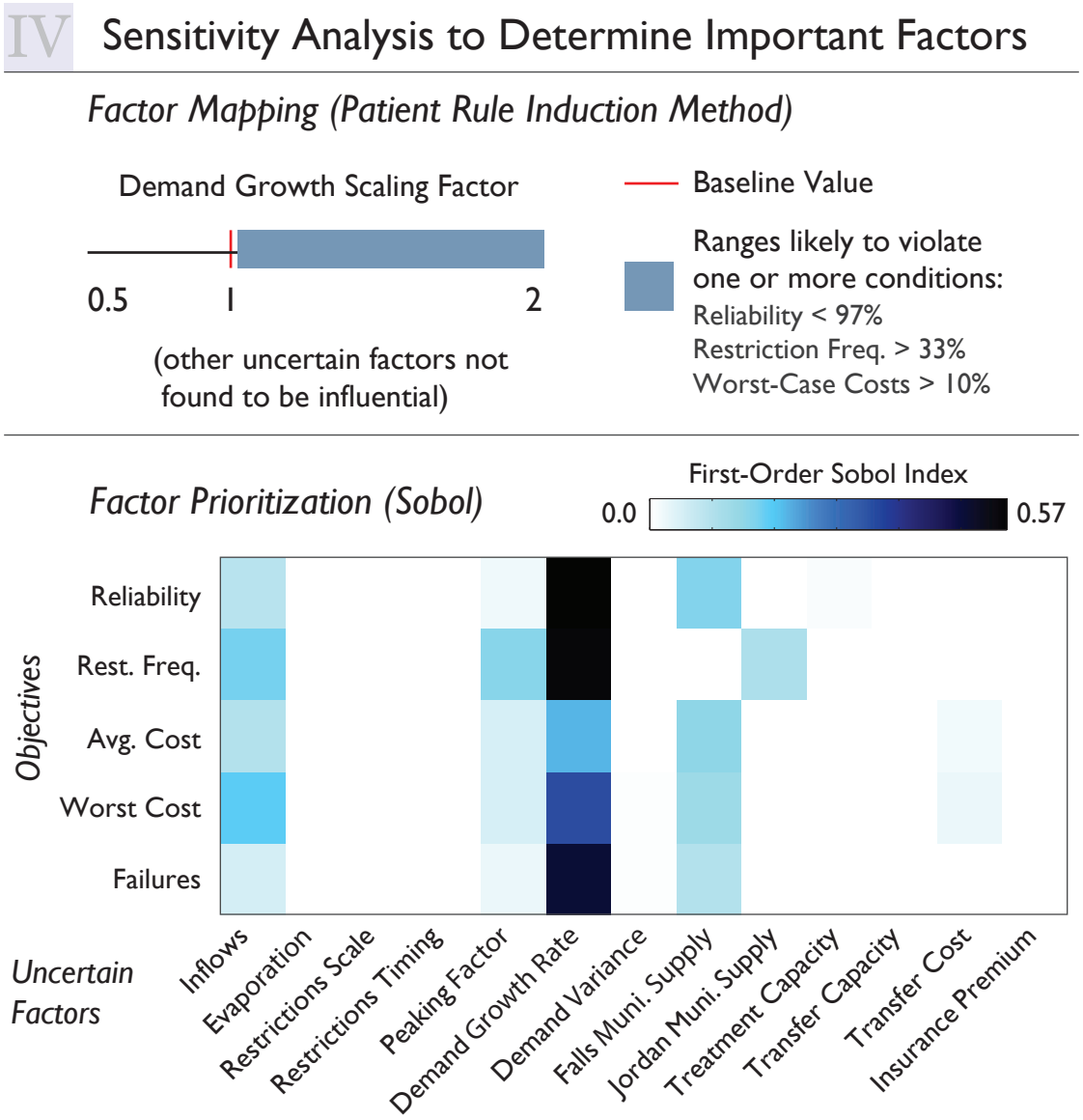

Figure 6.8: Results from a factor mapping sensitivity analysis (PRIM, top) compared to a factor prioritization (Sobol, bottom). PRIM identifies the ranges of one or more sensitive factors likely to cause a particular outcome, while Sobol indices provide a full ranking of sensitive factors. The two approaches complement each other in robustness analyses.

6.6 Conclusions

The need for water systems planning in a deeply uncertain future has challenged many classical approaches to decision support, including discrete prespecified alternatives, expected cost minimization, and probabilistic treatment of risk. Several robustness frameworks have evolved to meet this challenge, offering different approaches to identifying alternatives, sampling states of the world, quantifying robustness measures, and isolating robustness controls via sensitivity analysis. This study examines and classifies the key ideas in each of these frameworks, providing a common vocabulary to promote interchangeability of ideas and illustrating the consequences for decision making that emerge from each choice. Based on data from a recent study in multiobjective urban water supply portfolio planning, results indicate that methodological choices in the robustness analysis may lead to significantly different recommendations from the Pareto set of candidate alternatives. Alternatives prespecified by stakeholders rather than searched may prove suboptimal, both in the projected future and in other plausible states of the world (i.e., under both well-characterized and deep uncertainty). Similarly, an *a priori* restriction of the space of uncertain factors—for example, an analysis only considering climate factors—may prevent stakeholders from identifying the true vulnerabilities in the system. Decision alternatives and uncertain factors are best evaluated *a posteriori*, following optimization and sensitivity analysis, respectively. Finally, a multivariate satisficing measure of robustness allows stakeholders to achieve their problem-specific performance requirements.

Many open questions remain in the area of "bottom-up" or vulnerability-based robustness analyses. First, analyses must solidify the link to "top-down"

projections such as climate forecasts to better constrain the range of plausible future states of the world, including estimates of likelihoods where appropriate. Related to this point, it may be possible to predict failure points in the space of plausible futures to reduce the computational expense of a naive sampling. There remains the potential to better integrate the concept of deep uncertainty into optimization and the study of decision makers' risk attitudes. Given the often wide range of plausible futures, it may be prudent to move toward providing decision recommendations as a function of observed states or signposts. This study contributes an analysis of the benefits of *a posteriori* decision support and the importance of careful problem formulation, drawing concepts from multiple innovative frameworks to guide water systems planning under deep uncertainty.

CHAPTER 7

CONTRIBUTIONS & FUTURE WORK

7.1 Conclusions & Contributions

The value of computational models for decision making is challenged by uncertain assumptions regarding climate, demand, land use, and other factors essential to water resources planning and management. This dissertation has advanced diagnostic methods for water resources prediction and planning models to identify (1) time-varying dominant processes driving modeled hydrologic predictions in flood forecasting, and (2) tradeoffs and vulnerabilities to changing climate and demand in regional urban water supply systems planning for drought. Bridging these two areas is increasingly valuable, as hydrologic modelers begin to incorporate anthropogenic influences on the water cycle (e.g., Sivapalan et al., 2012), and water systems planners facing potential climate and land use change begin to explore uncertainty in hydrologic process representation (e.g., Steinschneider et al., 2012, 2014). In doing so, this work has sought to strengthen model-based decision support in water resources systems in the presence of conflicting objectives and uncertain extreme events.

(1) Dominant uncertainties vary in space and time, and inform model-based scientific inference as well as decision making.

One of the primary goals of model diagnostics is to identify the uncertain assumptions that dominate model performance. This dissertation contributes a novel approach to time-varying sensitivity analysis of a spatially distributed flood forecasting model, first advancing methods capable of efficiently analyz-

ing a high-dimensional uncertainty space (Chapter 3), then providing conceptual support to explore the dynamic model processes driving the prediction of flood events (Chapter 4). Results indicate that the method of Morris (1991), a screening approach, can identify sensitive and insensitive parameters with accuracy similar to the substantially more expensive Sobol sensitivity analysis (Sobol, 2001; Saltelli, 2002). The time-varying approach provides a foundation for model-based inference of hydrologic processes. For example, flood events in the model explored in Chapter 4 are controlled primarily by upper zone fluxes near the watershed outlet, where data collection could provide justification to falsify the internal processes of the model and improve their representation by calibrating against the new observations. Furthermore, the approach presents a valuable opportunity to overcome the complexity of distributed parameter identification by restricting search to only those parameters which are active at a specific time and location. This work contributes to the ongoing discussion of reconciling models with observations (e.g., Gupta et al., 2008), and represents a step toward the use of process-focused diagnostic methods to inform predictions in the absence of observations (such as ungauged basins or changing hydrologic conditions), where fidelity is vital (Wagener et al., 2010).

This work has also demonstrated the benefit of diagnostic methods to identify dominant uncertainties in model-based water systems planning. Traditional decision theory (e.g., von Neumann and Morgenstern, 1944; Raiffa, 1969) requires prespecification of policy alternatives, likelihood of future scenarios, and stakeholder preference over those scenarios, making the decision process vulnerable to error or contention if these prove inaccurate. Such methods are particularly challenged by the potential for long-term changes in climate (Milly et al., 2008; IPCC, 2013) and demand (Frederick and Major, 1997). Consider-

ing the deep uncertainties involved in predicting future states of the world (Lempert, 2002; Brown and Wilby, 2012), many water systems planning studies have moved toward bottom-up decision support frameworks, which focus on exploratory modeling to identify vulnerabilities prior to assessing the likelihood of specific scenarios. Chapter 6 of this dissertation has provided a thorough survey and comparison of the methods in this area. Chapter 5 advanced the Many-Objective Robust Decision Making (MORDM) framework (Kasprzyk et al., 2013) to account for multiple stakeholders, exploring tradeoffs between the robustness of four water utilities in the Research Triangle region of North Carolina, USA. Results indicated that the rate of demand growth could place far more pressure on this system through 2025 than changes in water availability, a finding which simplifies the utilities' planning process. This work represents a type of diagnostic method, though it is not commonly framed as such in the literature. The identification of vulnerabilities developed in this dissertation relates to the idea of sensitivity auditing (Saltelli et al., 2013) and seeks to strengthen model-based inference in water systems planning despite deeply uncertain parameters and forcing variables.

(2) Constructive learning and a posteriori decision support identify solutions and decision instruments that may be obscured by single-objective deterministic analysis.

Constructive learning, reviewed by Tsoukias (2008), refers to the iterative refinement of the problem definition and model structure based on elicited stakeholder feedback. This approach is consistent with multi-objective *a posteriori* decision support, in which stakeholders are presented with a set of approximately non-dominated alternatives prior to imposing preferences (Cohon and Marks, 1975). The MORDM framework combines constructive learning with

an *a posteriori* approach to identify solutions that may be obscured by decision frameworks that are non-iterative, single-objective, require *a priori* specification of preferences, or a combination thereof (Kasprzyk et al., 2013; Herman et al., 2014). This dissertation employs multi-objective evolutionary algorithms to discover promising portfolios of drought mitigation strategies, balancing objectives such as volumetric reliability and cost across multiple water utilities in a region (Chapters 5 and 6). A constructive learning approach is used to iteratively improve the set of decision instruments considered in the portfolio, building on the utilities' status quo policy to improve expected performance and robustness. Specifically, while the utilities had been considering only water use restrictions to mitigate the effects of drought, this work demonstrated the potential for a combination of water transfers and financial insurance to improve robustness substantially beyond what could be achieved with restrictions alone (Herman et al., 2014). This work has advanced the MORDM framework to incorporate multiple stakeholders, incorporating elements of Many-Objective Visual Analytics (Woodruff et al., 2013) with elicited stakeholder feedback.

The term *a posteriori* decision support traditionally refers to the discovery of solutions prior to imposing preference. This dissertation has broadened the scope of this approach to include the discovery of sensitive uncertainties and vulnerable scenarios via computational modeling rather than assuming them *a priori*. Chapter 4 demonstrated the ability of time-varying sensitivity analysis at sub-daily timescales to highlight dominant model processes relevant to flood forecasting applications that were obscured by temporal aggregation at the event scale (i.e., a prior assumption). Chapters 5-6 showed that the rate of demand growth was the primary vulnerability for the Research Triangle water supply system, which contrasts the potential prior assumption that climate fac-

tors would dominate. Each of these applications underscore the idea that sensitive uncertain factors should be discovered via computational modeling rather than assumed at the outset of an analysis. By incorporating model diagnostics as an element of *a posteriori* decision support under uncertainty, this dissertation seeks to elevate the role of sensitivity analysis in water systems planning for extreme events, informing policy decisions with model-based inference of dominant processes and vulnerable scenarios.

The contributions of this dissertation have been disseminated through several peer-reviewed journal articles. Chapters 3 and 4 were drawn from Herman et al. (2013a,b), published in *Hydrology & Earth System Sciences*. Chapter 5 was drawn from Herman et al. (2014), following the initial problem formulations developed by Zeff et al. (2014), both published in *Water Resources Research*. Chapter 6 was adapted from Herman et al. (2015), forthcoming in *Journal of Water Resources Planning & Management*. In addition, results from the Research Triangle water supply study have been presented to representatives from the Raleigh, Durham, Cary, and OWASA water utilities over a series of meetings. Finally, this work has contributed to ten conference presentations, including: six presentations at the American Geophysical Union Fall Meeting (four poster, two oral); three presentations at the ASCE World Environmental & Water Resources Congress; and one presentation at the International Congress on Environmental Modelling & Software.

Other publications external to this dissertation have supported and contributed to the concepts explored here. Chapters 3 and 4 build on Herman et al. (2013c), published in *Water Resources Research*, which considers the parameter sensitivity of three conceptual hydrologic models across a hydroclimatic gra-

dient of twelve U.S. watersheds. The findings confirm that the models represent processes differently under the same forcing, and that the changes to these dominant processes in wet versus dry catchments (for example) are difficult to predict prior to the analysis. The evolutionary optimization work presented in Chapters 5 and 6 is supported by Reed et al. (2013) (to which the author contributed), a comprehensive diagnostic assessment of multi-objective evolutionary algorithms for a suite of test problems including watershed model calibration, groundwater monitoring, and water supply planning. Finally, Giuliani et al. (2014) (to which the author contributed) advance the many-objective *a posteriori* approach for reservoir management problems on a sub-daily timescale, using a novel Direct Policy Search approach to optimize the storage-release curves for the Conowingo Reservoir in the Susquehanna River Basin. The collaborative work external to this dissertation provides a methodological foundation in sensitivity analysis, high-performance computing, and multi-objective systems planning to inform the work presented here, bridging the fields of hydrology and water supply planning.

7.2 Future Work

This dissertation ties to several burgeoning ideas across the fields of hydrology and water systems planning. This section highlights three promising areas in particular: process-based modeling of coupled human-environmental systems; interactive multi-objective decision support systems; and dynamic adaptive policies to improve robustness.

7.2.1 Process-Based Modeling of Coupled Human-Environmental Systems

There remains a need to improve the process representation not only of hydrologic models, but of coupled human-environmental systems. To the extent that water systems planning models consider hydrology, it usually serves as the forcing to the planning problem without a feedback loop to the environment. Likewise, hydrologic and ecological models often consider anthropogenic influences as inputs, but may ignore return effects. The field of sociohydrology (Sivapalan et al., 2012; Thompson et al., 2013) aims to incorporate these feedback loops in model development, tracing its lineage to studies of human-water interaction such as Falkenmark (1977). Studies in this field have drawn on system dynamics (Forrester, 1968) to better understand processes and feedback loops (e.g., Elshafei et al., 2014). Sociohydrologic models are generally descriptive with a focus on processes and behavior, which serves as a useful complement to optimization-focused models of human water systems.

The diagnostic methods put forth in this dissertation, with a combined focus on hydrology and water systems planning, can better explore the dominant processes in sociohydrologic models over a range of parameterizations and model structures. This relates to ongoing efforts in which improved process representation and the commensurate reduction of epistemic uncertainty are addressed with an information theoretic approach (e.g., Gong et al., 2013, 2014). This work also ties to the idea that the primary value of models is to understand dynamics and behavior rather than prediction at a particular time and place (Gupta and Nearing, 2014). Importantly, while most diagnostic methods (including those explored here) focus on uncertain parameters of a model, recent studies have

emphasized uncertainty in model structure (e.g., Fenicia et al., 2006; Clark et al., 2008), which can be thought of as a preconditioning that affects the sensitivity of model parameters (Gupta and Nearing, 2014). This dissertation can provide a foundation for diagnostic methods to explore process consistency in models of coupled human-environmental systems.

7.2.2 Interactive Multi-Objective Decision Support Systems

A posteriori decision support for multi-objective problems relies on the determination of stakeholder preferences after discovering the set of approximately non-dominated alternatives. This process greatly benefits from software for visual analytics (Thomas and Cook, 2006; Thomas and Kielman, 2009), or interactive data exploration, to navigate the set of alternatives. In the water resources area, Kollat and Reed (2007b) propose the VIDEO framework (Visually Interactive Decision-Making and Design using Evolutionary Optimization), demonstrated for a groundwater monitoring problem to allow stakeholders to forecast the value of investments in new observations for multiple objectives simultaneously. Visual analytics have been further developed in the water resources field by Reed and Kollat (2013) in combination with high performance computing to discover promising solutions and explore the efficiency and reliability with which they can be identified by evolutionary algorithms. There remains a need for visual analytics software capable of tracking the performance of candidate alternatives across uncertain states of the world. Hadka et al. (2015) seek to address this challenge with the OpenMORDM framework, an open-source library written in the R programming language to support Many-Objective Robust Decision Making. This work, in review in *Environmental Modelling & Software*, rep-

resents one of the first efforts to simplify and standardize the components of bottom-up decision frameworks under deep uncertainty into a single software library. OpenMORDM provides visualizations of multi-objective tradeoffs as well as tools for sensitivity analysis and model diagnostic methods to understand which candidate solutions are likely to be robust to uncertainty. The software and accompanying publication (to which the author contributed) is supported by the conceptual classification of methods proposed by Herman et al. (2015) and reviewed in Chapter 6 of this dissertation.

7.2.3 Dynamic Adaptive Policies to Improve Robustness

The current state of the art in model-based optimization and sensitivity analysis involves structuring policy instruments and parameterizing them prior to the simulation. In other words, the policy or decision variables are expected to remain static over the course of the simulation. This may not be a reasonable assumption for long-term simulations of water resources systems, because decision makers will respond to observed conditions as they change. Several recent studies have acknowledged that for solutions to be robust, they must be flexible (DiFrancesco and Tullos, 2014) or adaptive (Walker et al., 2001). However, it remains an open question how to achieve this in practice. To model adaptive responses to changing conditions, stakeholder action must be treated as a function of observed system states, whether continuous variables or binary trigger variables (i.e. "signposts", Dewar, 2002; Lempert et al., 2006). This approach has recently led to the Dynamic Adaptive Policy Pathways framework (Haasnoot et al., 2013; Hamarat et al., 2013, 2014), which seeks to model triggers that lead to a change in policy. Notably, the goal of designing adaptive policies—

to model and plan human response to changes in the environment—is only a slightly more prescriptive variation on the coupled feedback modeling pursued by sociohydrology, as described in Section 7.2.1. The common components of these approaches include system dynamics modeling and the presence of "load-bearing assumptions" (Dewar, 2002), or the processes and parameters assumed in the model that may significantly influence the outcome. The role of sensitivity analysis as described in this work is to computationally identify these important assumptions and, in the case of adaptive policy design, attempt to design trigger values for a particular action such as infrastructure investment. Methods for model diagnostics are likely to expand their role in decision support as hydrologists and water planners seek to represent the complex interactions between human stakeholders and the environment.

This dissertation advances model diagnostics as a key element of *a posterori* decision support for multi-objective water resources systems under uncertainty. The value of computational models for decision making is limited by uncertain assumptions regarding climate, demand, land use, and other factors essential to water resources planning and management; it is the task of diagnostic methods to locate vulnerable assumptions and design strategies to mitigate them. This work contributes a novel approach to time-varying sensitivity analysis of a spatially distributed flood forecasting model, identifying dominant processes responsible for model error with the goal of improving predictions in the absence of observations, such as ungauged basins or under changing hydrologic conditions. This work has also demonstrated the benefit of diagnostic methods to identify dominant uncertainties in model-based water systems planning, specifically, changing climate and demands in regional urban water supply systems planning for drought. This represents a broader class of bottom-up deci-

sion support frameworks, which focus on exploratory modeling to identify vulnerabilities under deep uncertainty prior to assessing the likelihood of specific scenarios. As computational models of increasing complexity are called upon to inform policy recommendations, understanding the relative importance of uncertain assumptions will continue to be an essential component of water resources planning. Although no model can be considered complete or correct, a structured approach to interrogate its credibility, vulnerabilities, and trade-offs can strengthen model-based decision support in the presence of conflicting objectives and uncertain extreme events.

BIBLIOGRAPHY

Abbott, M., Bathurst, J., Cunge, J., O'Connell, P., and Rasmussen, J. (1986). "An introduction to the European Hydrological System (SHE), 2: Structure of a physically-based, distributed modelling system." *Journal of Hydrology*, 87(1), 61–77.

Alton, P., Mercado, L., and North, P. (2006). "A sensitivity analysis of the land-surface scheme JULES conducted for three forest biomes: Biophysical parameters, model processes, and meteorological driving data." *Global Biogeochemical Cycles*, 20(1), GB1008.

Archer, G., Saltelli, A., and Sobol, I. (1997). "Sensitivity measures, ANOVA-like techniques and the use of bootstrap." *J. Statist. Comput. Simul.*, 58, 99–120.

Balling, R. J. and Richard, J. (2000). "Pareto sets in decision-based design." *Journal of Engineering Valuation and Cost Analysis*, 3(2), 189–198.

Bankes, S. C. (2002). "Tools and techniques for developing policies for complex and uncertain systems." *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3), 7263–7266.

Bankes, S. C., Lempert, R. J., and Popper, S. W. (2001). "Computer-assisted reasoning." *Computing in Science & Engineering*, 3(2), 71–77.

Banzhaf, H. S. (2009). "Objective or multi-objective? two historically competing visions for benefit-cost analysis." *Land Economics*, 85(1), 3–23.

Bastidas, L., Hogue, T., Sorooshian, S., Gupta, H., and Shuttleworth, W. (2006). "Parameter sensitivity analysis for different complexity land surface models using multicriteria methods." *Journal of Geophysical Research*, 111(D20), 20101.

Ben-Haim, Y. (2004). "Uncertainty, probability and information-gaps." *Reliability Engineering & System Safety*, 85(1), 249–266.

Ben-Haim, Y. (2012). "Why risk analysis is difficult, and some thoughts on how to proceed." *Risk Analysis*, 32(10), 1638–1646.

Beven, K. (1989). "Changing ideas in hydrology: The case of physically–based models." *Journal of Hydrology*, 105(1–2), 157–172.

Beyer, H.-G. and Sendhoff, B. (2007). "Robust optimization: A comprehensive survey." *Computer Methods in Applied Mechanics and Engineering*, 196(33-34), 3190–3218.

Bond, S. D., Carlson, K. A., and Keeney, R. L. (2008). "Generating objectives: Can decision makers articulate what they want?." *Management Science*, 54(1), 56–70.

Bond, S. D., Carlson, K. A., and Keeney, R. L. (2010). "Improving the generation of decision objectives." *Decision Analysis*, 7(3), 238–255.

Bonder, S. (1979). "Changing the future of operations research." *Operations Research*, 27(2), 209–224.

Borgonovo, E. and Peccati, L. (2008). "Sensitivity analysis in decision making: a consistent approach." *Advances in Decision Making Under Risk and Uncertainty*, Springer-Verlag, Berlin, 65–89.

Brill, E. D. (1979). "The use of optimization models in public-sector planning." *Management Science*, 25(5), 413–422.

Brown, C. (2010). "Decision-scaling for robust planning and policy under cli-

mate uncertainty." *Expert Perspectives Series Written for the World Resources Report*, 2011 Washington, DC.

Brown, C. and Carriquiry, M. (2007). "Managing hydroclimatological risk to water supply with option contracts and reservoir index insurance." *Water Resources Research*, 43(11), W11423.

Brown, C., Ghile, Y., Laverty, M., and Li, K. (2012). "Decision Scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector." *Water Resources Research*, 48(9), W09537.

Brown, C. and Wilby, R. L. (2012). "An alternate approach to assessing climate risks." *Eos, Transactions American Geophysical Union*, 93(41), 401–402.

Bryant, B. P. (2009). *sdtoolkit: Scenario Discovery Tools to Support Robust Decision Making*, <http://CRAN.R-project.org/package=sdtoolkit>. R package version 2.31.

Bryant, B. P. and Lempert, R. J. (2010). "Thinking inside the box: A participatory, computer-assisted approach to scenario discovery." *Technological Forecasting and Social Change*, 77(1), 34–49.

Burnash, R. and Singh, V. (1995). "The NWS river forecast System–Catchment modeling." *Computer Models of Watershed Hydrology*, Water Resour. Publ., Littleton, Colorado, 311–366.

Caldwell, C. and Characklis, G. W. (2014). "Impact of contract structure and risk aversion on inter-utility water transfer agreements." *Journal of Water Resources Planning and Management*, 140(1), 100–111.

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). "An effective screening

design for sensitivity analysis of large models." *Environmental Modelling & Software*, 22(10), 1509–1518.

Campolongo, F., Saltelli, A., and Cariboni, J. (2011). "From screening to quantitative sensitivity analysis. a unified approach." *Computer Physics Communications*, 182(4), 978–988.

Carpenter, T. M., Georgakakos, K. P., and Sperfslagea, J. A. (2001). "On the parametric and NEXRAD-radar sensitivities of a distributed hydrologic model suitable for operational use." *Journal of Hydrology*, 253(1–4), 169–193.

Chamberlin, T. C. (1890). "The method of multiple working hypotheses." *Science*, 15(366), 92–96.

Clark, M., Slater, A., Rupp, D., et al. (2008). "Framework for understanding structural errors (FUSE): a modular framework to diagnose differences between hydrological models." *Water Resour. Res*, 44, W00B02.

Clemen, R. T. (1996). *Making hard decisions: an introduction to decision analysis*. Duxbury.

Clímaco, J. C. (2004). "A critical reflection on optimal decision." *European Journal of Operational Research*, 153(2), 506–516.

Cloke, H., Pappenberger, F., and Renaud, J. (2008). "Multi–method global sensitivity analysis (MMGSA) for modelling floodplain hydrological processes." *Hydrological processes*, 22(11), 1660–1674.

Coello Coello, C. A. (2007). *Evolutionary Algorithms for Solving Multi-Objective Problems*. Genetic and Evolutionary Computation. Springer, New York, 2 edition.

Cohon, J. and Marks, D. (1973). "Multiobjective screening models and water resource investment." *Water Resources Research*, 9(4), 826–836.

Cohon, J. and Marks, D. (1975). "A review and evaluation of multiobjective programming techniques." *Water Resources Research*, 11(2), 208–220.

Crossley, W. A., Cook, A. M., and Fanjoy, D. W. (1999). "Using the two-branch tournament genetic algorithm for multiobjective design." *AIAA journal*, 37(2), 261–267.

Cuo, L., Giambelluca, T. W., and Ziegler, A. D. (2011). "Lumped parameter sensitivity analysis of a distributed hydrological model within tropical and temperate catchments." *Hydrological Processes*, 25(15), 2405–2421.

Deb, K. and Agrawal, R. B. (1994). "Simulated binary crossover for continuous search space." *Report No. IITK/ME/SMD-94027*, Indian Institute of Technology, Kanpur.

Deb, K. and Gupta, H. (2006). "Introducing robustness in multi-objective optimization." *Evolutionary Computation*, 14(4), 463–494.

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

Deb, K., Mohan, M., and Mishra, S. (2005). "Evaluating the epislon-domination based multiobjective evolutionary algorithm for a quick computation of pareto-optimal solutions." *Evolutionary Computation Journal*, 13(4), 501–525.

Demaria, E., Nijssen, B., and Wagener, T. (2007). "Monte Carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model." *J. Geophys. Res*, 112, D11113.

Dessai, S., Browne, A., and Harou, J. J. (2013). "Introduction to the special issue on adaptation and resilience of water systems to an uncertain changing climate." *Water Resour Manage*, 27(4), 943–948.

Dessai, S. and Hulme, M. (2007). "Assessing the robustness of adaptation decisions to climate change uncertainties: A case study on water resources management in the East of England." *Global Environmental Change*, 17(1), 59–72.

Dewar, J. A. (2002). *Assumption-based planning: A tool for reducing avoidable surprises*. Cambridge University Press.

DiFrancesco, K. N. and Tullos, D. D. (2014). "Flexibility in water resources management: Review of concepts and development of assessment measures for flood management systems." *JAWRA Journal of the American Water Resources Association*, 50(6), 1527–1539.

Dooge, J. C. (1957). "The rational method for estimating flood peaks." *Engineering*, 184(1), 311–313.

Dougherty, D. E. and Marryott, R. A. (1991). "Optimal groundwater management: 1. simulated annealing." *Water Resour. Res.*, 27(10), 2493–2508.

Efron, B. (1979). "Bootstrap methods: another look at the jackknife." *The annals of Statistics*, 1–26.

Efron, B. and Stein, C. (1981). "The jackknife estimate of variance." *The Annals of Statistics*, 9(3), 586–596.

Ehret, U., Gupta, H. V., Sivapalan, M., et al. (2013). "Advancing catchment hydrology to deal with predictions under change." *Hydrol. Earth Syst. Sci. Discuss.*, 10(7), 8581–8634.

Elshafei, Y., Sivapalan, M., Tonts, M., and Hipsey, M. (2014). "A prototype framework for models of socio-hydrology: identification of key feedback loops and parameterisation approach." *Hydrology and Earth System Sciences*, 18(6), 2141–2166.

Ewen, J. and Parkin, G. (1996). "Validation of catchment models for predicting land–use and climate change impacts. 1. method." *Journal of Hydrology*, 175(1–4), 583–594.

Falkenmark, M. (1977). "Water and mankind: A complex system of mutual interaction." *Ambio*, 3–9.

Fenicia, F., Savenije, H., Matgen, P., and Pfister, L. (2006). "Is the groundwater reservoir linear? Learning from data in hydrological modelling." *Hydrology and Earth System Sciences*, 10(1), 139–150.

Forrester, J. (1968). *Principles of Systems*. Wright-Allen Press, Cambridge, MA.

Franchini, M., Wendling, J., Obled, C., and Todini, E. (1996). "Physical interpretation and sensitivity analysis of the TOPMODEL." *Journal of Hydrology*, 175(1–4), 293–338.

Frederick, K. D. and Major, D. C. (1997). "Climate change and water resources." *Climatic Change*, 37(1), 7–23.

Frederick, K. D. and Schwarz, G. (1999). "Socioeconomic impacts of climate change on U.S. water supplies." *Journal of the American Water Resources Association*, 35(6), 1563–1583.

Freer, J., Beven, K., and Ambroise, B. (1996). "Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach." *Water Resources Research*, 32(7), 2161–2173.

Friedman, J. H. and Fisher, N. (1999). "Bump hunting in high-dimensional data." *Stat. Comput.*, 9(2), 123–143.

Friedman, M. (1976). *Price Theory*. Aldine Publishing Company, Chicago.

Gerst, M., Wang, P., and Borsuk, M. (2013). "Discovering plausible energy and economic futures under global change using multidimensional scenario discovery." *Environmental Modelling & Software*, 44, 76–86.

Ghile, Y., Taner, M., Brown, C., Grijsen, J., and Talbi, A. (2014). "Bottom-up climate risk assessment of infrastructure investment in the Niger River Basin." *Climatic Change*, 122(1-2), 97–110.

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

Gleick, P. H. (2002). "Soft water paths." *Nature*, 418, 373.

Gleick, P. H. (2003). "Global freshwater resources: soft-path solutions for the 21st century." *Science*, 302(5650), 1524–1528.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

Gong, W., Gupta, H. V., Yang, D., Sricharan, K., and Hero, A. O. (2013). "Estimating epistemic and aleatory uncertainties during hydrologic modeling: An information theoretic approach." *Water Resources Research*, 49(4), 2253–2273.

Gong, W., Yang, D., Gupta, H. V., and Nearing, G. (2014). "Estimating information entropy for hydrological data: One-dimensional case." *Water Resources Research*, 50(6), 5003–5018.

Griffin, M. T., Montz, B. E., and S Arrigo, J. (2013). "Evaluating climate change induced water stress: A case study of the Lower Cape Fear Basin, NC." *Applied Geography*, 40, 115–128.

Groves, D., Bloom, E., Lempert, R., et al. (2014). "Developing key indicators for adaptive water planning." *Journal of Water Resources Planning and Management*, 0(0), 05014008.

Groves, D. G. and Lempert, R. J. (2007). "A new analytic method for finding policy-relevant scenarios." *Global Environmental Change*, 17(1), 73–85.

Gupta, Wagener, T., and Liu, Y. (2008). "Reconciling theory with observations: elements of a diagnostic approach to model evaluation." *Hydrological Processes*, 22(18), 3802–3813.

Gupta, H., Sorooshian, S., and Yapo, P. (1998). "Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information." *Water Resources Research*, 34, 751–763.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. (2009). "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling." *Journal of Hydrology*, 377(1), 80–91.

Gupta, H. V. and Nearing, G. S. (2014). "Debates—the future of hydrological sciences: A (common) path forward? using models and data to learn: A systems theoretic perspective on the future of hydrological science." *Water Resources Research*, 50(6), 5351–5359.

Guse, B., Reusser, D. E., and Fohrer, N. (2013). "How to improve the representation of hydrological processes in SWAT for a lowland catchment–temporal

analysis of parameter sensitivity and model performance." *Hydrological Processes*.

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

Hadjigeorgalis, E. (2008). "Managing drought through water markets: Farmer preferences in the Rio Grande Basin." *Journal of the American Water Resources Association*, 44(3), 594–605.

Hadka, D., Herman, J. D., Reed, P. M., and Keller, K. (2015). "Openmordm: An open source framework for many-objective robust decision making." *Environmental Modelling & Software*, In Press.

Hadka, D. and Reed, P. (2012). "Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization." *Evolutionary Computation*, 20(3), 423–452.

Hadka, D. and Reed, P. (2013). "Borg: An auto-adaptive many-objective evolutionary computing framework." *Evolutionary Computation*, 21(2), 231–259.

Hadka, D., Reed, P. M., and Simpson, T. W. (2012). "Diagnostic assessment of the borg moea for many-objective product family design problems." *Evolutionary Computation (CEC), 2012 IEEE Congress on*, IEEE, 1–10.

Haimes, Y. Y. and Hall, W. A. (1974). "Multiobjectives in water resource systems analysis: The surrogate worth trade off method." *Water Resources Research*, 10(4), 615–624.

Haimes, Y. Y. and Hall, W. A. (1977). "Sensitivity, responsivity, stability, and

irreversability as multiple objectives in civil systems." *Advances in Water Resources*, 1(2), 71–81.

Hall, J., Tarantola, S., Bates, P., and Horritt, M. (2005). "Distributed sensitivity analysis of flood inundation model calibration." *Journal of Hydraulic Engineering*, 131, 117–126.

Hall, J. W., Lempert, R. J., Keller, K., et al. (2012). "Robust climate policies under uncertainty: A comparison of robust decision making and info-gap methods." *Risk Analysis*, 32(10), 1657–1672.

Hamarat, C., Kwakkel, J. H., and Pruyt, E. (2013). "Adaptive robust design under deep uncertainty." *Technological Forecasting and Social Change*, 80(3), 408–418.

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

Hashimoto, T., Loucks, D. P., and Stedinger, J. R. (1982). "Robustness of water resources systems." *Water Resources Research*, 18(1), 21–26.

Hawkins, E. and Sutton, R. (2009). "The potential to narrow uncertainty in regional climate predictions." *Bulletin of the American Meteorological Society*, 90(8), 1095–1107.

Hawkins, E. and Sutton, R. (2011). "The potential to narrow uncertainty in projections of regional precipitation change." *Climate Dynamics*, 37(1-2), 407–418.

Herman, J., Kollat, J., Reed, P., and Wagener, T. (2013a). "Technical note: Method of Morris effectively reduces the computational demands of global sensitiv-

ity analysis for distributed watershed models." *Hydrology and Earth System Sciences*, 17(4), 2983–2903.

Herman, J. D., Kollat, J. B., Reed, P. M., and Wagener, T. (2013b). "From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models." *Hydrology and Earth System Sciences*, 17(12), 5109–5125.

Herman, J. D., Reed, P. M., and Wagener, T. (2013c). "Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior." *Water Resources Research*, 49(3), 1400–1414.

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

Herman, J. D., Zeff, H. B., Reed, P. M., and Characklis, G. W. (2014). "Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty." *Water Resources Research*, 50(10), 7692–7713.

Hine, D. and Hall, J. W. (2010). "Information gap analysis of flood model uncertainties and regional frequency analysis." *Water Resources Research*, 46(1).

Hipel, K. W. and Ben-Haim, Y. (1999). "Decision making in an uncertain world: Information-gap modeling in water resources management." *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 29(4), 506–517.

Hitch, C. J. (1960). "On the choice of objectives in systems studies." *Report No. P-1955*, The RAND Corporation.

Hornberger, G. and Spear, R. (1981). "Approach to the preliminary analysis of environmental systems." *Journal of Environmental Management*, 12(1), 7–18.

Hughes, J. and Leurig, S. (2013). "Assessing water system revenue risk: Considerations for market analysts." *Report no.*, Ceres Organization.

Hwang, C.-L., Masud, A. S. M., Paidy, S. R., and Yoon, K. P. (1979). *Multiple objective decision making, methods and applications: a state-of-the-art survey*, Vol. 164. Springer Berlin, <http://library.wur.nl/WebQuery/clc/155578>.

Iooss, B. and Lemaître, P. (2014). "A review on global sensitivity analysis methods." *arXiv:1404.2405 [math, stat]* arXiv: 1404.2405.

IPCC (2013). "Climate change 2013. the physical science basis. working group I contribution to the fifth assessment report of the intergovernmental panel on climate change." Cambridge University Press, Cambridge UK and New York, NY, USA.

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, The Pennsylvania State University.

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

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

Kasprzyk, J. R., Reed, P. M., Kirsch, B. R., and Characklis, G. W. (2009). "Managing population and drought risks using many-objective water portfolio planning under uncertainty." *Water Resources Research*, 45(12), W12401.

Keim, D., Mansmann, F., Schneidewind, J., and Ziegler, H. (2006). "Challenges in visual data analysis." *Proceedings of Information Visualization*, IEEE Computer Society, London, UK, 9–16.

Kirchner, J. W. (2006). "Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology." *Water Resources Research*, 42(3), W03S04.

Kirsch, B. R., Characklis, G. W., and 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(4), 396–406.

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

Knight, F. H. (1921). *Risk, uncertainty and profit*. Hart, Schaffner and Marx, New York.

Kollat, J. B. and Reed, P. M. (2007a). "A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications." *Advances in Water Resources*, 30(3), 335–353.

Kollat, J. B. and Reed, P. M. (2007b). "A framework for Visually Interactive Decision-making and Design using Evolutionary Multiobjective Optimization (VIDEO)." *Environmental Modelling & Software*, 22(12), 1691–1704.

Koren, V., Reed, S., Smith, M., Zhang, Z., and Seo, D.-J. (2004). "Hydrology laboratory research modeling system (HL-RMS) of the US national weather service." *Journal of Hydrology*, 291(3–4), 297–318.

Korteling, B., Dessai, S., and Kapelan, Z. (2013). "Using information-gap decision theory for water resources planning under severe uncertainty." *Water resources management*, 27(4), 1149–1172.

Kwakkel, J. H. and Pruyt, E. (2013). "Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty." *Technological Forecasting and Social Change*, 80(3), 419–431.

Lane, M. E., Kirshen, P. H., and Vogel, R. M. (1999). "Indicators of impacts of global climate change on U.S. water resources." *Journal of Water Resources Planning and Management*, 125(4), 194–204.

Langlois, R. N. and Cosgel, M. M. (1993). "Frank Knight on risk, uncertainty, and the firm: A new interpretation." *Economic Inquiry*, 31(3), 456–465.

Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). "Combining convergence and diversity in evolutionary multiobjective optimization." *Evolutionary computation*, 10(3), 263–282.

Lempert, R. J. (2002). "A new decision sciences for complex systems." *Proceedings of the National Academy of Sciences*, 99(suppl. 3), 7309–7313.

Lempert, R. J., Bryant, B. P., and Bankes, S. C. (2008). "Comparing algorithms for scenario discovery." *Report No. WR-557-NSF*, RAND.

Lempert, R. J. and Collins, M. (2007). "Managing the risk of an uncertain threshold response: Comparison of robust, optimimum, and precautionary approaches." *Risk Analysis*, 27(4), 1009–1026.

Lempert, R. J. and 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(6), 960–974.

Lempert, R. J., Groves, D. G., Popper, S. W., and Bankes, S. C. (2006). "A general, analytic method for generating robust strategies and narrative scenarios." *Management Science*, 52(4), 514–528.

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

Leung, L., Collins, B., and Famiglietti, J. (2013). "Community modeling and long-term predictions of the integrated water cycle." Washington, D.C., U.S. Department of Energy Office of Science.

Li, K., Qi, J., Brown, C., and Ryan, J. (2014). "Effect of scenario assumptions on climate change risk estimates in a water resource system." *Climate Research*, 59(2), 149–160.

Liebman, J. C. (1976). "Some simple-minded observations on the role of optimization in public systems decision-making." *Interfaces*, 6(4), 102–108.

Lins, H. F. and Stakhiv, E. Z. (1998). "Managing the nation's water in a changing climate." *JAWRA Journal of the American Water Resources Association*, 34(6), 1255–1264.

Liu, Y., Gupta, H., Springer, E., and Wagener, T. (2008). "Linking science with environmental decision making: Experiences from an integrated modeling

approach to supporting sustainable water resources management." *Environmental Modelling & Software*, 23(7), 846–858.

Lownsbery, K. E. (2014). "Quantifying the impacts of future uncertainties on the Apalachicola-Chattahoochee-Flint basin." M.S. thesis, University of Massachusetts Amherst, University of Massachusetts Amherst.

Lund, J. R. and Israel, M. (1995). "Optimization of transfers in urban water supply planning." *Journal of Water Resources Planning and Management*, 121(1).

Maass, A., Hufschmidt, M. M., Dorfman, R., et al. (1962). *Design of Water-Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning*. Harvard University Press, Cambridge.

Mahmoud, M., Liu, Y., Hartmann, H., et al. (2009). "A formal framework for scenario development in support of environmental decision-making." *Environmental Modelling and Software*, 24(7), 798–808.

Maier, H. R., Kapelan, Z., Kasprzyk, J., et al. (2014). "Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions." *Environmental Modelling & Software*, 62, 271–299.

Matalas, N. C. (2012). "Comment on the announced death of stationarity." *Journal of Water Resources Planning and Management*, 138(4), 311–312.

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

Matrosov, E. S., Padula, S., and Harou, J. J. (2013a). "Selecting portfolios of water supply and demand management strategies under uncertainty: Contrasting

economic optimisation and robust decision making approaches." *Water Resources Management*, 27(4), 1123–1148.

Matrosov, E. S., Woods, A. M., and Harou, J. J. (2013b). "Robust decision making and info-gap decision theory for water resource system planning." *Journal of Hydrology*, 494, 43–58.

McCuen, R. (1973). "The role of sensitivity analysis in hydrologic modeling." *Journal of Hydrology*, 18(1), 37–53.

McInerney, D., Lempert, R., and Keller, K. (2012). "What are robust strategies in the face of uncertain climate threshold responses?." *Climatic Change*, 112(3-4), 547–568.

McKinney, D. C. and Lin, M.-D. (1994). "Genetic algorithm solution of groundwater management models." *Water Resour. Res.*, 30(6), 1897–1906.

Mendoza, P. A., Clark, M. P., Barlage, M., et al. (2015). "Are we unnecessarily constraining the agility of complex process-based models?." *Water Resources Research*, 51.

Messac, A. and Mattson, C. A. (2002). "Generating well-distributed sets of Pareto points for engineering design using physical programming." *Optimization and Engineering*, 3(4), 431–450.

Milly, P. C. D., Betancourt, J., Falkenmark, M., et al. (2008). "Stationarity is dead: Whither water management?." *Science*, 319, 573–574.

Moody, P. and Brown, C. (2013). "Robustness indicators for evaluation under climate change: Application to the upper Great Lakes." *Water Resources Research*, 49(6), 3576–3588.

Moreda, F., Koren, V., Zhang, Z., Reed, S., and Smith, M. (2006). "Parameterization of distributed hydrological models: learning from the experiences of lumped modeling." *Journal of Hydrology*, 320(1–2), 218–237.

Morris, M. D. (1991). "Factorial sampling plans for preliminary computational experiments." *Technometrics*, 33(2), 161–174.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., et al. (2010). "The next generation of scenarios for climate change research and assessment." *Nature*, 463(7282), 747–756.

Muleta, M. and Nicklow, J. (2005). "Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model." *Journal of Hydrology*, 306(1–4), 127–145.

Mulvey, J. M., Vanderbei, R. J., and Zenios, S. A. (1995). "Robust optimization of large-scale systems." *Operations research*, 43(2), 264–281.

Nazemi, A. A. and Wheater, H. S. (2014). "Assessing the vulnerability of water supply to changing streamflow conditions." *Eos, Transactions American Geophysical Union*, 95(32), 288–288.

Nicklow, J., Reed, P., Savic, D., et al. (2010). "State of the art for genetic algorithms and beyond in water resources planning and management." *Journal of Water Resources Planning and Management*, 136(4), 412–432.

Nossent, J. and Bauwens, W. (2012a). "Application of a normalized Nash-Sutcliffe efficiency to improve the accuracy of the Sobol sensitivity analysis of a hydrological model." *EGU General Assembly Conference Abstracts*, Vol. 14, 237.

Nossent, J. and Bauwens, W. (2012b). "Optimising the convergence of Sobol sensitivity analysis for an environmental model: Application of an appropriate estimate for the square of the expectation value and total variance." *Proceedings of the International Congress on Environmental Modelling and Software.*, R. Seppelt, A. Voinov, S. Lange, and D. Bankamp, eds., 1080–1087.

Nossent, J., Elsen, P., and Bauwens, W. (2011). "Sobol sensitivity analysis of a complex environmental model." *Environmental Modelling & Software*, 26(12), 1515–1525.

Palmer, R. N. and Characklis, G. W. (2009). "Reducing the costs of meeting regional water demand through risk-based transfer agreements." *Journal of Environmental Management*, 90.

Pappenberger, F., Beven, K. J., Ratto, M., and Matgen, P. (2008). "Multi–method global sensitivity analysis of flood inundation models." *Advances in Water Resources*, 31(1), 1–14.

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

Paton, F. L., Maier, H. R., and Dandy, G. C. (2014). "Including adaptation and mitigation responses to climate change in a multiobjective evolutionary algorithm framework for urban water supply systems incorporating GHG emissions." *Water Resour. Res.*, 50(8), 6285–6304.

Perelman, L., Housh, M., and Ostfeld, A. (2013). "Robust optimization for water distribution systems least cost design." *Water Resources Research*, 49(10), 6795–6809.

Plischke, E., Borgonovo, E., and Smith, C. L. (2013). "Global sensitivity measures from given data." *European Journal of Operational Research*, 226(3), 536–550.

Polasky, S., Carpenter, S. R., Folke, C., and Keeler, B. (2011). "Decision-making under great uncertainty: environmental management in an era of global change." *Trends in Ecology and Evolution*, 26(8), 398–404.

Popper, K. R. (1963). *Conjectures and refutations*, Vol. 28. Routledge & Kegan Paul London.

Postel, S. L. (2000). "Entering an era of water scarcity: the challenges ahead." *Ecological applications*, 10(4), 941–948.

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

Pujol, G., Iooss, B., and Janon, A. (2013). *Sensitivity Analysis Package*, <http://cran.r-project.org/web/packages/sensitivity/index.html>. R package version 1.7-0.

Raiffa, H. (1969). "Preferences for multi-attributed alternatives." *Report No. RM-5868-DOT/RC*, RAND, Santa Monica, CA.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., and 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. and Kasprzyk, J. R. (2009). "Water resources management: The myth, the wicked, and the future." *Journal of Water Resources Planning and Management*, 135(6), 411–413.

Reed, P. M. and Kollat, J. B. (2013). "Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example." *Advances in Water Resources*, 56, 1–13.

Reed, S., Koren, V., Smith, M., et al. (2004). "Overall distributed model intercomparison project results." *Journal of Hydrology*, 298(1–4), 27–60.

Renwick, M. E. and Green, R. D. (2000). "Do residential water demand side management policies measure up? An analysis of eight California water agencies." *Journal of Environmental Economics and Management*, 40(1), 37–55.

Reuss, M. (2003). "Is it time to resurrect the Harvard water program?." *Journal of Water Resources Planning and Management*, 129(5), 357–360.

Reusser, D., Blume, T., Schaefli, B., and Zehe, E. (2009). "Analysing the temporal dynamics of model performance for hydrological models." *Hydrology and Earth System Sciences*, 13(7), 999–1018.

Reusser, D., Buytaert, W., and Zehe, E. (2011). "Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test." *Water Resources Research*, 47(7), W07551.

Reusser, D. and Zehe, E. (2011). "Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity." *Water Resources Research*, 47(7), W07550.

Reynoso-Meza, G., Sanchis, J., Blasco, X., and García-Nieto, S. (2014). "Physical programming for preference driven evolutionary multi-objective optimization." *Applied Soft Computing*, 24, 341–362.

Rittel, H. W. J. and Webber, M. M. (1973). "Dilemmas in a general theory of planning." *Policy Sci*, 4(2), 155–169.

Rosenblueth, A. and Wiener, N. (1945). "The role of models in science." *Philosophy of science*, 12(4), 316–321.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Gonçalves, L. G. G., and Zeng, X. (2012). "Towards a comprehensive approach to parameter estimation in land surface parameterization schemes." *Hydrological Processes*, 27, 2075–2097.

Roy, B. (1971). "Problems and methods with multiple objective functions." *Mathematical Programming*, 1(1), 239–266.

Roy, B. (1990). "Decision-aid and decision-making." *European Journal of Operational Research*, 45(2–3), 324–331.

Ruano, M., Ribes, J., Seco, A., and Ferrer, J. (2012). "An improved sampling strategy based on trajectory design for application of the Morris method to systems with many input factors." *Environmental Modelling & Software*, 37, 103–109.

Saltelli, A. (2002). "Making best use of model evaluations to compute sensitivity indices." *Computer Physics Communications*, 145(2), 280–297.

Saltelli, A., Annoni, P., Azzini, I., et al. (2010). "Variance based sensitivity analysis of model output: Design and estimator for the total sensitivity index." *Computer Physics Communications*, 181(2), 259–270.

Saltelli, A. and Funtowicz, S. (2013). "When all models are wrong." *Issues in Science and Technology*, 30(2), 79–85.

Saltelli, A., Guimarães Pereira, Â., Van der Sluijs, J. P., and Funtowicz, S. (2013). "What do i make of yourLatinorum? Sensitivity auditing of mathematical modelling." *International Journal of Foresight and Innovation Policy*, 9(2), 213–234.

Saltelli, A., Ratto, M., Andres, T., et al. (2008). *Global Sensitivity Analysis: The Primer*. Wiley, West Sussex, England.

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

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

Schoemaker, P. J. (1982). "The expected utility model: Its variants, purposes, evidence and limitations." *Journal of economic literature*, 20(2), 529–563.

Scudder, T. (2005). *The Future of Large Dams*. Cromwell Press, London.

Seager, R., Tzanova, A., and Nakamura, J. (2009). "Drought in the southeastern United States: causes, variability over the last millennium, and the potential for future hydroclimate change." *Journal of Climate*, 22(19), 5021–5045.

Sieber, A. and Uhlenbrook, S. (2005). "Sensitivity analyses of a distributed catchment model to verify the model structure." *Journal of Hydrology*, 310(1–4), 216–235.

Simon, H. A. (1959). "Theories of decision-making in economics and behavioral science." *The American Economic Review*, 49(3), 253–283.

Sivapalan, M., Savenije, H. H., and Blöschl, G. (2012). "Socio-hydrology: A new science of people and water." *Hydrological Processes*, 26(8), 1270–1276.

Smith, M. B., Koren, V., Reed, S., et al. (2012). "The distributed model intercomparison project – phase 2: Motivation and design of the oklahoma experiments." *Journal of Hydrology*, 418–419, 3–16.

Smith, M. B., Seo, D.-J., Koren, V. I., et al. (2004). "The distributed model intercomparison project (DMIP): motivation and experiment design." *Journal of Hydrology*, 298(1–4), 4–26.

Sobol, I. (2001). "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates." *Mathematics and Computers in Simulation*, 55(1–3), 271–280.

Sobol, I. M. (1993). "Sensitivity estimates for nonlinear mathematical models and their Monte Carlo estimation." *Math. Model. Comput. Exp.*, 1(4), 407–417.

Starr, M. K. (1962). *Product Design and Decision Theory*. Prentice-Hall, Englewood Cliffs, N.J.

Steinschneider, S., Polebitski, A., Brown, C., and Letcher, B. H. (2012). "Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change." *Water Resources Research*, 48(11).

Steinschneider, S., Wi, S., and Brown, C. (2014). "The integrated effects of climate and hydrologic uncertainty on future flood risk assessments." *Hydrological Processes*, In Press.

Storn, R. and Price, K. (1997). "Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces." *Journal of Global Optimization*, 11(4), 341–359.

Tang, Y. (2007). "Advancing hydrologic model evaluation and identification using multiobjective calibration, sensitivity analysis, and parallel computation." Ph.D. thesis, The Pennyslvania State University, The Pennyslvania State University.

Tang, Y., Reed, P., Van Werkhoven, K., and Wagener, T. (2007a). "Advancing the identification and evaluation of distributed rainfall–runoff models using global sensitivity analysis." *Water Resources Research*, 43(6), W06415.

Tang, Y., Reed, P., Wagener, T., and Van Werkhoven, K. (2007b). "Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation." *Hydrology and Earth System Sciences*, 11(2), 793–817.

Thomas, J. and Cook, K. A. (2006). "A visual analytics agenda." *IEEE Computer Graphics and Applications*, 26(1), 10–13.

Thomas, J. and Kielman, J. (2009). "Challenges for visual analytics." *Information Visualization*, 8(4), 309–314.

Thompson, S., Sivapalan, M., Harman, C., et al. (2013). "Developing predictive insight into changing water systems: use-inspired hydrologic science for the anthropocene." *Hydrology and Earth System Sciences*, 17(12), 5013–5039.

Tingstad, A. H., Groves, D. G., and Lempert, R. J. (2014). "Paleoclimate scenarios to inform decision making in water resource management: Example from southern California's Inland Empire." *Journal of Water Resources Planning and Management*, 140(10), 04014025.

Tsoukias, A. (2008). "From decision theory to decision aiding methodology." *European Journal of Operational Research*, 187(1), 138–161.

Tsutsui, S., Yamamura, M., and Higuchi, T. (1999). "Multi-parent recombination with simplex crossover in real coded genetic algorithms." *Genetic and Evolutionary Computation Conference (GECCO 1999)*.

Turner, S. W., Marlow, D., Ekström, M., et al. (2014). "Linking climate projections to performance: A yield-based decision scaling assessment of a large urban water resources system." *Water Resources Research*, 50(4), 3553–3567.

van Griensven, A., Meixner, T., Grunwald, S., et al. (2006). "A global sensitivity

analysis tool for the parameters of multi-variable catchment models." *Journal of Hydrology*, 324(1–4), 10–23.

Van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2008a). "Characterization of watershed model behavior across a hydroclimatic gradient." *Water Resources Research*, 44(1), W01429.

Van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2008b). "Rainfall characteristics define the value of streamflow observations for distributed watershed model identification." *Geophys. Res. Lett*, 35, L11403.

Van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2009). "Sensitivity–guided reduction of parametric dimensionality for multi–objective calibration of watershed models." *Advances in Water Resources*, 32(8), 1154–1169.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Game and Economic Behavior*. Princeton University Press, Princeton.

Vorosmarty, C. J. (2000). "Global water resources: Vulnerability from climate change and population growth." *Science*, 289(5477), 284–288.

Vrugt, J. A. and Robinson, B. A. (2007). "Improved evolutionary optimization from genetically adaptive multimethod search." *Proceedings of the National Academy of Science*, 104(3), 708–711.

Vrugt, J. A., Robinson, B. A., and Hyman, J. M. (2009). "Self-adaptive multimethod search for global optimization in real-parameter spaces." *Evolutionary Computation, IEEE Transactions on*, 13(2), 243–259.

Wagener, T., Boyle, D., Lees, M., et al. (2001). "A framework for development and application of hydrological models." *Hydrology and Earth System Sciences*, 5(1), 13–26.

Wagener, T., McIntyre, N., Lees, M., Wheater, H., and Gupta, H. (2003). "Towards reduced uncertainty in conceptual rainfall–runoff modelling: Dynamic identifiability analysis." *Hydrological Processes*, 17(2), 455–476.

Wagener, T., Reed, P., van Werkhoven, K., Tang, Y., and Zhang, Z. (2009a). "Advances in the identification and evaluation of complex environmental systems models." *Journal of Hydroinformatics*, 11(3/4), 266–281.

Wagener, T., Sivapalan, M., Troch, P., et al. (2010). "The future of hydrology: An evolving science for a changing world." *Water Resources Research*, 46(5), W05301.

Wagener, T., Werkhoven, K. v., Reed, P., and Tang, Y. (2009b). "Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling." *Water Resources Research*, 45(W02501).

Walker, W. E., Haasnoot, M., and Kwakkel, J. H. (2013). "Adapt or perish: a review of planning approaches for adaptation under deep uncertainty." *Sustainability*, 5(3), 955–979.

Walker, W. E., Rahman, S. A., and Cave, J. (2001). "Adaptive policies, policy analysis, and policy-making." *European Journal of Operational Research*, 128(2), 282–289.

Watkins, D. W. and McKinney, D. C. (1997). "Finding robust solutions to water resources problems." *Journal of Water Resources Planning and Management*, 123(1), 49–58.

Weaver, C. P., Lempert, R. J., Brown, C., et al. (2013). "Improving the contribution of climate model information to decision making: the value and de-

mands of robust decision frameworks." *Wiley Interdisciplinary Reviews: Climate Change*, 4(1), 39–60.

Wilchfort, O. and Lund, J. R. (1997). "Shortage management modeling for urban water supply systems." *Journal of the American Water Resources Association*, 123(4), 250–258.

Woodruff, M. J., Reed, P. M., and Simpson, T. (2013). "Many objective visual analytics: Rethinking the design of complex engineered systems." *Structural and Multidisciplinary Optimization*, 48(1), 201–219.

Yang, J. (2011). "Convergence and uncertainty analyses in Monte–Carlo based sensitivity analysis." *Environmental Modelling & Software*, 26(4), 444–457.

Yatheendradas, S., Wagener, T., Gupta, H., et al. (2008). "Understanding uncertainty in distributed flash flood forecasting for semiarid regions." *Water Resources Research*, 44(W05S19).

Zaehle, S., Sitch, S., Smith, B., and Hatterman, F. (2005). "Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics." *Global Biogeochemical Cycles*, 19(3), GB3020.

Zeff, H., Kasprzyk, J., Herman, J., Reed, P., and Characklis, G. (2014). "Navigating financial and supply reliability tradeoffs in regional drought portfolios." *Water Resources Research*, 50(6), 4906–4923.

Zeff, H. B. and Characklis, G. W. (2013). "Managing water utility financial risks through third-party index insurance contracts." *Water Resources Research*, 49(8), 4939–4951.