ROBUST WATER RESOURCES MANAGEMENT UNDER CLIMATIC AND ENVIRONMENTAL UNCERTAINTY

by
Julie Shortridge

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
April, 2016

ABSTRACT

Climate change is expected to have dramatic impacts on the water resources sector, and there is increasing concern that some degree of adaptation will be required to ensure sustainable water provision in many regions of the world. However, adapting to future climatic conditions is challenged by the considerable uncertainty and disagreement surrounding projections of future hydrologic conditions, particularly at local scales relevant for decision making. Furthermore, many argue that these impacts cannot be confidently represented probabilistically, resulting in uncertainty that confounds traditional approaches for decision support under uncertainty.

In the face of these challenges, a number of methods have been developed to better characterize and make decisions in the face of climatic uncertainty. The objective of this dissertation is to critically evaluate methods for impact assessment and decision support in the water resource sector, with a particular emphasis on deep uncertainty surrounding climatic and environmental conditions. This issue is explored through the evaluation of four research questions:

1. How does the choice of modeling approach for empirical streamflow simulation contribute to bias and uncertainty when predicting climate change impacts?
2. How does Robust Decision Making (RDM), a method largely developed in the water resource and climate adaptation field, compare to other methods for risk assessment under deep uncertainty that have been developed in the risk analysis field?
3. How does the method used to aggregate multiple criteria impact the results of the scenario discovery process within the RDM framework?

4. How can methods such as RDM, which generally still rely on complex simulation models and detailed climate model projections, be adapted to data-scarce regions where these models and projections may not be available?

By providing a systematic and thorough evaluation of novel methods for climate change impact assessment and adaptation, this dissertation ultimately aims to improve our ability to create robust, sustainable water infrastructure in the face of highly uncertain future climate conditions. Additionally, the use of the Lake Tana basin in Ethiopia as a case study for three of the above questions has led to important applied contributions to infrastructure planning in data-scarce regions of the developing world.

Readers:

Dr. Seth Guikema

Dr. Benjamin Zaitchik

Dr. Benjamin Hobbs

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Seth Guikema, for his support and guidance throughout my time as a PhD student. Not only has he provided a wealth of practical guidance in terms of developing and pursuing my research interests, but his unwavering confidence and encouragement have been truly helpful in seeing this work through to its finish. Seth's enthusiasm for research, teaching, and mentoring students is inspiring and contagious, and working with him has truly been a pleasure.

I have also benefited greatly from the broader community of faculty and students at JHU. In particular, I would like to acknowledge Ben Zaitchik, who introduced me to the Lake Tana region and the issues surrounding water resources and climate there, and has provided invaluable advice and guidance on my work in the area. Ben Hobbs has given me thoughtful feedback on my work and it has been always been a pleasure to talk with him about research ideas and projects, particularly through the weekly systems seminars. I have also learned a lot through conversations with Ciaran Harman and Grace Brush, and appreciate their interest and willingness to discuss and provide feedback on my research. Finally, I would like to acknowledge the other students in the Guikema research group and the IGERT Water, Climate and Health program; working with such a great group of fellow students has made the last few years a fun and enriching experience.

Chapter three of this dissertation was a collaboration with Dr. Terje Aven of the University of Stavanger. I learned a lot from his perspectives and experience in risk analysis, and appreciated his insights into fundamental questions associated with risk that I may not have otherwise considered. The remaining chapters are based on work conducted in the Lake Tana region of Ethiopia. This work would not have been possible without Belay

Simane of Addis Ababa University. Not only did his enthusiasm and support for the project allow me to meet a number of people and organizations who were instrumental in planning and conducting this research, but he was also my first introduction to his beautiful country. For this I am very grateful. Finally, I would like to acknowledge the Ethiopian Ministry of Water and Energy, the Tana Sub-basin Organization, and the International Water Management Institute for providing data, models, and feedback that this work relied on.

This work was funded by a National Defense Science and Engineering Graduate Fellowship and the Johns Hopkins Water, Climate and Health IGERT program (NSF Grant 1069213). This funding is gratefully acknowledged. Additionally, the IGERT program helped me better understand the broader, interdisciplinary context of my work, and provided an opportunity to share my research with external board members, resulting in valuable feedback. I'm very thankful for this unique opportunity.

Finally, I would like to sincerely thank my family for their love and support throughout the years. In particular, my husband Ehren has been the steadiest, most supportive partner I could ask for, and I feel so lucky to have someone who can make me laugh and feel better after even the most stressful days. Last but certainly not least, a huge thank you to my Mom. She has not only provided me with so much practical support, guidance and advice, but has truly been a role model and inspiration for my entire life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Climate change adaptation in the water resources sector

Water resource management is a sector that is highly vulnerable to climate change, and there seems to be a consensus that traditional planning paradigms are insufficient for dealing with the deep uncertainty surrounding climatic projections. Water managers have always had to consider aleatory uncertainty associated with natural variability in precipitation and streamflow, which was typically addressed by reviewing historical records and designing infrastructure to meet reliability and cost/benefit requirements. However, climate change makes this assumption of stationarity invalid (Milly et al., 2008), meaning that the historical planning paradigm of reliability[1] based on historic records is no longer sufficient (Brown, 2010). In situations where a long historic record isn't available, safety factors are often used to account for epistemic uncertainty in water availability and demand (Kundzewicz and Stakhiv, 2010). However, the indiscriminate use of safety factors could result in highly inefficient allocation of resources, and in situations where the direction of climate-induced changes is uncertain, it may be difficult to determine what these safety factors should even be protecting against. Because of these issues, it has been claimed that decision rules and evaluation principles used to justify projects will need to be improved to meet the challenge posed by climate change (Stakhiv, 2011).

Efforts to incorporate climate change information into water resource planning has most often been conducted in a "predict-then-act" framework, which aims to predict the

[1] Reliability refers to the probability of failure, and is commonly used in water resource engineering to account for aleatory uncertainty in hydrologic conditions. For example, flood protection infrastructure is often required to protect against 100-year or 500-year floods.

hydrologic impacts of climate change in a given region and design water infrastructure to accommodate those impacts. However, development of these predictions is a complex process fraught with challenges. Generating climate projections at the local scales relevant for water planning generally requires an integrated approach, where projections from a global circulation model (GCM) are downscaled to regional projections, which can then be used as inputs for a hydrologic or water resource model. However, GCMs often result in varying projections due to different assumptions regarding boundary conditions, parameterization, and model structure (Tebaldi and Knutti, 2007). This is particularly evident in hydrologic projections, as GCMs disagree about even the direction of changes in precipitation in some regions of the world (See Figure 1.1 as an example). Furthermore, GCMs are notoriously limited in their ability to reproduce observed hydrologic climatology at regional levels (Kundzewicz and Stakhiv, 2010), meaning that confidence in their projections may be quite low. Using these projections as inputs to regional climate models and hydrologic models results in a "cascade of uncertainty" (Mitchell and Hulme, 1999; Wilby and Dessai, 2010), in which uncertainties at each stage of the modeling process influence outcomes at subsequent levels. The optimal choice of infrastructure can vary greatly depending on the model used to generate projections (Nassopoulos et al., 2012), creating a risk of maladaptation where adaptation measures actually increase vulnerability to climate change. For example, flood protection measures in Ho Chi Minh City that were designed to protect against projections of climate change available in 1999 are already expected to be insufficient for new projections of precipitation change and sea level rise, potentially worsening flood risks in many parts of the city (Lempert et al., 2013).

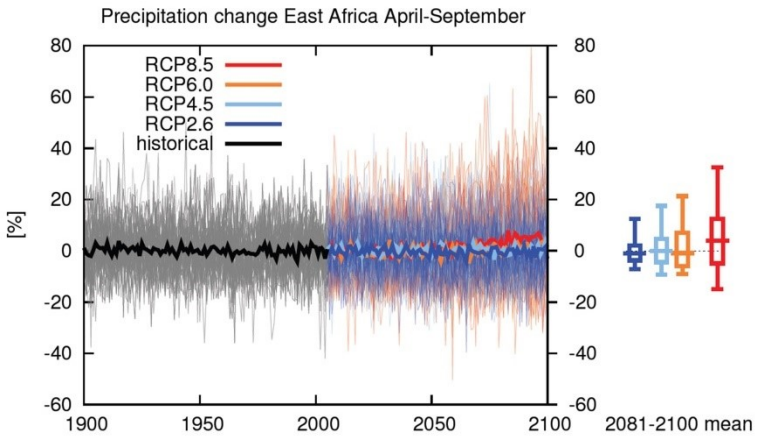

Figure 1.1: Projected percentage change in wet season precipitation for East Africa from the Coupled Model Intercomparison Project Phase 5 (CMIP5) for multiple representative concentration pathway (RCP) scenarios. Thin lines represent individual model simulations, and bold lines represent ensemble means for a given RCP. Boxplots to the right show the distribution of model projections for the percentage change in precipitation in the period 2081-2100 relative to 1986-2005 (van Oldenborgh et al., 2013)

1.2 Probabilistic and scenario-based climate adaptation

The uncertainty surrounding climate projections has resulted in calls for a risk management approach to adaptation, wherein adaptation options are implemented based on the relative likelihood of different climate impacts (National Research Council, 2009). This is particularly true in sectors such as water resources with a rich tradition of probabilistic planning (Dessai and Hulme, 2004), and a significant body of work has focused on generating probabilistic projections of climate change. This approach generally relies on using multi-model ensembles (MMEs) where individual model projections are used to develop a probability density function (PDF) for the outcome of interest. Models are often weighted so that those with a low bias relative to observed climate and high agreement with the ensemble average, which tends to outperform single models (Tebaldi and Knutti, 2007), have a greater influence on the PDF (Giorgi and Mearns, 2002; Tebaldi et al., 2005).

Extensive research has been conducted on this topic, and in the past few years this approach has begun to be put in practice for adaptation planning, particularly in the UK and New Zealand (Hall et al., 2012b; New Zealand Climate Change Centre, 2010).

However, there are a number of challenges associated with developing probabilistic projections based on MMEs. There is not yet a clear consensus on the physical and statistical interpretation of MME projections (Stephenson et al., 2012), and the width and shape of a distribution of MME results is more a measure of model agreement rather than certainty regarding future projections (Tebaldi et al., 2005). Multiple methods exist for generating a PDF from a given set of MME projections, and the resulting distributions can be highly sensitive to assumptions and methodology used (Tebaldi et al., 2005; Tebaldi and Knutti, 2007). While it seems logical to assign greater influence to the best GCMs, model weighting schemes are still subject to considerable debate. Low bias relative to observations is generally considered a necessary but not sufficient requirement for accurately predicting future climate, and may be the result of model tuning and reuse of datasets rather than accurate physical representation (Tebaldi and Knutti, 2007). Assigning weights based on agreement with the model average is based on the assumption that models are independent predictors. However, the commonalities in structure and parameterization that many models share make it unlikely that they are truly independent, and empirical evaluation of model outcomes indicates that this is not likely the case (Knutti et al., 2010). Assuming independence and assigning low weights to outlying projections can thus result in overprecision of probabilistic projections and the sampling process by which models are included in an ensemble likely represents a minimum rather than full range of uncertainty (Tebaldi and Knutti, 2007).

Recognizing these issues, scenario-based adaptation has been proposed as an alternative approach that does not rely on assigning probabilities to different climate impacts. In this approach multiple contrasting descriptions of potential future conditions are evaluated and used to inform planning efforts. Scenarios may be based on established storylines and assumptions, such as those developed in the IPCC Special Report on Emissions Scenarios (SRES), or assumptions and storylines relevant to the specific decision being evaluated. Scenario planning is often promoted as an example of process-focused decision support aimed exploring a wide range of future conditions and building consensus, rather than prescriptive determination of an optimal strategy (Lempert, 2013). However, the degree to which scenarios can inform climate adaptation decision-making has been questioned. It is noted that scenarios may be better suited to informing small groups of decision makers, rather than public debates with highly diverse stakeholders (European Environmental Agency, 2009). Evaluated scenarios may be selected in an arbitrary manner (often including a central estimate, worst case, and best case scenario), and may neglect surprises or discontinuities, focusing instead on extrapolation of current trends (Lempert, 2013). Finally, in many situations each scenario will point towards a different policy implication and serve as proxies for a need to take action, providing little guidance if the relative likelihood of each scenario is unknown (Parson, 2008).

1.3 Robust decision making

These issues have led many to conclude that climate change is an example of "deep uncertainty", a term commonly used to refer to situations where probabilistic models of uncertainty cannot be confidently determined or agreed upon (Cox, 2012) or where frequentist probabilities based on repeatable events cannot be developed (Aven, 2013). To address this uncertainty, there has been increasing interest in so-called "robust decision

frameworks" (Weaver et al., 2013) to support infrastructure planning in the face of climate change. These frameworks typically distinguish themselves from traditional "predict-then-act" frameworks in two ways. The first is that they aim to identify strategies that are robust, or that perform well over many possible conditions that may be encountered, rather than strategies that are optimal for a specific set of assumed conditions. The second is that they do not focus on predicting what future conditions may be, but instead focus on identifying conditions that cause the system of interest to fail (Weaver et al., 2013). A number of novel methodologies fall into this general family, including robust decision making (RDM; Lempert et al., 2006), decision scaling (Brown et al., 2012), and info-gap decision theory (Ben-Haim, 2000). In addition to providing decision support in situations where deeply uncertain situations, they can also be useful in situations characterized by poorly understood nonlinear or threshold responses (Lempert and Collins, 2007) or many stakeholders with conflicting values and beliefs about the future (Hallegatte and Rentschler, 2015).

RDM is one framework that has been applied to a number of climate adaptation problems (Groves et al., 2013a; Groves and Bloom, 2013; Lempert et al., 2013; Lempert and Groves, 2010). It is a multi-step, iterative approach that includes both analytical and deliberative components (Lempert et al., 2006). The analytical components of the process simulate how a system or policy alternatives will perform in many plausible future states of the world, and then use the results of these simulations to 1) compare the robustness of different alternatives and to 2) identify the conditions under which a preferred alternative will perform poorly (Lempert et al., 2006). Robustness is typically evaluated based on some measure of the regret of different alternatives across simulations (such as maximum or upper quartile) or on an evaluation of which alternatives meet pre-specified performance criteria in the largest number of scenarios. The identification of conditions where an alternative

performs poorly uses the Patient Rule Induction Method (PRIM; Friedman and Fisher, 1999) to identify regions of a multidimensional input variable space that result in undesirable values of the output variable. These regions are defined by quantitative logical conditions involving individual input variables. This process is referred to as scenario discovery because it identifies the conditions that represent vulnerabilities for a proposed policy and thus the conditions under which an alternative solution would be preferred (Lempert, 2013). By identifying these conditions, the scenario discovery process can identify which uncertainties are most important for a given decision problem (and thus potentially inform research activities) and specify the vulnerable conditions for which decision-makers may want to prepare.

The key contribution of the RDM methodology is that it provides a systematic approach for identifying and developing strategies that are robust to non-probabilistic uncertainty. Traditional decision analysis requires that uncertain parameters be characterized probabilistically, which may not be possible in problems that must consider input from multiple stakeholders with different, often conflicting, beliefs about the future. While sensitivity analysis can be conducted to see how optimal strategies change when different probability distributions are assumed, this provides no guidance on choosing between several strategies that are optimal for different assumptions. Scenario planning has also been suggested as a better strategy for situations where uncertainties cannot be characterized probabilistically, particularly when multiple stakeholders are involved. However, the choice of scenarios to consider is often arbitrary, and it provides no systematic way to compare and choose between the optimal strategies for each scenario. RDM addresses this by generating scenarios that represent key vulnerabilities of candidate strategies and outlining a systematic, quantitative way to compare the strategies' performance under these scenarios. Finally, it

doesn't require any specific mathematical formalism in the underlying model, as would be required, for instance, in robust optimization.

While RDM has been positively received in the water resources and climate change literature and used to support climate change adaptation in a number of water systems (Fischbach et al., 2015; Groves et al., 2013a; Groves and Bloom, 2013; Lempert et al., 2013; Lempert and Groves, 2010), there are questions and issues with the approach that remain to be addressed. One issue is that RDM has been developed in relative isolation from other techniques aimed at addressing non-probabilistic uncertainty. This topic has a rich history in the risk analysis field, and a number of methodologies have been developed to address situations where Bayesian probabilistic representation of uncertainty is insufficient (Dubois, 2010). The differences and similarities between RDM and these methods have not been fully explored. Additionally, existing applications of RDM generally focus on a single outcome (Hall et al., 2012a; Lempert and Groves, 2010) or a very simplistic treatment of multiple criteria (Matrosov et al., 2013a, 2013b). Finally, RDM has been proposed as a method for climate adaptation in developing countries (Hallegatte et al., 2012; Lempert and Kalra, 2011), but most applications to date have been in relatively well understood hydrologic systems in the developed world. The process often relies on sophisticated simulation models and downscaled GCM projections that may be subject to considerable uncertainty or even unavailable in developing world regions. Further investigation of these issues will strengthen our ability to apply RDM and other robust decision frameworks in a variety of contexts and ultimately improve our ability to effectively incorporate climate change into water resource planning today.

1.4 Scenario Discovery

The RDM framework is a multi-step, iterative approach to decision support under uncertainty that contains both quantitative analysis and deliberation. The process includes two analytical components based on simulation model results. When multiple alternatives or policy options are available for a given system, the first analytical component of the approach identifies the most robust alternatives based on regret minimization or satisficing criteria (Lempert et al., 2006). The second analytical component, termed "scenario discovery," aims to identify the conditions which cause unsatisfactory performance in a preferred alternative. In this work, we focus on the scenario discovery process for two reasons. The first is that the measurement of robustness as described above is sensitive to the distribution assumed in generating samples of uncertain input parameters, and could be contentious in situations of deep uncertainty where there is disagreement or uncertainty surrounding these distributions (Whateley et al., 2014). The scenario discovery process provides a description of robustness and vulnerability (characterized by the conditions where an alternative is able and unable to achieve satisfactory performance) that is less sensitive to these input distributions. Secondly, the scenario discovery process provides additional important information that can inform decision making by both identifying the uncertain parameters that have the greatest impact on system performance (and thus suggest research priorities), and by highlighting the vulnerabilities that decision makers may want to address to make their system more robust. Because scenario discovery is a focus of three of the four chapters in the dissertation, it is described here in more detail.

The scenario discovery process runs hundreds to thousands of simulations to assess system performance under different combinations of input variables. The patient rule induction method (PRIM) bump-hunting algorithm (Friedman and Fisher, 1999) is then

applied to the simulation results. The objective of the PRIM algorithm is to find regions of the multivariate input variable space $\boldsymbol{X}$ that result in unacceptable values of the output variable $Y = f(\boldsymbol{X})$. This region R is made up of one or more "boxes" B that can be defined by simple logical conditions involving the value of individual input variables (Friedman and Fisher, 1999). For instance, in one study a regional water plan was found to result in unacceptably high costs when precipitation declined by more than 10%, groundwater recharge decreased by over 3%, and a water recycling program failed to meet its goals (Lempert, 2013; Lempert and Groves, 2010).

To identify these boxes, the algorithm uses top-down successive refinement, referred to as "peeling," followed by bottom-up successive expansion or "pasting" (Friedman and Fisher, 1999). The peeling phase begins with a box B_0 containing all of the data. At each iteration, a small sub-box b^* is removed, resulting in a smaller box equal to B-b^*. The sub-box b^* chosen for removal is selected from a set of candidate sub-boxes, each of which is defined by a single input variable x_j, to maximize the percentage of simulations in the resulting smaller box B-b^* with unacceptable values of Y (Bryant and Lempert, 2010; Friedman and Fisher, 1999). This process is continued until the size of the box falls below a pre-specified value. The pasting process then readjusts the boundaries of this box by essentially reversing the peeling algorithm. In this stage, a small box b^* is added to the existing box B from a set of candidate sub-boxes to capture more simulations with unacceptable performance in the new larger box B+b^*. This process continues until the density of simulations with unacceptable values of Y within the larger boxes starts to decrease. The peeling and pasting algorithm can be repeated on remaining subsets of the data to obtain a set of boxes that collectively include a sufficiently high portion of the input

space where the output Y assumes unacceptable values (Bryant and Lempert, 2010; Friedman and Fisher, 1999).

In this work, the PRIM algorithm is implemented using the SD toolkit package in R (Bryant, 2014). This package provides an interactive implementation of the PRIM algorithm on a binary output variable Y. The package generates a tradeoff-curve showing the sequence of boxes identified during the peeling and pasting process. Boxes are scored on the basis of 1) box density, which describes the percentage of points within the box where Y is below the threshold, 2) box coverage, which describes the percentage of points where Y is below the threshold that are described by the box, and 3) restricted variables, which describes the number of input variables x_j used to define the box(Bryant and Lempert, 2010). Ideally a box would have coverage and density equal to 1 while being described by only a small number of variables, but this will rarely be the case when applying the algorithm to complex, real-world systems. Generally, as the density of the boxes increases, the coverage decreases and the numbers of variables needed to describe the box go up. By presenting a tradeoff curve showing these three parameters, the user can compare and select boxes that have sufficiently high coverage and density for their purposes while remaining interpretable.

1.5 Study area – Lake Tana, Ethiopia

Three of the four chapters in this dissertation use the Lake Tana basin in Ethiopia as a case study. Lake Tana is the source of the Blue Nile River, located in the highlands of northwest Ethiopia at an elevation of approximately 1790 meters. The lake has a surface area of approximately 3000 square kilometers, and the catchment draining to the lake encompasses approximately 12,000 square kilometers (Figure 1.1). The four main tributaries providing water to the lake are the Gilgel Abbay, Ribb, Gumara, and Megech Rivers, which

collectively account for 93% of the inflow to the lake.(Alemayehu et al., 2010) The basin's climate is characterized by distinct wet and dry seasons, with approximately 90% of rainfall and steamflow occurring during the wet period from May until October. Rainfall in the basin exhibits significant interannual variability, ranging from below 1000 mm/yr to over 1800 mm/year (Achenef et al., 2013). Population growth and expansion of agricultural and pastoral land use in the region have resulted in substantial deforestation and land degradation (Garede and Minale, 2014; Gebrehiwot et al., 2010; Rientjes et al., 2011).

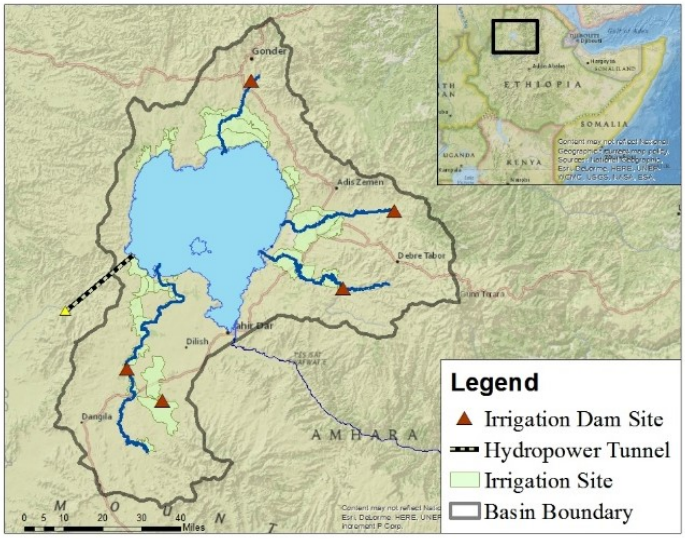

Figure 1.2: Lake Tana Study Area

The basin's population of 2.6 million is largely located in rural areas and reliant on rainfed subsistence agriculture, making the region quite vulnerable to climate variability and change. To help address this vulnerability and promote economic development, the basin has seen extensive investment in planning and construction of water resources infrastructure in recent years. The Tana-Beles hydropower tunnel was completed in 2012, and is currently the largest hydropower facility in the country with a capacity of 460 MW. The 12-km tunnel collects water from Lake Tana and transfers it to the adjacent Beles River basin, taking advantage of a 350-meter difference in elevation. A reservoir with 83 million cubic meters

(MCM) of capacity was also constructed on the Koga River in 2010 to provide irrigation to a command area of approximately 7000 hectares. There are five other reservoirs being planned for construction in the basin, ranging in volume from 80 to 220 MCM (Alemayehu et al., 2010). These reservoirs are generally designed to store water from the rainy season to support a second growing period during the dry season. Finally, there are three projects under consideration that would pump water directly from the lake to provide irrigation to surrounding areas. Planning documents for the basin describe three development levels that are planned to be progressively pursued in the coming decades (Achenef et al., 2013). Development level 0 (D0) consists only of existing infrastructure in the basin, including the Koga River irrigation reservoir and Tana-Beles hydropower transfer tunnel. Development Level 1 (D1) consists of existing infrastructure as well as four additional irrigation reservoirs and two pumped irrigation schemes. Development level 2 (D2) consists of all of the projects included in Development Level 1 as well as the Gilgel Abbay and Jema reservoirs and a pumped irrigation scheme from the southwestern portion of the lake. These development levels, and details on the projects comprising them, are summarized in Table 1.1.

Because of the long-lived nature of these projects, there is understandable concern about how climate change and other future conditions may impact their long-term performance. However, quantifying and planning for these changes presents a number of challenges. Projections of climate change in Ethiopia are highly uncertain, with climate models disagreeing on even the direction of precipitation change (van Oldenborgh et al., 2013). Efforts to reduce this uncertainty by identifying the best performing GCMs for the region have been unsuccessful, with little consistency between which models are best able to replicate the historical amount, seasonality, and variability of precipitation (Bhattacharjee and Zaitchik, 2015). Additionally, land cover in the basin has changed dramatically over the past

fifty years, with agricultural and pastoral land cover replacing native vegetation over large portions of the basin (Garede and Minale, 2014; Gebrehiwot et al., 2010; Rientjes et al., 2011), resulting in changes to surface water hydrology and increased suspended sediment concentrations (Gebremicael et al., 2013; Rientjes et al., 2011). While it is reasonable to assume that land cover will continue to evolve in response to increasing agricultural development or expansion of conservation efforts, it is impossible to say how exactly it will change.

Project Name	Type	Annual demand (MCM)		Reservoir capacity (MCM)	Average annual flow into dam site (MCM)	Catchment area (km^2)	Irrigable Area (km^2)
		Min	Max				
Development Level 0: Existing Projects							
Koga	Irrigation reservoir	62	86	83	114	185	70
Tana-Beles	Hydropower tunnel	2681	2681	NA	NA	NA	NA
Development Level 1: Planned projects							
Gumara	Irrigation reservoir	115	161	60	236	385	140
Megech	Irrigation reservoir	63	98	182	172	424	73
Ribb	Irrigation reservoir	172	220	234	210	677	199
NE Lake	Pumped irrigation	50	50	NA	NA	NA	57
NW Lake	Pumped irrigation	54	54	NA	NA	NA	67
Development Level 2: Planned projects							
Gilgel Abbay	Irrigation reservoir	104	142	563	1883	2044	103
Jema	Irrigation reservoir	57	80	200	128	218	78
SW Lake	Pumped irrigation	42	42	NA	NA	NA	51

Table 1.1: Existing and proposed water resources infrastructure in the Lake Tana basin.

Even if future climate and land cover conditions were known, predicting the impact that these changes would have on surface water availability and sediment loads through the use of hydrologic models presents additional challenges. Data limitations hinder our understanding and quantification of hydrologic processes in the basin even today. Estimates of sediment loading rates in the five proposed reservoirs are generally based on a small number of samples or sediment rating curves from other rivers (Acres International Limited and Shawel Consult International, 1995; Water Works Design & Supervision Enterprise (WWDSE), 2008; Water Works Design & Supervision Enterprise (WWDSE) and Tahal Group, 2009a, 2009b, 2009c, 2009d), despite the fact that high sediment loads have had significant negative impacts on the performance of other reservoirs in the region (Haregeweyn et al., 2012). Estimates of evaporative losses of the proposed reservoirs were based on weather station data 20 to 50 miles away from the proposed reservoir sites, where meteorological conditions may be quite different due to microclimatic effects associated with the Lake and surrounding topography (Haile et al., 2009). These data limitations also hinder efforts to develop reliable hydrologic models for the region. Limited spatially explicit data on climatic, soil, and vegetation conditions means that many hydrologic models are heavily calibrated and sometimes based on physically unrealistic parameterization schemes (van Griensven et al., 2012). A number of models also rely on empirical relationships (such as the Universal Soil Loss Equation, curve numbers, and Hargreave's equation for potential evapotranspiration) that were developed for temperate regions and may not be accurate in a highly seasonal climates. This results in considerable uncertainty surrounding the performance and downstream impacts of the proposed infrastructure even over the short term.

1.6 Research objectives and scope

The objective of this dissertation is to conduct a critical evaluation of methodologies for climate change impact assessment and adaptation in the water resources sector, with an emphasis on the deep uncertainty surrounding climate change projections. In particular, this dissertation aims to answer the following research questions:

1. How does the choice of modeling approach for empirical streamflow simulation contribute to bias and uncertainty when predicting climate change impacts?
2. How does RDM, a method largely developed in the water resource and climate adaptation field, compare to other methods for risk assessment under deep uncertainty that have been developed in the risk analysis field?
3. How does the method used to aggregate multiple criteria impact the results of the scenario discovery process within the RDM framework?
4. How can methods such as RDM, which generally still rely on complex simulation models and detailed climate model projections, be adapted to data-scarce regions where these models and projections may not be available?

Chapter 2 addresses the first research question by applying multiple regression and machine-learning approaches to simulate monthly streamflow in the tributaries of Lake Tana. The methods are compared in terms of their predictive accuracy, error structure and bias, model interpretability, and uncertainty when faced with extreme climate conditions. While the relative predictive performance of models differed across basins, data-driven approaches were able to achieve reduced errors when compared to physical models developed for the region. Methods such as random forests and generalized additive models may have advantages in terms of visualization and interpretation of model structure, which

can be useful in providing insights into physical watershed function. However, the uncertainty associated with model predictions under climate change should be carefully evaluated, since certain models (especially generalized additive models, multivariate adaptive regression splines, and random forest) can became highly variable or biased when faced with high temperatures. Because the majority of research into empirical streamflow simulation to date has focused on the predictive accuracy of a small number of model types, this broader comparison demonstrates the value in comparing multiple methodologies for a given streamflow simulation problem and improves our understanding of what makes models suitable for planning and management decisions.

Chapter 3 addresses the second research question by comparing RDM with two other methods (uncertainty factors and probability bounds analysis) that have been developed in the risk analysis field to support risk assessment in deeply uncertain conditions. The three methodologies are applied to a simple example problem related to flood risk under climate change and compared in terms of their representation of uncertain quantities, analytical output, and implications for risk management. By comparing three methodologies that take very different approaches to dealing with the issue of deep uncertainty, this work builds upon previous research that have generally focused on relatively similar methodologies, such as robustness-based approaches or alternative uncertainty representations. While each methodology aims to assess and describe risks in a manner that is more reflective of the uncertainties and assumptions underlying the assessment, we find that the analytical output and implications for decision making are not necessarily consistent between approaches. This suggests the potential value in additional comparative research to better understand the sources of these deviations, as well as the need for analysts to consider the ways in which the choice of methodology might impact analytical results.

Chapter 4 addresses the third research question by applying the RDM scenario discovery process to proposed water resource infrastructure in Lake Tana and evaluating system performance in terms of five performance criteria related to water provision to different economic sectors and environmental conditions. In this analysis, multiple methods for multi-criteria aggregation, including multiplicative and additive utility functions with varying weight schemes, are used to identify failure scenarios that cause poor performance of the proposed infrastructure. These failure scenarios are then compared to those that are identified when each performance criterion is evaluated separately. We find that failure scenarios may vary depending on the method used to aggregate multiple criteria, and that common aggregation methods can obscure connections between failure scenarios and system performance, limiting the information provided to support decision making. Applying scenario discovery over each performance metric separately provides more nuanced information regarding the relative sensitivity of the objectives to different uncertain parameters, leading to clearer insights on measures that could be taken to improve system robustness and areas where additional research might prove useful.

Chapter 5 addresses the final research question, again by applying the RDM process to proposed infrastructure in Lake Tana. This chapter demonstrates a modified application of the robust decision making methodology that is specifically tailored for application in data-scarce situations and makes two contributions that build on previously conducted RDM studies. The first is an emphasis on characterizing the relative contribution of uncertainty stemming from data limitations and model simplifications relative to uncertain future conditions, aimed at identifying priority areas for additional research and evaluation. The second contribution is a novel method for generating transient climate change sequences that does not rely on detailed GCM projections but accounts for potential dependencies

between uncertain parameters. By modifying the RDM methodology to account for model uncertainty and focus on the most valuable areas for additional research and model improvement, this work improves our ability to apply such methodologies in data-scarce regions where more complex modeling and analysis may be impractical.

2 EMPIRICAL STREAMFLOW SIMULATION FOR WATER RESOURCE MANAGEMENT IN DATA-SCARCE SEASONAL WATERSHEDS[2]

2.1 Introduction

Hydrologists and water managers have made use of observed relationships between rainfall and runoff to predict streamflow ever since the creation of the rational method in the 19th century (Beven, 2011). However, the development of increasingly sophisticated machine learning techniques, combined with rapid increases in computational ability, has prompted extensive research into advanced methods for data-driven streamflow prediction in the past decade. Artificial neural networks (ANNs), regression trees, and support vector machines have been shown to be powerful tools for predictive modeling and exploratory data analysis, particularly in systems that exhibit complex, non-linear behavior (Abrahart and See, 2007; Solomatine and Ostfeld, 2008).

While distributed physical models that accurately represent hydrologic processes can still be considered the gold standard for rainfall runoff modeling, empirical models can be a useful tool in contexts where there is limited data on physical watershed processes but long time-series of precipitation and streamflow (Iorgulescu and Beven, 2004). The development of historical data centers and more recent efforts to merge satellite data with *in situ* observations to monitor climate and hydrology has made acceptable climate and streamflow data more widely available in data poor regions. Because obtaining measurement-based estimates of soil hydraulic parameters or details on hydrologically-relevant land management

[2] This chapter is based on the following manuscript: Shortridge, J.E., Guikema, S.D., and Zaitchik, B.F. Empirical streamflow simulation for water resource management in data-scarce seasonal watersheds. Submitted to *Hydrology and Earth System Science*s in September 2015. Currently under second round of review.

activities can be more difficult, empirical models may be particularly useful in these locations.While many criticize these approaches as "black boxes" with no relationship to underlying physical processes (See et al., 2007), a number of studies have demonstrated how empirical approaches can be used to gain insights about physical system function (Galelli and Castelletti, 2013a; Han et al., 2007). Additionally, improvements in interpretation and visualization methods can make complex models more easily interpretable (Jain et al., 2004; Sudheer and Jain, 2004). Finally, data-driven models can be useful in identifying situations where observed data disagree with what would be predicted based on conceptual models, and thus identify assumptions regarding runoff generation processes that may be incorrect (Beven, 2011).

While there have been some applications of alternative machine learning methods, such as support vector machines (Asefa et al., 2006; Lin et al., 2006) and regression-tree based approaches (Galelli and Castelletti, 2013a; Iorgulescu and Beven, 2004) for streamflow simulation, the vast majority of research has focused on artificial neural networks (Solomatine and Ostfeld, 2008). While they have demonstrated impressive predictive accuracy in a number of different contexts, excessive parameterization of ANNs can result in overfit models that are not generalizable to unseen data (Gaume and Gosset, 2003; Iorgulescu and Beven, 2004). While methods exist to avoid overfitting, such as cross validation and bootstrapping, these methods are not always employed (Solomatine and Ostfeld, 2008). Relatively few studies have evaluated model performance based on parameters such as Akaike information criterion that would lead to parsimonious models that are likely to be more generalizable and interpretable (Maier et al., 2010). This can lead to complex models that only result in modest improvements (or no improvements at all) over much simpler approaches (Gaume and Gosset, 2003; Han et al., 2007).

Even outside of a hydrology context, it has been argued that ANNs are better suited for problems aimed at prediction without any need for model interpretation, rather than those where understanding the process generating predictions and the role of input variables is important (Hastie et al., 2009). Given the importance that this interpretation plays in understanding the contexts in which a hydrologic model is appropriate and reliable, the strong opinions surrounding the use of ANNs for water resources management are perhaps not surprising. To address this issue, a number of studies have focused on highlighting the structure and mechanism by which machine learning models make predictions to confirm their physical realism and gain insight into physical watershed function. For example, some studies have demonstrated how internal ANN structure corresponds to physical hydrologic processes (Jain et al., 2004; Sudheer and Jain, 2004; Wilby et al., 2003), while others have shown how variable selection and importance can be used to gain insights about model structure and runoff generating processes (Galelli and Castelletti, 2013a, 2013b). While these studies demonstrate that a number of methods exist for characterizing model structure, they generally focus on a single model type and thus provide little insight into the comparative ease with which different model types can be interpreted.

While a number of comparison studies exist that apply multiple empirical models to a given problem, finding generalizable insights from these studies is hindered because of the limited number of models and datasets evaluated. Perhaps the most comprehensive comparison to date is that of Elshorbagy et al. (2010a, 2010b), who compared six methods for data-driven modeling of daily discharge in the Ourthe River in Belgium. This work found that linear models were able to perform comparably to much more complex methods when the data content of the models were limited, or when system input-output behavior was close to linear. However, other studies have demonstrated the value of using more complex

approaches when modeling more complex rainfall-runoff behavior (e.g., Abrahart and See, 2007; Asefa et al., 2006). The differing results obtained across these studies indicate that no single method is likely to be suitable for all basins, timescales, or applications.

However, it is important to recognize that predictive accuracy alone is not necessarily sufficient justification for applying a model to a given problem. Models should not only be accurate, but also be fit-for-purpose (Beven, 2011; van Griensven et al., 2012). For instance, accurate representation of low return period flows is more important in a flood forecasting model than one aimed at predicting average amounts of water available for withdrawal and human consumption. Similarly, the ability to provide insights into physical watershed function may be more important in basins where land-use change could alter the hydrologic regime, compared to a basin that is heavily urbanized and expected to remain so. The use of multiple objective functions in training data-driven models can address this to some degree by identifying models that provide sufficient balance between different performance objectives, such as accurate representation of different portions of the flow hydrograph (de Vos and Rientjes, 2008). However, more refined model training procedures will not necessarily address other aspects of model performance that make it suitable for planning purposes, such as interpretability (Solomatine and Ostfeld, 2008). More comprehensive consideration of model strengths and limitations should be standard practice in model development and selection, rather than simply evaluating global error metrics.

In this work, we compare six methods for empirical streamflow prediction (linear models, generalized additive models, multivariate adaptive regression splines, random forests, M5 model trees and ANNs) in their ability to predict monthly streamflow in five rivers in the Lake Tana basin in Ethiopia. This study region was selected as it provides

insights into the use of data-driven models for streamflow simulation in tropical regions of the world that are underrepresented in existing studies; for instance, a review of 210 articles on water resource applications of ANNs found that over three quarters of the studies evaluated were conducted in North America, Europe, Australia, or temperate East Asia (Maier et al., 2010). Existing studies conducted in tropical regions generally apply a single methodology to the basin of interest and evaluate predictive accuracy alone (see for instance, Antar et al., 2006; Aqil et al., 2007; Chibanga et al., 2003; Machado et al., 2011), making it difficult to find generalizable insights into the relative advantages of different modeling approaches in these regions. Better development of data-driven models for these regions has the potential to be particularly valuable because data limitations and complex hydrodynamic processes often hinder the use of physical watershed models, but relatively long time series of streamflow, precipitation and temperature may be available at a monthly timescale. These data, combined with information on relevant landscape change (in particular, the expansion of agricultural land cover), can be leveraged to create reasonably accurate empirical models.

Models are compared not only in terms of their predictive accuracy, but also in terms of model error structure and the implications that this structure may have for water resource applications. Additionally, we evaluate the methods by which model structure and predictor variable influence can be evaluated to gain insights into physical system function for each model type. Finally, we assess the suitability of using different model types for climate change impact assessment by comparing model uncertainty in projections made for increasingly extreme climate conditions. The overall objective of this research is not to identify a single "best" model, but rather to highlight some of the strengths and limitations of different approaches, as well as demonstrate important issues that should be kept in mind for model comparisons in the future.

2.2 Data and Methods

2.2.1 Study Area

This study used the Lake Tana basin in Ethiopia, described in Section 1.5. A summary of basin characteristics for the evaluation period of 1960-2004 is presented in Table 2.1, and Figure 2.1 shows a map of the study area with stream gauge locations and their contributing areas.

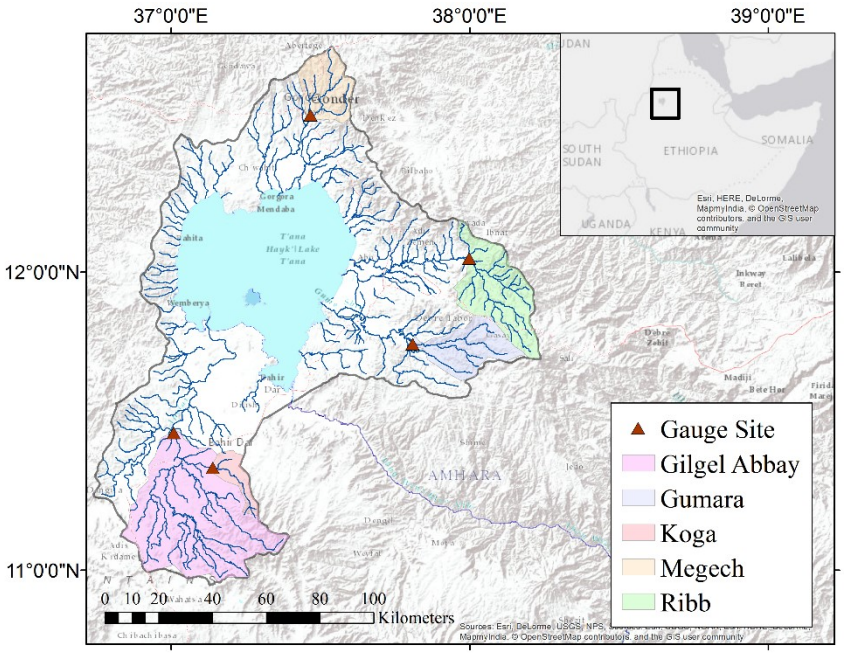

Figure 2.1: Map of Lake Tana, stream gauge locations, and upstream contributing areas

Basin	Drainage area above gauge (km^2)	Average annual streamflow at gauge (MCM)	Standard deviation of annual streamflow (MCM)	Coefficient of variation of annual streamflow	Average temp	Average monthly rainfall [mm]	
					(°C)	May-Oct	Nov-Apr
Gilgel Abbay	2664	1883	217	0.12	15.7	206	39.3
Gumara	385	236	71	0.30	17.7	186	29
Koga	200	114	31	0.27	15.7	206	39.3
Megech	424	172	66	0.31	20.6	234	41.4
Ribb	677	210	83	0.36	18.2	263	45.8

Table 2.1: Study basin characteristics over the evaluation period of 1961 to 2004

To better understand the potential implications of proposed water resources infrastructure construction in the region this development, extensive effort has been put towards developing rainfall-runoff models for the Lake Tana basin, as well as other areas of the Ethiopian highlands with similar characteristics (van Griensven et al., 2012). Many of these studies rely on Soil and Water Assessment Tool (SWAT) models, although there are some that use water balance approaches (van Griensven et al., 2012). While these models have in some cases demonstrated reasonably high accuracy, previous evaluations were largely based on Nash-Sutcliffe Efficiency (NSE) which can be a flawed performance metric in highly seasonal watersheds (Legates and McCabe Jr, 1999; Schaefli and Gupta, 2007). More importantly, the limited data available for physical parameterization of these models required a heavy reliance on model calibration, which sometimes resulted in parameterization schemes that are inconsistent with physical understanding of the region's hydrology (van Griensven et al., 2012; Steenhuis et al., 2009). Furthermore, a number of studies relied on empirical relationships such as curve numbers and the Hargreave's equation that were developed for temperate regions (e.g., Mekonnen et al., 2009; Setegn et al., 2010). While these limitations are likely to introduce considerable uncertainty into model projections, particularly in situations where climatic or environmental conditions differ from those experienced in the calibration period, few studies from this region of Ethiopia include any sort of uncertainty analysis in model predictions. Empirical models could provide a useful complement to physical models developed for the region by providing insights into physical system function and allowing for more comprehensive uncertainty analysis.

2.2.2 Data and Model Development

Models were developed using monthly streamflow, climate, and land cover data for the period from 1961 to 2004, resulting in 528 monthly observations. In each of the five major rivers in the basin, we developed empirical models that estimated monthly streamflow as a function of climate conditions and agricultural land cover in each basin. Monthly streamflow data were taken from historic stream gauge records for each basin, as reported in feasibility studies developed for proposed irrigation projects (Alemayehu et al., 2010). Historic data for monthly average temperature, monthly total precipitation, and monthly wet days in each river basin were derived from the University of East Anglia Climate Research Unit (CRU) TS3.10 gridded meterological fields (Harris et al., 2014), which are based on meteorological station observations. Historic estimates of rainfall intensity were also calculated by dividing monthly total precipitation by CRU TS3.10 records of the number of wet days in that month, but was found to be highly correlated with monthly precipitation and did not result in significant improvements to the predictive accuracy of tested models. Thus, it was not included in the final model formulations. Finally, to account for historic increases in agricultural and pastoral land cover that have occurred in the basin, the percentage of land cover used for any crop or grazing was estimated from historic land cover analyses described by Rientjes et al. (2011), Gebrehiwot et al. (2010), and Garede and Minale (2014). These studies used historic aerial photos and satellite images to estimate land cover changes in the Ribb, Gilgel Abbay, and Koga basins from the periods of 1957 to 2011. The percentage of agricultural land cover was interpolated for years when data weren‘t available, and the value of agricultural land cover in the two basins without data was assumed to be equal to average agricultural land cover in the basins with data. Land cover was assumed to change on an annual, rather than monthly basis. While this approach is prone to errors that

could stem from differing rates of land use change through time and between basins, it does provide a mechanism for capturing the long-term trend of expanding agricultural land cover that has been observed throughout the Ethiopian highlands when detailed land-cover data are unavailable. Including this data improved out-of-sample predictive accuracy of the models, further suggesting that it was a valuable addition.

Two general formulations for the empirical models were evaluated. The first (referred to below as the standard model formulation) was

Equation 2.1:

$$\log(Q_{b,t}) = f\left(P_{b,t}, P_{b,t-1}, P_{b,t-2},\ T_{b,t}, T_{b,t-1}, T_{b,t-2},\ AgLC_{b,t}\right) +\ \varepsilon_{b,t}$$

where $Q_{b,t}$ is the monthly streamflow in river b at time period t, $P_{b,t}$ and $T_{b,t}$ are the monthly total precipitation and average temperature in river basin b at time period t, $AgLC_{b,t}$ is the total percentage of agricultural land cover in basin b at time t, and $\varepsilon_{b,t}$ is the model error. The subscripts *t-1* and *t-2* indicate lagged measurements from one and two months prior, and were included to roughly account for storage times longer than one month that could impact streamflow in each river. While the exact time of concentration is not known in each basin, the minor influence of climate conditions at two months prior suggest that climate conditions from beyond this time period do not contribute significantly to flow variability. The function *f* represents a general function that differed between the specific models assessed and is discussed in more detail below. The logarithm of monthly streamflow was used as a response variable to keep model predictions positive.

In the second formulation, streamflow and climate anomalies were used as the response and predictor variables to better account for the highly seasonal nature of

streamflow and precipitation in the region. Streamflow anomalies were calculated for each observation by subtracting the long-term average streamflow for that month (*m*) from the observed value and dividing this number by the long-term standard deviation of that month's streamflow as in Equation 2.2. This procedure was repeated for precipitation and temperature, and these values were then used to fit models of the form described in Equation 2.3. It should be noted that although this formulation uses long-term averages and standard deviations to convert anomaly values to flow volumes, the anomaly values themselves are calculated based on climatic and land cover conditions that are nonstationary through time.

Equation 2.2:

$$Q_{b,t}^{AN} = \frac{Q_{b,t} - \bar{Q}_{b,m}}{sd(Q_{b,m})}$$

Equation 2.3:

$$Q_{b,t}^{AN} = f\left(P_{b,t}^{AN}, P_{b,t-1}^{AN}, P_{b,t-2}^{AN}, T_{b,t}^{AN}, T_{b,t-1}^{AN}, T_{b,t-2}^{AN}\, AgLC_{b,t}\right) + \varepsilon_{b,t}$$

Six different types of models were compared using each formulation in each basin:

1. A Gaussian linear regression model (GLM) using the basic stats package in the R statistical computing software (R Development Core Team, 2014).
2. Gaussian generalized additive model (GAM): GAMs are a semi-parametric regression approach where the response variable is estimated as the sum of smoothing functions applied over predictor variables. These functions allow the model to capture non-linear relationships between the predictor and response variables without *apriori* assumptions about the form (eg., quadratic, logarithmic) of these functions, and are fit using

penalized likelihood maximization to prevent model overfitting (Hastie and Tibshirani, 1986). GAMs were fit using the mgcv package in R (Wood, 2011).

3. Multivariate adaptive regression splines (MARS): MARS are a non-parametric regression approach where the response variable is estimated as the sum of basis functions fit to recursively partitioned segments of the data (Friedman, 1991). MARS models were fit using the earth package in R (Milborrow, 2015).
4. Artificial neural network (ANN): ANNs are a non-parametric regression approach represented by a network of nodes and links that connects predictor variables to the response variable. Each link in the network represents a function that maps the input nodes into the output node (Ripley, 1996). ANN models were fit using the nnet package in R (Venables and Ripley, 2013).
5. Random forest (RF): Random forests are a rule-based, non-parametric regression approach where the model prediction is created by averaging the predicted value from multiple regression trees which are trained on separate bootstrapped resamples of the data. Each tree is fit using a small, randomly selected subset of predictor variables, resulting in reduced correlation between trees (Breiman, 2001). Random forest models were fit using the randomForest package in R (Liaw and Wiener, 2002).
6. M5 model: M5 models are a rule-based, non-parametric regression approach that fits a linear regression model to each terminal node of a regression tree (Quinlan, 1992). M5 models were fit using the Cubist package in R (Kuhn et al., 2014).
7. Climatology model: A climatology model that simply predicted each month's streamflow as equivalent to the long-term average streamflow for that month was included for comparison purposes.

Model type	R package	Parameters defined in model formulation	Parameters selected through cross validation
GLM	stats	family = Gaussian	NA
GAM	mgcv	family = Gaussian method = generalized cross validation variable selection = true basis dimension k = 3 epsilon = 10^{-7} maxit = 200	
MARS	earth	nk = 21 thresh = 0.001 fast.k = 20 pmethod = backward	degree = {1, 2, 3} nprune = {5, 10, 15, 20, 25}
ANN	nnet	weights = 1 rang = 0.7 maxit = 100 maxNWts = 1000 abstol = 10^{-4} reltol = 10^{-8}	size = {1, 2, 4, 8, 20} decay = {0.0, 0.1, 0.5, 1.0, 2.0}
RF	randomForest	ntree = 500 sampsize = 528 nodesize = 5 nPerm = 1	mtry = {2, 3, 4, 5, 6, 7}
M5	Cubist	rules = 100 extrapolation = 100 sample = 0	committees = {10, 50, 100} neighbors = {0, 5, 9}

Table 2.2: Model parameters predefined and evaluated through cross validation

2.2.3 Model Evaluation

When using non-parametric regression approaches, it is important to avoid overfitting a model to a given dataset because this can result in large errors in out-of-sample predictions (Hastie et al., 2009). To avoid model overfit, the caret package in R (Kuhn, 2015) was used to determine model parameters for the MARS, ANN, RF and M5 models. This package uses resampling to evaluate the effect that model parameters have on the model's predictive performance and chooses the set of parameters that minimizes out-of-sample error (Kuhn, 2015). In this evaluation, 25 bootstrap resamples of the training dataset were generated for each parameter value to be assessed. A model was fit using each bootstrap sample and used to predict the remaining observations, and the parameter values that minimized average RMSE across all resamples. Details on the specific parameters evaluated for each model are

presented in Table 2.2. While the development of more complex structures are possible for some models, this process can result in over-parameterization and poor model performance (Gaume and Gosset, 2003; Han et al., 2007). Additionally, the use of a standardized parameterization procedure allows for a more even comparison between different model types.

The predictive ability of each model was assessed using 50 random holdout cross-validation samples. In each sample, a random selection of years were chosen, and observations from these years were removed ("held-out") from the dataset. The size of the held-out sample ranged from 1 to 9 years. Each model was then fit to the remaining portion of the data, using the caret package described above to determine model parameters for the MARS, ANN, RF and M5 models. These models were then used to predict streamflow for the held-out portion of the data, and both the mean absolute error (MAE) and NSE were calculated after transforming model predictions after back to the original streamflow units. Mean MAE and NSE were calculated for each model across the 50 cross-validation samples and used to choose the model with the highest predictive accuracy in each basin. This cross-validation procedure provides a mechanism for evaluating how well a model will generalize to an unseen set of data while avoiding some of the problems that can arise from the use of a single calibration and validation dataset (Elshorbagy et al., 2010a; Han et al., 2007).

MAE was included as an error metric because it provides a simple and easily interpretable measure of error on the same scale as observed flow volumes. While NSE values are acknowledged to be a flawed performance metric in highly seasonal watersheds where seasonal fluctuations contribute to a substantial portion of flow variability (Legates and McCabe Jr, 1999; Schaefli and Gupta, 2007), this metric was included to provide a rough

comparison of how empirical model performance compared to the performance of physical models developed for the region. The use of alternative error metrics has been discussed extensively in the literature (for instance, Criss and Winston, 2008; Mathevet et al., 2006; Pushpalatha et al., 2012), and could provide additional insights into what contributes to predicitive capabilities of different model formulations. However, this work examined predicitve accuracy based on MAE and NSE alone to allow for greater focus on how models differ in terms of error structure and uncertainty.

As a rough point of comparison for the statistical models developed in this research, we also evaluated discharge estimates derived from a process-based hydrological model. The model used in this application is the Noah Land Surface Model version 3.2 (Noah LSM; Chen et al., 1996; Ek et al., 2003). Noah LSM was implemented for offline simulations of the Lake Tana basin at a gridded spatial resolution of 5km for the period 1979-2010 using a time step of 30 minutes. Meteorological forcing was drawn from the Princeton 50-year reanalysis dataset (Sheffield et al., 2006), downscaled to account for Ethiopia's steep terrain using MicroMet elevation correction equations (Liston and Elder, 2006). The Princeton reanalysis was selected because it provides relatively high resolution meteorological fields, including all variables required to run a water and energy balance LSM like Noah, for the period 1948-present. While higher resolution and possibly higher quality datasets are available for recent years, this longer dataset was utilized to compare the process-based model to statistical models developed for a long historical period. Soil parameters for the Noah simulation were drawn from the FAO global soil database, land use was defined according to the United States Geological Survey (USGS) global 1km land cover product, and vegetation fraction was derived from MODerate Imaging Spectroradiometer (MODIS) imagery. Land cover was

treated as a static parameter over the full length of the simulation, as spatially complete estimates of historical land use were not available at the required resolution and specificity.

The highest performing model in each basin based on MAE was retained for more detailed evaluation of model error structure, covariate influence, and uncertainty in climate change sensitivity analysis. To generate a complete time-series of out-of-sample model predictions for error analysis, the holdout cross validation procedure was repeated for the highest performing standard-formulation and anomaly-formulation models for each basin, but this time holding out a single year of observations in each iteration. The predictions from this cross validation were used to evaluate the how model error structure might impact model predictions used for water resource applications. The influence of different predictor variables on model predictions was also assessed for the highest performing model in each basin after being fit to the complete dataset. Each predictor variable was assessed using metrics for covariate importance and influence that are unique to that model type, demonstrating how models could be used to gain physical insights about data-scarce regions and the mechanisms for generating these insights for each type of model. Partial dependence plots (Hastie et al., 2009) were also generated for each covariate for the highest performing model in each basin to provide insights about how covariate influence compared across different basins and model types.

Finally, two evaluations were conducted to assess uncertainty in model projections of streamflow under increasingly extreme climate conditions to better understand the implications of using different model formulations for climate change impact studies. Model projections of streamflow in different climate conditions are likely to be accompanied by considerable uncertainty, particularly when climate conditions exceed those experienced

historically. To assess this uncertainty, the best performing model in each basin was used to generate streamflow predictions for 1) changes in temperature from 0 to 5° C, 2) changes in precipitation from -30 to +30%, 3) an increase in temperature to 5° C combined with a decrease in precipitation to -30%, and 4) an increase in temperature to 5° C combined with an increase in precipitation to +30%. For each of the four assessments, the models generated predictions for the 45-year historic climate record adjusted for a given degree of climate change using the delta-change method (Gleick, 1986), while holding agricultural land cover constant at 60%. In this method, monthly temperature values are simply added to the temperature change value, and monthly precipitation values are multiplied by the precipitation change percentage. Model predictions for the altered climate record were then used to calculate the average annual streamflow in each river. This process was repeated 100 times for models fit on random bootstrap resamples of the historic dataset to generate uncertainty bounds surrounding model predictions and evaluated how the uncertainty in these predictions increased as climate conditions became more extreme. It is important to recognize that these should not be interpreted as a prediction or assessment of actual climate change impacts, but rather a measurement of the sensitivity of modeled streamflow in the basin to different climate conditions. Since one of the key motivations for using rainfall-runoff models is to understand how climate change may impact water resources, it is important to understand how model formulation contributes to this sensitivity and uncertainty.

2.3 Results

2.3.1 Model Accuracy and Error Structure

Table 2.3 shows the out-of-sample cross validation errors for each model assessed in each basin. The random forest model had the lowest mean absolute error for the standard-

formulation model in four of the five basins, with the M5 model performing best for the Koga basin. These models outperformed the Noah LSM simulations in all basins assessed. The Noah LSM errors are for a single period of analysis and thus don't present an exact corollary to the cross validation performed for the empirical models. Nevertheless, the significant increases in errors associated with the Noah LSM model demonstrates the difficulty associated with the use of process-based models in the region, particularly when relying on global datasets that may be unreliable at the spatial and temporal resolutions required for physical modeling. Physical models developed for monthly streamflow prediction in other basins within the Ethiopian highlands have reported NSE values ranging from 0.53 to 0.92 (van Griensven et al., 2012), compared to values ranging from 0.71 to 0.87 for the random forest models developed here. If this measure alone was used for model evaluation, these empirical models would generally be classified as having good performance based on the guidelines suggested by Moriasi et al. (2007). However, the climatology model outperforms the best standard formulation models in all basins except Megech, indicating that in the majority of basins the errors from the fitted empirical models are higher than those that result from simply using the long-term monthly average for each month's prediction. This is due to the fact that seasonality accounts for such a large portion of the variability in monthly flow values, and demonstrates how high NSE values can be quite easy to obtain in seasonal basins.

Standard Formulation		GLM	GAM	MARS	RF	M5	ANN	Climatology	Noah LSM
MAE	GA	30.78	18.54	16.75	14.89	15.11	17.22	10.42	28.11
	G	4.29	3.41	3.28	2.67	2.96	3.15	2.57	3.95
	K	1.50	1.30	1.38	1.20	1.17	1.23	1.06	1.97
	M	4.45	2.64	2.83	2.37	2.53	3.04	2.54	4.09
	R	4.69	2.98	3.50	2.97	3.27	3.17	2.81	7.01
NSE	GA	-0.02	0.81	0.83	0.87	0.86	0.84	0.95	0.59
	G	0.04	0.51	0.61	0.80	0.66	0.70	0.81	0.48
	K	0.45	0.71	0.65	0.76	0.77	0.76	0.83	0.25
	M	-1.85	0.63	0.46	0.73	0.65	0.52	0.71	0.41
	R	-1.14	0.71	0.39	0.71	0.31	0.67	0.73	-0.75
Anomaly Formulation		GLM	GAM	MARS	RF	M5	ANN	Climatology	Noah LSM
MAE	GA	9.73	9.82	10.10	10.12	9.94	9.79	10.42	28.11
	G	2.22	2.25	2.43	2.23	2.16	2.22	2.57	3.95
	K	1.03	1.06	1.08	1.09	1.05	1.05	1.06	1.97
	M	2.49	2.48	2.63	2.66	2.69	2.50	2.54	4.09
	R	2.79	2.76	2.84	2.70	2.78	2.77	2.81	7.01
NSE	GA	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.59
	G	0.85	0.85	0.82	0.85	0.86	0.86	0.81	0.48
	K	0.83	0.82	0.81	0.81	0.82	0.82	0.83	0.25
	M	0.73	0.72	0.65	0.66	0.61	0.72	0.71	0.41
	R	0.73	0.75	0.72	0.75	0.73	0.74	0.73	-0.75

Table 2.3: Cross validation errors for each assessed model in the Gilgel Abbay (GA), Gumara (G), Koga (K), Megech (M), and Ribb (R) river basins.

Evaluation of anomaly model errors indicates that the models using this formulation achieve better predictive accuracy than those using the standard formulation, and are able to outperform the climatology model based on both NSE and MAE in all basins. However, the highest performing models in each basin varies more when the anomaly formulation is used, with the GLM, GAM, random forest, and M5 models all minimizing MAE in different

basins. In all basins except Koga, the highest performing model significantly outperformed the climatology model based on paired Wilcoxon rank-sum tests (Bonferroni-corrected p-value < 0.01).

Further exploration of model residuals indicates another important advantage of using the anomaly model formulation. In the standard model formulation, model residuals appear to be non-random. Example autocorrelation plots are shown for the Gilgel Abbay and Ribb Rivers in Figure 2.2, and demonstrate that a positive autocorrelation exists at the 12 month time lag. For brevity, only plots for two rivers are shown, although this autocorrelation existed in the standard-formulation models for all basins except Megech (Table 2.4). This autocorrelation occurs because the standard-formulation models consistently underestimate wet-season streamflow while overestimating dry-season flows, as is apparent in hydrographs of observed and predicted streamflow (Figure 2.3). Because wet-season flows contribute such a large portion of the total annual flow volume, this results in regular underestimation of aggregate values such as mean annual flow (Table 2.4). This autocorrelation is reduced in the anomaly-formulation models, meaning that they are better able to capture the peak flow volumes experienced in the wet season and do not underestimate mean annual flow to the same degree that the standard formulation models do.

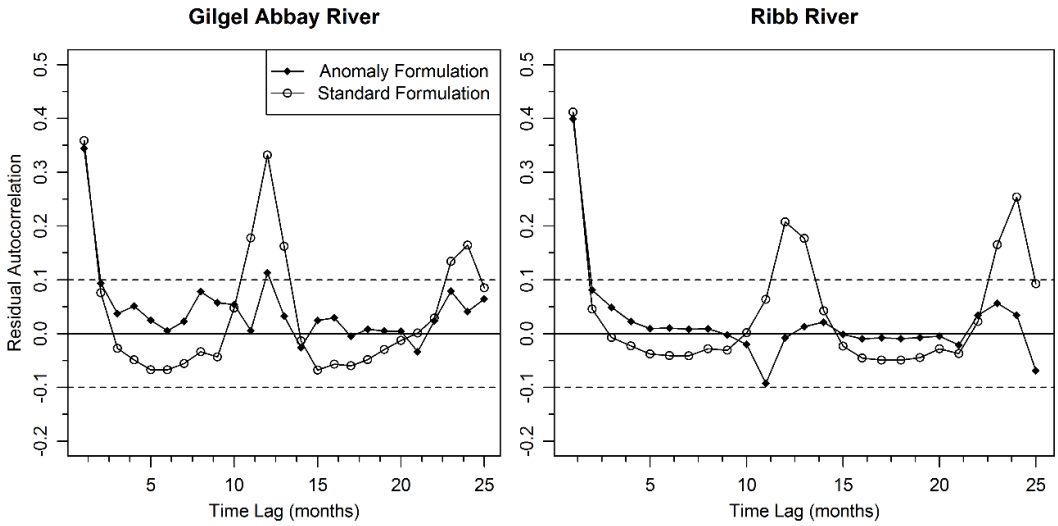

Figure 2.2: Autocorrelation in model residuals for the Gilgel Abbay and Ribb Rivers

	Autocorrelation Factors		Mean Annual Flow (MCM)		
	Standard	Anomaly	Observed	Standard	Anomaly
Gilgel	0.33	0.11	22,925	20,703	22,958
Gumara	0.29	0.07	2,870	2,392	2,734
Koga	0.04	0.10	1,383	1,333	1,386
Megech	0.05	0.04	2,035	1,637	2,028
Ribb	0.21	-0.01	2,575	1,969	2,615

Table 2.4: Residual autocorrelation factors at a 12-month lag for the standard formulation and anomaly formulation models, and resulting mean annual observed and predicted flow.

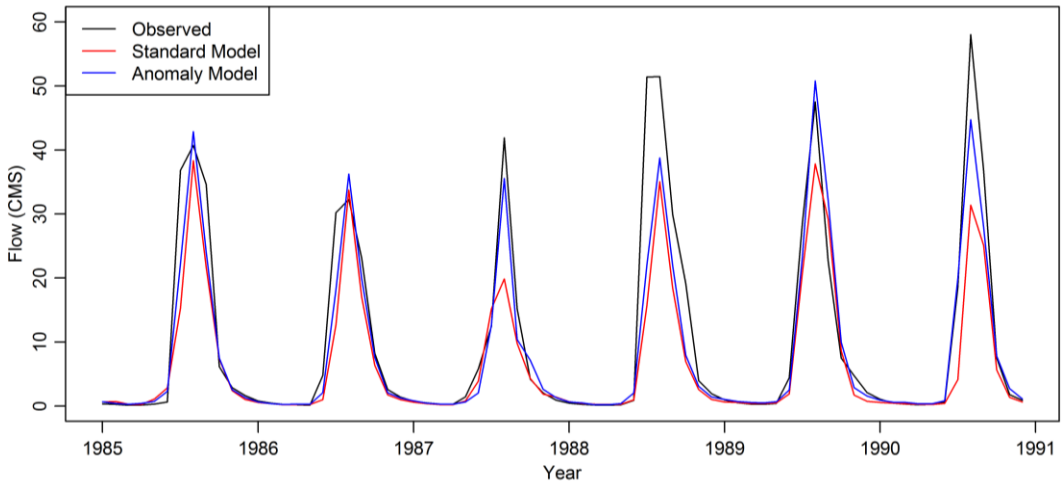

Figure 2.3: Example observed and predicted streamflow for Gumara River from 1985 to 1990

2.3.2 Model Structure and Covariate Influence

Evaluating the relationship between predictor covariates and streamflow response can lend insight into the physical processes underlying runoff generation in each basin. There are two components of this relationship that can be evaluated: how much each covariate contributes to model accuracy (covariate importance), and the direction and nature of the relationship between covariate values and model response (covariate influence). In many machine-learning models, complete description of the all of the mathematical relationships within the model (for instance, through description of each tree comprising a random forest model) is infeasible, requiring the use of other mechanisms for understanding covariate importance and influence. However, because each model type is structured in a different way, these mechanisms differ. This section first describes the mechanisms available for obtaining insights about covariate influence in each of the highest performing models. To provide a mechanism for comparing results across different basins, each basin model is then assessed using the general approach of partial dependence plots.

Model type	Linear model				Generalized additive model		M5 model tree		Random forest
Measure of influence	Linear regression coefficients and associated p-values				Estimated degrees of freedom (EDF) and associated p-values		Covariate usage in tree rules and model coefficients		Increase in MSE when covariate is randomly permuted
Basin	Gilgel Abbay		Koga		Megech		Gumara		Ribb
Covariate	Coefficient estimate	P-value	Coefficient estimate	P-value	EDF	P-value	Tree rules	Model coefficients	Percent increase in MSE
Prec	0.22	< 0.01	0.24	< 0.01	1.346	< 0.01	5%	58%	7.71%
Prec (lag 1)	0.10	0.03	0.16	< 0.01	0.624	0.08	0%	19%	2.79%
Prec (lag 2)	0.01	0.74	0.05	0.26	0	0.29	0%	0%	1.10%
Temp	-0.09	0.08	-0.07	0.17	1.023	0.07	0%	47%	12.74%
Temp (lag 1)	-0.04	0.49	-0.06	0.22	0	0.32	0%	46%	4.97%
Temp (lag 2)	-0.01	0.81	-0.09	0.08	0	0.56	0%	0%	8.16%
Agr. LC	0.00	0.33	0.02	0.01	1.986	< 0.01	86%	73%	15.21%

Table 2.5: Covariate importance measurements from each basin's model

In the Gilgel Abbay and Koga basins, the highest performing model was a simple linear regression model. These models can be evaluated by reviewing model coefficients and associated p-values, as shown in Table 2.5. In a standard linear regression, model coefficients can be interpreted as the mean change in the response variable that results from a unit change in that covariate when all others are held constant. These coefficients are for streamflow anomalies rather than raw values, making their immediate interpretation less intuitive. For instance, in the Gilgel Abbay model an increase of one standard deviation in precipitation results in an increase of 0.22 standard deviations in flow. The associated p-value for each coefficient evaluates a null hypothesis that the true coefficient value is equal to zero given the other covariates in the model, and thus has no influence on the response variable.

Evaluating model structure based on regression coefficients is appealing due to their simplicity and familiarity. However, it is important to keep in mind that the above interpretations rely on specific assumptions regarding model error distributions. Examination of fitted model residuals from both basins indicate that errors are autocorrelated in the Koga basin and not normally distributed due to the presence of outliers in both basins. Non-normality and autocorrelation both impact the t statistics and f statistics used to test for the significance of model coefficients, and thus the p-values for these models are likely biased (Montgomery et al., 2012).

Interpretation of variable influence in GAMs is based on the estimated degrees of freedom (EDF) a covariate's smoothing function $s(X_i)$ uses within a model (Hastie and Tibshirani, 1986). An EDF value of one or below indicates a linear function relating the response variable to that covariate, while values greater than one represent a non-linear

smoothing function. An EDF value of zero indicates that the covariate smoothing function is penalized to zero (meaning it has no influence on model predictions). In the model for the Megech River, the terms for lagged temperature at one and two months, as well as precipitation lagged at two months were all smoothed to zero. Of the remaining covariates, lagged precipitation has a linear impact on model response, while precipitation, temperature and land cover have non-linear impacts. Smoothing functions can be plotted to gain more insight about these relationships (Figure 2.4). The functions for precipitation anomaly, lagged (one month) precipitation anomaly, and agricultural land cover show a positive relationships with streamflow, while the function for temperature anomaly predicts low streamflow at both high and low anomalies.

P-values test the null hypothesis that a covariate's smoothing function is equal to zero, but rest on the assumption that model residuals are homoscedastic and independent (Wood, 2012). Similar to the linear models, residuals in the Megech GAM model appear to be both autocorrelated and heteroscedastic, meaning that a formal statistical interpretation of this value may be inappropriate and that confidence bounds around smoothing functions might be misleading.

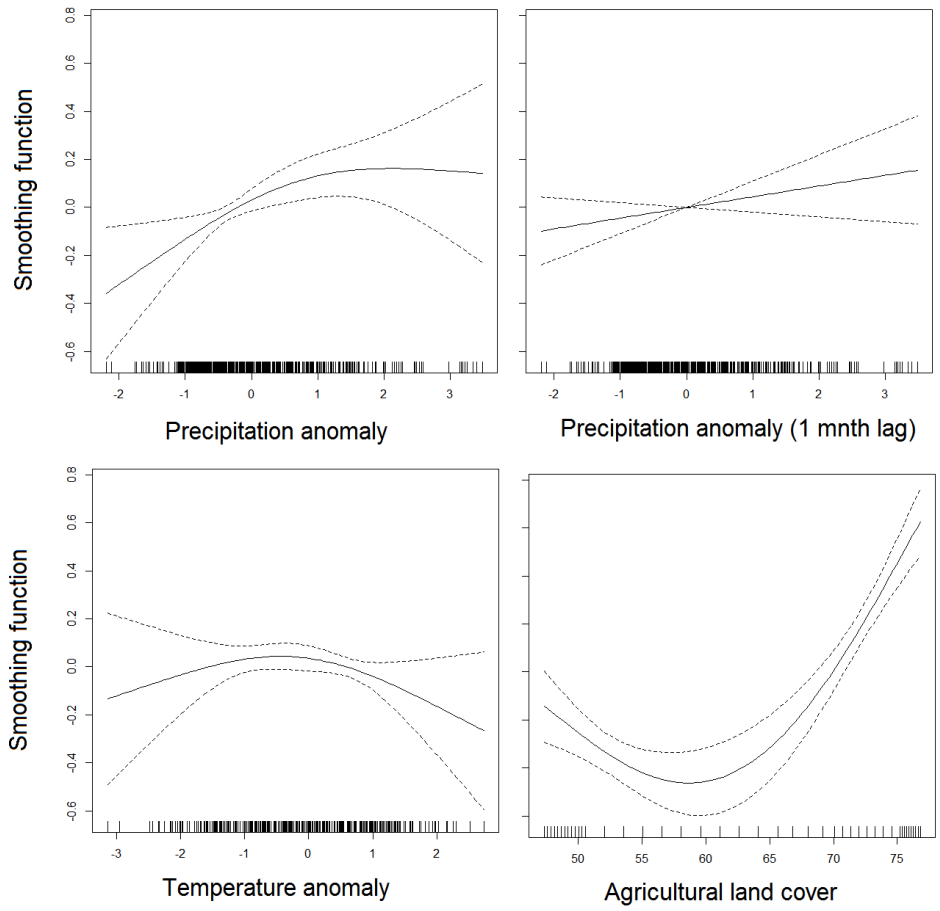

Figure 2.4: Plots of the smoothing functions used in the Megech River GAM. Hash marks along the x-axis indicate observation values of each covariate

The M5 cubist model fit for the Gumara basin is an ensemble of 100 small M5 regression trees. In each tree, the model splits observations based on logical rules related to one or more covariates and fits a linear regression model to each set of observations. The final model prediction is the average across all of the individual trees. Using this sort of ensemble approach can reduce model variance and improve accuracy if the individual trees are unbiased, uncorrelated predictors (Breiman, 1996). This can be useful in avoiding models that are overfit to the data, but can reduce model interpretability since direct visualization of model structure becomes impractical as the number of trees increases. However, the

frequency with which individual covariates are used as splitting points within trees and as regression coefficients can provide some insights about covariate importance (Table 2.5; note that because multiple covariates can be used for rules and linear models, these don't necessarily add to 100%). Model rules were largely based on land cover, with some rules based on precipitation. These two covariates were also used most frequently in linear regressions at model nodes, followed by temperature (current and 1-month lag) and 1-month lagged precipitation. Notably, climate data from 2 months lagged were not used at all. While this can be useful in identifying which covariates have the largest impact on model predictions, it doesn't provide any information regarding the nature or direction of that influence.

Similarly, the random forest model developed for the Ribb basin is an ensemble of regression trees in which the final model prediction is the average of the predictions from each individual tree. However, random forests use standard regression trees that do not incorporate linear regression models at terminal nodes. Variable importance within the final model is measured by recording the increase in out-of-sample MSE that results when a covariate is randomly permuted for each tree in the ensemble. This increase in error is then averaged across all trees in the ensemble. In our model, the largest increases in error resulted from permutation of land cover and temperature, followed by 2-month lagged temperature and precipitation. Covariate influence can be evaluated through the use of partial dependence plots (Figure 2.5), which measure the change in model predictions that result from changing the value of one parameter while leaving all other covariates constant (Hastie et al., 2009). Partial dependence plots indicate that model predictions of streamflow are higher when the percent of agricultural land cover is greater than approximately 75%, when temperatures anomalies are low, and when precipitation anomalies are high. However, it

appears that the plot for lagged temperature might be sensitive to outliers at high temperature anomalies as evidenced by the large increase that occurs above an anomaly of +2, in a region where very few data points are present.

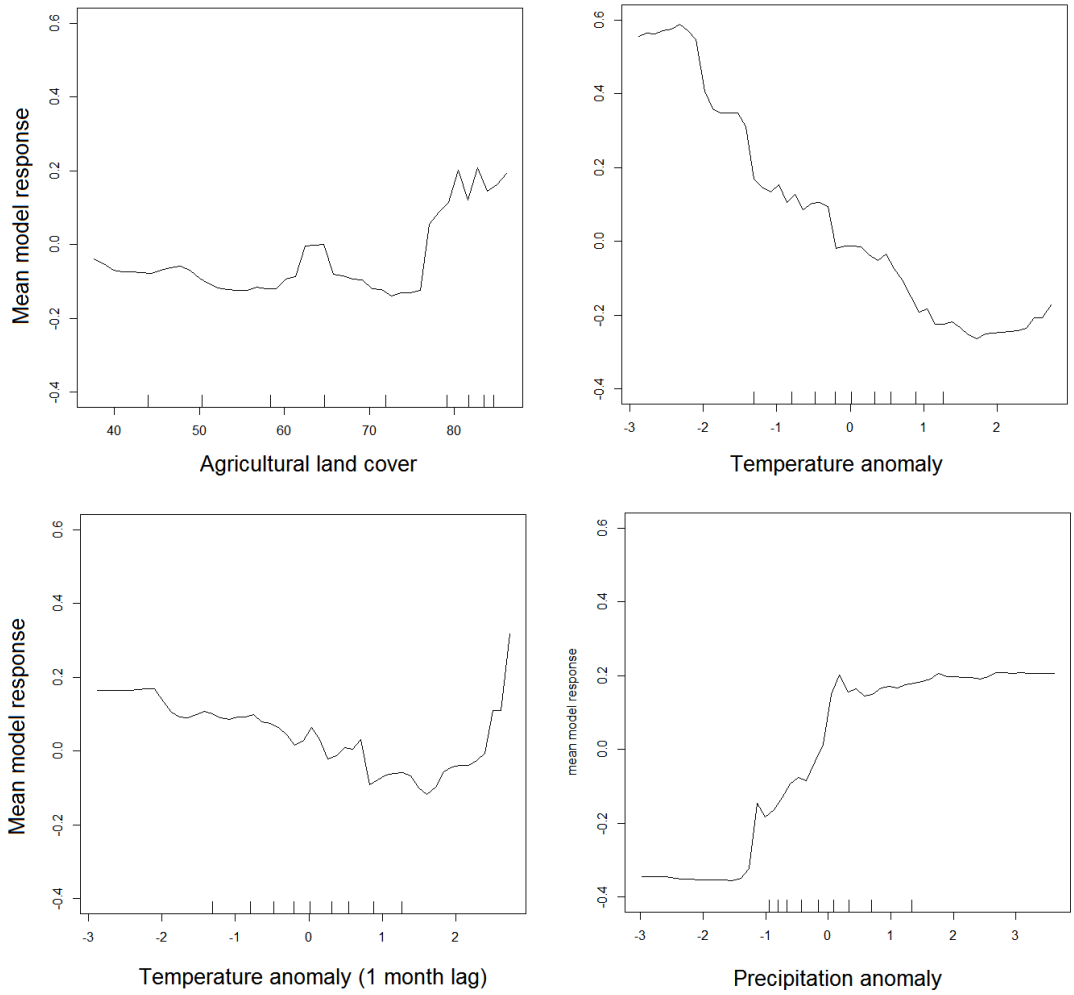

Figure 2.5: Partial dependence plots for the Ribb River random forest model. Hash marks along the x-axis show covariate sample decile values

Many of the measures used to evaluate covariate importance and influence are model specific, making inter-basin and inter-model comparisons difficult. However, the partial dependence plots used in the randomForest R package can be developed for any model and provide a mechanism for comparing the influence that covariates have in the different

models and basins (Shortridge et al., 2015). Partial dependence plots were generated for each basin's best performing model and results are shown for climatic variables in Figure 2.6. As expected, models generally respond positively to increases in precipitation and negatively to increases in temperature, with the greatest influence in the current month and decreasing influence at one and two months prior. The influence of the current month's precipitation is linear in three of the five basins; while this is constrained to the be the case in the Gilgel Abbay and Koga basins due to the use of a linear model, the linear response in Gumara is not required from the M5 model structure. Interestingly, both Megech and Ribb demonstrate a linear response to negative precipitation anomalies, but little response to positive anomalies. Streamflow response to temperature is strongest in the Gumara basin; interestingly, this is the basin with the smallest response to precipitation.

The partial dependence plots for the percentage of the basin classified as agricultural land cover indicates a positive relationship between agricultural land cover and streamflow in all basins except for the Gilgel Abbay (Figure 2.7). This would be expected if deforestation had contributed to a decrease in evapotranspiration in the contributing watersheds. The exact nature of this response differs across the different rivers, with the relatively minor responses in Koga and Ribb, and much stronger responses in the Gumara and Megech basins. However, this plot also demonstrates some of the limitations associated with different model structures. The plot for Gumara is highly erratic, indicating that the M5 model might be overfit to the training dataset, despite the use of model averaging to reduce model variance. Additionally, the GAM used in the Megech basin was only trained on agricultural land cover values up to 77%; while this model may be accurately representing the impact of land cover changes within this range, extrapolating this relationship to higher values leads to predictions that may not be physically realistic.

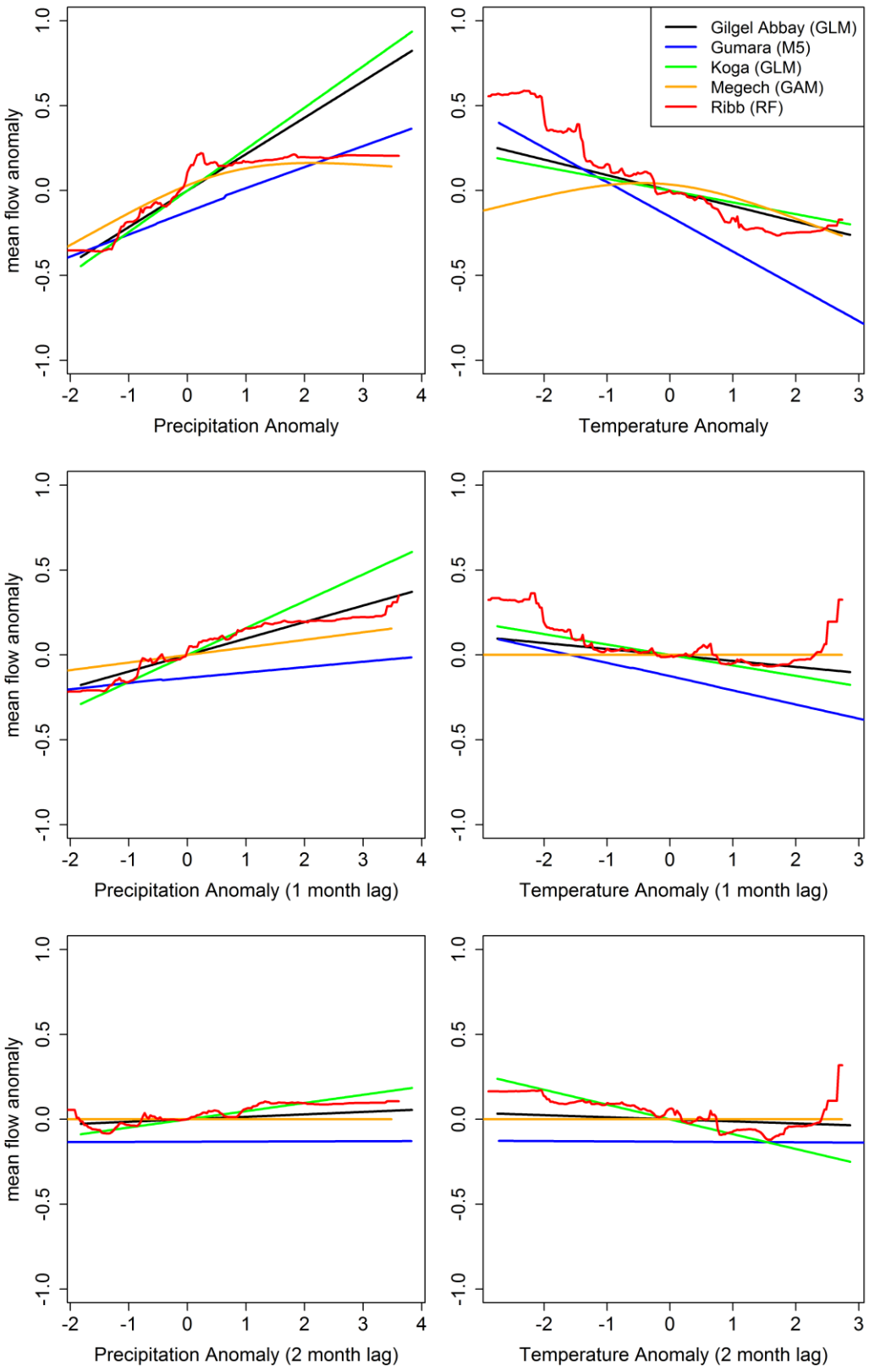

Figure 2.6: Partial dependence plots for the climate covariates in the highest performing model in each basin. Model type is indicated in parentheses.

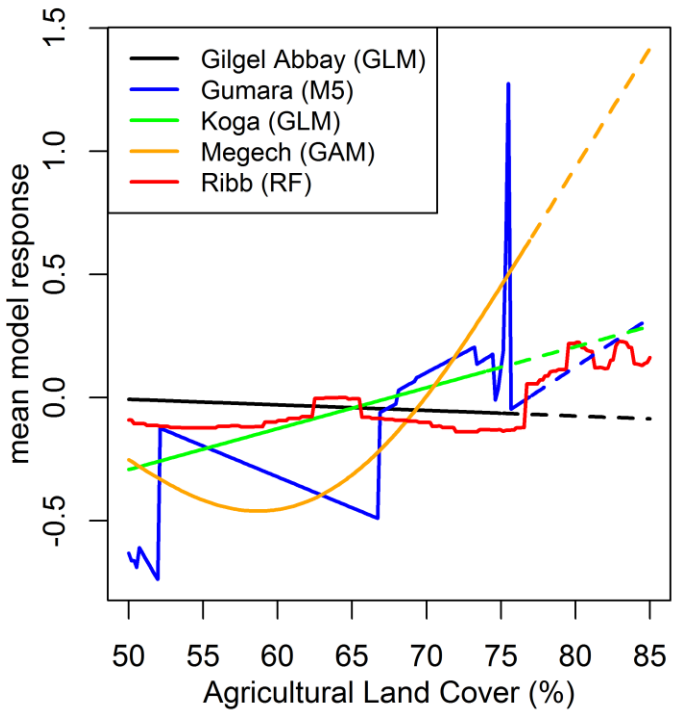

Figure 2.7: Partial dependence plot for agricultural land cover in the highest performing model in each basin. Model type is indicated in parentheses for each basin. Dashed lines indicate values that exceed historic levels of agricultural land cover experienced in that basin.

2.3.3 Climate Change Sensitivity and Uncertainty Assessment

Figure 2.8 shows the results of the climate change sensitivity analysis for total flow from all five tributaries, with dashed lines representing 95% confidence intervals obtained through 100 bootstrapped resamples of the data set. As would be expected, increasing temperature independently of precipitation results in decreasing total flows while increasing precipitation results in higher flows. However, the uncertainty surrounding temperature sensitivity increases at higher changes in temperature, while the uncertainty surrounding precipitation sensitivity remains relatively constant, even at extreme changes in annual precipitation. The bottom panels of the figure show the sensitivity of total inflows to concurrent changes in temperature and precipitation. Unsurprisingly, decreasing precipitation combined with higher temperatures results in greater decreases in total flow

than when temperature and precipitation are varied independently. However, even if temperature increases are combined with higher precipitation, total flows decline in the majority of bootstrap resamples.

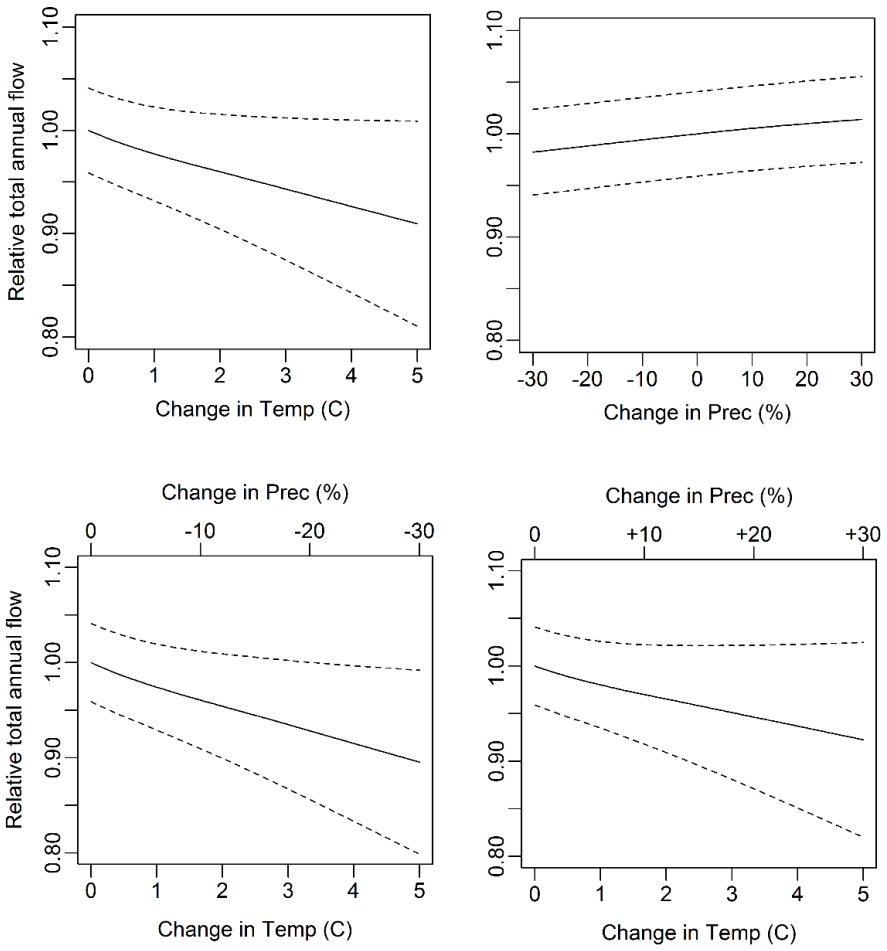

Figure 2.8: Projected changes in total streamflow (relative to current long-term average) under changing climate conditions. The top two panels show the sensitivity to changes in temperature and precipitation when they are varied independently. The bottom panel shows sensitivity to changing temperature in conjunction with decreasing (left panel) and increasing (right panel) precipitation. Dashed lines represent 95% confidence bounds from bootstrap resampling.

The uncertainty surrounding temperature sensitivity is a key limitation to using data-driven approaches for climate impact assessment. To better understand which models and basins are contributing to this uncertainty, Figure 2.9 shows how the coefficient of variation (the standard deviation of predictions from all bootstrap samples divided by the mean of

these predictions) varies as a function of temperature change in each basin. From this figure, it is apparent that the Megech model is by far the largest contributor to model uncertainty; however, it is not clear whether this contribution is due to model structure (the GAM model used for the Megech River) or characteristics associated with the basin itself. To investigate how different model structures contributed to this uncertainty, the bootstrap resampling procedure was used to assess uncertainty in streamflow predictions in the Gumara River from all model types. This basin was chosen because all six models were able to outperform the climatology model, and thus could be considered good choices for model selection based on predictive accuracy alone. The results indicate that the increase in uncertainty is highest, and increases non-linearly, in the GLM, GAM, and MARS models. Uncertainty increases more slowly in the ANN and M5 models, and no noticeable increase in uncertainty is apparent in the random forest model.

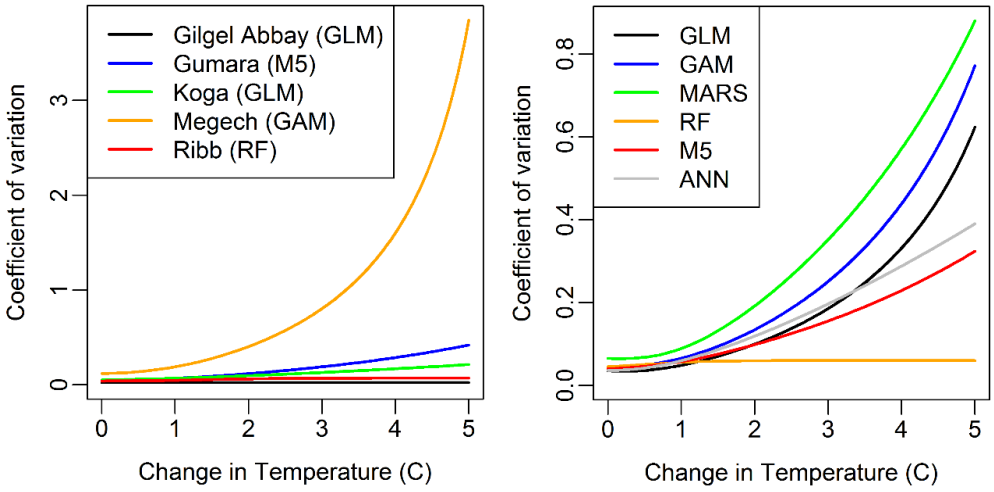

Figure 2.9: Changes in the coefficient of variation across bootstrap resamples from the highest performing model in each basin (left panel) and multiple models all applied to the Gumara basin (right panel).

2.4 Discussion

The objective of this study was not to identify the "best" approach for empirical rainfall-runoff modeling, as this is likely to be highly specific to the basin and problem to which a model is applied. However, we hope that the comparison conducted here can highlight some of the strengths and limitations of different approaches, as well as demonstrate some important issues that should be kept in mind for model comparisons in the future. One important finding was the limitation with using NSE as an error metric. Our results confirm previous studies that found that even uninformative models able to capture basic seasonality are able to achieve high NSE values (Legates and McCabe Jr, 1999; Schaefli and Gupta, 2007), and provide further evidence indicating that high NSE values should be considered a necessary but not sufficient requirement for model usage in planning situations. For instance, the simple climatology model used for comparison purposes here is able to achieve high NSE values, but would be unsuitable for planning since it does not account for any interannual variability nor the possibility for non-stationary conditions caused by changing climate and land cover. In particular, understanding error structure can be valuable in evaluating whether model biases might undermine the model's suitability for management activities. In our example, the autocorrelation present in the standard-formulation models meant that these models were consistently underestimating wet-season flows, resulting in low estimates of the total annual flow in the rivers. Since multiple reservoirs are planned for construction on these rivers to support irrigation activities, this bias could lead to poor estimates of how much water is available for agricultural use in the short term (ie., seasonal forecasting) and long-term (due to climate change). Interestingly, difficulties in accurately capturing high flows has been observed in physical hydrologic models for Ethiopia (e.g., Mekonnen et al., 2009; Setegn et al., 2011) and more generally (e.g., Wilby, 2005). The

implications of this limitation should be carefully evaluated before using models for water resource planning or (more importantly) flood risk evaluation.

Depending on the model type used, different mechanisms are available to evaluate covariate importance and influence within the model. This evaluation can be useful in confirming that the model is replicating physically realistic relationships between input and output variables. While the relationships identified in this evaluation are fairly straightforward (for example, increasing runoff with higher precipitation and lower temperatures), these simple relationships are still important in highlighting the mechanisms by which the models make predictions so that they are not "black boxes." For instance, Han et al. (2007) explore how ANN flood forecasting models responds to a double-unit input of rain, finding that some formulations respond in a hydrologically meaningful way to increased rainfall intensity, while others do not. Similarly, Galelli and Castelletti (2013a) describe how input variable importance can be used to highlight differences in hydrologic processes between an urbanized and forested watershed. The easy manner in which covariate relationships within the GAM and random forest models can be visualized using a single command within their respective R packages is a strong advantage to these approaches compared to methods such as M5 model trees and artificial neural networks. Of course, partial dependence plots can be developed for any model type (as was done in this research), but code must be written by the user and thus requires a higher degree of effort than is necessary for in-package functions. A downside to most machine-learning models is that they do not support the statistical formalism in assessing variable importance that is possible when linear models and GAMs are used. However, this formalism often rests on assumptions regarding model residuals that are unlikely to be met in many hydrologic models (Sorooshian and Dracup, 1980).

Within the Lake Tana basin, evaluation of covariate influence indicates that each basin's model is performing in a physically realistic manner, with a runoff increasing with higher precipitation levels and decreasing with higher temperatures. The influence of precipitation and temperature is greatest in the current month, and progressively declines to a very small influence after two months. This suggests that long-term (multi-month) storage does not significantly contribute to variability in flow volumes. One interesting finding is the non-linear relationship between concurrent month precipitation and runoff that exists in the Megech and Ribb basins, which suggests that above a certain point increasing rainfall does not result in a commiserate increase in streamflow. Other studies have noted the dampening effect that wetlands and floodplains have had on river flows in the region (Dessie et al., 2014; Gebrehiwot et al., 2010); this phenomenon could explain the non-linear relationship identified in this work. The clearly negative relationship between temperature and runoff demonstrates the degree to which upstream evapotranspiration impacts streamflow and suggests that evapotranspiration is largely energy-limited, rather than water-limited. Increasing agricultural land-use appears to be associated with higher runoff in all rivers except for Gilgel Abbay (where no clear relationship between land cover and runoff was observed), and suggests that agricultural expansion at the expense of forest cover has reduced the evaporative component of the water balance in these basins. Finally, the relative performance of different model formulations themselves can also be informative. For instance, the improved performance of the anomaly-formulation models indicates that the relationship between precipitation and runoff varies throughout the year and could point towards differences in runoff-generating mechanisms in the wet and dry seasons that have been observed in other case studies (Wilby, 2005).

One limitation with data-driven approaches for streamflow prediction is that the relationships they model can only generate reliable predictions for conditions that are comparable to those experienced historically. Using these models to generate predictions for conditions that exceed historic variability is likely to introduce considerable uncertainty into their projections. Our results indicate that uncertainty in projections of streamflow under changing precipitation is relatively constant, whereas uncertainty increases markedly in projections of streamflow under increasing temperature. This result is not surprising when one considers the basin's climate, which is characterized by highly variable rainfall but fairly consistent temperatures (Table 2.6). A temperature increase of 3° C equates to almost two standard deviations beyond the historic mean, whereas a change in precipitation of 30% is well within the range of conditions experienced historically. One would expect that in other climates (for example, temperate watersheds with only minor changes in rainfall throughout the year), this relationship could be reversed. Despite the uncertainty that exists in projections of streamflow under changing temperature, total annual flow appears to be quite sensitive to increasing temperatures. In fact, the decreases in streamflow due to increasing temperature appears likely to be more than enough to counteract any increases in streamflow resulting from higher precipitation that is projected for the region in some global circulation models (GCMs). This is consistent with the work of Setegn et al. (2011), who used projections from multiple GCMs as input for a SWAT model developed for the region and found that streamflow decreased in the majority of emissions scenarios and models, even when precipitation increased. Unfortunately, this suggests that any hopes for a "windfall" of additional water to support agriculture and hydropower in the region under climate change may be unfounded.

	Temperature (°C)		Wet season rainfall (mm/month)		Dry season rainfall (mm/month)	
	Mean	SD	Mean	SD	Mean	SD
Gilgel Abbay	15.7	1.54	206	145	39.3	56.5
Gumara	17.7	1.55	186	137	29.0	43.6
Koga	15.7	1.54	206	145	39.3	56.5
Megech	20.6	1.75	234	118	41.4	60.9
Ribb	18.2	1.61	263	115	45.8	57.0

Table 2.6: Mean and standard deviation values for temperature, wet-season rainfall, and dry-season rainfall in each basin

Repeating the climate change sensitivity experiment with multiple models fit to the Gumara watershed indicated that the MARS, GAM, and linear models all result in the largest increase in uncertainty at high temperatures. This indicates that when models are fit to slightly different bootstrap resamples of the historic dataset, the projected changes in streamflow at high temperature changes can be highly erratic. This is likely due to the fact that extrapolating the relationships that are observed between historic temperature and streamflow to higher temperatures can lead to very large changes in streamflow. Fitting the models to bootstrap resamples of the data results in minor changes to these relationships that can result in widely varying projections when the models are used to predict streamflow at higher temperatures, particularly when these relationships are nonlinear (as in the GAM). At the other end of the spectrum, the random forest model exhibits almost no increase in uncertainty at high temperatures, meaning that projections of streamflow at high temperatures are consistent across the bootstrap resamples. This is likely the result of the random forest model structure. The predicted value for each of a regression tree's terminal nodes is the average of all observations that meet the conditions described for that node. Thus, the model will not predict values beyond those experienced historically, even if

covariate values exceed those contained within the historic dataset. Thus, this model is likely to underestimate the change in streamflow that results from increasing temperatures.

2.5 Conclusions

In this work, we compared multiple methods for data-driven rainfall-runoff modeling in their ability to simulate streamflow in five highly-seasonal watersheds in the Ethiopian highlands. Despite the popularity of ANNs in research on streamflow prediction to date, ANNs were not found to be the most accurate model in any of the five basins evaluated. Other methods, in particular GAMs and random forests, are able to capture non-linear relationships effectively and lend themselves to simpler visualization of model structure and covariate influence, making it easier to gain insights on physical watershed functions and confirm that the model is operating in a physically realistic manner. However, it is important to carefully evaluate model structure and residuals, as these can contribute to biased estimates of water availability and uncertainty in estimating sensitivity to potential future changes in climate. In particular, autocorrelation in model residuals can result in underestimation of aggregate metrics such as annual flow volumes, even in models with high NSE performance. Uncertainty in GAM projections was found to rapidly increase at high temperatures, whereas random forest projections may be underestimating the impact of high temperatures on river flows. Thorough consideration of this uncertainty and bias is important any time that models are used for water planning and management, but especially crucial when using such models to generate insights about future streamflow levels. By considering multiple model formulations and carefully assessing their predictive accuracy, error structure and uncertainties, these methods can provide an empirical assessment of watershed behavior and generate useful insights for water management and planning. This makes them a valuable complement to physical models, particularly in data-scarce regions

with little data available for model parameterization, and warrants additional research into their development and application.

3 RISK ASSESSMENT UNDER DEEP UNCERTAINTY: A METHODOLOGICAL COMPARISON[3]

3.1 Introduction

The use of probabilities in describing uncertainty is a foundational pillar of risk analysis. Quantifying the likelihood of undesired consequences in complex systems requires a mechanism for drawing inference and quantifying uncertainty in situations where frequentist data is limited. The usefulness of Bayesian probability in meeting these needs is clear and unparalleled. Nevertheless, it has long been acknowledged that Bayesian probabilities are conditioned on underlying knowledge (Mosleh and Bier, 1996) and that low levels of underlying knowledge can present issues for probabilistic representation of uncertainty. Using a single probability (or probability distribution) to describe uncertainty masks information about what portion of the overall uncertainty is epistemic versus aleatory (Dubois, 2010; Paté-Cornell, 1996) as well as the strength of underlying knowledge supporting that probability (Aven, 2008). Conditions of deep uncertainty, such as situations where probabilities of different outcomes are unknown, previous data is deemed insufficient for estimating future consequences, and experts disagree on the consequences of different policies, present particularly difficult challenges (Cox, 2012). These issues have led to an ongoing, lively discussion regarding the theoretical and practical basis for alternative approaches to uncertainty representation in risk assessment – see, for instance, the discussion by Dubois (2010) and others in *Risk Analysis* Vol. 30, No. 3.

[3] This chapter is based on the following manuscript: Shortridge, J.E., Aven, T., and Guikema, S.D. Risk assessment under deep uncertainty: a methodological comparison. Submitted to *Reliability Engineering and System Safety* in January 2016. Currently under first round of review.

Attempts to address high-profile contemporary issues such as climate change have also raised a number of practical issues associated with applying probabilistic risk assessment to these problems. These problems are often characterized by multiple experts, stakeholders, and decision-makers who may all have dramatically different beliefs regarding future uncertain events. A probabilistic analysis may be met with resistance in situations with multiple stakeholders who disagree with the likelihood and consequences assigned by selected experts (Aven and Zio, 2011). Some research has found that experts themselves may be hesitant to assign subjective probabilities that may be perceived as unreliable or untrustworthy (Chao et al., 1999). It has been argued that probabilistic projections of climate change could mislead decision-makers by obscuring the real range of possible futures and implying a greater degree of certainty than actually exists (Clark and Pulwarty, 2003), resulting in "disguised subjectivity" (Reid, 1992) when the underlying assumptions and uncertainties are not made clear. This has led to some organizations promoting the use of probabilistic analysis only in very limited cases. For instance, IPCC guidance on reporting climate impacts requires high confidence (based on robust sources of evidence that are in general agreement with each other) for authors to characterize uncertainties probabilistically (Mastrandrea et al., 2010).

Given these concerns, a number of methodologies have been proposed to provide a more comprehensive treatment of non-probabilistic uncertainty, including "frequency of probability" approaches (Kaplan and Garrick, 1981); numerical alternatives to probabilities such as imprecise probabilities (Walley, 1991), probability bounds analysis (Ferson and Ginzburg, 1996), and possibility theory (Dubois et al., 1988); qualitative measures for describing the weight of evidence on which probability assessments are based (Aven, 2008); and robustness-based decision support frameworks that do not rely on probabilities such as

RDM (Lempert et al., 2006) and Info-Gap Theory (Ben-Haim, 2000). Existing research on these approaches largely focuses on their development, debate on their practicality and theoretical foundations, and application to specific problems. However, many of these methodologies have been developed in relative isolation from each other, making the advantages, limitations, assumptions and practical implications of each approach relative to others unclear. This limits the degree to which researchers and practitioners can build upon previous research in this field and apply these methods to problems where probabilistic analysis is considered insufficient or inappropriate.

Systematic comparisons between different approaches could serve to address some of these issues. However, relatively few comparisons between different methods exist, and those that do tend to focus on numerical alternatives to probability without considering semi-quantitative and robustness-based approaches. For instance, Dubois and Prade (1992) compare Bayesian probabilities, belief functions, and possibility theory in their ability to combine multiple expert opinions, finding that all methods can be subject to numerical instability when faced with strongly conflicting information. Soundappan et al. (2004) apply Bayesian theory and evidence theory to a series of algebraic challenge problems related to expert disagreement and imprecision, comparing their underlying assumptions and treatment of different sources of uncertainty and imprecision. Aven and Zio (2011) critically review multiple uncertainty representations as applied to a simplified nuclear reactor failure risk problem and present a broad framework for uncertainty analysis that is compatible with these representations. Similarly, Hall et al.'s (2012a) comparison of RDM with Info-Gap theory to evaluate greenhouse-gas emission policies focuses on two robustness-based approaches.

The objective of this work is to systematically compare three diverse approaches to risk assessment under deep uncertainty in terms of their representation of uncertain quantities, analytical output, and information provided for risk management and decision making. This work builds on previous comparisons that have generally focused on relatively similar methodologies (for example, alternative uncertainty representations in Aven and Zio (2011) or robustness-based approaches in Hall et al. (2012a)) by comparing three methodologies (semi-quantitative uncertainty factors, probability bounds analysis, and Robust Decision Making) that address epistemic uncertainty in very different ways. A comparison of these methodologies also has the added benefit of evaluating two approaches (uncertainty factors and RDM) that are relatively new to the risk and reliability field. The methods selected for this comparison should not be considered a judgement on which methods are most appropriate or promising, and future work comparing additional methodologies would be a valuable extension of this research.

We use a simple, stylized flood risk example to evaluate the informational requirements, underlying assumptions, and information provided to decision-makers in each case, and also evaluate how each approach relates to fundamental issues associated with risk assessment and description. While the example problem includes a number of simplifications that would make it unsuitable for evaluating flood risk in a real city, the use of a simple problem for methodological comparison does serve a number of purposes that have been highlighted by previous studies employing this approach (Aven and Zio, 2011; Flage et al., 2013; Hall et al., 2012a; Lempert and Collins, 2007). One main advantage of using a stylized example problems is to avoid situations where computational challenges overwhelm and obscure fundamental differences between the evaluated methods(Flage et al., 2013). A simple problem "allows clear focus on the essential comparisons among decision

approaches,"(Lempert and Collins, 2007) rather than the intricacies of the example problem. The use of a stylized example problem is a logical first step in comparing three very divergent approaches to risk assessment under deep uncertainty. Once the fundamental differences between approaches have been highlighted and clarified, then comparisons using more sophisticated examples would be a valuable area for additional research. By comparing these methodologies in a clear and comprehensive manner, this paper aims to improve understanding of how the choice of methodology may impact implications for risk management and suggest contexts in which certain approaches may be more suitable than others.

3.2 Framework for Comparison

To provide a clear framework for comparing non-probabilistic and quasi-probabilistic approaches to risk assessment, it is important to be clear about what exactly risk assessment entails. In this analysis, risk assessment refers to an analytical process that aims to identify and describe possible hazards, their causes and consequences, and the uncertainty surrounding their occurrence. In a probabilistic risk assessment, this process includes hazard identification, cause and consequence analysis, and a probabilistic analysis that describes the likelihood of occurrence for different scenarios and their consequences (Aven et al., 2013). This process ultimately results in a risk description, which can be used alongside managerial review and judgment to inform risk management decisions.

Traditionally, risk has been described using the "triplets" definition of risk introduced by Kaplan and Garrick (Kaplan and Garrick, 1981), which included possible scenarios or events A, the consequences that result from these events C, and the associated probability of an event P. Depending on the phenomena of interest, this probability P can

represent the frequency of a repeatable event or the Bayesian (subjective) probability of a unique event (Aven, 2012; Bedford and Cooke, 2001). However, this is not a perfect tool for describing uncertainty, particularly because it provides no information on the background knowledge on which *P* is based [23]. To more explicitly account for this element of uncertainty, we use Aven et al.'s (2013) more general risk description which includes specific events *A'*, a measurement of the quantities of interest *C'* that represent consequences *C*, a measurement *Q* of the uncertainty in *A'* and *C'*, and the knowledge *K* on which *A'*, *C'*, and *Q* are based. Figure 3.1 illustrates this conceptualization.

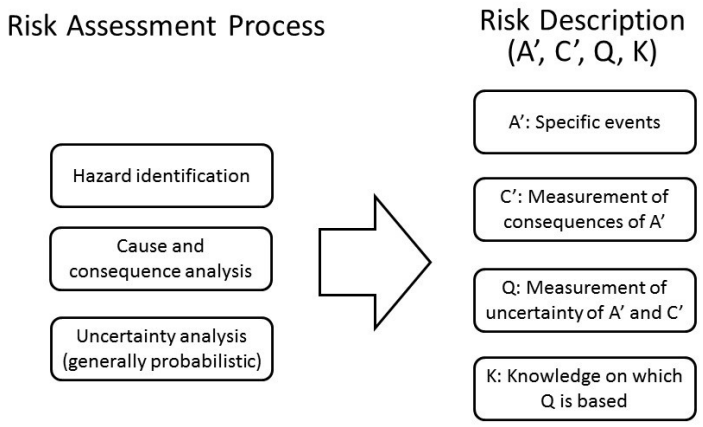

Figure 3.1: Conceptualization of risk assessment process

This comparison will assess how each of the three methods relate to the process and resulting risk description described above through the use of a stylized climate change adaptation problem related to flood risks in a riverfront city. The city is considering upgrading its existing floodwall due to concerns that its reliability has decreased with age and that climate change could result in floods occurring more frequently. The outcome quantity of interest is total cost over the next 30 years, which is the sum of damage costs from flooding and construction costs should the floodwall be upgraded. A number of simplifying

assumptions are made so that complexities in the example problem do not obscure or confuse the methodological comparison. Both floods and floodwall failures are assumed to be binary events; thus this analysis will not address varying magnitudes of flooding or flood damage. Should a flood occur, the floodwall will either hold or fail completely, resulting in either no damage or complete damage of all assets in the floodplain. Additionally, the analysis will be conducted assuming constant conditions over the 30 year period and not consider non-stationary conditions or discount rates. Finally, the comparison will only evaluate damage and construction costs, and will not consider other impacts such as lives lost or secondary economic losses.

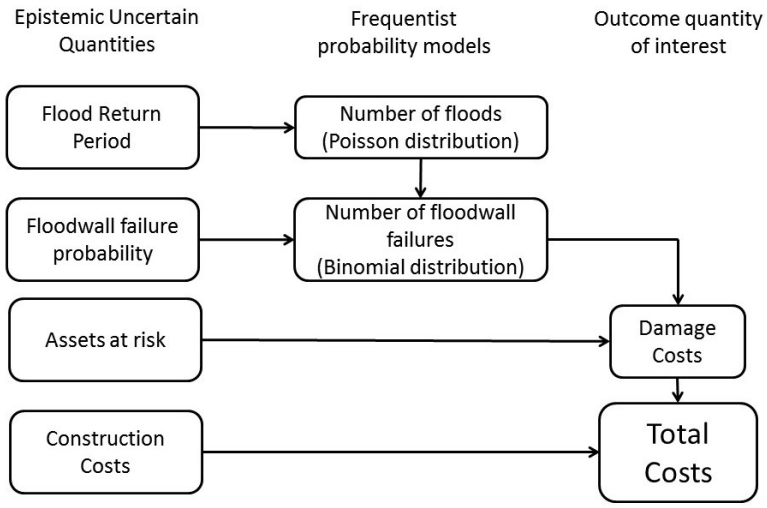

Figure 3.2: Characterization of uncertain input quantities for flood risk example problem

There are four uncertain factors that impact total costs. The first uncertain factor is the flood return period. Historical data suggest a flood return period of 100 years, but climate change could alter precipitation amounts for the region, in turn impacting the frequency of flooding. Climate model projections of future precipitation in the region differ, with some models predicting an increase in rainfall while others predict a decrease. Because of this, there is significant uncertainty about what the flood return period will be in the

future. The second uncertain factor is the failure probability of the existing floodwall. The floodwall was installed many years ago and city engineers disagree about its failure probability should a flood occur. However, if the upgrades are installed, they will result in a floodwall reliability of 99%; thus, only the reliability of the existing floodwall is uncertain. The third uncertain factor is the value of assets that will be present in the floodplain in the future (including the costs to repair the floodwall should it fail), as this depends on the rate of development in the city in years to come. The fourth uncertain factor is construction costs. While the city has obtained preliminary cost estimates from multiple contractors, there is disagreement about how much upgrades are likely to cost. For simplicity, these four uncertain quantities are all assumed to be independent of each other.

It is important to recognize that each of these four quantities (flood return period, floodwall failure probability, assets at risk, and construction costs) are all epistemically uncertain. They are not inherently variable or repeatable events; thus any probabilistic description of their values represents a Bayesian degree of belief. However, two quantities (the flood return period and number of floodwall failures) are parameters that describe frequentist models of varying phenomena (Figure 3.2). For a given flood return period, the distribution of the number of floods that will occur over the 30-year period of analysis is modeled using a Poisson distribution. Given a floodwall failure probability and number of floods occurring, the number of floodwall failures that will occur over the 30-year period is modeled as a binomial distribution. Based on this, the total costs are modeled as:

Equation 3.1:

$$C_{Total} = C_C + C_D = C_C + A * N_{fail}$$

where C_{total}, C_C and C_D are the total, construction, and damage costs, respectively, A is assets at risk, and N_{fail} is the number of floodwall failures that occur over the 30-year period of analysis. Note that the number of floodwall failures cannot exceed the number of floods that occur. If the upgrades are not installed, then construction costs are equal to zero. We assume all dollar values are in present-year dollars.

3.3 Methodological Comparison

3.3.1 Probabilistic Analysis with Uncertainty Factors

Semi-quantitative uncertainty factors have been proposed as an additional component of risk assessment aimed at communicating the level of underlying knowledge supporting the assessment, thus providing a more "comprehensive risk picture" than expected values and probabilities alone (Aven, 2008). The term "uncertainty factors" refers to limitations in background knowledge that can be hidden in the assumptions made to conduct the assessment. This method aims to identify these assumptions and assess the strength of knowledge on which they are based, as well as the degree to which their violation would impact the quantitative results of the assessment (sensitivity). For example, an assumption has strong supporting knowledge if the involved phenomena are well understood and there is reliable data and expert consensus supporting the assumption (Flage and Aven, 2009). An assumption has a high sensitivity if its violation results in dramatic changes to the quantitative results of the risk assessment.

For this approach, each of the four uncertain input quantities can be described probabilistically based on expert elicitation. However, because the flood return period and floodwall failure probability are parameters used to describe inherently variable phenomenon, allowing for uncertainty in these quantities would result in multiple

distributions of the frequency of floods and floodwall failures. While this corresponds to the "probability of frequency" analysis described by Kaplan and Garrick (1981) and the highest level of treatment of uncertainty outlined by Pate-Cornell (1996), we avoid this treatment to highlight the role of uncertainty factors and be consistent with typical risk assessment practices. Table 3.1 shows how each of the four uncertain input quantities are represented, along with key assumptions on which those representations rest:

- The flood return period is estimated to be 50 years. This is based on the assumption that average annual precipitation in the region will increase by 15% relative to historic levels. A precipitation increase of 15% is the average projection from multiple climate models.
- Multiple engineers have inspected the existing floodwall, and their elicited judgments on the probability of failure range from 0.1 to 0.2. The probabilistic assessment assumes a failure probability of 0.15, equal to the central value.
- The value of assets at risk in the floodplain depends on economic development and population growth. Based on ranges of growth experienced in the past, the value of assets that will be present in the future is represented by a uniform distribution from $500M to $600M. This assessment is based on the assumption that growth rates will not exceed those experienced historically.
- A distribution of possible costs is elicited from the lead city engineer based on preliminary cost estimates and their previous experience with similar projects. Their beliefs are found to roughly approximate a lognormal distribution with a mean of $45M and standard deviation of $7M, so this distribution is used for sampling and carrying out the probabilistic analysis. However, this distribution assumes that prices for land and materials remain consistent with current levels.

These quantities are used to determine a distribution of total costs for the existing and upgraded floodwall using 10,000-fold paired Monte Carlo simulations. The distribution of costs for each alternative is shown in Figure 3.3, and summary statistics are shown in Table 3.2. The expected total costs are similar for both alternatives ($53.08M and $47.99M for no-action and upgrades, respectively). However, the full distributions of cost for the two alternatives differ substantially. If no-action is taken, the probability of incurring no costs is 0.91, with a 0.09 probability of incurring costs over $500M. If the upgrades are constructed, then there is a 95% probability that the costs will be between $32.85M and $61.32M.

Uncertain Input Quantity	**Probabilistic Representation**	**Underlying Assumption**
Flood return period	50 years	Rainfall increases by 15%
Floodwall failure probability	0.15	Actual probability is close to average value from city engineers
Assets at risk	Uniform ~ ($500M, $600M)	Asset growth rates remain within historical range
Construction Costs	Lognormal ~ ($45M, $7M)	Prices for land and materials remain consistent with current levels

Table 3.1: Representation of uncertain input quantities in probabilistic assessment

	No Action	Upgrades
Expected value	$53.08	$47.99
0.025 percentile	$0	$32.85
0.975 percentile	$576	$61.32
P(cost = $0)	0.91	Negligible

Table 3.2: Summary statistics from probabilistic assessment

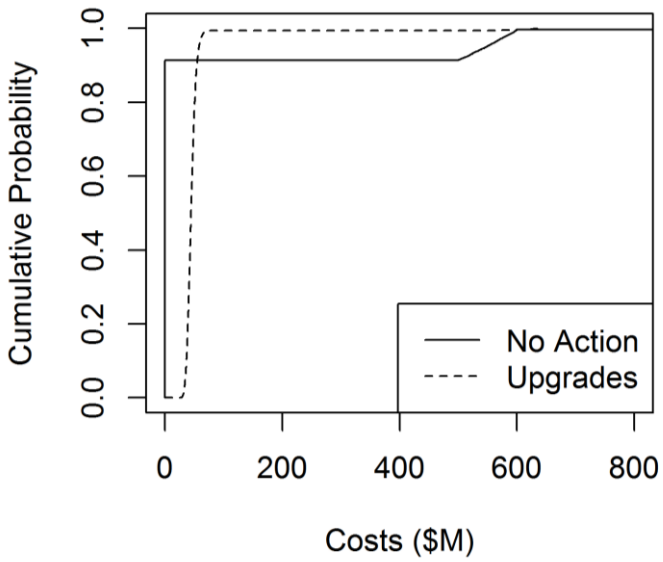

Figure 3.3: Probability distribution of total costs for no-action and upgrades alternatives

Table 3.3 shows the qualitative assessment of each of the assumptions used to probabilistically represent the uncertain input quantities. The strength of knowledge supporting each assumption is the same regardless of which alternative is being evaluated, but the sensitivity of each alternative to a given assumption can vary. For example, the distribution of costs for upgrades relies on the assumption that land and material costs remain consistent with current levels, but the distribution of costs for the no-action alternative does not rely on this assumption.

The assumption that rainfall increases by 15% is based on the average projection of changing precipitation from multiple climate models. This means that some models project increases greater than 15% and some models project increases less than 15% or perhaps even decreases. There is no reason to believe that this average is more likely than any single model's projection, nor that the models evaluated capture the complete range that could occur. Thus, it is entirely possible that the change in rainfall could be something other than 15%. Therefore, the strength of knowledge supporting this assumption is judged to be weak.

The assumption regarding failure probability is assigned a moderate strength of knowledge. While the majority of engineers believe the failure probability to be close to the central value, the possibility for more extreme values cannot be ruled out. The assumption about asset growth rates is assigned a high strength of knowledge since extensive long-term historical records exist on growth rates in the city and include periods of very rapid and even negative growth. Finally, the assumption about land and material costs is assigned a weak level of knowledge, since these prices have historically fluctuated and there is no reason to believe that they would not do so in the future.

Assumption	Quantity impacted	Strength of Knowledge	Sensitivity	
			No Action	Upgrades
Rainfall increases by 15%	Flood return period	Weak	Moderate	Low
Failure probability is equal to central value from engineers	Floodwall failure probability	Moderate	Moderate	None
Asset growth rates remain within historical range	Assets at risk	Strong	Moderate	Low
Land and material costs remain consistent	Upgrade construction costs	Weak	None	High

Table 3.3: Qualitative assessment of uncertainty factors in probabilistic example

The no-action alternative is judged to be moderately sensitive to the assumptions regarding rainfall, floodwall failure probability, and asset growth rates. The violation of these assumptions could change the probability of incurring a certain level of damage costs, but does not automatically result in damage costs since the number of floods and floodwall failures are subject to aleatory uncertainty. The upgrades alternative is judged to have a low sensitivity to the rainfall and assets assumptions, since the low failure probability of the upgraded floodwall means that even if more floods occur and assets at risk are high, it is still unlikely that damage will occur. The upgrades alternative is not impacted by the assumption

regarding the existing floodwall failure probability, just as the no-action alternative is not impacted by the assumption about land and material costs. However, the upgrades alternative is highly sensitive to the assumption about land and material costs, since construction costs are the main contributor to total costs for this alternative and could be significantly higher if this assumption is violated.

These qualitative assessments can be combined into an overall judgment on the strength of knowledge supporting the probabilistic analysis as a whole. For the no-action alternative, the three assumptions that could impact total costs vary in terms of strength of knowledge but could all have a moderate impact on the quantitative results. Thus, the strength of knowledge supporting the evaluation of the no-action alternative is judged to be moderate. The evaluation of the upgrades alternative is also assigned a moderate strength of knowledge since there is only one assumption that could greatly impact it but the strength of knowledge supporting that assumption is low.

3.3.2 Probability Bounds Analysis

Probability bounds analysis (PBA) has been proposed as method to distinguish between aleatory and epistemic uncertainty (Ferson and Ginzburg, 1996). This method employs imprecise probability distributions to describe uncertain parameters, which in the simplest case can arise from aleatory uncertainty being represented using probability distributions, and epistemic uncertainty being represented using intervals. These two forms of uncertainty can be combined and propagated using standard mathematical procedures, resulting in bounds on a cumulative distribution function, referred to as p-boxes.

Uncertain Quantity	Representation
Flood return period	20 to 200 years
Existing floodwall failure probability	0.1 to 0.2
Assets at risk	$500M to $600M
Construction costs	Lognormal: mean $40M to $50M standard deviation $5M to $10M

Table 3.4: Representation of uncertain input quantities in probability bounds assessment

In the example problem, the flood return period, existing floodwall failure probability, and the assets at risk are all epistemically uncertain, and thus represented as intervals (Table 3.4). The representation of these quantities in such a matter allows us to relax some of the assumptions that were required in the fully probabilistic analysis. For example, the probabilistic analysis assumed that annual rainfall would increase by 15% in the future, which was the average projection from multiple models whose individual projections ranged from decreases of 5% to increases of 25%. Instead of assuming that actual rainfall change will be equivalent to this model average, we can represent the flood return period as a range (20 years in the case of rainfall increase of 25% to 200 years in the case of a rainfall decrease of 5%). However, it is important to recognize that this analysis is still based on other (albeit weaker) assumptions. In the case of the flood return period, we still assume that the actual change in rainfall is captured within the model range.

In some cases, an uncertain quantity could be representative of both epistemic and aleatory uncertainty. For instance, the construction costs could be a function of two uncertain values: the total quantity of materials required, and the unit costs for these materials. The total quantity of materials required is epistemically uncertain, whereas the unit costs for materials may fluctuate through time and thus exhibit aleatory variability. Explicitly representing each of these quantities in the PBA framework would require combination of an interval encompassing possible material quantities and a probability distribution

representing unit costs. Combining these two parameters would result in an imprecise probability distribution, as shown in Figure 3.4. One can envision other situations where an imprecise probabilistic distribution for construction costs could arise as well. For example, an expert may be uncomfortable assigning precise values during probability elicitation procedures and prefer to express his degree of belief as ranges of probabilities, or elicitations may have been conducted on multiple experts, resulting in multiple probability distributions. In this paper we do not quantitatively explore all of the situations that could result in this imprecise distribution, but will instead simply assume that possible distributions of construction costs are found to approximate an imprecise lognormal distribution with a mean ranging from $40M to $50M and standard deviation ranging from $5M to $10M. However, understanding the rationale for imprecise representation of this quantity will be important when interpreting the final results of the analysis, and will be revisited in subsequent sections.

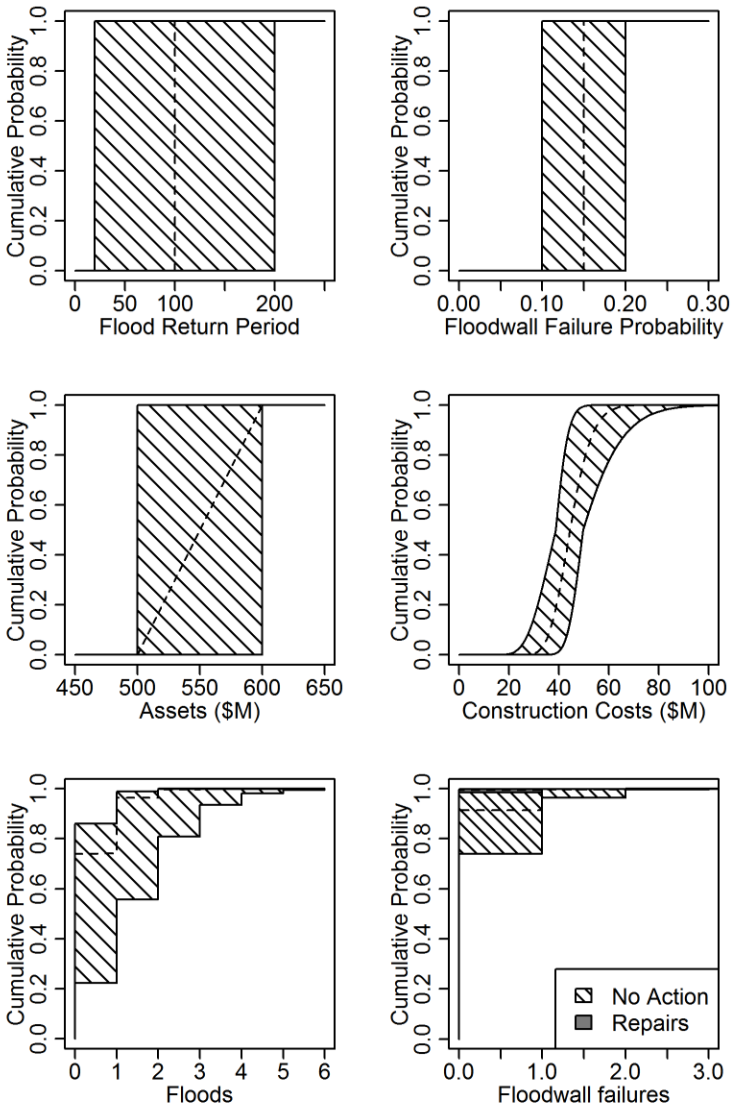

Figure 3.4: P-boxes for the uncertain input quantities used in the probability bounds assessment. The bottom two figures show p-boxes for the number of floods and the number of floodwall failures for each alternative. Dashed lines show the cumulative distribution function used in the probabilistic assessment.

Probability bounds for the two alternatives were computed using the Williamson and Downs (Williamson and Downs, 1990) algorithm as described by Tucker and Ferson (Tucker and Ferson, 2003). Instead of point values for summary measures such as expected value, we now have ranges for each of these values (Table 3.5). Whereas the expected values for each alternative were fairly close in the probabilistic assessment, the range in expected

value for the no-action alternative is much wider than the range for the upgrades. The interpretation of this range depends on the rationale for using imprecise probabilities. In the case of the no-action alternative, the total costs are a function of three uncertain quantities: the flood return period, floodwall failure probability, and assets at risk. Because each of these are epistemically uncertain, they were represented as intervals, while variability in the number of floods and floodwall failures that will occur is modeled using probability distributions. In the resulting p-box for total costs, epistemic uncertainty regarding the three input quantities contributes to the width of the box, while aleatory uncertainty in the number of floods and floodwall failures contributes to the tilt. This is consistent with Ferson and Ginzburg's (Ferson and Ginzburg, 1996) interpretation of uncertainty representation in PBA. Similarly, the range for expected value can be interpreted as resulting from lack of knowledge regarding the flood return period, floodwall failure probability, and assets at risk.

	No Action	Upgrades
Expected value	\$7 - \$182	\$37.12 - \$63.95
0.025 Percentile	\$0 - \$0	\$23.95 - \$41.05
0.975 Percentile	\$0 - \$1200	\$47.18 - 89.08
P(total cost = \$0)	0.74 to 0.985	Negligible

Table 3.5: Summary measures for the probability bounds analysis

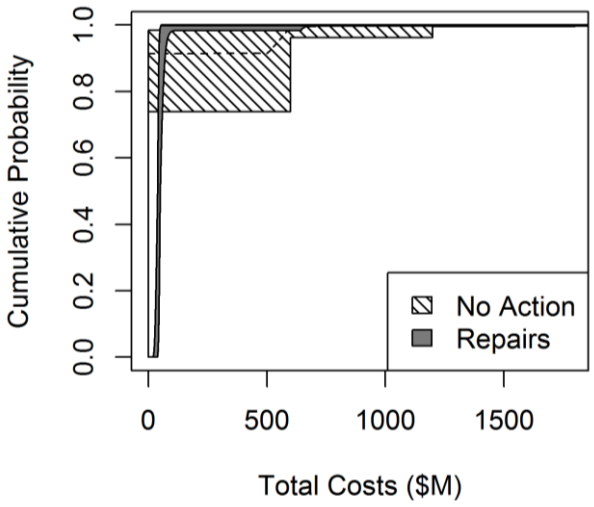

Figure 3.5: P-boxes for total costs of each alternative

The interpretation for the total cost of upgrades is different because this depends on construction costs, as well as the flood return period and assets at risk. In this case, the construction costs are characterized by epistemic uncertainty in the amount of materials required and aleatory uncertainty regarding future material costs. Therefore, the range of expected values can be interpreted as resulting from lack of knowledge regarding the flood return period, assets at risk, and the quantity of materials required. However, if probability bounds were used for other reasons, such as imprecision in elicited probabilities, this range would result from epistemic uncertainty regarding the flood return period and assets at risk, as well imprecision in the expert's elicited beliefs regarding construction costs. Similarly, if the bounds on the distribution of construction costs resulted eliciting probabilities from multiple experts, the range in expected values would partly result from disagreement in expert's elicited beliefs.

3.3.3 Robust Decision Making

Robust decision making (Lempert et al., 2006) is a multi-step, iterative process aimed at identifying and designing robust strategies, where robustness implies satisfactory performance in conditions other than those for which the system was designed. The process consists of both quantitative analysis and qualitative deliberation and review. In this comparison, we evaluate one of the key analytical components of the process, referred to as "scenario discovery." This process is motivated by the assumption that in highly uncertain situations, it is unlikely that any single alternative will be robust to all conditions it might encounter. For example, in our flood risk problem, the no-action alternative could result in large damage costs if the flood return period is low and the floodwall failure probability is high. However, the upgrades are likely to result in unnecessary construction costs if the flood return period is very high. Therefore, the scenario discovery process aims to identify regions in the input variable space that result in undesirable outcomes. These regions are used to create a quantitative description of scenarios where an alternative will fail to meet its goals (Lempert, 2013). This process is based on the PRIM algorithm and described in Section 1.4.

Input Quantity	**Representation**
Flood return period	Uniform (20 to 200 years)
Floodwall failure probability	Existing: Uniform ~ [0.1 to 0.2]
Assets at risk	Uniform ~ [$500M to $600M]
Upgrade Construction Costs	Lognormal ~ mean: $45M, sd = $10M

Table 3.6: Representation of uncertain input quantities in the RDM assessment

For the flood risk example, we apply scenario discovery to identify the conditions where the upgrades result in lower costs than the no-action alternative. Because the objective of running the simulations is not to develop a distribution of total costs but instead to create diverse combinations of input quantities, 10,000 Latin hypercube samples were created using

relatively wide distributions for each of the four uncertain input quantities (Table 3.6). The flood return period, floodwall failure probability, and assets at risk were all sampled from uniform distributions over the intervals used in the probability bounds analysis. The construction costs were sampled from a lognormal distribution with a mean of $45M (the same value use in the probabilistic analysis) and standard deviation of $10M (the largest standard deviation used in the probability bounds analysis).

No action results in lower total costs than conducting the upgrades in 95% of the simulations. However, this result should not be interpreted as a probabilistic statement since we make no assumptions that each simulation is equally likely. There were 532 simulations where the upgrades had lower total costs than no-action. The SDtoolkit package in R (Bryant and Lempert, 2010) was used to run the PRIM algorithm on the simulation results and identify boxes that describe the 532 simulations where no-action results in higher costs than the repairs (for brevity, we refer to these as "high-cost simulations"). The first step of the process creates a plot called a "peeling trajectory" which shows the sequence of boxes created through the peeling process (Figure 3.6). Coverage refers to the percentage of high-cost simulations captured by a box, while density refers to the percentage of simulations in the box that are high-cost. The peeling process starts at the lower right hand corner of the figure, with a box (represented by an unfilled circle) that includes all of the simulations. In this case, the coverage is 1.0 (it contains all of the high-cost simulations) but the density is only 0.05 (since only 5% of these simulations were high-cost). As the peeling process continues from right to left, the boxes become progressively smaller, resulting in decreases in coverage but increases in density. The number of restricted variables also increases as the peeling process continues, so that the boxes identified earlier in the algorithm are only defined by one variable, while the boxes identified later on are defined by multiple variables.

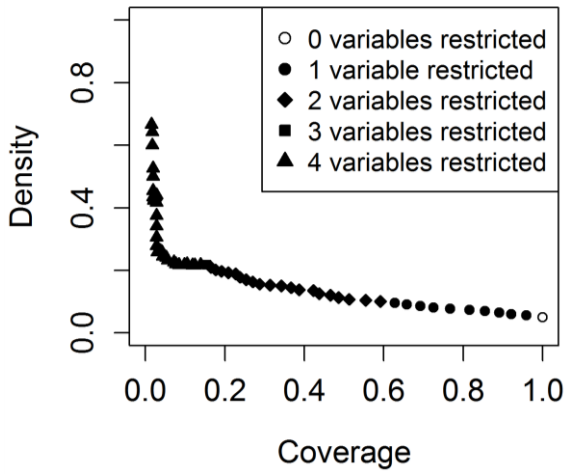

Figure 3.6: Boxes identified during scenario discovery process for no-action alternative

Through visual inspection of this figure, it is evident that the majority of boxes have a very low density. For example, the circular 1-dimensional box at the far right hand end of the plot represents all simulations where the flood return period was less than 182 years. This box's coverage is 100% but its density is only 5.3%; that is, it describes 100% of the high-cost simulations, but only 5.3% of the simulations with a flood return period of less than 182 years result in high-costs. Moving from right to left along the curve results in increasingly complex boxes with sizable decreases in coverage but only small increases in density. This result makes intuitive sense when one considers that there is a considerable amount of random variation that determines whether or not damage costs are incurred. For example, one could consider a "worst-case" scenario for the no-action alternative where the flood return period is low, the floodwall failure probability is high, and the assets at risk are high. Even under this scenario it is still possible that no damage costs will be incurred since the number of floods and floodwall failures are subject to aleatory variation.

The input variables used to define the boxes can be informative. For example, the circular one-dimensional boxes are all described by the flood return period. The diamond-shaped two-dimensional boxes are described by flood return period and floodwall failure probability. Because the objective of the PRIM algorithm is to identify combinations of input quantities that most effectively predict whether the no-action alternative will result in higher costs than the upgrades, this provides an indication of which uncertain input quantities drive the decision. Ultimately, a user would select a box, or combination of boxes, that provides a satisfactory balance between coverage, density, and interpretability. This choice is of course subjective, and it is possible (particularly in a case like our example) that none of the boxes will be deemed sufficiently informative. For discussion purposes, we select the two-dimensional box with the greatest coverage: this box includes all simulations where the flood return period is less than 83 years and the floodwall failure probability is greater than 0.11. It has a coverage of 60%, but a density of only 9%. We could then infer that there are two scenarios that would drive our decision about whether or not the upgrades should be installed. The first scenario would be that the flood return period is less than 83 years and the failure probability is greater than 0.11, and the second scenario would be if these conditions were not met. One could envision this informing a subsequent probabilistic analysis, where instead of developing probabilistic (or probability-bounds) distributions for the entire range of outcomes for each uncertain input quantity, expert elicitation would focus only on the relative likelihood of those two scenarios.

3.4 Comparison Results

The above sections describe how three diverse methods can be applied to the same simple example problem related to flood risk under deep uncertainty. In this section, we compare the approaches in terms of their representation of uncertain quantities, their

analytical output and its interpretation, and their implications for practical decision making. A summary of this comparison is presented in Table 3.7.

Method	Uncertainty Factors	Probability Bounds	Robust Decision Making
Representation of uncertain quantities	First order precise probability distributions	Imprecise probability distributions	Precise but uninformative probability distribution
Analytical output	- Precise distribution of costs for both alternatives - Key assumptions supporting probabilistic description, their strength of knowledge and sensitivity	- Imprecise distribution of costs for both alternatives	- Scenario describing conditions where upgrades result in lower costs than no-action - Relative cost of each alternative across ensemble of simulations
Risk management implications	- Expected costs of upgrades is slightly lower than no-action, but with a lower probability of incurring very high costs - Moderate strength of knowledge - Key assumptions are flood return period and land/material costs	- Expected costs could be lower for upgrades or no action - The probability of very high costs (above $500M) could be significant for the no-action alternative - Epistemic uncertainty has much larger impact on no-action than on upgrades	- Key uncertainties driving decision are flood return period and floodwall failure probability - No action results in lower costs than upgrades in majority (95%) of simulations

Table 3.7: Summary of comparison results

3.4.1 Representation of Uncertain Quantities

Each methodology takes a different approach in the manner in which it represents uncertain quantities. In the uncertainty factors analysis, all uncertain quantities were represented using first order precise probabilistic distributions. However, second order probability models could also be used if probability distributions were assigned to represent the analysts' beliefs regarding the flood return period and floodwall failure probability. In general, the approach can accommodate one or two probability levels so long as frequentist models based on repeatable events can be justified to represent aleatory variation, as

discussed by Aven (2012). In either case, the method focuses less on exact quantitative description of all uncertain quantities that may impact assessment results, and more on identifying underlying assumptions on which probabilistic descriptions of uncertainty (whether they represent the analysts' beliefs or repeatable phenomena) are based.

The probability bounds analysis represents uncertain parameters using imprecise probabilistic distributions. In the simplest case, these imprecise distributions can result from treating quantities that are epistemically uncertain as intervals and treating quantities that are inherently variable as probability distributions, although they can also arise from other circumstances as described by Tucker and Ferson (2003). This has the advantage of allowing for imprecision in probabilistic judgements when underlying data to support precise distributions may be limited. However, it has been pointed out that in cases where there is little data or underlying knowledge to describe a probabilistic model for aleatory phenomenon (such as the number of floods in a 30 year period) there may be even less basis to describe upper and lower bounds on the parameters that describe that phenomena, such as the flood return period (Berner and Flage, 2015). Thus, even the use of intervals to describe epistemically uncertain quantities may be questioned in situations where appropriate bounds on those quantities is subject to disagreement or uncertainty.

In the RDM analysis, uncertain quantities can be represented as intervals (which are effectively treated as uniform distributions for sampling purposes) or probabilistic distributions, and the method does not distinguish between epistemic and aleatory uncertainty. Because our example was focused on the scenario discovery process of identifying scenarios that result in poor outcomes for an alternative, uncertain quantities in our example were represented using relatively non-informative probabilistic distributions

that were used to generate samples across the uncertainty space. However, it is important to remember that these samples should not be interpreted in a probabilistic manner. Instead, each sample is interpreted as a "plausible future states of the world" (Lempert et al., 2006) which could conceivably occur, but does not assume that each sample is equally likely as would be the case in a Monte Carlo simulation.

These different representations of uncertainty shed some light on the types of situations where one approach might be more suitable than the others. The uncertainty factors approach represents uncertain parameters in the same manner they are represented in traditional PRA, and is likely to be appealing in situations where decision makers would like to conduct a probabilistic assessment but with a more comprehensive understanding of the underlying knowledge on which that assessment is based. The manner in which probability bounds analysis can provide a measure of discernment between aleatory variability and epistemic uncertainty may be valuable in contexts where the underlying phenomena can be meaningfully separated into these two categories and where decisions makers are concerned about what degree of risk stems from reducible versus non-reducible uncertainties. However, it is important to keep in mind that this separation is only meaningful if the use of frequentist probability models based on repeatable events can be justified to describe aleatory uncertainty (Aven, 2012). Furthermore, this method may become contentious in situations where there are not obvious upper and lower bounds to describe epistemic uncertainties. Finally, while the RDM methodology can use probabilistic distributions to generate samples across the uncertainty space, its analytical focus is not on determining the probability of different outcomes but instead on understanding the conditions where one alternative might be preferred over the other. This focus may be valuable when there is strong disagreement regarding the distribution of uncertain parameters or the likelihood of

different conditions which are unlikely to be resolved through the use of imprecise probabilities.

3.4.2 Analytical Output

The analytical output for the uncertainty factors assessment includes both the quantitative description of the distribution of total costs associated with each alternative, as well as the qualitative description of the underlying knowledge supporting this assessment. The quantitative output indicates that the expected costs between the two alternatives are relatively comparable, with the no-action and upgrades alternatives having expected costs of $53.08M and $47.99M, respectively. However, the distribution of costs for the no-action alternative is much wider than the upgrades, with a 95% uncertainty interval of $0 to $576M, compared to an interval of $32.9M to $61.3M for the upgrades. The strength of knowledge supporting these judgments is considered to be moderate for each alternative, suggesting that the phenomena may be well understood but modeled in a simple or crude way, or that only some reliable data to characterize the phenomena are available (Flage and Aven, 2009). Explicit description of the impact that each assumption could have on the different alternatives could be particularly useful when the strength of knowledge supporting the assessment of one alternative is significantly higher than the other. The systematic evaluation of each individual assumption also helps illuminate which assumptions are most critical to the validity of the quantitative assessment. Because it has a weak strength of knowledge and a moderate impact on no-action costs, the assumption regarding future rainfall levels appears most critical to the quantitative assessment of the no-action alternative. Meanwhile, the assumption regarding land and material costs is most critical to the upgrade alternative, because it has a weak strength of knowledge and a high impact on upgrade costs.

In the probability bounds analysis, the probabilistic distributions of costs from each alternative are now represented by imprecise p-boxes, and probabilistic summary statistics are represented by an interval of possible values rather than a precise value. The expected costs associated with the no-action alternative is represented by the interval from $7M to $182M, while the expected costs associated with the upgrades are represented by the interval from $37.12M to $63.95M. From this description, it is apparent that the expected cost of the upgrades may be higher or lower than the expected cost for no-action, but that the no-action alternative has the potential for a much higher expected cost than the upgrades. Visual examination of the p-boxes for each alternative can provide an understanding of the relative contribution of epistemic and aleatory uncertainty associated with each alternative as well. In comparing the two p-boxes, it is apparent that the total costs associated with the upgrades are associated with both less aleatory uncertainty (represented by a steeper cumulative distribution function) and less epistemic uncertainty (represented by the thickness of the p-box). While this can provide some insight into the cumulative effect of epistemic uncertainty on assessment outcomes, it does not provide any information about the relative contribution of each input quantity. Thus, the decision maker may find that the resulting p-boxes are unsatisfactorily wide, but have no insights into which uncertainties have the greatest impact on this width and thus possible ways for additional research to reduce this width. Obtaining this information would require additional analysis, for example through a sensitivity analysis as outlined by Ferson and Tucker (2006).

The analytical output from the RDM assessment takes a very different form. The results do not describe the likelihood of all outcomes that may occur, under the assumption that the limited knowledge available to support such an assessment would undermine its validity (Lempert et al., 2006). Instead, it provides an evaluation of the conditions where the

upgrades might be preferred over the no-action alternative, in the form of a scenario defined by a flood return period below 83 years and a floodwall failure probability greater than 0.11. While this is not as informative as a full distribution of the likelihood of different outcomes, this serves to reduce the complexity of a given problem to something more understandable. It would be possible to combine this approach with a later probabilistic analysis where the relative likelihood of different scenarios are evaluated, providing information more in line with that of a traditional risk assessment. This also provides insights into the uncertain inputs that have the most influence on the relative performance of the two alternatives, suggesting areas of additional research that may be particularly valuable. Additionally, the simulation process identified that the no-action alternative resulted in lower costs in the majority of simulations (95%) but this output cannot be interpreted in a probabilistic manner unless each simulation can assumed to be equally likely.

3.4.3 Implications for Risk Management and Decision Making

It has been acknowledged that risk assessment, particularly in cases of deep uncertainty, is conducted within a context of deliberative review and managerial judgment (Aven, 2013), and that decisions related to the risk in question should be considered "risk-informed", rather than "risk-based" (Apostolakis, 2004). This is particularly true in cases of highly uncertain or ambiguous risk problems, which require a management approaches based on precaution and discourse that are aimed at producing collective understanding and identifying mutually acceptable solutions (Klinke and Renn, 2002). For this reason, it is useful to consider how each approach contributes to improving understanding of the risks in a manner that could support and inform decision-making. However, it is important to recognize that none of the assessed methods are purely prescriptive; instead of identifying "an optimal decision", they instead aim to characterize risk and uncertainty in a way that can

inform decision making. While each methodology could be used in a prescriptive manner when certain decision rules are assumed, these decisions rules are likely to be contentious in deeply uncertain conditions, particularly when multiple viewpoints and values must be considered. While these conditions make objectively "correct" risk management solutions elusive, risk assessment can inform decision making in other ways; for example, by identifying which aspects of a system contribute most significantly to risk and which measures offer the greatest risk reduction potential (Amundrud and Aven, 2015). For this reason, we avoid a prescriptive interpretation of analytical results here and instead focus on the insights that each approach would provide to support managerial review and judgement.

In the uncertainty factors assessment, the expected value of the upgrades is slightly lower than the no-action alternative, but has a much smaller probability of incurring very high costs. This would suggest to many decision makers that the upgrades are a worthwhile investment, but the caveat that the strength of knowledge supporting this assessment is only deemed to be moderate might make planners less confident in that decision. However, one useful insight to arise from the qualitative component of the assessment is the identification of which assumptions are most critical for the quantitative risk assessment, as this can point towards assumptions that should be given additional consideration through either further research or more sophisticated treatment within the risk assessment. For instance, because the assumptions regarding future rainfall rates and land and material costs appear to be the most critical for comparison of the two alternatives, that might suggest that additional research aimed at understanding the validity of those assumptions would be a useful course of action. By providing improved knowledge and clarification about the risk context, this insight could be used to support an ongoing process of adaptive risk management (Bjerga and Aven, 2015). Alternatively, the risk assessment process could be repeated, but with these

assumptions relaxed through the use of intervals or imprecise distributions to describe the flood return period and land and material costs, as suggested by Berner and Flage (2015). This ability to refine the risk assessment or focus research activities could be one reason that including information on the background information that supports quantitative risk descriptions improves the perceived usefulness of risk assessments (Lin et al., 2015). However, one limitation with this approach may be the timing of the assessment. Ideally, if the strength of knowledge supporting an assessment is low, this could be identified and communicated prior to conducting a possibly resource-intensive quantitative assessment.

In the probability bounds assessment, it becomes apparent that epistemic uncertainty makes it unclear which of the two alternatives actually has a lower expected costs. However, the probability of very high costs (for instance, above $500M) could be as high as 0.8 for the no-action alternative. It should be noted that one common critique of alternative uncertainty representations is that they might result in bounds that are too wide to support decision-making (Aven, 2010), but this example demonstrates that this will not always be the case. While the bounds for the no-action alternative in our example are quite wide, this demonstrates that the impact of epistemic uncertainty on the no-action alternative appears much larger than the impact of epistemic uncertainty on the upgrades. This suggests that the possible costs associated with the upgrades are well characterized compared to the costs of doing nothing, and puts the upgrades in a relatively favorable light. The potential for very high costs associated with the no-action alternative due to both epistemic and aleatory uncertainty would likely suggest to decision makers that the upgrades would be a worthwhile investment.

In the RDM assessment, the analytical output provides more insight regarding the key uncertainties driving the decision, and thus the most valuable areas for further research and evaluation, rather than the comparison between the two alternatives. In this case, the key uncertainties were the flood return period and the floodwall failure probability. In particular, the no-action alternative appears particularly vulnerable to a scenario where the flood return period is less than 83 years and the floodwall failure probability is greater than 0.11. This would suggest that decision makers should base their decision around the likelihood of these specific conditions, rather than worrying about the full distribution of all uncertain parameters. If there is strong disagreement or uncertainty surrounding the likelihood of this scenario, it would suggest that additional research in this area would be valuable. The analysis also found that the no-action alternative resulted in lower costs in 95% of simulations, which at first glance might make the upgrades appear unnecessary. However, one must exercise caution in basing decisions on these results since they do not consider the likelihood or magnitude of costs in each case. While the magnitude of costs could be evaluated by using alternative performance measures based on regret-based or deviation-based metrics as in Lempert et al. (2006) and Kasprzyk et al. (2013), these measures are still reliant on the assumption that each simulation is equally likely. If there is strong disagreement surrounding the distributions used to generate the samples, this assumption is likely to be contentious.

It is interesting to note that the implications for risk management (both in terms of the choice between alternatives and suggested areas for additional research or refinement of the risk assessment process) are not necessarily consistent between the three approaches. The uncertainty factors assessment and probability bounds analysis both made the upgrades alternative appear quite favorable, while the RDM assessment (and particularly the fact that

the majority of simulations resulted in higher costs for the upgrades) could suggest to some decision makers that these upgrades are not necessary. While the probability bounds analysis suggested that the costs associated with the no-action alternative were much more sensitive to epistemic uncertainty than the upgrades, the uncertainty factors assessment suggested that the strength of knowledge supporting the assessment of each alternative was equal. This is likely due to the fact that the upgrade costs were judged to be highly sensitive to the assumption regarding land and material costs, which had weak supporting knowledge. While a violation of this assumption may result in high upgrade costs compared to the baseline quantitative risk assessment, these costs may still be quite small compared to the distribution of costs for the no-action alternative. Finally, the RDM assessment suggested that land and material costs were not actually a very important parameter in choosing between the two alternatives, and instead suggested that the flood return period and floodwall failure probability were most important. The divergence between risk management implications that occurred even within this very simple example suggest that further research comparing these methodologies using a more realistic problem could be very valuable.

3.5 Discussion

Relating each process back to the risk assessment framework described in Section 3.2 allows us to clarify how each methodology contributes to this risk assessment process and resulting risk description. Uncertainty factors were specifically developed to clearly demonstrate the underlying strength of knowledge for a risk assessment, and thus provide each of the components of the risk description used here. In our example, *A'* and *C'* refer to a certain number of failures, combined with a specific realization of asset value and construction cost, and the total costs that result. The uncertainty associated with each of

these realizations is measured probabilistically, and *K* is the qualitative description of the underlying knowledge on which these probability measurements are based.

The PBA provides the same description of *A'* and *C'*, but the uncertainty measurement *Q* is now the bounds on the probability associated with *A'* and *C'*. The underlying knowledge *K* is not explicitly described, but could be inferred by the shape of the resulting p-box for total costs in each case. This is a frequent interpretation of p-boxes, where the slant of the box represents variability while the width represents epistemic uncertainty. However, this interpretation depends on what led to the use of imprecise probability distributions in the first place. The width of the p-box for total costs of the no-action alternative can be interpreted as the epistemic uncertainty associated with this alternative, since the imprecision in this distribution resulted entirely from lack of knowledge regarding the flood return period, floodwall failure probability, and assets at risk. However, the width of the p-box for the repairs alternative could arise from multiple sources, including epistemic uncertainty, imprecision in elicited probabilities, or disagreement between experts. Thus, the precise interpretation of the PBA results requires careful consideration of the various sources of uncertainty and imprecision leading to that result.

The RDM analysis is the most challenging to relate back to the risk assessment process and description, but could be interpreted in one of two ways. The first could be that it serves as an alternative method for hazard identification which identifies specific combinations of uncertain conditions that lead to (or are relatively likely to lead to) undesirable outcomes. Another interpretation could be that it provides an alternative description of *A'* and *C'*, where instead of considering any possible outcome *A'* and its consequences *C'*, an "event" is simply a box that results in particularly high consequences *C'*.

In this interpretation, the set of outcomes *A'* includes only Scenario 1 (flood return period is less than 83 years and the failure probability is greater than 0.11) and Scenario 2 (the above conditions are not met), rather than all possible outcomes. This interpretation could be useful in situations where decision-makers are overwhelmed by having to consider all outcomes that could occur, and prefer to consider a smaller number of scenarios that would drive their decision. However, one limitation with this interpretation is that there is not a clear set of consequences associated with each of those two scenarios due to the presence of aleatory uncertainty in the number of floods and floodwall failures that will occur. Additionally, it provides no measurement of the relative likelihood or uncertainty associated these events (*Q'*), and thus no description of the underlying knowledge *K* for that measurement.

One advantage of the uncertainty factors methodology is its explicit description of the assumptions underlying the analysis, as well as the likelihood and consequences of their violation. While explicit description of underlying assumptions should be a component of any risk assessment, it could be argued that this is often treated as an afterthought to the quantitative analysis, if it receives any treatment at all. By creating an explicit framework for describing these assumptions and their impacts, uncertainty factors could make presentation of these assumptions more systematic and effective. While the assumptions underlying the probability bounds and RDM analyses are relaxed relative to the probabilistic analysis, they will always be present since complete modeling of all uncertainties and complexities is impossible in all but the most trivial of systems. Developing systematic and transparent methods for description and assessment of underlying assumptions for different technical approaches could be a valuable area of research for any method of risk and uncertainty

assessment, and could lead to important steps forward in improving the transparency of these analyses.

3.6 Conclusions

In this paper, we compare three methods that have been proposed for risk assessment under deep uncertainty and critically evaluate how these approaches contribute to the risk assessment process and resulting risk description. By applying each method to a stylized example problem related to flood risks under climate change, we are able to compare each approach's representation of uncertain quantities, analytical output, and implications for risk management. While each methodology aims to assess and describe risks in a manner that is more reflective of the uncertainties and assumptions underlying the assessment, the analytical output and implications for decision making are not necessarily consistent between approaches. This suggests the potential value in additional comparative research to better understand the sources of these deviations, as well as the need for analysts to consider the ways in which the choice of methodology might impact analytical results. However, the methodologies also demonstrate the ways that risk assessment can inform decision making in conditions where uncertainty and ambiguity make prescriptive approaches inappropriate. In particular, the identification of epistemic uncertainties that most contribute to uncertainty in the resulting risk description or choice of alternatives can provide useful insights into places where additional research or more sophisticated representation could most benefit the assessment. This can ultimately inform more effective responses to deeply uncertain risks such as climate change and support adaptive, deliberative and precautionary approaches to risk management and governance.

4 SCENARIO DISCOVERY WITH MULTIPLE CRITERIA: AN EVALUATION OF THE ROBUST DECISION MAKING FRAMEWORK FOR CLIMATE CHANGE ADAPTATION[4]

4.1 Introduction

In recent years, there has been increasing concern and discussion over deep uncertainty in the risk analysis field (Cox, 2012). The term "deep uncertainty" is commonly used to refer to situations where probabilistic models of uncertainty cannot be confidently determined or agreed upon (Cox, 2012) or where frequentist probabilities based on repeatable events cannot be developed (Aven, 2013). Concerns over deep uncertainty have been particularly strong in the climate change adaptation field, with some arguing that traditional approaches to risk management, such as maximization of expected utility, are poorly suited to climate policy and adaptation problems (Kunreuther et al., 2013). This has led to interest in robust decision frameworks (Weaver et al., 2013), which include methods such as robust decision making (RDM; Lempert et al., 2006), decision scaling (Brown et al., 2012), and info-gap decision theory (Ben-Haim, 2000). These methods are commonly contrasted with so-called "predict-then-act" frameworks by focusing on the identification of robust rather than optimal solutions, and by using analytics to first identify conditions where plans or strategies may fail, rather than first predicting what an uncertain future will look like (Weaver et al., 2013). These frameworks can be particularly useful in situations characterized by poorly understood nonlinear or threshold responses (Lempert and Collins, 2007) or many

[4] This chapter is based on the following manuscript: Shortridge, J.E., and Guikema, S.D. Scenario discovery with multiple criteria: an evaluation of the robust decision making framework for climate change adaptation. In press at *Risk Analysis*. Early view published online in February 2016. DOI: 10.1111/risa.12582

stakeholders with conflicting values and beliefs about the future (Hallegatte and Rentschler, 2015).

RDM is one such framework that has been applied to a number of climate adaptation problems (Groves et al., 2013a; Groves and Bloom, 2013; Lempert et al., 2013; Lempert and Groves, 2010). It is a multi-step, iterative approach that includes both analytical and deliberative components (Lempert et al., 2006). The analytical components of the process simulate how a system or policy alternatives will perform in many plausible future states of the world, and then use the results of these simulations to 1) identify robust alternatives (those that perform relatively well in many states of the world) and to 2) identify the conditions under which a preferred alternative will perform poorly (Lempert et al., 2006). This second objective has been referred to as scenario discovery, as it identifies the conditions that represent vulnerabilities for a proposed policy and thus the conditions under which an alternative solution would be preferred (Lempert, 2013). Scenario discovery uses the Patient Rule Induction Method (PRIM; Friedman and Fisher, 1999) to identify regions of a multidimensional input variable space that result in undesirable values of the output variable. These regions are defined by quantitative logical conditions involving individual input variables. For instance, in one study a regional water plan was found to result in unacceptably high costs when precipitation declined by more than 10%, groundwater recharge decreased by over 3%, and a water recycling program failed to meet its goals (Lempert, 2013; Lempert and Groves, 2010). By identifying these conditions, the scenario discovery process can identify which uncertainties are most important for a given decision problem (and thus potentially inform research activities) and specify the vulnerable conditions for which decision-makers may want to prepare.

The PRIM algorithm was developed for problems where multiple input variables influence the value of a single response variable, and does not contain a mechanism for incorporating multiple response variables or outcome criteria. Because of this, existing RDM literature incorporates multiple criteria in a number of different ways. Some studies have conducted scenario discovery over a single outcome metric, such as cost (Lempert et al., 2012; Lempert and Groves, 2010), system reliability (Groves et al., 2014), expected utility (Hall et al., 2012a), or a single aggregated performance score (Lempert et al., 2006). A number of evaluations that do consider multiple criteria apply scenario discovery over each criterion separately (Groves et al., 2013a, 2013b, 2014; Popper, 2009). By identifying the conditions that are likely to cause failure for each individual objective, this process can be highly informative but may be impractical for problems with a large number of performance metrics. Finally, some studies apply scenario discovery across multiple criteria where failure on any single criterion is equivalent to failure overall (Herman et al., 2014, 2015; Kasprzyk et al., 2013; Lempert et al., 2013). Collectively, these studies demonstrate that there are multiple methods that can be incorporated to conduct scenario discovery in a problem characterized by more than one performance metric. However, they provide little insight into how the choice of method used to incorporate multiple criteria might impact the scenarios identified by the PRIM algorithm and what methods may be the most informative for decision makers.

In this study, we compare different methods for incorporating multiple objectives into the scenario discovery process to evaluate how the treatment of multiple criteria can impact the vulnerable scenarios identified within the RDM framework. We use the Lake Tana basin in Ethiopia (described in Section 1.5) as a case study, where multiple long-lived water infrastructure projects are planned for construction but whose effectiveness could be impacted by climatic and environmental uncertainties. The scenario discovery process is

used to identify the conditions which are likely to cause unacceptable performance of this infrastructure with regard to multiple criteria, including provision of water to different economic sectors and downstream environmental conditions. We first identify failure scenarios by assessing each performance metric individually, and the implications that these scenarios have for the design of system improvements and research efforts focused on key uncertainties. We then compare these to failure scenarios identified using different methods for aggregating the metrics into a single performance score. By evaluating the sensitivity of the scenario discovery process to the treatment of multiple criteria, this works aims to support more effective application of robust decision frameworks in contexts where performance across multiple economic and environmental metrics must be balanced.

4.2 Methods

4.2.1 Simulation Model

A two-component simulation model was developed to assess how changes in climatic and environmental conditions would impact water resources in the basin. The first component consisted of empirical rainfall-runoff models that predicted monthly streamflow in each of the five rivers with proposed reservoirs (Gilgel Abbay, Gumara, Koga, Megech and Ribb) based on monthly temperature, rainfall, rainfall intensity, and agricultural land cover. The models were each fit by regressing a 40-year monthly time series of streamflow in that river against historic climate data taken from Climate Research Unit (CRU) gridded datasets (Harris et al., 2014) and agricultural land cover as reported by data taken from Rientjes et al. (2011), Gebrehiwot et al. (2010) and Garede and Minale (2014). Multiple regression and machine-learning algorithms were compared in their predictive ability through random hold-out cross validation. The highest performing models based on out-of-sample mean absolute error were used to generate streamflow predictions using climate and

land cover data. These included a linear model, M5 model (Quinlan, 1992), artificial neural network (Ripley, 1996), generalized additive model (Hastie and Tibshirani, 1986) and random forest model (Breiman, 2001). Each basin's model was compared to a null model which predicted streamflow in each month as simply the mean historic streamflow for that month. The models were able to achieve statistically significant reductions in predictive error based on bonferroni-corrected Wilcoxan signed rank tests. Additional details on model development is included in Chapter 2 and discussed by Shortridge et al. (2016).

The second component of the simulation model was a Water Evaluation and Planning (WEAP; Sieber and Purkey, 2015) water allocation model developed for the basin by Alemayehu et al. (2010). This model simulates natural hydrologic processes such as streamflow and evaporation, as well as human extraction and use of water. In each month, the model performs a mass balance to account for both extraction and inflows, allocating water to different demand nodes in order of user-defined priorities (Sieber and Purkey, 2015). The monthly streamflow sequences derived from the empirical rainfall-runoff model for each river, as well as time series of evaporation from the lake and each reservoir, were used as model inputs. The model then calculated the amount of water allocated and coverage (percent of demand delivered) for different demand nodes, as well as lake elevation and downstream flows. Additional information on WEAP model development, calibration and validation is discussed by Alemayehu et al. (2010).

4.2.2 RDM Evaluation

In the first step of the RDM evaluation, a range of feasible values was identified for each of the uncertain parameters that could impact infrastructure performance in the future (Table 4.1). Because the objective of the scenario discovery process is to find conditions that

result in unsatisfactory performance of the infrastructure, we used wide ranges of values to better identify the thresholds and tipping points that would result in poor performance.

Uncertain parameter	Symbol	Range of values
Change in temperature	ΔT	0.5 to 5.5° C
Change in rainfall	ΔP	-20% to +35%
Change in rainfall intensity	ΔInt	0% to +20%
Specific sediment loads	SedRate	80 to 2400 tons/km^2 annually
Agricultural land cover	AgLC	50% to 90%
Evaporation coefficient	EtC	0.8 to 1.2

Table 4.1: Uncertain parameters

Possible impacts of climate change were represented by a change in temperature ranging from 0.5 to 5.5° C and a change in annual precipitation ranging from -20% to positive 35%. These values were taken from IPCC multi-model ensemble projections for the East Africa region for the period 2081-2100 under all representative concentration pathways (van Oldenborgh et al., 2013). Additionally, there is concern that climate change could result in an intensification of precipitation, even when overall amounts of precipitation decrease (Barnett et al., 2006; Easterling et al., 2000; Kharin and Zwiers, 2005). For this reason, we also considered increases in rainfall intensity (defined as the total amount of rainfall in a month divided by the days where rainfall occurs) from 0 to 20%. Specific sediment yield is the amount of sediment deposited in the reservoir normalized by the upstream area contributing sediment. A range of values for specific sediment yield were taken from sampling results from various rivers in the basin (WWDSE, 2008; WWDSE and Tahal Group, 2009b) while future agricultural land cover was assumed to range from 50 to 90% based on values experienced over the past 50 years (Garede and Minale, 2014; Gebrehiwot et al., 2010; Rientjes et al., 2011). Finally, evaporation estimates were multiplied by a factor ranging from 0.8 to 1.2 to account for uncertainty arising from the limited meteorological data available to estimate evaporation from the reservoirs and lake. This parameter

represents the degree to which actual evaporation differs from our estimates, with any value over 1.0 implying underestimation of evaporation.

To assess how the proposed projects would perform in various possible future states of the world, 5000 random combinations of the six uncertain parameters were generated to be used as inputs for the simulation model described above. Samples were generated using Latin Hypercube sampling across a uniform distribution for the range of possible values for change in temperature, sedimentation rate, agricultural land cover, and the evaporation coefficient. While Latin hypercube sampling is often used to generate multivariate probabilistic distributions, here it is only used as a mechanism for generating a diverse sample of future conditions that could feasibly occur. These samples are used as input for exploratory modeling (Bankes, 1993) that evaluates how the system responds to different multivariate conditions while making no inference regarding the likelihood of those states. Other methods for sample generation, including full combinatorial sampling across discrete uncertain parameters and GCM ensemble projections (Groves et al., 2013a; McJeon et al., 2011), have been used in RDM evaluations and the application of further sample generation methods could be a valuable area for future research. Changes in rainfall and rainfall intensity are likely to be correlated with changes in temperature, as greater climate forcing is expected to result in more extreme changes to both temperature and precipitation. To account for this, a correlation was induced between temperature and the rainfall and rainfall intensity parameters. For each of the 5000 samples, the change in precipitation was randomly selected to be either positive or negative with an equal probability. For the *n*'th sample, a parameter $\overline{\Delta P_n}$ was calculated as in Equation 4.1 and a parameter $\overline{\Delta Int_n}$was calculated as in Equation 4.2. The change in rainfall and rainfall intensity for sample *n* were

then randomly sampled from normal distributions with means equal to $\overline{\Delta P_n}$ and $\overline{\Delta Int_n}$, a coefficient of variation equal to 0.5.

Equation 4.1:

$$\overline{\Delta P_n} = \begin{cases} \dfrac{(\Delta T_n - \Delta T_{min})}{(\Delta T_{max} - \Delta T_{min})} \times \Delta P_{max} & if\ \Delta P > 0 \\ \dfrac{(\Delta T_n - \Delta T_{min})}{(\Delta T_{max} - \Delta T_{min})} \times \Delta P_{min} & if\ \Delta P < 0 \end{cases}$$

Equation 4.2:

$$\overline{\Delta Int_n} = \frac{(\Delta T_n - \Delta T_{min})}{(\Delta T_{max} - \Delta T_{min})} \times \Delta Int_{max}$$

Each of the 5000 samples could be thought of as a possible future state of the world under which the infrastructure might have to operate. The simulation model was then used to assess how well the infrastructure would be able to meet the multiple objectives required of it under each of the 5000 possible futures. For each possible future, the change in temperature, rainfall, and rainfall intensity was used to adjust the 40-year historic climate record in each basin using the delta-change method (Gleick, 1986). These adjusted climate scenarios were then used, along with estimates of agricultural land cover, to generate streamflow sequences for each river. Evaporation from Lake Tana and each reservoir was calculated using Penman's equation (Penman, 1948). These estimates used the adjusted temperature values reflective of climate change and historic monthly average values for wind speed, relative humidity and solar radiation from the Bahir Dar meteorological station as reported by Kebede et al. (2006). These evaporation estimates were then multiplied by the EtC parameter to account for uncertainty stemming from the use of historic average values for calculating evaporation rates under future climates. The capacity of each reservoir

diminished annually based on the specific sediment loading rate assumed for that possible future.

This resulted in 40-year sequences of monthly streamflow, evaporation, and reservoir capacity for each possible future. These sequences were then used as inputs to the WEAP model of the basin, which allocated water to agricultural and hydropower demand nodes and calculated the resulting downstream flows and lake levels. Five performance metrics identified based on stakeholder discussions were calculated to assess how well the infrastructure performed in each possible future (Table 4.2). Previous studies have identified 1784.75 meters as the minimum elevation that Lake Tana can reach before negative impacts to the navigation and fishing industries begin to occur (SMEC International, 2008). Alemayhu et al. (2010) calculated flow requirements needed to support tourism at the Tis Issat waterfall downstream of the lake, as well as environmental flow requirements for each of the tributaries to the lake. These were used to calculate the average percentage of flow requirement met as a measurement of impacts on tourism and environmental conditions. Table 4.2 shows baseline results for each metric, assuming that the infrastructure was operated under historic climate conditions. For each metric, an acceptable performance threshold was identified based on the project design documents (in the case of irrigation water delivery and reliability and hydropower delivery) or baseline performance levels (in the case of lake levels, Tis Issat flows, and environmental flows). These thresholds represent the minimum performance level for each metric that can be considered acceptable.

Objective	Metric	Units	Baseline Results	Acceptable Performance Threshold
Maximize irrigation water reliability	Percentage of years when minimum demand is met	%	98%	90%
Maximize hydropower water delivered	Average water delivered annually	MCM	2699	2681
Minimize percent of time where lake elevation is below minimum acceptable level	Percent of months where lake is above 1784.75 amsl	%	100%	90%
Maximize flows over Tis Issat waterfall	Average flow requirement met for Tis Issat	%	33%	30%
Maximize environmental flows	Average flow requirement met for all rivers	%	78%	70%

Table 4.2: Performance metrics. Baseline performance is based on historic climate conditions, an annual specific sediment yield of 1000 tons/km2, 50% agricultural land cover and an evaporation coefficient of 1.0.

4.2.3 Scenario Discovery

The RDM framework is a multi-step, iterative approach to decision support under uncertainty that contains both quantitative analysis and deliberation. The process includes two analytical components based on simulation model results. When multiple alternatives or policy options are available for a given system, the first analytical component of the approach identifies the most robust alternatives based on regret minimization or satisficing criteria (Lempert et al., 2006). The second analytical component, termed "scenario discovery," aims to identify the conditions which cause unsatisfactory performance in a preferred alternative. In this work, we use the scenario discovery process to identify the conditions which cause unsatisfactory performance for the proposed infrastructure associated with the D2 development level shown in Table 1.1 and discussed in Section 1.5.

The scenario discovery process uses the results of the 5000 simulations described above to identify specific combinations of uncertain input parameters that are likely to result in poor performance. It is based on the patient rule induction method (PRIM) bump-

hunting algorithm (Friedman and Fisher, 1999). The PRIM algorithm was implemented using the SD toolkit package in R (Bryant, 2014). Details on the PRIM algorithm as implemented in the SD toolkit package are discussed in Section 1.4.

The PRIM algorithm does not include a method for considering multiple output variables, and requires that multiple outcomes be either separately evaluated, or aggregated into an overall performance score. In this work, we first apply the PRIM algorithm to each of the five performance metrics separately. This identifies the specific scenarios that are likely to result in unsatisfactory performance for each individual performance metric. We then use five different methods to aggregate the performance metrics into a single overall performance score, and apply the PRIM algorithm to these aggregated results. The aggregation methods are shown in Table 4.3. In the first method, if the infrastructure fails to meet any of the six performance criteria in a given possible future, that is considered a failure overall. This approach is similar to a multiplicative multi-attribute utility function applied to binary performance scores, as a score of zero for any single attribute results in a score of zero overall. This is the approach used previously by Kasprzyk et al. (2013), Herman et al. (2014 and 2015) and Lempert et al. (2013). This method is demonstrated in Equation 4.3, where $y_{i,n}$ is the binary performance score (1 for acceptable performance and 0 for unacceptable performance) for individual metric i in possible future n, and Y_n is the overall performance score for possible future n. In the other four methods, an additive performance score is calculated, with the weights between different attributes varied to reflect different priorities. In this approach, the scores for each metric are normalized across the range of outcomes experienced in the 5000 possible futures and a weighted sum is calculated as in Equation 4, where $u_{i,n}$ is the normalized performance score on attribute i in possible future n, and w_i is the weight assigned to attribute i.

Equation 4.3:

$$Y_n = \prod_{i=1}^{I} y_{i,n}$$

Equation 4.4:

$$Y_n = \sum_{i=1}^{I} w_i u_{i,n}$$

	Aggregation Method	Multiplicative	Additive – equal weighting	Additive – agricultural priority	Additive – hydropower priority	Additive – environmental priority
Weights	Irrigation water reliability	NA	0.2 (1)	0.5 (1)	0.33 (2)	0.05 (4)
	Hydropower water delivery	NA	0.2 (1)	0.33 (2)	0.5 (1)	0.05 (4)
	Lake levels	NA	0.2 (1)	0.06 (3)	0.06 (3)	0.4 (1)
	Tis Issat falls coverage	NA	0.2 (1)	0.06 (3)	0.06 (3)	0.2 (3)
	Environmental flow coverage	NA	0.2 (1)	0.06 (3)	0.06 (3)	0.3 (2)
	Acceptable performance threshold	1.0	0.75	0.87	0.88	0.69

Table 4.3: Weighting schemes used to calculate aggregate performance scores. Numbers in parenthesis are the importance rankings of each criteria for a given weighting scheme.

Normalized weights were calculated using the rank sum weighting procedure based on four different possible rankings of attribute importance (Stillwell and Edwards, 1979). A summary of the four weighting schemes evaluated is presented in Table 4.3. For the additive aggregation schemes, an aggregated minimum acceptable performance threshold is calculated using Equation 4.4 and the performance thresholds presented in Table 4.2.

4.3 Results

A summary of the simulation results for each individual performance metric is presented in Table 4.4. It is apparent that accounting for uncertainty in the parameters listed in Table 4.1 have the potential to result in dramatic ranges in performance, particularly with

regard to irrigation reliability, the amount of water provided for hydropower, and the elevation of Lake Tana. The performance thresholds are not met in 36% to 66% of the simulated futures, depending on the metric assessed. It is important to note that these percentages should not be interpreted as a statement regarding the likelihood of failure, since that would implicitly assume that each simulated possible future was equally likely. However, it can provide information about the relative sensitivity of the different metrics to the uncertain parameters listed in Table 4.1. For instance, the Lake Tana elevation metric appears relatively robust to this uncertainty (failing in only 36% of futures) whereas the hydropower delivery metric fails in 66% of them.

Metric	**Acceptable performance level**	**Minimum**	**Maximum**	**Futures where threshold is unmet**
Irrigation reliability	0.90	0.02	1.00	57%
Hydropower water delivered	2681	271	2855	66%
Lake Tana elevation	0.95	0.10	1.00	36%
Tis Issat Falls coverage	0.30	0.26	0.39	46%
Environmental flow coverage	0.76	0.64	0.83	48%

Table 4.4: Simulation results. Baseline results assume historic climate conditions, a specific sediment yield of 1000 tons/km2, 50% agricultural land cover, and EtC = 1.0.

The scenario discovery process was used to identify combinations of uncertain input parameters that best described the simulations where performance thresholds were not met. These combinations can be interpreted as scenarios to which the proposed infrastructure is vulnerable (termed "failure scenarios" from here forward). Table 4.5 shows the results of the scenario discovery process when it was run on each metric separately. Two failure scenarios were identified for each metric, and the box coverage and density are described for each individual scenario, as well as the ensemble as a whole, for each metric. When multiple

conditions are listed on a single line, this describes conditions which must simultaneously occur for performance to drop below the threshold. Conversely, when conditions are listed in separate failure scenarios for a given metric, this implies that either of those conditions will cause failure. For instance, irrigation reliability can fail if the change in precipitation is less than -3.8% or if EtC is greater than 1.09, whereas coverage for the Tis Issat falls tends to fail if both EtC is greater than 1.08 and the change in precipitation is less than +16.4%. These scenarios are shown graphically in Figure 4.1.

Unsurprisingly, precipitation plays a role in the failure scenarios for each metric, but the way in which it combines with other uncertain parameters differs. While a decrease in precipitation must be combined with certain conditions regarding temperature and evaporation estimates to cause failure for the lake elevation metric, it is enough to cause failure for the irrigation, hydropower, Tis Issat and environmental metrics on its own. Additionally, the relative sensitivity of the different metrics to changes in precipitation is apparent, with the Tis Issat metric vulnerable to any decrease beyond approximately 2% while environmental flow coverage is only vulnerable to decreases beyond approximately 8%. Another important insight is that both irrigation reliability and hydropower are vulnerable to underestimation of evaporation, even if climate conditions are favorable. Interestingly, the only metrics that appeared sensitive to changes in temperature were lake elevation and environmental flows. This could be due to the large role that evaporation off of Lake Tana plays in the basin's water balance, which would be expected to increase with higher temperatures.

Metric	Percent failures	Failure Scenarios	Box density	Box coverage	Ensemble density	Ensemble coverage
Irrigation reliability	57%	1. ΔP < -3.8%	0.86	0.56	0.83	0.79
		2. EtC > 1.09	0.77	0.25		
Hydropower water delivery	66%	1. EtC > 0.99	0.91	0.73	0.91	0.93
		2. ΔP < -5.5%	0.91	0.93		
Lake elevation	36%	1. ΔP < 1.4%, EtC > 0.94, ΔT >1.16°	0.85	0.68	0.82	0.77
		2. ΔT > 4.1°, ΔP < 6.6%	0.65	0.1		
Tis Issat Falls coverage	46%	1. ΔP < -2.2%	0.8	0.71	0.82	0.9
		2. EtC > 1.08, ΔP < 16.4%	0.9	0.19		
Environmental flow coverage	48%	1. ΔP < -7.8%	0.7	0.38	0.61	0.78
		2. ΔT > 2.6°	0.55	0.40		

Table 4.5: Failure scenarios for individual performance metrics

Aggregation Scheme	Percent failures	Failure Scenarios	Box density	Box coverage	Ensemble density	Ensemble coverage
Multiplicative	77%	1. EtC > 0.96	0.91	0.7	0.91	0.86
		2. ΔP < -5.5%	0.92	0.15		
Additive – equal weighting	52%	1. ΔP < -4.6%	0.87	0.57	0.84	0.79
		2. EtC > 1.11	0.77	0.22		
Additive – irrigation priority	59%	1. ΔP < -3.7%	0.88	0.55	0.87	0.78
		2. EtC > 1.10	0.86	0.23		
Additive – hydropower priority	60%	1. EtC > 1.03	0.85	0.61	0.86	0.86
		2. ΔT > 2.8°, ΔP < 7.0%	0.92	0.25		
Additive – environmental priority	46%	1. ΔP < -0.34%, ΔT > 1.96°	0.85	0.64	0.84	0.83
		2. EtC > 1.1, ΔP < 16.9%	0.81	0.19		

Table 4.6: Failure scenarios for aggregated performance scores

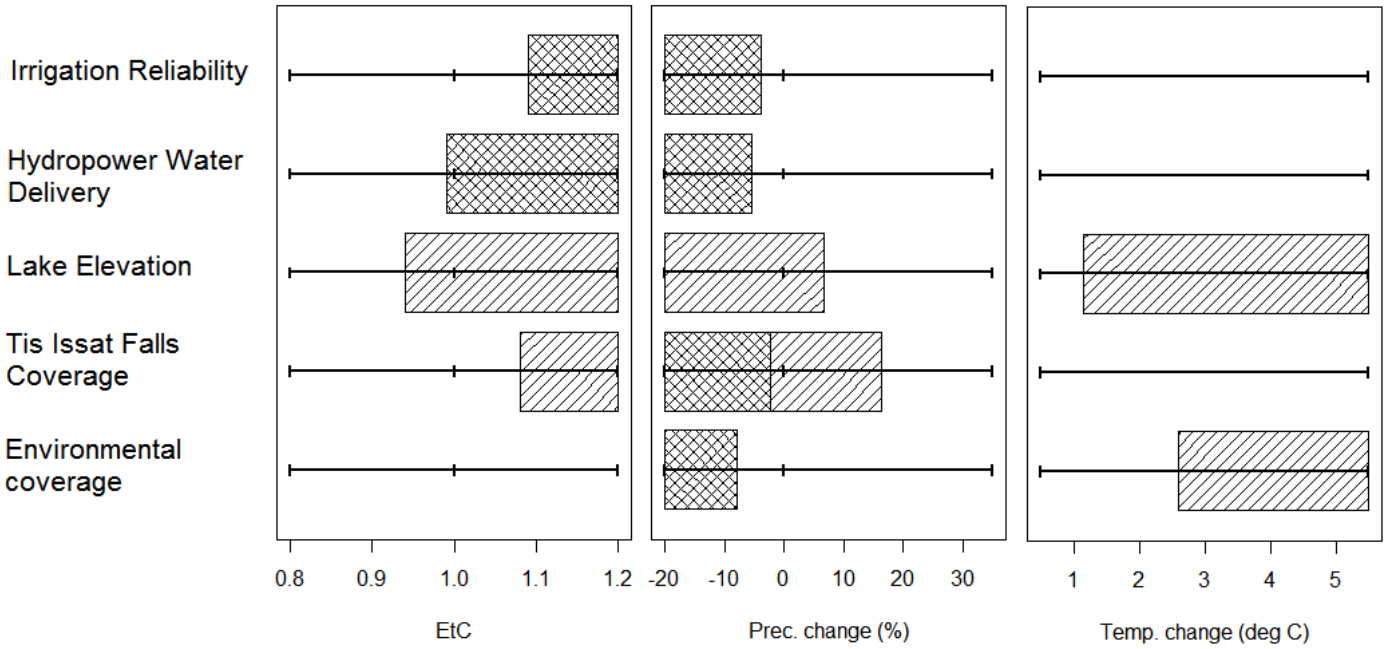

Figure 4.1: Failure scenarios for individual performance metrics. Diagonal lines indicate a condition that has to occur in conjuction with specific conditions regarding the other parameters identified by diagonal lines. Boxes with hash marks indicate conditions that are sufficient to cause failure on their own, regardless of the values taken on by other parameters.

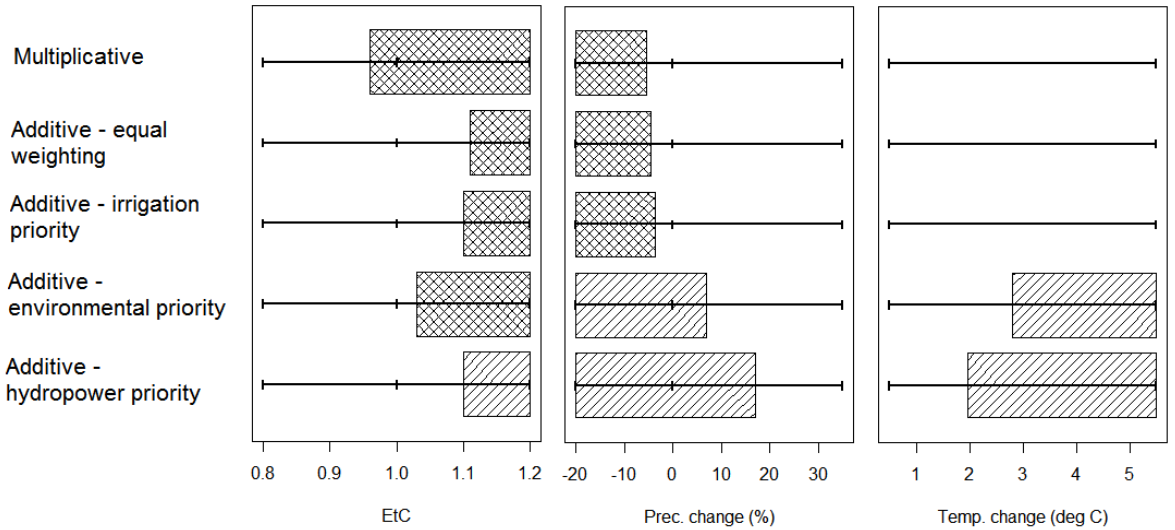

Figure 4.2: Failure scenarios for aggregated performance scores. Diagonal lines and hash marks are as in Figure 4.1.

Table 4.6 and Figure 4.2 show the failure scenarios identified for the aggregated multiattribute performance measures. From looking at the percentage of simulations classified as failures based on each aggregation scheme, it is apparent that they give different pictures of overall system robustness. The multiplicative aggregation scheme is very strict when implemented in a binary fashion, since unsatisfactory performance on any single metric will result in failure overall. This results in a high percentage of simulations that were classified as failures when compared to the additive approaches, where poor performance on one metric can be compensated for by good performance on another. Because the additive aggregation methods are less strict than the multiplicative method, they provide a more optimistic view of system performance, with failure occurring in a smaller percentage of simulations. However, they do not provide any insight into which individual performance thresholds are being satisfied and which are not. While this method ensures that at least one performance threshold will be satisfied for the multi-criteria performance threshold to be met, it cannot ensure that any single metric (e.g., hydropower provision) is achieved.

The failure scenarios for the multiplicative scheme closely mirror those for the hydropower water delivery, which was the most sensitive individual metric. This indicates that when such an aggregation scheme is used, it is possible for the resulting failure scenarios to be dominated by a single metric. When the additive method with a priority on hydropower delivery is used, the failure scenarios still indicate a vulnerability to evaporation overestimates, but do not indicate a vulnerability to decreases in precipitation unless combined with an increase in temperature. While three aggregation schemes (multiplicative, additive with equal weights and additive with a priority on irrigation) result in relatively consistent failure scenarios, the threshold values identified for the EtC and ΔP parameters differ between them. For instance, the additive method with an irrigation priority appears the

most sensitive to even small decreases in precipitation, while the multiplicative scheme is most sensitive to evaporation underestimates.

Additional investigation into the conditions that cause failure for a given metric demonstrate how some insights and nuances about system performance can be lost when performance metrics are combined into a single score. The left hand side of Figure 4.3 shows a scatterplot demonstrating how changing precipitation and evaporation estimates impact hydropower performance. Filled in dots represent simulations where the threshold for hydropower water delivery was not met, and hollow dots represent those simulations where it was. A fairly distinct linear divide is apparent, demonstrating how the system's tolerance for higher rates of evaporation relates to the level of precipitation experienced. While the hyper-rectangles identified by the PRIM algorithm are unable to capture this sort of relationship precisely (although orthogonal transformations have been used to address this issue (Dalal et al., 2013)), the identification of precipitation and EtC as the key uncertainties driving performance, combined with a simple visualization, makes it apparent. However, when the same scatterplot is generated using the multiplicative performance metric, this relationship is no longer discernable.

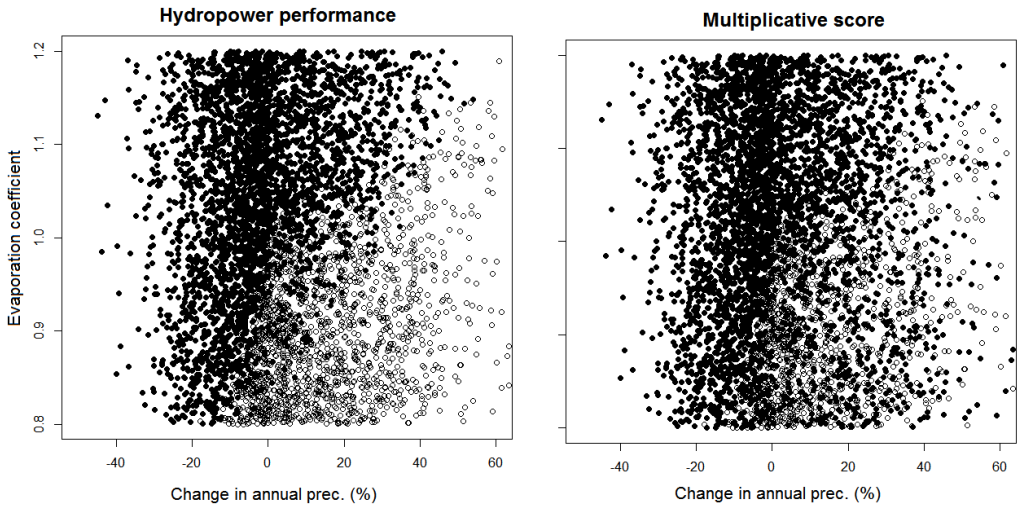

Figure 4.3: Scatterplots showing simulations with hydropower performance and multiplicative aggregated perfomrance above their respective thresholds. Filled in circles represent simulations where the threshold was not met, and empty circles indicate simulations where it was.

4.4 Discussion

To understand the potential implications of these results, it is important to consider the different ways in which such scenarios might be used to support decision making. One useful outcome of the scenario discovery process is that it can identify the uncertain parameters that have the greatest impact on system response and thus, the areas where a reduction in uncertainty could be the most valuable. It also may provide useful insights by identifying the parameters that are not as influential over system performance and thus don't warrant as much concern (Hallegatte and Rentschler, 2015). In our analysis the parameters identified as important were generally consistent over different metrics and aggregation methods, with precipitation and evaporation uncertainty being the strongest drivers of vulnerability while precipitation intensity, future land cover, and sedimentation rates were not identified as influential. However, there were some notable differences. One interesting result was that uncertainty in evaporation estimates could result in unacceptable levels of irrigation reliability and hydropower water delivered even in favorable climate conditions.

This indicates that even without the impacts of climate change, the proposed infrastructure might be unable to meet its goals if current estimates of evaporation prove to be too low. While uncertainty surrounding future projections of climate change is unlikely to be reduced in the coming years (Kunreuther et al., 2013), additional meteorological monitoring, combined with the development of remote sensing products, could be used to refine evaporation estimates and gain a better sense of likely system performance. However, this sensitivity to evaporation alone is not apparent when the environmental priority additive weighting scheme is used, indicating that this insight could be lost if individual metrics aren't separately assessed.

Another useful aspect of the scenario discovery approach is that it not only identifies which uncertain parameters are most influential, but can also determine threshold levels beyond which performance levels are unacceptable. This is one of the main advantages of the approach when compared to variance-based methods for global sensitivity analysis such as Sobol indices, which identify variables to which an outcome is most sensitive but not necessarily thresholds within that variable space (Herman et al., 2015). These thresholds can highlight the relative sensitivity of different performance metrics; for instance, Tis Issat flow coverage is more sensitive to decreases in precipitation than environmental coverage. These precipitation thresholds can also be informative when considering interannual variability in performance, even under current climate conditions. For example, during the 20-year period from 1977 to 1996 the basin experienced lower than average rainfall (Figure 4.4), and these decadal-scale dry periods would be expected to occasionally occur even without the impact of climate change. The average annual precipitation during this period was 1360 mm, which is approximately 8% less than the long-term average of 1470 mm and thus below the threshold for hydropower performance. The amount of water provided for hydropower thus

appears sensitive not only to long-term climate change, but also to interannual variability experienced currently. However, if one were to assess performance using the additive weighting scheme with a priority on hydropower, this vulnerability would not be apparent. Finally, these thresholds could be used in additional probabilistic analysis to determine the relative likelihood of the different scenarios identified, as has been done by Lempert et al. (2012). Because the probability of these scenarios is contingent on their quantitative definition, this could in turn impact the expected value and probability of failure.

A third way in which the scenario discovery process can help inform decision making is by highlighting the vulnerabilities that decision makers may want to address to make their system more robust. For instance, after recognizing that water supply costs were vulnerable to a decrease in the amount of groundwater recharge, Lempert and Groves (2010) proposed additional investment in stormwater capture and groundwater replenishment facilities to help address this vulnerability. In this regard, the scenarios identified for the aggregated performance scores are much less informative than those identified for the individual metrics. In our example, the two metrics that are most sensitive to climatic and environmental uncertainty based on the number possible futures resulting in failure are the irrigation and hydropower metrics. This is despite the fact that these are the two objectives driving the large infrastructure investments in the region. Thus, decision makers may see this information and try to adopt policies or adapt the proposed infrastructure to make its performance with regard to those metrics more robust, particularly given their economic importance. For instance, the irrigation drainage systems could be adjusted to improve irrigation efficiency, or water allocation rules could be adapted to provide more water for hydropower. When the aggregated performance scores are used, these avenues for system improvement are not apparent.

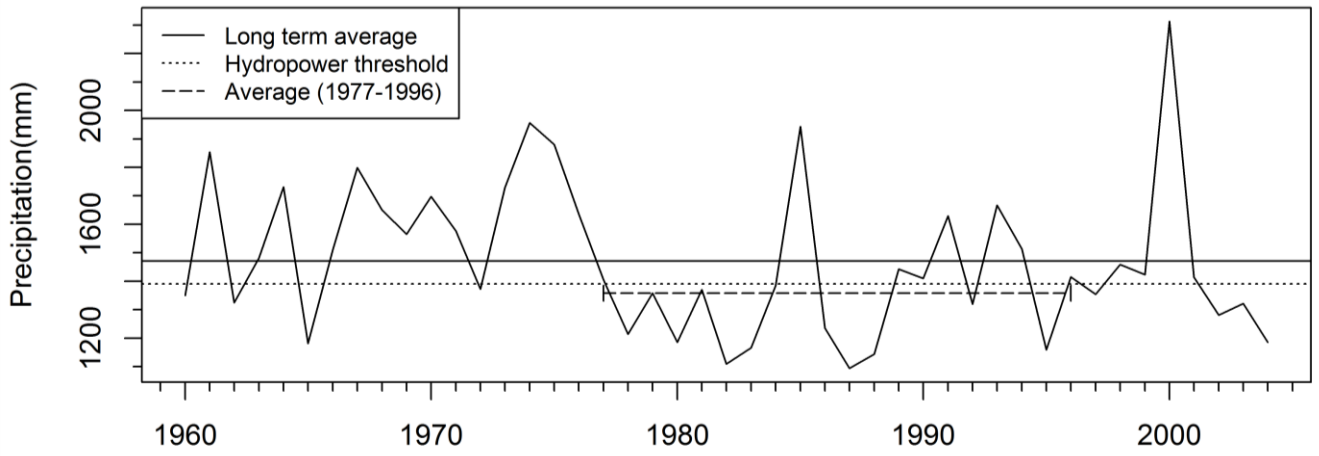

Figure 4.4: Total annual precipitation (Gilgel Abbay). Horizontal lines indicate the thresholds for annual average precipitation identified by the PRIM algorithm.

Based on these results, the insights that can be obtained through a process like scenario discovery appear to be compromised when multiple performance objectives are combined into a single score. While the uncertain parameters driving vulnerability were relatively consistent across the scenarios identified for different metrics and aggregation schemes, the more subtle ways in which uncertain parameters interact with each other to impact different objectives were not always apparent when the aggregated scores were used. It is also important to note that performance across the objectives in our example were relatively correlated, since a low availability of water impacts all of the objectives negatively. It is quite possible that the discrepancies in failure scenarios for aggregated metrics would be larger if other objectives were included that were impacted in the opposite direction, such as flood risk. Regardless of the aggregation method used, the information provided to decision makers using aggregated criteria cannot match the information provided through assessment of criteria individually. While we specifically evaluated the impact of aggregating objectives through multiplicative and additive utility functions, this result is likely to also occur when other methods, such as conversion of metrics to monetary flows through cost benefit analysis, are used. Admittedly, performing a separate scenario discovery on each of our metrics was made easier in our example problem due to the relatively small number of performance metrics assessed, and repeating this process may become increasingly impractical as the number of objectives under consideration increases, as may be the case in participatory processes involving many stakeholders. One potentially promising way to address this issue could be by identifying groups of objectives that are vulnerable to similar conditions and grouping them together so that failure for performance objectives are

described by the same scenarios. This would likely reduce the coverage and density of the failure scenarios for some objectives, but would make the evaluation's results more interpretable and avoid the need to weigh and aggregate the objectives of competing groups early in the analysis.

4.5 Conclusions

Robust decision frameworks are becoming increasingly popular in both research and practice, particularly in the climate adaptation field. By identifying the conditions to which a given system or policy is vulnerable, these tools can provide valuable insights in situations with multiple deeply uncertain parameters that could impact the system of interest. These methods are increasingly being applied in sectors that have to balance performance across multiple criteria, such as water resource management, infrastructure protection, and energy policy. This research demonstrates that common methods used to aggregate multiple criteria into a single utility score can lead to inconsistent failure scenarios and obscure the relationship between key uncertainties and system performance. Applying scenario discovery over each performance metric separately provides more nuanced information regarding the relative sensitivity of the performance objectives and the ways in which they are impacted by different uncertain parameters. This in turn can provide insights on measures that could be taken to improve system robustness, as well as areas where additional research might prove useful. Because the RDM framework was designed to provide quantitative decision support in contexts where there may be conflicting beliefs about what the future will look and contentious disagreements about the best course of action, it is important that the steps of the process remain as transparent as possible. To this end, the additional effort required to apply scenario discovery to each metric separately provides valuable benefits by identifying

failure scenarios that inform a more complete picture of system performance and provide more detailed guidance for vulnerability-reduction efforts.

5 ROBUST DECISION MAKING IN DATA SCARCE CONTEXTS: ADDRESSING DATA AND MODEL LIMITATIONS FOR INFRASTRUCTURE PLANNING UNDER TRANSIENT CLIMATE CHANGE[5]

5.1 Introduction

Water resources is one of the sectors expected to be most vulnerable to climate change, and there is increasing concern that water systems require some degree of adaptation now to avoid severe consequences due to climate change in the future (Field et al., 2014; Melillo et al., 2014). This is particularly true when considering hard infrastructure such as reservoirs, canals, and levees, as these expensive systems generally have operating lives that will stretch decades into the future. However, incorporating climate change into infrastructure planning today is hindered by the considerable uncertainty surrounding general circulation model (GCM) projections of future rainfall. GCM projections disagree about even the direction of changes in precipitation in many regions of the world, and are notably limited in their ability to reproduce observed hydrologic climatology at regional levels (Kundzewicz and Stakhiv, 2010). Using these projections as inputs to regional climate models and hydrologic models results in a "cascade of uncertainty", in which uncertainties at each stage of the modeling process influence outcomes at subsequent levels (Mitchell and Hulme, 1999; Wilby and Dessai, 2010). While efforts have been made to represent climate change uncertainty probabilistically using multi-model ensembles, these distributions can be highly sensitive to the assumptions and methodology used (Tebaldi et al., 2005; Tebaldi and

[5] This chapter is based on the manuscript "Robust decision making in data scarce contexts: addressing data and model limitations for infrastructure planning under transient climate change" submitted to *Climatic Change* in March 2016. Currently under first round of review.

Knutti, 2007), and there is not yet a clear consensus on the physical and statistical interpretation of MME projections (Stephenson et al., 2012). In situations where probabilistic climate projections have been developed, their use in adaptation decision making has been hindered by their complexity and inability to represent variables most important for adaptation planning, such as extreme events (Tang and Dessai, 2012). These issues have led many to conclude that climate change is an example of "deep uncertainty" where probabilistic models of uncertainty cannot be confidently determined or agreed upon (Cox, 2012), and prompted a number of water managers to argue that climate change projections are not yet a suitable basis for infrastructure design and planning (Kundzewicz and Stakhiv, 2010).

These challenges impact infrastructure planning efforts the world over, but they are particularly acute in the developing world. While the construction of large water storage and transfer projects has slowed in developed countries due to environmental concerns and an increased reliance on managerial and conservation-based approaches to water management (Gleick, 2000), expanding hard infrastructure is still viewed as an essential path to economic growth in many developing countries. For instance, the African Development Bank's Programme for Infrastructure Development in Africa currently calls for expanding hydropower capacity across major river basins in the country by approximately 600% and irrigation capacity by up to 700% in some basins (Cervigni et al., 2015). It is estimated that dam development in the Organization for Economic Co-operation and Development (OECD) countries has already reached 70% of economically feasible potential, but Africa has only exploited 10% of this potential (Wang et al., 2013). As developing countries are consistently recognized to be amongst the most vulnerable to climate change (Field et al., 2014), there is understandable interest in improving climate change resilience through

methods such increased water storage and irrigation capacity. However, the overwhelming majority of climate research is undertaken in and focuses on developed countries (Washington et al., 2006), and GCM disagreement in terms of the direction of future precipitation changes is particularly prevalent in tropical regions of the world (van Oldenborgh et al., 2013). Attempts to address this lack of consensus by comparing GCM performance to identify the "best models" for a given region can be highly sensitive to the method used for evaluation and may not always be an effective way to reduce uncertainty (Bhattacharjee and Zaitchik, 2015). Finally, the use of ensemble projections such as those compiled through the Coupled-Model Intercomparison Project (CMIP) for adaptation planning in developing countries is likely to be hindered by the limited availability of computing facilities and expertise needed to process these large, complex datasets (McSweeney et al., 2010).

Even without considering the impacts of climate change, data limitations in many developing countries hinder the ability to confidently evaluate the future performance of planned water infrastructure projects. The density of hydrological monitoring networks is low and even decreasing in many countries, and issues of missing and unreliable data plague many stations (Hughes, 2006). While satellite data and modeled reanalysis products can help fill these gaps, their accuracy in tropical regions of the world is not always well demonstrated, particularly over the seasonal and geographic scales most relevant for planning purposes (Hughes, 2006; Poccard et al., 2000). Limited data on other factors such as soil and land cover conditions can also lead to considerable uncertainty in the parameterization of hydrologic models used to project the impacts of infrastructure development and climate change on basin-level hydrology (van Griensven et al., 2012; Schuol and Abbaspour, 2006). These limitations can result in poor estimates of the actual

benefits that will accrue from the construction of water resource infrastructure projects even over the relatively short term. For instance, the Angereb dam reservoir in Ethiopia was built in 1994 to supply water to the town of Gondar for a period of 25 years, but sedimentation rates have been 50% higher than expected and effectively halved the functional life of the reservoir (Haregeweyn et al., 2012). More broadly, the World Commission on Dams reports that approximately half of large irrigation dams fail to provide irrigation to their planned command areas, with one quarter of projects reaching less than 35% of planned command areas (World Commission on Dams, 2000). When combined with miscalculation of actual project costs, this can result in significantly poorer *ex post* benefit cost ratios than *ex ante* estimates (Flyvbjerg, 2009).

To support infrastructure planning under climatic uncertainty, there has been increasing interest in "robust decision frameworks" (Weaver et al., 2013) to support infrastructure planning in the face of non-probabilistic uncertainty. These frameworks distinguish themselves from traditional "predict-then-act" frameworks in two ways. The first is that they aim to identify strategies that are robust, or that perform well over many possible conditions that may be encountered, rather than strategies that are optimal for a specific set of assumed conditions. The second is that they do not focus on predicting what future conditions may be, but instead focus on identifying conditions that cause the system of interest to fail (Weaver et al., 2013). A number of novel methodologies fall into this general family, including robust decision making (RDM; Lempert et al., 2006), decision scaling (Brown et al., 2012), and info-gap decision theory (Ben-Haim, 2000). In addition to providing decision support under deep uncertainty, they can also be useful in situations characterized by poorly understood nonlinear or threshold responses (Lempert and Collins,

2007) or many stakeholders with conflicting values and beliefs about the future (Hallegatte and Rentschler, 2015).

It has been pointed out that the comprehensive manner in which robust decision frameworks address uncertainty may be particularly well-suited to developing country contexts (Lempert et al., 2013). However, the majority of applications of robust decision frameworks to date have been in developed countries and often rely on sophisticated simulation models and projections that might not be available to planners in the developing world. For instance, applications of robust decision making to water planning problems in the United States have generally taken advantage of sophisticated water system simulation models forced by downscaled GCM projections that had already been developed for the region under evaluation (Fischbach et al., 2015; Groves et al., 2013; Groves and Bloom, 2013). Applications in developing countries have generally required dramatic simplifications to this process that could potentially undermine their results. For instance, physically-based, well-established hydrologic models may be replaced by simple relationships or assumptions about how climate conditions impact streamflow (Brown, 2011; Kalra et al., 2015), or uncertain parameters related to climate change may be randomly sampled as independent variables even though they are likely to be dependent on each other (Lempert et al., 2013). While these are understandable simplifications that are likely necessary in poorly-studied, data-scarce regions, understanding the implications they may have on analytical results is crucial if these analyses are to inform expensive investment decisions.

The objective of this paper is to demonstrate a modified application of the robust decision making methodology that is specifically tailored for application in data-scarce situations. Specifically, the approach outlined here makes two contributions that build on

previously conducted RDM studies. The first is an emphasis on characterizing the relative contribution of uncertainty stemming from data limitations and model simplifications relative to uncertain future conditions such as climate change. This can be informative because uncertainty in current conditions may be addressed through additional research and evaluation, whereas (at least for water managers and practitioners) uncertainty surrounding future climate conditions is largely irreducible. Thus, the identification of data-limitations that have the greatest impact on proposed infrastructure developments could provide valuable prioritization of research activities. The second contribution is a novel method for generating transient climate change sequences that do not rely on GCM projections but account for potential dependencies between uncertain parameters. The advantage of this approach is that managers can see how system vulnerability changes through time, rather than at an arbitrary period some years in the future. This methodology is demonstrated using the Lake Tana basin in Ethiopia (Section 1.5), where multiple water resource infrastructure projects have been recently constructed or proposed to support hydropower and irrigation in the basin. By demonstrating how established methodologies for infrastructure planning under uncertainty can be modified to better suit developing-country contexts, this work aims to ultimately improve the long-term effectiveness and sustainability of infrastructure investments in these countries.

5.2 Methods

5.2.1 Simulation Model

An integrated simulation model was used to assess how changes in climatic and environmental conditions would impact water resources in the basin (Figure 5.1). The first component consisted of empirical rainfall-runoff models developed by Shortridge et al. (2016) that predicted monthly streamflow in each of the five rivers with proposed reservoirs

as a function of climate conditions and agricultural land cover. Monthly streamflow data was taken from historic stream gauge records for each basin (Alemayehu et al., 2010). In each river, the 40-year monthly streamflow series was regressed against historic monthly climate data taken from University of East Anglia Climate Research Unit (CRU) TS3.10 gridded meterological fields (Harris et al., 2014) and the percentage of agricultural land cover in the basin estimated from historic aerial photographs and satellite images reported by Rientjes et al. (2011), Gebrehiwot et al. (2010), and Garede and Minale (2014). To account for the strong seasonality in the region's climate and hydrology, monthly flow anomalies and climate anomalies were used as the response variable and predictor variables, respectively, using the anomaly formulation discussed in Chapter 2. Flow anomalies in each basin were then regressed against climate anomalies and the percentage of agricultural land cover using a Gaussian regression model. While more sophisticated machine-learning models resulted in slightly lower errors than the linear model in some of the basins assessed, the simpler linear models were chosen for this evaluation for ease of interpretation and because it allowed for a consistent model formulation and measure of model uncertainty to be used in each river basin. Nevertheless, there are some aspects of hydrologic behavior in the region that are lost through the use of a linear model, such as non-linear relationships between agricultural land cover and streamflow anomalies that were observed in some basins (Shortridge et al., 2016).

Evaporation off of Lake Tana and each of the proposed reservoirs was calculated using Penman's equation (Penman, 1948). These estimates used the adjusted temperature values reflective of climate change and historic monthly average values for wind speed, relative humidity and solar radiation from the Bahir Dar meteorological station as reported by Kebede et al. (2006). To estimate the sediment load to each proposed reservoir, sediment rating curves developed by Guzman et al. (2013) for the Ethiopian highlands were used.

Suspended sediment concentrations in the region have been observed to exhibit a seasonal shift as the rainy season progresses, with high sediment concentrations early in the season and lower concentrations at the end (Guzman et al., 2013). To account for this shift, as well as the observed relationship between agricultural land cover and sediment concentrations, sediment rating curves took the form:

Equation 5.1:

$$C_W = A_C a R^b$$

where C_W is the suspended sediment concentration in kg/m^3, A_C is the fractional cropland area in the watershed (ranging from 0.0 to 1.0), R is the runoff volume in mm/day, and a and b are regression parameters fit based on data collected from multiple basins in the Ethiopian highlands (Guzman et al., 2013). The two regression parameters a and b varied depending on the cumulative rainfall that had fallen by that point in the rainy season, as in Table 5.1. Assuming that runoff accounted for 40% of rainy season streamflow (Easton et al., 2010), this resulted in average annual specific sediment yields that ranged from 853 to 2072 tons/km^2 across the five rivers assessed. This appears comparable to the range used for the reservoir feasibility studies (289 to 2037 tons/km^2), and thus suggests that this method provides a conservative but reasonable estimate of the sediment loads that these reservoirs could experience given the limited suspended sediment data available in the watersheds above the proposed reservoirs.

Period	Cumulative Rainfall (mm)	a	b
Early Rainy Season	< 150	75	0.45
Mid Rainy Season	150 – 700	13	0.4
Late Rainy Season	> 700	9	0.4

Table 5.1: Parameters of seasonal sediment rating curve equations

The final component of the simulation model was a Water Evaluation and Planning (WEAP; Sieber and Purkey, 2015) water allocation model developed for the basin by Alemayehu et al. (2010). This model simulates natural hydrologic processes such as streamflow and evaporation, as well as human extraction and use of water. In each month, the model performs a mass balance to account for both extraction and inflows, allocating water to different demand nodes in order of user-defined priorities (Sieber and Purkey, 2015). The monthly sequences of streamflow and sediment load in each river, as well as sequences of evaporation from the lake and each reservoir, were used as model inputs. The model then calculated the amount of water allocated and coverage (percent of demand delivered) for different demand nodes, as well as lake elevation and downstream flows. Additional information on WEAP model development, calibration and validation is discussed by Alemayehu et al. (2010).

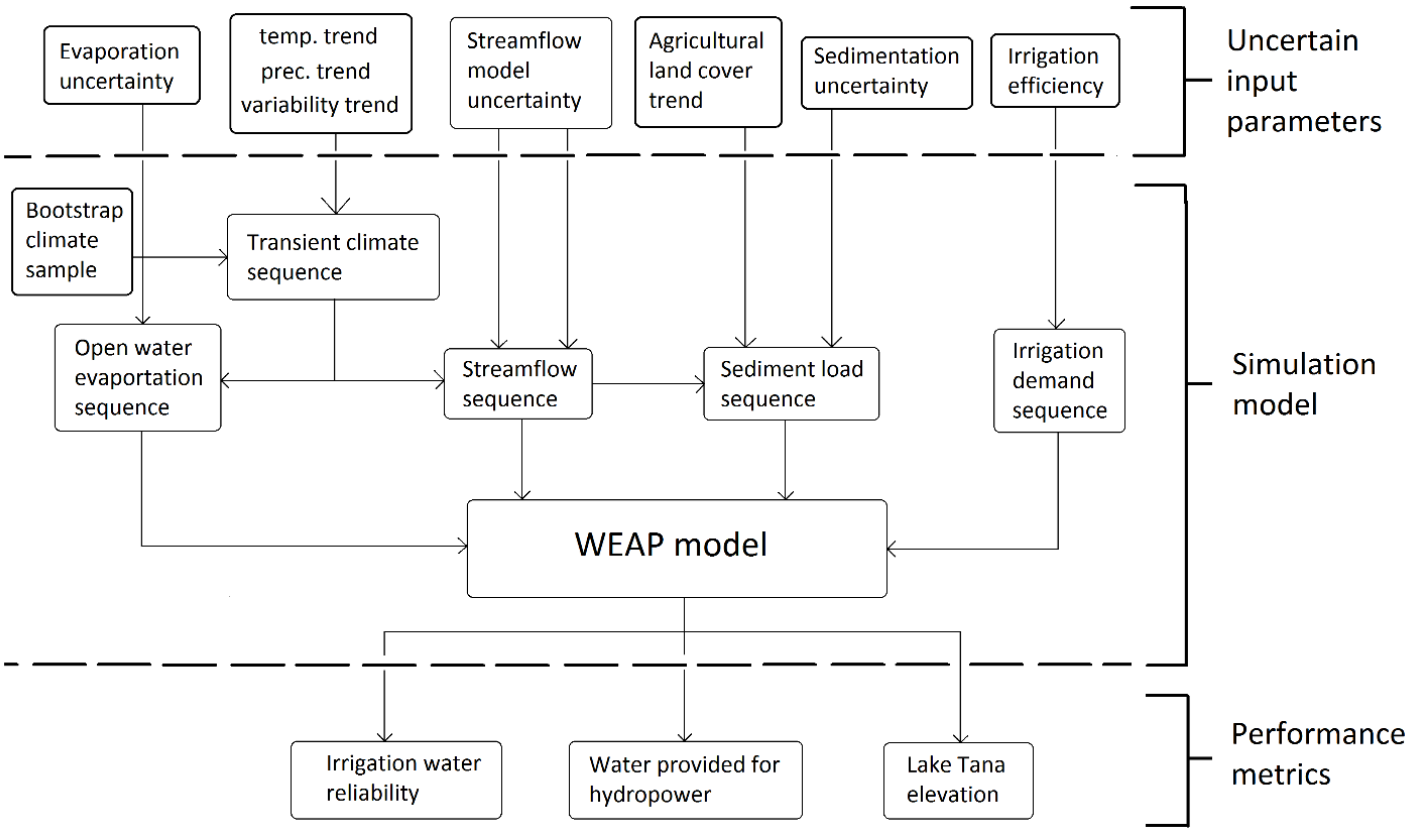

Figure 5.1: Simulation Framework

5.2.2 RDM Evaluation

5.2.2.1 Development Alternatives

The RDM evaluation initially considered the three different development levels based on current plans for infrastructure development in the region that are shown in Table 1.1 (Achenef et al., 2013). Development level 0 (D0) consists only of existing infrastructure in the basin, including the Koga River irrigation reservoir and Tana-Beles hydropower transfer tunnel. Development Level 1 (D1) consists of existing infrastructure as well as four additional irrigation reservoirs and two pumped irrigation schemes. Development level 2 (D2) consists of all of the projects included in Development Level 1 as well as the Gilgel Abbay and Jema reservoirs and a pumped irrigation scheme from the southwestern portion of the lake. This development level would entail full construction of all proposed water-resource infrastructure projects in the basin.

Additionally, three modifications to Development Level 2 were also assessed to better understand the degree to which measures aimed at improved efficiency and land conservation might impact water resources management in the basin more broadly. In all of these instances, the quantitative impact that each modification would have on streamflow and sediment loading rates in the basin were estimated based on field studies and evaluations conducted in the Blue Nile highlands. However, these estimates are highly uncertain, given the heteregenous conditions across the basin and the potential for impacts to vary when implemented on a watershed level rather than a field or plot-scale. For this reason, the results should be interpreted as an initial estimate of the impact that the measures could have

on water resources in the basin more broadly if the quantitative impacts described here were achieved.

The first modification measure (referred to as M1) was lined irrigation canals. The five proposed reservoirs for the basin and the existing Koga reservoir are all designed to utilize unlined canals to convey water from the reservoir to the irrigation areas. Five of the six reservoirs were designed assuming a conveyance efficiency of 81%, with one reservoir (Gumara) assuming a conveyance efficiency of 62.5%. However, evaluation of existing irrigation projects in the country suggest that actual conveyance efficiencies could be much less than this (Awulachew and Ayana, 2011). To evaluate the impact of installing lined canals at the time of reservoir construction, this modification assumed that irrigation demands from each reservoir would be 10% lower than the D2 development level for the full period of analysis.

The second modification measure (M2) consisted of improved on-farm irrigation efficiency. The feasibility studies for the proposed irrigation systems assume that gravity-fed furrow or paddy irrigation will be used, resulting in field efficiencies of 60% to 80%. However, evaluations of other irrigation schemes in Ethiopia have often found that field efficiencies in practice can be closer to 50%. The adoption of more sophisticated methods for field irrigation, either through improved timing and management or through the use of pressurized irrigation, could reduce losses to runoff and deep percolation by an up to 30% (Bekele and Tilahun, 2006; WWDSE and Tahal Group, 2009a, 2009c). While these methods and technologies are likely to be impractical in the short-term due to limitations in financial capital and technical capacities, they could be progressively promoted and implemented through time as economic conditions improve. Thus, this modification assumes that

efficiency improvements gradually reduce total irrigation demands by 33% over the 50-year period of analysis.

The third modification measure (M3) consisted of upstream soil and water conservation (SWC) measures. The expansion of agriculture in the basin, particularly on marginal lands, has resulted in extensive land degradation over the past few decades. This can degrade soil quality in farmed areas, reducing agricultural productivity, and harm downstream water quality by increasing rates of runoff and erosion. To combat these impacts, a number of agricultural practices aimed at improved soil and water conservation are being promoted throughout the region. These include measures such as the installation of soil bunds and terracing, gully treatment to reduce erosion, and the promotion of perennial crop production in marginal, sloped terrains (Simane et al., 2012). While the impact that these measures could have on a landscape scale are highly uncertain, field-scale studies suggest that their installation could reduce sediment loads to surface water by up to 50% (Adimassu et al., 2014; Tamene and Vlek, 2007) and reductions in runoff of up to 30% (Adimassu et al., 2014). Because the implementation and effectiveness of these measures is likely to increase through time, this modification assumes that the promotion of these methods will result in a gradual reduction in sediment loads to 50% of their current levels, and a gradual reduction in streamflow to 88% of its current level (based on the assumption that runoff contributes 40% of streamflow and the SWC measures reduce runoff by 30%).

5.2.2.2 Representation of uncertainty in model simulations, future conditions, and system operation

There are a number of uncertainties that could impact future performance of the proposed infrastructure, not only related to unknown future conditions, but also to system operation and simulation model simplifications. However, most RDM evaluations to date

have focused primarily on uncertainty related to future conditions, with little consideration of the impact that data limitations and model simplifications could have on infrastructure performance. To understand how these different sources of uncertainty could impact system performance, a range of feasible values was identified for eight uncertain parameters related to model limitations, future conditions, and infrastructure operation efficiency. Because the main objective of the analysis is to find conditions that result in unsatisfactory performance of the infrastructure, we used wide ranges of values to better identify the thresholds that would result in poor performance. The uncertain parameters evaluated are presented in Table 5.2 and described in more detail below.

Source of Uncertainty	**Uncertain parameter**	**Symbol**	**Range of values**
Simulation model	Streamflow model prediction interval	Qmod	0.05 to 0.95
	Evaporation coefficient	EtC	0.8 to 1.2
	Sedimentation coefficient	SedC	0.7 to 1.3
Future conditions	Change in temperature	ΔT	0.5 to 5.5° C
	Change in rainfall	ΔP	-20% to +35%
	Change in interannual variability	ΔVar	0% to +20%
	Agricultural land cover	AgLC	-1% to +1% per year
System operation	Irrigation efficiency	IrrEf	30% to 64%

Table 5.2: Uncertain input parameters

As discussed above, data limitations in the basin hinder the development of hydrologic models needed to relate climate and land cover conditions with outcomes such as streamflow and sediment loads. Because of this, the simulation of hydrologic processes such as streamflow generation, sediment loading and evaporation are highly simplified and subject to considerable uncertainty. Three parameters were included to account for uncertainty in the degree to which the proposed simulation model can accurately represent the system response to changing climate and land cover conditions. The Qmod parameter, which

ranged from 0.05 to 0.95, was used to incorporate uncertainty in the empirical streamflow models described in Chapter 2. The regression models used to predict streamflow as a function of climate and agricultural land cover are clearly very simple representations of the region's hydrology and not able to capture system behavior with complete accuracy. As in any regression model, the fact that the model is built using only a sample of historic data (the period from 1960 to 2004) rather than the basin's full history, as well as the limited number of explanatory variables included, introduces uncertainty and error. To account for this, the simulated streamflow for each simulation was taken from a prediction interval encompassing the 5^{th} to 95^{th} percentile of predicted values for each new observation in accordance with the randomly sampled Qmod parameter.

Similarly, the equations used to estimate evaporation off of the lake and reservoirs, as well as sediment loads to the reservoirs, are based on limited data and fairly simplistic representation of complex landscape-scale processes. Because of this, they are likely to only provide a rough estimate of the true value of evaporative losses and sediment loads that will occur. For this reason, two coefficients (EtC and SedC) were introduced to estimate how actual evaporation and sedimentation rates could compare to the estimates. Monthly evaporation sequences were multiplied by the parameter EtC, which ranged from 0.8 to 1.2, prior to being used as inputs to the WEAP model. This parameter represents the degree to which actual evaporation might differ from our estimate, with any value over 1.0 implying that actual evaporation is higher than our estimates. Similarly, sediment loading rates to each reservoir were multiplied by a SedC parameter ranging from 0.7 to 1.3, with any value greater than 1.0 indicating sediment loads are higher than those estimated using the rating curve in Equation 5.1.

Possible impacts of climate change on average temperature for the region were represented by a change in temperature ranging from 0.5 to 5.5° C based on IPCC multi-model ensemble projections for the East Africa region for the period 2081-2100 under all representative concentration pathways (van Oldenborgh et al., 2013). These same ensembles project changes in annual precipitation ranging from -20% to positive 35%. Furthermore, there is also concern that climate change could result in increasing interannual variability, regardless of changes in average annual rainfall (Barnett et al., 2006). To account for this, climate change was assumed to have the potential to increase interannual variability in rainfall by up to 20%. In addition to future climate conditions, changing land cover in the region and uncertain levels of irrigation efficiency could also impact system performance in the future. The possible change in the percentage of agricultural land cover in each basin was assumed to range from -1% to +1%, and continued to increase or decrease at this rate until a minimum value of 35% or a maximum value of 90% was reached. An increase of 1% per year would be roughly equivalent to a continuation of the maximum rate of agricultural land use expansion observed over the past 40 years in the basin, whereas a decrease of 1% per year would be a complete reversal of this trend.

Finally, overall irrigation efficiency (the product of conveyance efficiency and field efficiency) was assumed to be uncertain. Feasibility studies prepared for the basin's reservoirs each assume a specific value for overall irrigation efficiency ranging from 50% to 57%. However, evaluations of existing irrigation projects in the country suggest that conveyance efficiency in unlined canals could be as low as 60%, and field efficiency for gravity-fed irrigation could be as low as 50%, which would result in an overall efficiency of approximately 30% (Awulachew and Ayana, 2011). In a best-case scenario, a conveyance efficiency of 80% combined with a field efficiency of 80% would result in an overall

efficiency of 64%. Thus, the evaluation assumed that actual baseline irrigation efficiency for the projects (without implementation of the modification measures described in Section 5.2.2.1) could range from 30% to 64%.

5.2.2.3 Generation of transient climate sequences

One limitation with many climate change impact assessments is that they estimate the severity of impacts at a discrete period in the future (for example, at the end of this century) rather than estimating transient impacts through time. Understanding how the severity of climate change impacts develops through time provides more valuable information for adaptation decision making, but has generally relied on using downscaled climate projections that can be difficult to obtain and highly uncertain in developing regions of the world. To provide insights into how gradual changes in climate conditions would impact system performance through time, transient climate change sequences were developed based on the range of feasible 50-year changes in climate shown in Table 5.2.

Transient climate sequences were developed first by generating correlated samples of changes in temperature, precipitation, and interannual variability because greater climate forcing is expected to result in more extreme changes to both temperature and precipitation. For each of the samples used to drive the simulation model, a value for the change in temperature was selected from a uniform distribution across the range of values presented in Table 5.2 and the change in precipitation was randomly selected to be either positive or negative with an equal probability. For the *n*'th sample, a parameter $\overline{\Delta P_n}$ was calculated as in Equation 5.2 and a parameter $\overline{\Delta Var_n}$ was calculated as in Equation 5.3, with the maximum and minimum values for ΔT, ΔP, and ΔVar as presented in Table 5.2. The change in rainfall and rainfall variability for sample *n* were then randomly sampled from normal distributions

with means equal to $\overline{\Delta P_n}$ and $\overline{\Delta Var_n}$ and a coefficient of variation equal to 0.5. This resulted in correlated samples of the ΔT, ΔP, and ΔVar parameters representing possible changes in temperature, precipitation, and interannual variability that could occur by 50 years in the future.

Equation 5.2:

$$\overline{\Delta P_n} = \begin{cases} \dfrac{(\Delta T_n - \Delta T_{min})}{(\Delta T_{max} - \Delta T_{min})} \times \Delta P_{max} & if\ \Delta P > 0 \\ \dfrac{(\Delta T_n - \Delta T_{min})}{(\Delta T_{max} - \Delta T_{min})} \times \Delta P_{min} & if\ \Delta P < 0 \end{cases}$$

Equation 5.3:

$$\overline{\Delta Var_n} = \frac{(\Delta T_n - \Delta T_{min})}{(\Delta T_{max} - \Delta T_{min})} \times \Delta Var_{max}$$

To account for gradual changes in climatic conditions, these parameters were then used to generate 50-year sequences of climate change perturbations with each parameter increasing linearly through time. For example, if the ΔT parameter was equal to 5°C for a given sample, then the ΔT perturbation sequence for that sample would gradually increase by 0.1°C per year from 0°C to 5°C by the end of the 50-year sequence. For each sample, a two-year blocked bootstrap sample of historic climate conditions was used to generate a random 50-year sequence of climate conditions that accounted for inherent climate variability while maintaining the moderate degree of autocorrelation that is observed in annual climate conditions. The bootstrapped climate sequences were then adjusted by these

perturbation sequences to account for gradually changing climatic conditions through time. The ΔT and ΔP perturbation sequences were applied using the delta-change method (Gleick, 1986). To account for increasing interannual variability, each year of the bootstrapped climate was classified as a wet or dry year based on whether total rainfall in that year was greater or less than the long-term annual average. In wet years, the monthly precipitation values were increased by the percentage in the perturbation sequence, while monthly precipitation values in dry years were decreased by this percentage.

5.2.2.4 Simulation and Evaluation Procedure

The simulation model described above was used to assess how the proposed infrastructure development alternatives and modifications performed in a wide variety of "potential futures" (Lempert et al., 2006) represented by combinations of the eight uncertain parameters described in Table 5.2. In the first step of the simulation procedure, 10,000 random samples were generated using Latin Hypercube sampling across a uniform distribution for the range of possible values for change in temperature (ΔT), agricultural land cover (AgLC), irrigation efficiency (IrrEf), sediment coefficient (SedC), evaporation coefficient (EtC), and streamflow prediction model interval (Qmod). For each sample, correlated values of change in precipitation (ΔP) and change in interannual variability (ΔVar) were calculated based on the randomly sampled ΔT value and Equations 5.2 and 5.3. While Latin hypercube sampling is often used to generate multivariate probabilistic distributions, here it is only used as a mechanism for generating a diverse sample of future conditions that could feasibly occur. These samples are used as input for exploratory modeling (Bankes, 1993) that evaluates how the system responds to different multivariate conditions while making no inference regarding the likelihood of those states. Other methods for sample generation, including full combinatorial sampling across discrete uncertain parameters and

GCM ensemble projections (Groves et al., 2013; McJeon et al., 2011), have been used in RDM evaluations and the application of further sample generation methods could be a valuable area for additional research.

The ΔT, ΔP, and ΔVar parameters were then used to develop transient climate change sequences as described in Section 5.2.2.3. These climate sequences were then used as inputs to the empirical streamflow prediction models to generate a distribution of predicted streamflow values for each month in the 50-year sequence. The parameter Qmod was used to select a percentile value from this distribution to be used as the predicted streamflow value; for instance, if Qmod equaled 0.95, then the 95th percentile from the prediction interval was used, whereas a Qmod value of 0.5 would entail that the median value from the prediction interval was used. These sequences of monthly streamflow were then used to generate a sequence of sediment loading to each reservoir using Equation 5.1, which was multiplied by the SedC parameter to account for uncertainty in the sediment rating equations. Monthly sequences of evaporation from Lake Tana and the proposed reservoirs were calculated using Penman's equation (Penman, 1948) and the perturbed temperature sequences and then multiplied by the EtC parameter.

These 50-year sequences of monthly streamflow, sedimentation loads, and evaporative losses were then used as inputs into the WEAP simulation model. The WEAP model calculated three performance metrics for each simulation as shown in Table 5.3. The first performance metric was agricultural coverage, which is defined as the percentage of years where 90% of agricultural demand across the basin was provided. The second performance metric was the average annual water provided for the Tana-Beles hydropower transfer. The third performance metric was the percentage of months where the water

elevation in Lake Tana was above 1784.75 meters above mean sea level (amsl). Previous studies have identified this elevation as the minimum level that the lake can reach before negative impacts to the navigation and fishing industries begin to occur (SMEC, 2008); this metric therefore can be considered a proxy for the impacts to the navigation sector as well as environmental conditions in the lake. For each performance metric, an acceptable performance level was identified by reviewing feasibility studies for the proposed projects and evaluating system performance under historic climate conditions (Shortridge and Guikema, 2016).

Objective	Metric	Units	Acceptable Performance Threshold
Maximize irrigation water reliability	Percentage of years when minimum demand is met	%	90%
Maximize hydropower water delivered	Average water delivered annually	MCM	2681
Minimize percent of time where lake elevation is below minimum acceptable level	Percent of months where lake is above 1784.75 amsl	%	95%

Table 5.3: Performance metrics

The robustness of each alternative is characterized in two ways. The first is by assessing the percentage of simulations where the alternative was able to achieve the acceptable performance threshold for an individual performance metric. This is consistent with the "satisficing" criterion employed by a number of prior RDM evaluations (for instance, Kalra et al., 2015; Lempert et al., 2013; Lempert and Groves, 2010). This metric was calculated over three time periods to demonstrate how system performance changes through time: early (years 1-15), all (years 1-50), and late (years 36-50). However, this measurement is sensitive to the distribution assumed in generating samples of uncertain input parameters, and could be contentious in situations of deep uncertainty where there is

disagreement or uncertainty surrounding these distributions (Whateley et al., 2014). Therefore, the robustness of each alternative was also characterized by evaluating the specific regions of the input variable space where an alternative was unlikely to meet the performance thresholds described in Table 5.3. These regions, referred to as failure scenarios in the remaining discussion, describe the specific conditions that are likely to result in unacceptable performance. This is similar to the robustness metric proposed by Whateley et al. (2014); however, we avoid quantifying the portion of the input variable space described by each failure scenario for two reasons. The first is that this quantification is sensitive to range of values assessed for each uncertain parameter, and there may not always be a strong basis for defining these endpoints for all parameters assessed. The second is that in an evaluation with multiple uncertain parameters, understanding the specific parameters and thresholds that define the failure scenarios is likely to be informative for decision makers, and characterizing robustness using a single numerical metric will not convey this information. The robustness of each alternative is evaluated over each performance metric separately to avoid biases or loss of information that can result from aggregation of multiple criteria in robust decision approaches (Shortridge and Guikema, 2016). The failure scenarios were determined using the RDM scenario discovery process in the SDtoolkit package in R (Bryant, 2014), which is based on the patient rule induction method (PRIM) bump hunting algorithm (Friedman and Fisher, 1999) described in Section 1.4.

5.3 Results

The percentage of simulations in which each performance metric was satisfied for each development alternative over the three time periods of evaluation is presented in Table 5.4. In comparing the three development alternatives currently under consideration by planners in the region (D0, D1, and D2), it is apparent that the performance of D1 and D2

are comparable to each other, but are substantially less robust than D0 with regard to each performance metric. This is not surprising, given that these development alternatives would entail an approximate 7-fold and 10-fold increase in the amount of water withdrawn for agriculture, respectively. However, even though development level D2 requires more irrigation water than D1, it actually results in better performance than D1 across all performance metrics evaluated. This is likely due to the large capacity of the Gilgel Abbay and Jema reservoirs relative to the amount of irrigation water they are designed to provide, allowing for more long-term storage and ability to meet demands in dry periods. The two modification measures aimed at improving irrigation efficiency (M1 and M2) both result in more robust performance in terms of irrigation reliability when compared to the unmodified D2 alternative, with little-to-no impact on hydropower provision or lake levels. The SWC measures (M3) result in slightly poorer performance than the unmodified D2 alternative for each performance metric, due to their impact on streamflow volumes.

Regardless of the alternative evaluated, irrigation water reliability appears highly sensitive to the uncertain parameters present in Table 3, with the minimum performance level being met in only 57% of simulations for D0 and between 8% and 24% of simulations for the other alternatives when evaluated over the full simulation time period. In comparison, the hydropower metric is met in anywhere from 58% to 74% of simulations when assessed over the full time period, and the lake level metric is met in 85% to 93% of simulations. For all metrics, there is a clear decline in performance through time, with the performance thresholds met more frequently in the early years of the simulation (years 1-15) when compared to the full 50-year simulation period and late years (years 36-50). This decline in performance is most dramatic for the irrigation water reliability metric, but is also apparent in the hydropower metric and (to a lesser extent) the lake level metric.

Development alternative	Irrigation water reliability			Water for Hydropower			Lake Levels		
	Early	Full	Late	Early	Full	Late	Early	Full	Late
D0	0.82	0.57	0.49	0.91	0.74	0.73	0.99	0.93	0.91
D1	0.30	0.08	0.06	0.80	0.62	0.63	0.98	0.87	0.87
D2	0.33	0.10	0.07	0.80	0.63	0.64	0.98	0.88	0.87
M1 - D2 with lined canals	0.54	0.20	0.15	0.82	0.66	0.66	0.98	0.89	0.88
M2 – D2 with field efficiency	0.42	0.24	0.22	0.80	0.65	0.67	0.98	0.88	0.88
M3 – D2 with SWC	0.32	0.08	0.05	0.79	0.58	0.57	0.98	0.85	0.84

Table 5.4: Percentage of simulations where performance thresholds were satisfied

Regardless of the alternative evaluated, irrigation water reliability appears highly sensitive to the uncertain parameters present in Table 3, with the minimum performance level being met in only 57% of simulations for D0 and between 8% and 24% of simulations for the other alternatives when evaluated over the full simulation time period. In comparison, the hydropower metric is met in anywhere from 58% to 74% of simulations when assessed over the full time period, and the lake level metric is met in 85% to 93% of simulations. For all metrics, there is a clear decline in performance through time, with the performance thresholds met more frequently in the early years of the simulation (years 1-15) when compared to the full 50-year simulation period and late years (years 36-50). This decline in performance is most dramatic for the irrigation water reliability metric, but is also apparent in the hydropower metric and (to a lesser extent) the lake level metric.

To characterize the conditions that cause unsatisfactory performance for each alternative, the scenario discovery process was used to find failure scenarios for both the irrigation and hydropower metrics (assessed over the full simulation time period) for each development alternative. Failure scenarios for lake level are not presented because each alternative satisfied this metric in the vast majority of simulations conducted, but in practice this metric could be evaluated using the same methodology. Failure scenarios for irrigation reliability are presented in Table 5.5 and shown graphically in Figure 5.2. For each alternative, between two and three failure scenarios were identified. Sometimes these failure scenarios are defined by a single variable; for example, the D0 alternative is unlikely to achieve satisfactory performance if the Qmod parameter is below 0.36 (meaning that actual flows are below the 36th percentile of values predicted by the empirical streamflow models). Failure scenarios can also be defined by the intersection of conditions regarding multiple variables, as is the case for the second failure scenario identified for the D0 alternative,

which is defined by an irrigation efficiency below 45% occurring in conjunction with a Qmod parameter below 0.64. Collectively, these two failure scenarios describe 83% of the simulations where D0 resulted in unacceptable performance with regard to irrigation reliability (coverage), and 74% of the simulations contained within these failure scenarios resulted in unacceptable performance (density).

Across all alternatives, it is apparent that the uncertain parameter with the greatest influence on irrigation water reliability is streamflow modeling uncertainty represented by the Qmod parameter. Although low values of this parameter are sufficient to cause failure for all alternatives evaluated, the specific threshold for failure differs between alternatives. For example the D1 alternative is likely to fail if the value of Qmod is below 0.64, whereas this parameter has to be substantially lower (below 0.52) for the D2 alternative to fail. This is consistent with the satisficing metrics presented in Table 5.4, which suggest that D2 is more robust to uncertainty than D1. Low irrigation efficiency is also sufficient to cause failure for the D1, D2, and M3 alternatives; however, the M1 and M2 alternatives (which were specifically designed to address irrigation losses) are no longer sensitive to this parameter. The other parameter that can cause failure is the change in temperature, suggesting that the D2 alternative will fail if the increase in average temperature over the next fifty years is over 2.84°C. In general, the M1 and M2 modification measures mostly improve the robustness of the D2 alternative by reducing its vulnerability to uncertainty in irrigation efficiency and temperature increases, and but only reduce its vulnerability to model uncertainty by a very slight margin (Qmod < 0.48 rather than 0.52). The M3 modification actually makes the D2 alternative more vulnerable to streamflow model uncertainty, but makes it less sensitive to temperature increases ($\Delta T > 3.34°C$).

Dev. Alternative	Percent failures	Irrigation Failure Scenarios	Box density	Box coverage	Ensemble density	Ensemble coverage
D0	0.43	1) Qmod < 0.36	0.78	0.63	0.74	0.83
		2) IrrEff < 0.45, Qmod < 0.64	0.65	0.2		
D1	0.92	1) Qmod < 0.64	0.98	0.7	0.98	0.92
		2) IrrEff < 0.46	0.99	0.18		
		3) $\Delta T > 4.35$	0.97	0.04		
D2	0.90	1) Qmod < 0.52	0.98	0.27	0.91	0.9
		2) IrrEff < 0.43	0.99	0.09		
		3) $\Delta T > 2.84$	0.97	0.58		
M1	0.80	1) Qmod < 0.48	0.97	0.58	0.96	0.79
		2) $\Delta T > 3.78$	0.92	0.2		
M2	0.76	1) Qmod < 0.48	0.96	0.61	0.94	0.79
		2) $\Delta T > 4.00$	0.88	0.19		
M3	0.92	1) Qmod < 0.64	0.98	0.7	0.98	0.94
		2) IrrEff < 0.44	0.99	0.16		
		3) $\Delta T > 3.34$	0.94	0.09		

Table 5.5: Failure Scenarios for Irrigation Reliability (full time period)

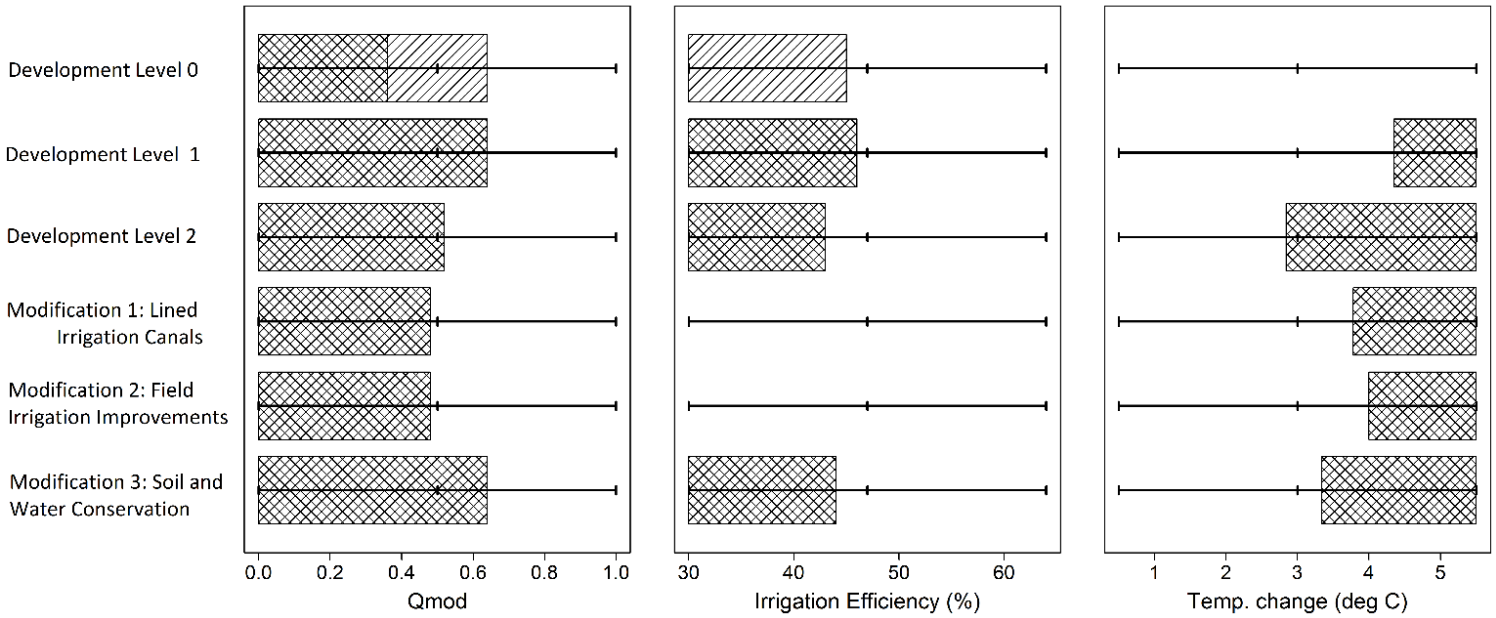

Figure 5.2: Failure scenarios for irrigation reliability. Hash marks indicate parameter values that are sufficient to cause failure regardless of other parameter values. Diagonal lines indicate parameter values that must be combined with certain values of other parameters to cause failure.

Dev. Alternative	Percent failures	Hydropower Failure Scenarios	Box density	Box coverage	Ensemble density	Ensemble coverage
D0	0.26	1) EtC > 1.03, Qmod < 0.52	0.75	0.66	0.83	0.71
		2) EtC > 0.98, ΔP < -9.5	0.6	0.17		
D1	0.38	1) EtC > 0.99, Qmod < 0.64	0.77	0.72	0.75	0.84
		2) EtC > 0.94, ΔP < -9.4	0.65	0.12		
D2	0.37	1) EtC > 0.99, Qmod < 0.64	0.76	0.73	0.83	0.75
		2) EtC > 0.94, ΔP < -10.9	0.65	0.11		
M3	0.34	1) EtC > 1.01, Qmod< 0.64	0.78	0.71	0.76	0.83
		2) EtC > 0.95, ΔP < -10.8	0.65	0.12		
M4	0.35	1) EtC > 0.99, Qmod < 0.64	0.74	0.75	0.73	0.85
		2) EtC > 0.93, ΔP < -12.1	0.65	0.1		
M5	0.42	1) EtC > 0.99, Qmod < 0.71	0.79	0.73	0.77	0.84
		2) EtC > 0.92, ΔP < -9.7	0.68	0.11		

Table 5.6: Failure Scenarios for Hydropower Water Provision (full time period)

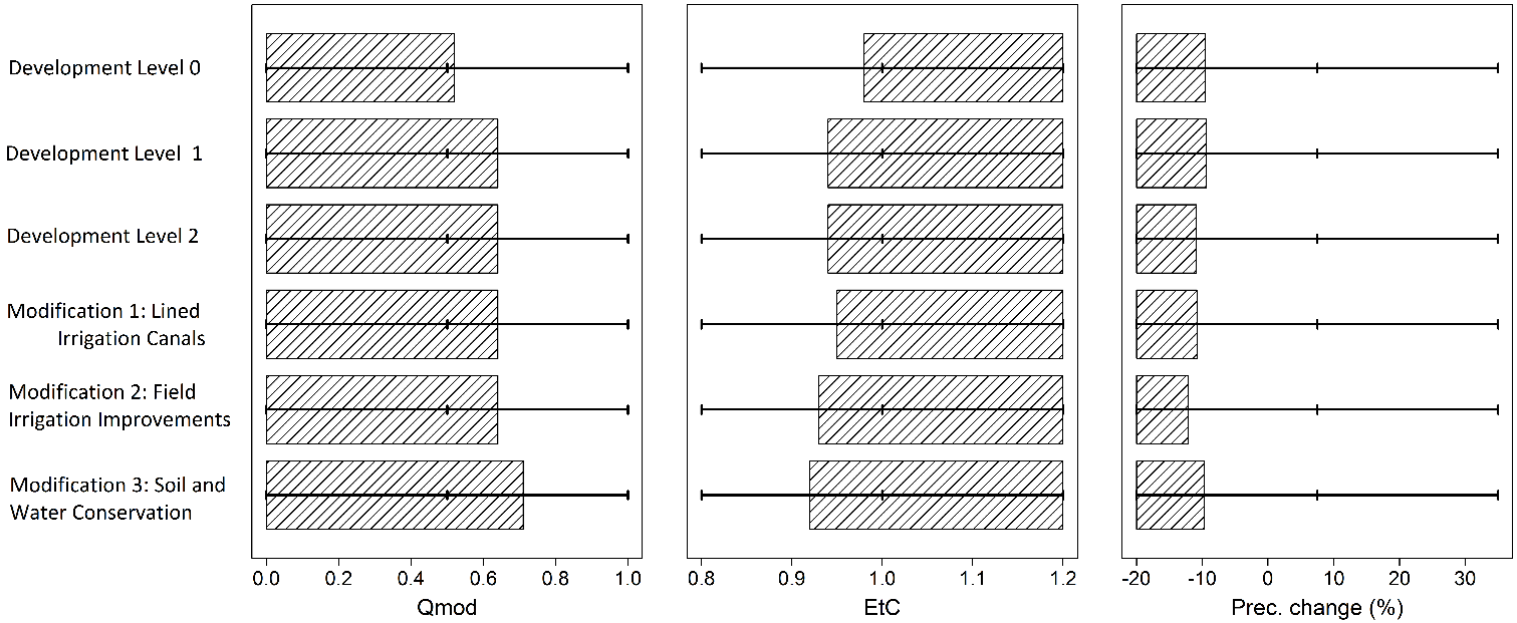

Figure 5.3: Failure scenarios for provision of water for hydropower. Hash marks indicate parameter values that are sufficient to cause failure regardless of other parameter values. Diagonal lines indicate parameter values that must be combined with certain values of other parameters to cause failure.

Failure scenarios for hydropower water provision are presented in Table 5.6 and shown graphically in Figure 5.3. In this case, uncertainty in evaporation estimates off of the lake and reservoirs appears to be the most influential uncertain parameter, appearing in every failure scenario identified. However, high values of EtC (which imply actual evaporation above current estimates) are not enough to cause failure on their own for any alternative evaluated, but have to occur in conjunction with either low values of Qmod or strong decreases in precipitation. For instance, the D2 alternative tends to provide insufficient water for hydropower if EtC is greater than 0.99 and Qmod is less than 0.64 (implying that actual evaporation is greater or equal to current estimates and actual streamflow is below the 64th percentile of model predictions), or if EtC is greater than 0.94 and precipitation decreases by over 10.9% in 50 years. With the exception of D0, which appears less sensitive to the Qmod and EtC parameters, there is only minor variation between the thresholds identified for the other development alternatives.

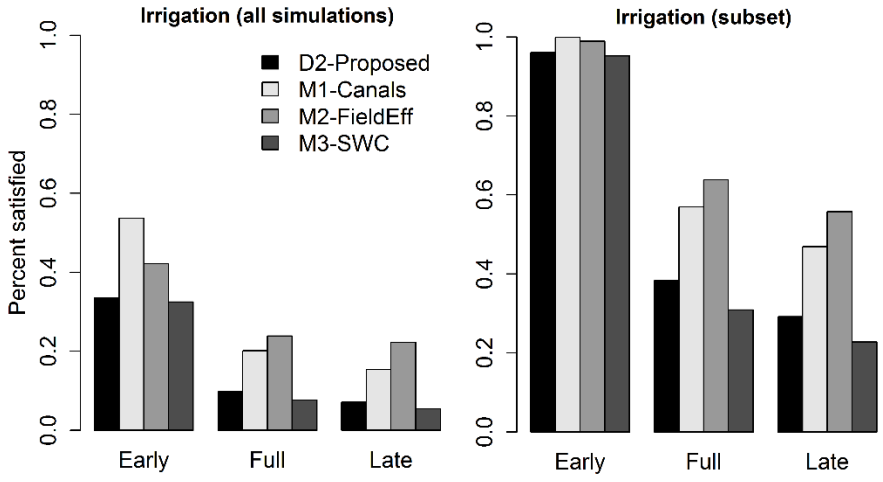

Figure 5.4: Percentage of simulations where irrigation reliability threshold is met under the full set of simulations (left) and optimistic subset of simulations (right).

The strong influence that model uncertainty, and to a lesser extent irrigation efficiency, exerts on irrigation water performance makes it difficult to understand the relative vulnerability of each alternative to other uncertain parameters that may also influence their results. To better understand the influence of other uncertain parameters, a subset of simulations where Qmod was at least 0.5 and irrigation efficiency was at least 50% (termed "optimistic" simulations from here forward for brevity) was evaluated in more detail. These simulations would represent conditions if the streamflow models were known to be accurate or conservative and irrigation efficiency achieved levels assumed in the project feasibility studies. The percentage of optimistic simulations where the irrigation water performance threshold was met for each time period under the D2 alternative and its modifications is shown in Figure 5.4. Under these conditions, the performance threshold is met in the early period for the vast majority (95-99%) of simulations. However, even under these optimistic conditions, the performance metric is only met in 31-64% of simulations when evaluated over the full time period, and 23-56% of simulations in the late period. The failure scenarios for this subset of simulations are presented in Table 5.7. It is apparent that changes in temperature are the dominant uncertainty driving performance in these simulations. Interesting, changes in interannual variability also contribute to vulnerability for the D2 alternative, but for the modifications the change in precipitation is more important. Again, the increase in vulnerability through time is apparent for all of the alternatives evaluated; for instance, when evaluated over the full time period the D2 alternative is vulnerable to any change in temperature above 3.86°C, whereas its performance in the late years is vulnerable to any change in temperature above 3.30°C.

Alternative and time period	Percent failures	Irrigation failure scenarios (optimistic simulations)	Box density	Box coverage	Ensemble density	Ensemble coverage
D2 Full	0.62	1) $\Delta T > 3.86$	0.95	0.49	0.84	0.84
		2) $\Delta T > 2.26$, $\Delta Var > 6.0$	0.73	0.35		
D2 Late	0.71	1) $\Delta T > 3.30$	0.97	0.59	0.91	0.81
		2) $\Delta T > 2.08$, $\Delta Var > 6.1$	0.79	0.22		
M3 Full	0.43	1) $\Delta T > 3.99$	0.85	0.57	0.8	0.76
		2) $\Delta T > 2.9$, $\Delta P < 5.5$	0.68	0.19		
M3 Late	0.53	1) $\Delta T > 3.50$	0.9	0.66	0.85	0.82
		2) $\Delta T > 2.63$, $\Delta P < 14.85$	0.69	0.16		
M4 Full	0.36	1) $\Delta T > 4.27$	0.79	0.52	0.75	0.73
		2) $\Delta T > 3.08$, $\Delta P < -5.97$	0.66	0.21		
M4 Late	0.44	1) $\Delta T > 4.04$	0.87	0.56	0.81	0.76
		2) $\Delta T > 2.91$, $\Delta P < 10.5$	0.67	0.2		
M5 full	0.69	1) $\Delta T > 3.50$	0.95	0.54	0.91	0.72
		2) $\Delta T > 2.36$, $\Delta P < 14.4$	0.81	0.19		
M5 late	0.77	1) $\Delta T > 1.421$	0.88	0.92	0.88	0.92

Table 5.7: Failure Scenarios for Irrigation Reliability (optimistic simulations)

5.4 Discussion

This analysis results in a number of insights that could prove useful for decision making, both in terms of the relative robustness of different development alternatives and areas where additional research may be most useful for infrastructure planning in the region. Regarding the choice between different alternatives, this analysis demonstrates that the expansion of irrigation withdrawls associated with D1 and D2 does result in increased vulnerability to hydrologic and climatic uncertainty when compared to existing infrastructure. However, assuming that the benefits of this expansion in terms of irrigated area are deemed acceptable, the proposed infrastructure under D2 actually results in better performance not just in terms of irrigation reliability, but hydropower and lake levels as well. This is likely due to the fact that the Gilgel Abbay and Jema reservoirs have very large capacities relative to the amount of water they are planned to provide, which may allow the system to take advantage of wet periods more effectively. Across all alternatives, there is a clear decline in performance through time; for instance, existing infrastructure is able to provide sufficient irrigation water in 82% of simulations when evaluated over the next 15 years, but only 49% of simulations when evaluated in years 36-50. This is not surprising when one considers the gradual onset of climate change and other issues such as reservoir sedimentation, but is a fact that can get lost in climate impact studies that focus on a discrete period of time decades in the future. Because of the gradual nature of these impacts, adaptable mitigation measures that can be modified based on changing conditions and new information are likely to be particularly valuable in addressing long-term risks to infrastructure performance (Hallegatte and Rentschler, 2015). While mechanisms for

developing adaptable infrastructure systems have been formally developed through methods such as adaptation pathways (Haasnoot et al., 2011; Ranger et al., 2013) and real options (Jeuland and Whittington, 2014; Woodward et al., 2014), this evaluation suggests that adaptability and learning in the non-engineered components of infrastructure systems (such as promoting improved irrigation efficiency) may be effective as well.

The ability of the D2 alternative to provide reliable water for irrigation is substantially improved by measures aimed at increasing irrigation efficiency, either through lining canals or promoting gradual improvements in field irrigation efficiency. Both measures reduce vulnerability to climatic and environmental uncertainty, but act across different time scales. While the improvements associated with canal lining are greater in the early years of system operation, these improvements are surpassed by those associated with field efficiency improvements over the long term. The feasibility of these modifications are likely to be highly dependent on issues related to financing and support for irrigation users; for example, lining canals requires a greater capital investment at the time of construction while improved field efficiency would require an ongoing efforts towards farmer education and support. Additionally, improved field efficiency has the potential to result in other benefits such as higher crop yields as well (Mintesinot et al., 2004; WWDSE and Tahal Group, 2009c), potentially making this a "win-win" solution that would be justified even if climate conditions didn't deteriorate (Hallegatte and Rentschler, 2015). The implementation of soil and water conservation measures upstream of the proposed reservoirs does have minor negative impacts on infrastructure performance since they reduce the amount of surface water flowing into the reservoirs; however, it is important to consider that these measures are likely to result in a number of benefits that weren't considered in this evaluation, such as improvements in water quality and agricultural productivity. While reduced surface water

downstream is one potentially negative impact of these measures, this evaluation suggests that this impact is likely to be relatively minor.

In addition to comparing the performance of different decision alternatives, the scenario discovery process also provides insights into which uncertainties have the greatest influence over infrastructure performance, and thus areas where additional research or understanding might be most valuable. The uncertain parameters that had the greatest influence over whether the infrastructure system was able to meet its objectives were related to uncertainty in streamflow simulation model accuracy, irrigation efficiency, and evaporation estimates. This suggests that uncertainty stemming from data limitations in hydrological processes today has a greater potential to negatively impact infrastructure performance than uncertainty regarding climate and land cover conditions in the future. From a risk management perspective, this is actually a positive outcome since focused research activities in the basin could help address and reduce some of these uncertainties. For instance, additional research on hydrologic processes in the region, combined with expansion of hydrologic monitoring networks and development of more sophisticated and realistic models for the region could help address the uncertainty associated with streamflow simulation modeling. Monitoring of water withdraws and usage from the Koga reservoir, which is already in place, could provide insights into likely values of irrigation efficiency. Finally, while predicting evaporation in the future due to climate change presents a number of challenges, the installation of additional meterological stations, combined with efforts to ground-truth satellite data products, could provide valuable information on evaporation rates under current climate conditions. This evaluation suggest that any of the above research is likely to be more valuable from an infrastructure planning perspective than a downscaled climate change impact assessment.

It should be noted that the framework that we used to account for streamflow model uncertainty is fairly conservative, as the prediction intervals associated with the streamflow model are quite wide. However, it is interesting to note that these wide bounds do not exert the same influence over all components of the water resource system. While irrigation reliability is very sensitive to this uncertainty, the provision of water for hydropower is significantly less so, and lake levels appear relatively robust even to very low values from the model prediction intervals. This suggests that the intervals are not so wide as to undermine learning about the relative vulnerability of different system components. The importance of hydrologic model uncertainty in determining infrastructure performance mirrors a number of other studies suggesting that hydrologic model uncertainty is likely to be a significant contributor to uncertainty in streamflow under changing climate. For example, Wilby (2005) found that model uncertainty stemming from the choice of training data period and non-uniqueness of parameter values was comparable to uncertainty due to greenhouse gas emission scenarios when predicting future surface water flows in the River Thames. The relative contribution of hydrologic model uncertainty can be particularly dramatic when considering specific aspects of future flow regimes; for example, some studies suggest that estimates of low-flow values are especially sensitive (Jung et al., 2012). Additionally, climate change has the potential to alter the physical processes that hydrologic models aim to represent, making models calibrated to historic conditions biased in their prediction of future conditions (Brigode et al., 2013; Coron et al., 2012; Merz et al., 2011). It should also be noted that the majority of studies comparing the relative contribution of emissions, climate model, and hydrologic model uncertainty in were conducted in Europe, North America or Australia; it is not unreasonable to suspect that hydrologic model uncertainty would be even more severe in data-scarce regions of the developing world.

Collectively, this work suggests that any evaluation considering water infrastructure performance that relies on hydrologic models should carefully consider the degree to which model uncertainty contributes to uncertainty in infrastructure performance. Some RDM studies conducted to date have considered hydrologic uncertainty through evaluating different parameters related to groundwater recharge (Lempert and Groves, 2010) or evapotranspiration (Fischbach et al., 2015), but this has generally been given relatively little consideration compared to the evaluation of uncertainty in future climate conditions. Our research suggests that more sophisticated evaluation of model uncertainty should be an important component of robust decision frameworks applied in data-scarce contexts. However, one practical challenge in doing so is representing uncertainty in a manner that can be evaluated through the scenario discovery algorithm, which requires uncertain input parameters to be represented by continuous numeric values. This would allow for comparison of simulation results across a range of values for a specific parameter, providing insights in situations where certain model parameters are known to have a limited empirical basis for their assigned values or which are observed to vary depending on the calibration procedure. For instance, a review of SWAT model applications in the Blue Nile suggested that a number of studies used calibration procedures that resulted in physically unrealistic values of parameters associated with loss of water to groundwater (van Griensven et al., 2012). Sampling a range of values for these parameters would be one way to characterize the contribution of model uncertainty to infrastructure performance. However, incorporating competing hydrologic or conceptual models or uncertainty stemming from the use of different calibration periods would likely be more difficult. Development of methods to incorporate hydrologic model uncertainty into the RDM framework would likely be a

valuable area of research to support robust planning of water resource systems in data-scarce regions.

While this evaluation demonstrates how an exploratory modeling framework can be combined with the RDM methodology to better understand infrastructure vulnerability over the near and long-term in data-scarce situations, it did include a number of simplifications that should be considered in interpreting its results and which suggest areas for further research. One simplification was that uncertain parameter values were assumed to take on the same values across the entire study area, even though in reality you would expect these to exhibit spatial variability. For instance, it is possible that existing methods for estimating evaporation result in underestimates in one region of the basin and overestimates in another, or that heterogeneous soil conditions mean that irrigation efficiency associated with certain projects is much higher than others. Incorporating this spatial heterogeneity into the RDM process without having to consider an overwhelming number of uncertain parameters could be valuable in understanding which uncertainties are most important in different regions. Additionally, the pre-defined development alternatives evaluated here may not be the most optimal selection of projects, and having a sense of which reservoirs are likely to provide the best performance could provide valuable insights into project prioritization. Methods that aim to identify high-performing combinations of projects and operating procedures, such as many-objective RDM (Kasprzyk et al., 2013) might prove useful in this regard. Finally, it should be acknowledged that the uncertain parameters evaluated in this study are not entirely separable; for instance, streamflow prediction uncertainty also impacts sedimentation rates and temperature change uncertainty also impacts evaporation rates. Understanding how these interactions may influence the failure scenarios identified by the PRIM algorithm would be a valuable area of additional research.

5.5 Conclusions

The development of robust decision frameworks provides a valuable tool for water infrastructure planning under deeply uncertain climate change. The need for these tools is particularly severe in the developing world, where climate model projections are especially uncertain and difficult to evaluate, and where the construction of expensive, long-lived hard infrastructure is largely seen as an important step in improving economic and living conditions. However, applying these methods in data-scarce regions presents its own set of challenges. Limited availability of computing facilities and expertise may hinder the use of data from multi-model climate ensembles needed to characterize GCM uncertainty, and data limitations present challenges in the development of reliable hydrologic and water resource models to understand the impacts of potential climate change. To address these challenges, this work presented a modified application of the RDM methodology that is specifically tailored for application in data-scarce situations. This modification includes a novel yet simple method for generating transient climate change sequences that account for potential variable dependence but do not rely on GCM projections, and an emphasis on identifying the relative importance of data limitations and model uncertainty that may be addressed through future research. Application of this methodology to the Lake Tana basin in Ethiopia leads to a number of insights regarding the relative vulnerability of different development alternatives, as well as the potential for efficiency improvements to address these vulnerabilities across different time scales. Additionally, we find that uncertainty in streamflow model accuracy, irrigation efficiency, and evaporation rates has the greatest potential to impact infrastructure performance, suggesting that additional research to address these uncertainties could provide valuable insights for long-term infrastructure planning. Ultimately, this demonstrates how established methodologies for infrastructure planning

under uncertainty can be modified to better suit developing-country contexts and improve the long-term effectiveness and sustainability of infrastructure investments in these countries.

6 CONCLUSIONS

6.1 Summary and Contributions

Climate change has the potential to dramatically impact the water resources sector, and it is crucial that we consider these impacts today to avoid negative consequences in the future. However, the considerable uncertainty surrounding climate change projections, particularly at the local and regional scales relevant for infrastructure planning, makes this a difficult challenge. In the face of these issues, a number of methods have been developed to better characterize and address climate change impacts in the water sector. The objective of this dissertation was to critically evaluate methods for impact assessment and decision support in the face of deeply uncertain climate change projections, particularly focusing on the RDM methodology and empirical methods for streamflow simulation.

Chapter 2 of this dissertation presented a comparison of multiple machine-learning methods for data-driven streamflow simulation in five seasonal rivers in the Lake Tana basin. To understand the relative suitability of different methods for planning and management decisions, this comparison evaluated each method not only in terms of its predictive accuracy, but error structure and bias, model interpretability, and uncertainty when faced with extreme climate conditions as well. Despite the popularity of ANNs in existing research on streamflow simulation, ANNs were not found to be the most accurate model in any of the five basins evaluated. Other methods such as GAMs and random forests are able to capture non-linear relationships effectively and lend themselves to simpler visualization of model structure and covariate influence, making it easier to gain insights on watershed function and confirm that the model is operating in a physically realistic manner. We also demonstrate how certain model formulations can lead to autocorrelation in model residuals and biased estimates of water availability, emphasizing the importance of considering error

structure when evaluating if a model is fit-for-purpose. Finally, we find that predictions from some model types (particularly GAMs, GLMs, and MARS) exhibit considerable variability when faced with extreme climate conditions, while others (such as random forests) may be biased in their predictions under these conditions. This work demonstrates the value of considering multiple model formulations for a given problem and the importance of evaluating multiple facets of model performance in determining if it is suitable for planning decisions.

Chapter 3 compares RDM with uncertainty factors and probability bounds analysis in an effort to bridge developments in the water resources and risk analysis fields aimed at assessing and managing risk in the face of deep uncertainty. By applying each approach to a stylized example problem related to flood risk under climate change, this evaluation demonstrates how these methods differ in their representation of uncertainty quantities, analytical output, and implications for risk management. This example problem demonstrates the ways that risk assessment can inform decision making in conditions where uncertainty and ambiguity make prescriptive approaches inappropriate. In particular, the identification of epistemic uncertainties that most contribute to uncertainty in the resulting risk description or choice of alternatives can provide useful insights into places where additional research or more sophisticated representation could most benefit the assessment. However, we find that the analytical output and implications for decision making, both in terms of preferred alternatives and key sources of uncertainty, are not necessarily consistent between approaches. This suggests the potential value in additional comparative research to better understand the sources of these deviations, as well as the need for risk analysts to consider the ways in which the choice of methodology might impact analytical results.

Water resource problems are generally characterized by the need to balance multiple conflicting objectives, but RDM applications to date have provided little insight into how the treatment of multiple criteria impacts the method's analytical results. To address this gap, Chapter 4 uses the proposed infrastructure in Lake Tana as a case study to evaluate how the method used to address multiple criteria impacts the process of system vulnerability identification within the RDM methodology. It demonstrates that common methods used to aggregate multiple criteria into a single utility score can lead to inconsistent failure scenarios and obscure the relationship between key uncertainties and system performance. Applying scenario discovery over each performance metric separately provides more nuanced information regarding the relative sensitivity of system objectives and the ways in which they are impacted by different uncertain parameters. This in turn can provide insights on measures that could be taken to improve system robustness, as well as areas where additional research might prove useful. Because the RDM framework was designed to provide quantitative decision support in contexts where there may be conflicting beliefs about what the future will look and contentious disagreements about the best course of action, it is important that the steps of the process remain as transparent as possible. To this end, the additional effort required to apply scenario discovery to each metric separately provides valuable benefits by identifying failure scenarios that inform a more complete picture of system performance and provide more detailed guidance for vulnerability-reduction efforts.

Decision frameworks such as RDM that provide a comprehensive treatment of non-probabilistic uncertainty have the potential to be particularly valuable in developing countries with extensive infrastructure needs. However, applying these methods in data-scarce regions presents its own set of challenge related to limitations in available data, impact models, and computing facilities. Chapter 5 demonstrates a modified application of the RDM

methodology that is specifically tailored for application in data-scarce situations. This modification includes a novel yet simple method for generating transient climate change sequences that account for potential variable dependence but do not rely on GCM projections, and an emphasis on identifying the relative importance of data limitations and model uncertainty that may be addressed through future research. In applying this methodology to the Lake Tana basin, we demonstrate how the approach can provide a number of insights regarding the relative vulnerability of different development alternatives and priorities for additional research. In particular, we find that uncertainty in streamflow model accuracy, irrigation efficiency, and evaporation rates has the greatest potential to impact infrastructure performance, suggesting that additional research to address these uncertainties should be a greater priority than downscaled climate impact studies. Ultimately, this demonstrates how established methodologies for infrastructure planning under uncertainty can be modified to better suit developing-country contexts and improve the long-term effectiveness and sustainability of infrastructure investments in these countries.

Taken as a whole, this work demonstrates a number of issues that should be taken into account when applying novel methods for climate change impact assessment and decision support in water resource systems and long-lived infrastructure systems more broadly. Choices about empirical model structure, risk assessment procedures, and the treatment of multiple criteria can all have important impacts on analytical results; it is crucial that these choices be given due consideration if analytical results are to inform infrastructure design and investment decisions. Ultimately, an improved understanding of how analytical steps in climate impact assessment and adaptation can be tailored to specific problems and objectives will improve our ability to create robust, sustainable infrastructure systems in the face of highly uncertain climate conditions.

6.2 Future Research

This work demonstrated some of the ways in which systematic evaluation of novel methodologies for climate change impact assessment and robust decision support can highlight the abilities and limitations of these methods. Additional research has the potential to result in additional insights for sustainable infrastructure planning based on these tools. The following paragraphs discuss priorities for additional research regarding the use of machine learning and robust decision making to support climate change impact assessment and adaptation.

Machine learning methods like those evaluated in Chapter 2 have the potential to be very useful in modelling complex systems characterized by nonlinearities and interactions between variables, making them well suited for complex hydrologic systems. This work compared a small number of model types in terms of their error structure, bias and uncertainty, but the inclusion of additional model structures could yield further insights into the benefits of using different model types. For instance, the autocorrelation present in model errors suggests that a persistence or time series model might be useful in simulating streamflow more accurately. While this has been done in the context of creating flood forecasting models based on machine learning methods (for instance, Galelli and Castelletti, 2013a; Han et al. 2007; and Yu et al., 2014), the use of such models for long-term simulations raises questions about how model errors and uncertainty might propagate through time. Additional comparative studies could be useful in understanding how accounting for persistence in hydrologic conditions (for instance, by incorporating a lagged flow term as an explanatory variable) might impact the accuracy, bias and uncertainty of model predictions in long-term simulations.

One well noted limitation to empirical models is that they can only simulate conditions that are comparable to those in the model training data, presenting a problem when models are needed to simulate system response under extreme climates not well represented in the historic record. However, the degree to which climate change is likely to differ from historic climate conditions varies; for instance, in Lake Tana projected temperature changes could very well exceed historic variability, whereas precipitation projections are largely within the range of historic conditions. This work demonstrated how bootstrap sampling could be used to quantify the variance in model predictions under extreme climate conditions and better understand the implications of using the models under changing temperatures and changing precipitation patterns. A similar evaluation conducted in differing climatic regions (for example temperate, arid, or snow-dominated) could be valuable in understanding the where extrapolation of empirical relationships to projected climates presents the greatest uncertainty and where is may be more suitable.

There are also a number of ways in which additional research could strengthen the RDM framework and suggest best practices for its application to climate adaptation problems. One issue that has not be thoroughly explored within the RDM framework is the degree to which the input parameter distributions used to generate samples might impact analytical results. It is somewhat concerning that many RDM applications to date focus heavily on identification of robust alternatives through satisficing or regret-based measures (e.g., Fischbach et al., 2015; Groves et al., 2013; Groves and Bloom, 2013) despite the fact that these robustness metrics are likely to be highly sensitive to the input distributions used to generate samples for simulation modelling. Since the framework was explicitly designed for use in situations where there might be disagreement or uncertainty surrounding these distributions, failure scenarios identified by PRIM (which are likely to be less sensitive to

these distributions) would likely be a more appropriate basis for evaluating robustness. However, a formal evaluation of how input parameter distributions impact the failure scenarios identified by the PRIM algorithm would be necessary to confirm that this was the case. Additionally, dependence and correlation between uncertain parameters could conceivably impact the results of the PRIM algorithm; for instance, if changes in temperature are correlated with changes in precipitation variability (as was the case in Chapter 5), it may be difficult to separate the influence of these two variables. Additional research would be useful in developing a better understanding of potential issues associated with input parameter distribution and dependencies, as well as mechanisms that could be used to address these issues.

One limitation with the scenario discovery process is the requirement that performance be classified in a binary satisfactory/unsatisfactory fashion. While this may make sense in some contexts where there are obvious thresholds for unsatisfactory performance (for example, if project costs exceed benefits), in many cases these thresholds may not be apparent and evaluation of performance across a continuous scale may be preferable. In cases where there are no obvious thresholds for acceptable performance, the development of better methods to demonstrate how sensitive failure scenarios are to the chosen acceptable performance threshold could be useful. Additionally, in many cases the impact of uncertain parameters on system performance may not be independent; for example, the left panel of Figure 4.3 suggests that the impact of evaporation and precipitation are additive in their impact on hydropower performance. Previous research has demonstrated how orthogonal rotations based on principle components analysis can be used to identify failure scenarios that better reflect these interactions (Dalal et al., 2013) and could be further developed through the use of other methods for dimensionality reduction, such as

discriminant analysis. However, care must be taken since this may also make the scenarios less interpretable (Parker et al., 2015).

Finally, the identification of the key uncertain parameters driving system performance is arguably one of the most valuable insights provided by RDM and other methods for risk and decision assessment under deep uncertainty. This leads naturally to the question of how valuable a reduction in uncertainty surrounding those parameters could be for the decision under consideration, and a useful area of additional research would be to develop methods for quantifying the value of reduction in uncertainty surrounding different parameters. This could conceivably be done by using the scenarios identified through RDM to create decision trees that could be used to calculate the value of perfect and imperfect information; however, some consideration would be needed to make these values robust to uncertainty surrounding the likelihood of different scenarios.

APPENDIX A: WEAP MODEL SUMMARY

Model system overview

The Water Evaluation and Planning (WEAP) system is a simulation modeling software designed to aid planners in water management and decision making. It integrates modeling of physical hydrologic processes with built infrastructure and water management operations to characterize how changes in system design, operation, and external conditions (such as climate or land use change) impact water availability and allocation at the basin scale. It uses a simple water balance accounting approach to represent both the supply and demand side of water resource systems. In this sense, it sacrifices physical realism when compared to distributed hydrologic models that simulate natural hydrologic processes in greater detail, but allows for more detailed representation of water allocation decisions. The model can be customized to support increasingly refined representation of both water supply and demand and has been used widely in water resources research; for instance, the Stockholm Environment Institute lists approximately 60 scholarly publications that used WEAP models in 2015[6].

In the simplest case, water supply within the WEAP modeling system can be represented by user-specified time series of streamflow volumes. However, the system can also use semi-distributed hydrologic models that link surface water flows and groundwater levels to climate and land cover conditions. Hydrologic modeling capabilities within WEAP are based on a semi-distributed water balance approach that partitions water introduced to a watershed as precipitation into runoff, infiltration to groundwater, and evaporative losses

[6] List of publications available at: http://www.weap21.org/index.asp?action=216

(Yates et al., 2005). The model can also incorporate interactions between groundwater and surface water through lumped modeling of aquifer storage levels and associated stream levels, or through connecting WEAP with a gridded MODFLOW groundwater flow model. It is also possible to include other sub-modules that provide more sophisticated representation of water quality and contaminant transport, if needed.

Water demand is represented by user defined demand nodes, which can represent municipal, irrigation, and industrial demand sites, as well as in-stream flow requirements. The model can support increasingly refined representation of water demand; for example, a user can simply represent municipal demand using a time series of demand for a whole city, or demand could be disaggregated into industrial and municipal use, which is then further disaggregated into single and multi-family residences and changes through time based on assumed population growth rates. Similarly, irrigation demands can be incorporated into the hydrologic modeling component (and thus account for water provided by precipitation or stored soil moisture) or be user-defined. These demand sites are linked with sources of water that can be used to provide supply, including rivers, groundwater, lakes and reservoirs. Users can define simple reservoir operating criteria to determine how much water can be released to satisfy downstream demands based on water levels within the reservoir. Demand sites are ranked in order of priority, ranging from 1 for highest priority to 99 for lowest priority, in case there are times when supply is insufficient to cover all demand nodes. These priorities can be adjusted through time to account for seasonal variations in water management needs (for example, the need to fill reservoirs during the rainy season so water can be released during the dry season).

To allocate available water to different demand nodes, the model first calculates a mass balance for each node and link within the system at each time step. Water is then allocated to demand nodes in decreasing order of priority. If multiple demand nodes have the same priority level, then a linear program is solved to maximize the coverage (defined as the percentage of demand supplied) of total demand at that priority level subject to mass balance and equity constraints (Yates et al., 2005). This process is repeated for subsequent priority levels until all demands have been satisfied to the greatest possible level, and then repeated for subsequent time steps. While water stored in reservoirs or groundwater is accounted for in subsequent time steps, the allocation of water in each time step is independent of allocation in previous time steps.

Lake Tana WEAP Model

The WEAP model for the Lake Tana basin was developed and calibrated by Alemayehu et al. (2010) to evaluate the impact that proposed water infrastructure development would have on downstream flows and lake levels under historic streamflow conditions. The model used historical streamflow sequences and externally estimated rates of evaporation off of the lake and proposed reservoirs as inputs, and the lake itself was simulated as a reservoir because flows out of the lake are regulated by the Chara Chara weir. The model was manually calibrated by comparing observed and simulated lake levels for the period 1960-2004, with the final model achieving a Nash-Sutcliffe efficiency of 0.74 (Alemayehu et al., 2010).

The simulations run for the research presented in this work did not rely on the internal hydrologic modeling capabilities of the WEAP software, and instead used external empirical models to estimate streamflow response to changing climate and land cover

conditions because spatially explicit data on vegetation and soil characteristics needed to meaningfully inform rainfall-runoff modeling within WEAP were not available. Similarly, evaporation off of Lake Tana and the proposed reservoirs was estimated externally using Penman's equation and input directly into WEAP. Similarly, sediment transport was not explicitly modeled within WEAP; instead, monthly sequences of sediment loads into the reservoir sites were estimated using sediment rating curves for the region and then used to estimate how the capacity of each reservoir changed through time. This method assumes that all sediment is retained behind the dams with no removal, which is consistent with observations from other reservoirs in the region and plans for the proposed reservoirs, which don't include any sedimentation abatement strategies.

Water demands represented within the Lake Tana model included irrigated agriculture, the filling of reservoirs and Lake Tana, the Tana-Beles hydropower transfer, environmental flows, and flow requirements at the Tis Issat waterfall located downstream of the lake. Demands for agriculture were split into a "base" demand, which represented the minimum irrigation demand for that site, and "extra" demand which represented the difference between this minimum demand and the maximum possible demand outlined in feasibility studies for the projects. This allowed additional water to be allocated to irrigation sites when it was available to augment agricultural production at those locations, but only if other water demands within the basin were also met. The priorities for different demand nodes are shown in Table A.1 and were based on the stated priorities from the Tana Sub-basin Authority. Priorities varied slightly between wet and dry season months to prioritize filling Lake Tana during the wet season so that dry season lake levels don't drop too low.

Dry Season		Wet Season	
Priority Rank	Demand Sites	Priority Rank	Demand Sites
1	Agriculture (base)	1	Agriculture (base)
2	Filling reservoirs, hydropower	2	Filling reservoirs, hydropower, filling lake
3	Filling lake	3	Environmental flows, Tis Issat waterfall flows
4	Environmental flows, Tis Issat waterfall flows	4	Agriculture (extra)
5	Agriculture (extra)		

Table A.1: Priority rankings for demand sites in Lake Tana WEAP model

Because no operating rule curves were available to specify operation of the proposed reservoirs, water was held and released from the reservoirs in accordance with the priority levels defined in Table A1. Thus, there was no consideration of flood-control or forecasts of multi-month flows or storage, although the high priority for reservoir filling inherently results in limited releases to ensure that enough storage is built up during the wet season to meet dry season demands. Releases from the lake were treated in the largely the same manner as releases from the reservoirs. However, a buffer zone was also specified for the lake, below which releases were constrained to only be a specified fraction of stored water (10-30%, depending on the season) to prevent the lake elevation from dropping below the threshold of 1784.75 meters amsl (identified as the point where negative impacts to environmental conditions, navigation, and fisheries in the lake are expected to occur; SMEC International, 2008). Environmental flow requirements in the lake's tributaries and the Blue Nile River downstream of the lake were determined using the desktop reserve method to account for seasonal and interannual variability of flow as described by McCartney et al (2009).

REFERENCES

Abrahart, R. J. and See, L. M.: Neural network modelling of non-linear hydrological relationships, Hydrol. Earth Syst. Sci., 11(5), 1563–1579, doi:10.5194/hess-11-1563-2007, 2007.

Achenef, H., Tilahun, A. and Molla, B.: Tana Sub Basin Initial Scenarios and Indicators Development Report, Tana Sub Basin Organization, Bahir Dar, Ethiopia., 2013.

Acres International Limited and Shawel Consult International: Feasibility Study of the Birr and Koga Irrigation Project., 1995.

Adimassu, Z., Mekonnen, K., Yirga, C. and Kessler, A.: Effect of soil bunds on runoff, soil and nutrient losses, and crop yield in the central highlands of Ethiopia, Land Degradation & Development, 25(6), 554–564, doi:10.1002/ldr.2182, 2014.

Alemayehu, T., McCartney, M. and Kebede, S.: The water resource implications of planned development in the Lake Tana catchment, Ethiopia, Ecohydrology & Hydrobiology, 10(2-4), 211–221, doi:10.2478/v10104-011-0023-6, 2010.

Amundrud, Ø. and Aven, T.: On how to understand and acknowledge risk, Reliability Engineering & System Safety, 142, 42–47, doi:10.1016/j.ress.2015.04.021, 2015.

Antar, M. A., Elassiouti, I. and Allam, M. N.: Rainfall-runoff modelling using artificial neural networks technique: a Blue Nile catchment case study, Hydrol. Process., 20(5), 1201–1216, doi:10.1002/hyp.5932, 2006.

Apostolakis, G. E.: How useful is quantitative risk assessment?, Risk analysis, 24(3), 515–520, 2004.

Aqil, M., Kita, I., Yano, A. and Nishiyama, S.: Neural Networks for Real Time Catchment Flow Modeling and Prediction, Water Resources Management, 21(10), 1781–1796, doi:10.1007/s11269-006-9127-y, 2007.

Asefa, T., Kemblowski, M., McKee, M. and Khalil, A.: Multi-time scale stream flow predictions: The support vector machines approach, Journal of Hydrology, 318(1-4), 7–16, doi:10.1016/j.jhydrol.2005.06.001, 2006.

Aven, T.: A semi-quantitative approach to risk analysis, as an alternative to QRAs, Reliability Engineering & System Safety, 93(6), 790–797, 2008.

Aven, T.: On the need for restricting the probabilistic analysis in risk assessments to variability, Risk analysis, 30(3), 354–360, 2010.

Aven, T.: Foundations of risk analysis, 2nd ed., Wiley, Hoboken, N.J., 2012.

Aven, T.: On how to deal with deep uncertainties in a risk assessment and management context, Risk Analysis, 33(12), 2082–2091, 2013.

Aven, T. and Zio, E.: Some considerations on the treatment of uncertainties in risk assessment for practical decision making, Reliability Engineering & System Safety, 96(1), 64–74, 2011.

Aven, T., Zio, E., Baraldi, P. and Flage, R.: Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-probabilistic Methods, John Wiley & Sons., 2013.

Awulachew, S. B. and Ayana, M.: Performance of irrigation: an assessment at different scales in Ethiopia, Experimental Agriculture, 47(S1), 57–69, doi:10.1017/S0014479710000955, 2011.

Bankes, S.: Exploratory Modeling for Policy Analysis, Operations Research, 41(3), 435–449, doi:10.1287/opre.41.3.435, 1993.

Barnett, D. N., Brown, S. J., Murphy, J. M., Sexton, D. M. H. and Webb, M. J.: Quantifying uncertainty in changes in extreme event frequency in response to doubled CO2 using a large ensemble of GCM simulations, Clim Dyn, 26(5), 489–511, doi:10.1007/s00382-005-0097-1, 2006.

Bedford, T. and Cooke, R.: Probabilistic risk analysis: foundations and methods, Cambridge University Press., 2001.

Bekele, Z. and Tilahun, K.: On-farm performance evaluation of improved traditional small-scale irrigation practices: A case study from Dire Dawa area, Ethiopia, Irrigation and Drainage Systems, 20(1), 83–98, 2006.

Berner, C. and Flage, R.: Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions, Reliability Engineering & System Safety, doi:10.1016/j.ress.2015.10.009, 2015.

Beven, K. J.: Rainfall-Runoff Modelling: The Primer, John Wiley & Sons., 2011.

Bhattacharjee, P. S. and Zaitchik, B. F.: Perspectives on CMIP5 model performance in the Nile River headwaters regions: CMIP5 in the Nile headwaters region, International Journal of Climatology, 35(14), 4262–4275, doi:10.1002/joc.4284, 2015.

Bjerga, T. and Aven, T.: Adaptive risk management using new risk perspectives – an example from the oil and gas industry, Reliability Engineering & System Safety, 134, 75–82, doi:10.1016/j.ress.2014.10.013, 2015.

Breiman, L.: Bagging predictors, Mach Learn, 24(2), 123–140, doi:10.1007/BF00058655, 1996.

Breiman, L.: Random forests, Mach Learn, 45(1), 5–32, 2001.

Brigode, P., Oudin, L. and Perrin, C.: Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?, Journal of Hydrology, 476, 410–425, doi:10.1016/j.jhydrol.2012.11.012, 2013.

Brown, C.: The end of reliability, Journal of Water Resources Planning and Management, 136(2), 143–145, 2010.

Brown, C.: Decision-scaling for Robust Planning and Policy under Climate Uncertainty, Washington, D.C. [online] Available from: http://www.wri.org/sites/default/files/uploads/wrr_brown_uncertainty.pdf (Accessed 5 February 2016), 2011.

Brown, C., Ghile, Y., Laverty, M. and Li, K.: Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector, Water Resour. Res., 48(9), W09537, doi:10.1029/2011WR011212, 2012.

Bryant, B. J.: sdtoolkit: Scenario Discovery Tools to Support Robust Decision Making. R package version 2.33-1., [online] Available from: http://CRAN.R-project.org/package=sdtoolkit (Accessed 1 January 2015), 2014.

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

Cervigni, R., Liden, R., Neumann, J. E. and Strzepek, K. M., Eds.: Enhancing the Climate Resilience of Africa's Infrastructure: The Power and Water Sectors, The World Bank. [online] Available from: http://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-0466-3 (Accessed 11 February 2016), 2015.

Chao, P. T., Hobbs, B. F. and Venkatesh, B. N.: How climate uncertainty should be included in Great Lakes management: modeling workshop results, Journal of the American Water Resources Association, 35(6), 1485–1497, doi:10.1111/j.1752-1688.1999.tb04232.x, 1999.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q. Y., Ek, M. and Betts, A.: Modeling of land surface evaporation by four schemes and comparison with FIFE observations, J. Geophys. Res., 101(D3), 7251–7268, doi:10.1029/95JD02165, 1996.

Chibanga, R., Berlamont, J. and Vandewalle, J.: Modelling and forecasting of hydrological variables using artificial neural networks: the Kafue River sub-basin, Hydrological Sciences Journal, 48(3), 363–379, doi:10.1623/hysj.48.3.363.45282, 2003.

Clark, M. P. and Pulwarty, R. S.: Devising resilient responses to potential climate change impacts, Ogmius: Newsletter of the Center for Science and Technology Policy Research, 5, 2–3, 2003.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M. and Hendrickx, F.: Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments, Water Resour. Res., 48(5), W05552, doi:10.1029/2011WR011721, 2012.

Cox, L. A. T.: Confronting Deep Uncertainties in Risk Analysis, Risk Analysis, 32(10), 1607–1629, 2012.

Criss, R. E. and Winston, W. E.: Do Nash values have value? Discussion and alternate proposals, Hydrol. Process., 22(14), 2723–2725, doi:10.1002/hyp.7072, 2008.

Dalal, S., Han, B., Lempert, R., Jaycocks, A. and Hackbarth, A.: Improving scenario discovery using orthogonal rotations, Environmental Modelling & Software, 48, 49–64, doi:10.1016/j.envsoft.2013.05.013, 2013.

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

Dessie, M., Verhoest, N. E. C., Admasu, T., Pauwels, V. R. N., Poesen, J., Adgo, E., Deckers, J. and Nyssen, J.: Effects of the floodplain on river discharge into Lake Tana (Ethiopia), Journal of Hydrology, 519, 699–710, doi:10.1016/j.jhydrol.2014.08.007, 2014.

De Vos, N. J. and Rientjes, T. H. M.: Multiobjective training of artificial neural networks for rainfall-runoff modeling, Water Resour. Res., 44(8), W08434, doi:10.1029/2007WR006734, 2008.

Dubois, D.: Representation, propagation, and decision issues in risk analysis under incomplete probabilistic information, Risk analysis, 30(3), 361–368, 2010.

Dubois, D. and Prade, H.: On the relevance of non-standard theories of uncertainty in modeling and pooling expert opinions, Reliability Engineering & System Safety, 36(2), 95–107, 1992.

Dubois, D., Prade, H. M., Farreny, H., Martin-Clouaire, R. and Testemale, C.: Possibility theory: an approach to computerized processing of uncertainty, Plenum press New York., 1988.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R. and Mearns, L. O.: Climate Extremes: Observations, Modeling, and Impacts, Science, 289(5487), 2068–2074, doi:10.1126/science.289.5487.2068, 2000.

Easton, Z. M., Fuka, D. R., White, E. D., Collick, A. S., Biruk Ashagre, B., McCartney, M., Awulachew, S. B., Ahmed, A. A. and Steenhuis, T. S.: A multi basin SWAT model analysis of runoff and sedimentation in the Blue Nile, Ethiopia, Hydrology and Earth System Sciences, 14(10), 1827–1841, doi:10.5194/hess-14-1827-2010, 2010.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res., 108(D22), 8851, doi:10.1029/2002JD003296, 2003.

Elshorbagy, A., Corzo, G., Srinivasulu, S. and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1:

Concepts and methodology, Hydrol. Earth Syst. Sci., 14(10), 1931–1941, doi:10.5194/hess-14-1931-2010, 2010a.

Elshorbagy, A., Corzo, G., Srinivasulu, S. and Solomatine, D. P.: Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application, Hydrol. Earth Syst. Sci., 14(10), 1943–1961, doi:10.5194/hess-14-1943-2010, 2010b.

European Environmental Agency: Looking back on looking forward: a review of evaluative scenario literature., 2009.

Ferson, S. and Ginzburg, L. R.: Different methods are needed to propagate ignorance and variability, Reliability Engineering & System Safety, 54(2), 133–144, 1996.

Ferson, S. and Tucker, W. T.: Sensitivity in risk analyses with uncertain numbers, Sandia National Laboratories Albuquerque, NM. [online] Available from: http://prod.sandia.gov/techlib/access-control.cgi/2006/062801.pdf (Accessed 15 January 2016), 2006.

Field, C. B., Barros, V. B., Dokken, D. J., Mach, K. J., Mastrandrea, M. D., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C., Girma, B., Kissel, E. S., Levy, A. N., MacCracken, S., Mastrandrea, P. R. and White, L. L.: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2014.

Fischbach, J. R., Lempert, R. J., Molina-Perez, E., Tariq, A. A., Finucane, M. L. and Hoss, F.: Managing water quality in the face of uncertainty: a robust decision making demonstration for EPA's National Water Program, Rand Corporation, Santa Monica, Calif., 2015.

Flage, R. and Aven, T.: Expressing and communicating uncertainty in relation to quantitative risk analysis, Reliability & Risk Analysis: Theory & Application, 2(13), 9–18, 2009.

Flage, R., Baraldi, P., Zio, E. and Aven, T.: Probability and Possibility-Based Representations of Uncertainty in Fault Tree Analysis: Comparison of Probabilistic/Possibilistic Computational Approaches, Risk Analysis, 33(1), 121–133, doi:10.1111/j.1539-6924.2012.01873.x, 2013.

Flyvbjerg, B.: Survival of the unfittest: why the worst infrastructure gets built--and what we can do about it, Oxford Review of Economic Policy, 25(3), 344–367, doi:10.1093/oxrep/grp024, 2009.

Friedman, J. H.: Multivariate adaptive regression splines, The annals of statistics, 1–67, 1991.

Friedman, J. H. and Fisher, N. I.: Bump hunting in high-dimensional data, Statistics and Computing, 9(2), 123–143, 1999.

Galelli, S. and Castelletti, A.: Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling, Hydrology and Earth System Sciences, 17(7), 2669–2684, doi:10.5194/hess-17-2669-2013, 2013a.

Galelli, S. and Castelletti, A.: Tree-based iterative input variable selection for hydrological modeling, Water Resour. Res., 49(7), 4295–4310, doi:10.1002/wrcr.20339, 2013b.

Garede, N. M. and Minale, A. S.: Land Use/Cover Dynamics in Ribb Watershed, North Western Ethiopia, Journal of Natural Sciences Research, 4(16), 9–16, 2014.

Gaume, E. and Gosset, R.: Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology?, Hydrology and Earth System Sciences Discussions, 7(5), 693–706, 2003.

Gebrehiwot, S. G., Taye, A. and Bishop, K.: Forest Cover and Stream Flow in a Headwater of the Blue Nile: Complementing Observational Data Analysis with Community Perception, AMBIO, 39(4), 284–294, doi:10.1007/s13280-010-0047-y, 2010.

Gebremicael, T. G., Mohamed, Y. A., Betrie, G. D., van der Zaag, P. and Teferi, E.: Trend analysis of runoff and sediment fluxes in the Upper Blue Nile basin: A combined analysis of statistical tests, physically-based models and landuse maps, Journal of Hydrology, 482, 57–68, doi:10.1016/j.jhydrol.2012.12.023, 2013.

Giorgi, F. and Mearns, L. O.: Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging"(REA) method, Journal of Climate, 15(10), 1141–1158, 2002.

Gleick, P. H.: Methods for evaluating the regional hydrologic impacts of global climatic changes, Journal of Hydrology, 88(1–2), 97–116, doi:10.1016/0022-1694(86)90199-X, 1986.

Gleick, P. H.: A Look at Twenty-first Century Water Resources Development, Water International, 25(1), 127–138, doi:10.1080/02508060008686804, 2000.

Groves, D. G. and Bloom, E.: Robust Water-Management Strategies for the California Water Plan Update 2013, Santa Monica, Calif.: RAND Corporation, RR-182-DWR. [online] Available from: http://www.waterplan.water.ca.gov/docs/swan/2013/RR182_FNLCompiled.pdf (Accessed 11 August 2015), 2013.

Groves, D. G., Fischbach, J. R., Bloom, E., Knopman, D. S. and Keefe, R.: Adapting to a changing Colorado River: making future water deliveries more reliable through robust management strategies, RAND, Santa Monica, CA., 2013a.

Groves, D. G., Bloom, E., Johnson, D. R., Yates, D. and Mehta, V.: Addressing climate change in local water agency plans: demonstrating a simplified Robust Decision Making approach in the California Sierra foothills, Rand Corporation. 2013.

Groves, D. G., Water Research Foundation, New York State Energy Research and Development Authority and WSAA (Association): Developing robust strategies for climate change and other risks: a water utility framework. 2014.

Guzman, C. D., Tilahun, S. A., Zegeye, A. D. and Steenhuis, T. S.: Suspended sediment concentration–discharge relationships in the (sub-) humid Ethiopian highlands, Hydrology and Earth System Sciences, 17(3), 1067–1077, doi:10.5194/hess-17-1067-2013, 2013.

Haasnoot, M., Middelkoop, H., van Beek, E. and van Deursen, W. P. A.: A method to develop sustainable water management strategies for an uncertain future, Sustainable Development, 19(6), 369, 2011.

Haile, A. T., Rientjes, T., Gieske, A. and Gebremichael, M.: Rainfall Variability over Mountainous and Adjacent Lake Areas: The Case of Lake Tana Basin at the Source of the Blue Nile River, Journal of Applied Meteorology and Climatology, 48(8), 1696–1717, doi:10.1175/2009JAMC2092.1, 2009.

Ben-Haim, Y.: Robust rationality and decisions under severe uncertainty, Journal of the Franklin Institute, 337(2), 171–199, 2000a.

Hallegatte, S. and Rentschler, J.: Risk Management for Development-Assessing Obstacles and Prioritizing Action: Risk Management for Development, Risk Analysis, 35(2), 193–210, doi:10.1111/risa.12269, 2015.

Hallegatte, S., Shah, A., Brown, C., Lempert, R. and Gill, S.: Investment decision making under deep uncertainty–application to climate change, World Bank Policy Research Working Paper, (6193), 2012.

Hall, J. W., Lempert, R. J., Keller, K., Hackbarth, A., Mijere, C. and McInerney, D. J.: Robust Climate Policies Under Uncertainty: A Comparison of Robust Decision Making and Info-Gap Methods, Risk Analysis: An International Journal, 32(10), 1657–1672, 2012a.

Hall, J. W., Watts, G., Keil, M., Vial, L. de, Street, R., Conlan, K., O'Connell, P. E., Beven, K. J. and Kilsby, C. G.: Towards risk-based water resources planning in England and Wales under a changing climate, Water and Environment Journal, 26(1), 118–129, 2012b.

Han, D., Kwong, T. and Li, S.: Uncertainties in real-time flood forecasting with neural networks, Hydrological Processes, 21(2), 223–228, doi:10.1002/hyp.6184, 2007.

Haregeweyn, N., Melesse, B., Tsunekawa, A., Tsubo, M., Meshesha, D. and Balana, B. B.: Reservoir sedimentation and its mitigating strategies: a case study of Angereb reservoir (NW Ethiopia), Journal of Soils and Sediments, 12(2), 291–305, doi:10.1007/s11368-011-0447-z, 2012.

Harris, I., Jones, P. d., Osborn, T. j. and Lister, D. h.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34(3), 623–642, doi:10.1002/joc.3711, 2014.

Hastie, T. and Tibshirani, R.: Generalized Additive Models, Statistical Science, 1(3), 297–310, 1986.

Hastie, T., Tibshirani, R. and Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference and Prediction, Second., Springer, New York., 2009.

Herman, J. D., Zeff, H. B., Reed, P. M. and Characklis, G. W.: Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty, Water Resour. Res., 50(10), 7692–7713, doi:10.1002/2014WR015338, 2014.

Herman, J. D., Zeff, H. B. and Characklis, G. W.: How Should Robustness Be Defined for Water Systems Planning under Change?, Journal of Water Resources Planning and Management, 0(0), 04015012, doi:10.1061/(ASCE)WR.1943-5452.0000509, 2015.

Hughes, D. A.: Comparison of satellite rainfall data with observations from gauging station networks, Journal of Hydrology, 327(3-4), 399–410, doi:10.1016/j.jhydrol.2005.11.041, 2006.

Iorgulescu, I. and Beven, K. J.: Nonparametric direct mapping of rainfall-runoff relationships: An alternative approach to data analysis and modeling?, Water Resour. Res., 40(8), W08403, doi:10.1029/2004WR003094, 2004.

Jain, A., Sudheer, K. P. and Srinivasulu, S.: Identification of physical processes inherent in artificial neural network rainfall runoff models, Hydrological Processes, 18(3), 571–581, doi:10.1002/hyp.5502, 2004.

Jeuland, M. and Whittington, D.: Water resources planning under climate change: Assessing the robustness of real options for the Blue Nile, Water Resour. Res., 50(3), 2086–2107, doi:10.1002/2013WR013705, 2014.

Jung, I.-W., Moradkhani, H. and Chang, H.: Uncertainty assessment of climate change impacts for hydrologically distinct river basins, Journal of Hydrology, 466–467, 73–87, doi:10.1016/j.jhydrol.2012.08.002, 2012.

Kalra, N. R., Groves, D. G., Bonzanigo, L., Molina Perez, E., Ramos, C., Carter, B. and Rodriguez Cabanillas, I.: Robust Decision-Making in the Water Sector, World Bank Group Climate Change Group. [online] Available from: https://wdronline.worldbank.org/handle/10986/22861 (Accessed 11 February 2016), 2015.

Kaplan, S. and Garrick, B. J.: On the quantitative definition of risk, Risk analysis, 1(1), 11–27, 1981.

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

Kebede, S., Travi, Y., Alemayehu, T. and Marc, V.: Water balance of Lake Tana and its sensitivity to fluctuations in rainfall, Blue Nile basin, Ethiopia, Journal of Hydrology, 316(1-4), 233–247, doi:10.1016/j.jhydrol.2005.05.011, 2006.

Kharin, V. V. and Zwiers, F. W.: Estimating Extremes in Transient Climate Change Simulations, J. Climate, 18(8), 1156–1173, doi:10.1175/JCLI3320.1, 2005.

Klinke, A. and Renn, O.: A New Approach to Risk Evaluation and Management: Risk-Based, Precaution-Based, and Discourse-Based Strategies1, Risk Analysis, 22(6), 1071–1094, 2002.

Knutti, R., Abramowitz, G., Collins, M., Eyring, V., Gleckler, P. J., Hewitson, B., Mearns, L., Stocker, T. and Dahe, Q.: IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections, 2010.

Kuhn, M.: caret: Classification and regression training, [online] Available from: http://CRAN.R-project.org/package=caret, 2015.

Kuhn, M., Weston, S., Keefer, C. and Coulter, N.: Cubist: Rule- and instance-based regression modeling, [online] Available from: http://CRAN.R-project.org/package=Cubist, 2014.

Kundzewicz, Z. W. and Stakhiv, E. Z.: Are climate models "ready for prime time" in water resources management applications, or is more research needed?, Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55(7), 1085–1089, 2010.

Kunreuther, H., Heal, G., Allen, M., Edenhofer, O., Field, C. B. and Yohe, G.: Risk management and climate change, Nature Climate Change, 3(5), 447–450, doi:10.1038/nclimate1740, 2013.

Legates, D. R. and McCabe Jr, G. J.: Evaluating the use of" goodness-of-fit" measures in hydrologic and hydroclimatic model validation, Water resources research, 35(1), 233–241, 1999.

Lempert, R.: Scenarios that illuminate vulnerabilities and robust responses, Climatic Change, 1–20, 2013.

Lempert, R., Sriver, R. L. and Keller, K.: Characterizing Uncertain Sea Level Rise Projections to Support Investment Decisions, California Energy Commission., 2012.

Lempert, R., Kalra, N., Peyraud, S., Mao, Z., Tan, S. B., Cira, D. and Lotsch, A.: Ensuring robust flood risk management in Ho Chi Minh City, World Bank Policy Research working paper, (6465) [online] Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2271955 (Accessed 11 August 2015), 2013.

Lempert, R. J. and Collins, M. T.: Managing the Risk of Uncertain Threshold Responses: Comparison of Robust, Optimum, and Precautionary Approaches, Risk Analysis, 27(4), 1009–1026, 2007.

Lempert, R. J. and Groves, D. G.: 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, 2010.

Lempert, R. J. and Kalra, N.: Managing climate risks in developing countries with robust decision making, 2011.

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

Liaw, A. and Wiener, M.: Classification and regression by randomForest, R News, 2(3), 18–22, 2002.

Lin, J.-Y., Cheng, C.-T. and Chau, K.-W.: Using support vector machines for long-term discharge prediction, Hydrological Sciences Journal, 51(4), 599–612, doi:10.1623/hysj.51.4.599, 2006.

Lin, L., Nilsson, A., Sjölin, J., Abrahamsson, M. and Tehler, H.: On the perceived usefulness of risk descriptions for decision-making in disaster risk management, Reliability Engineering & System Safety, 142, 48–55, doi:10.1016/j.ress.2015.04.012, 2015.

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), J. Hydrometeor, 7(2), 217–234, doi:10.1175/JHM486.1, 2006.

Machado, F., Mine, M., Kaviski, E. and Fill, H.: Monthly rainfall–runoff modelling using artificial neural networks, Hydrological Sciences Journal, 56(3), 349–361, doi:10.1080/02626667.2011.559949, 2011.

Maier, H. R., Jain, A., Dandy, G. C. and Sudheer, K. P.: Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions, Environmental Modelling & Software, 25(8), 891–909, doi:10.1016/j.envsoft.2010.02.003, 2010.

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R. and others: Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties, Intergovernmental Panel on Climate Change (IPCC), 2010.

Mathevet, T., Michel, C., Andreassian, V. and Perrin, C.: A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins, in Large Sample Basin Experiment for Hydrological Model Parameterization: Results of the Model Parameter Experiment – MOPEX, vol. 307, IAHS Publish., 2006.

Matrosov, E., Woods, A. and Harou, J.: Robust Decision Making and Info-Gap Decision Theory for Water Resource System Planning, Journal of Hydrology, 2013a.

Matrosov, E. S., Padula, S. and Harou, J. J.: 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, 2013b.

McCartney, M. P., Shiferaw, A. and Seleshi, Y.: Estimating environmental flow requirements downstream of the Chara Chara weir on the Blue Nile River, Hydrol. Process., 23(26), 3751–3758, doi:10.1002/hyp.7254, 2009.

McJeon, H. C., Clarke, L., Kyle, P., Wise, M., Hackbarth, A., Bryant, B. P. and Lempert, R. J.: Technology interactions among low-carbon energy technologies: What can we learn from a large number of scenarios?, Energy Economics, 33(4), 619–631, doi:10.1016/j.eneco.2010.10.007, 2011.

McSweeney, C., Lizcano, G., New, M. and Lu, X.: The UNDP Climate Change Country Profiles: Improving the Accessibility of Observed and Projected Climate Information for Studies of Climate Change in Developing Countries, Bulletin of the American Meteorological Society, 91(2), 157–166, doi:10.1175/2009BAMS2826.1, 2010.

Mekonnen, M. A., Wörman, A., Dargahi, B. and Gebeyehu, A.: Hydrological modelling of Ethiopian catchments using limited data, Hydrological Processes, 23(23), 3401–3408, doi:10.1002/hyp.7470, 2009.

Melillo, J. M., Richmond, T. C. and Yohe, G. H.: Climate Change Impacts in the United States: The Third National Climate Assessment, 2014.

Merz, R., Parajka, J. and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, Water Resour. Res., 47(2), W02531, doi:10.1029/2010WR009505, 2011.

Milborrow, S.: earth: Multivariate Adaptive Regression Splines, [online] Available from: http://CRAN.R-project.org/package=earth, 2015.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P. and Stouffer, R. J.: Stationarity Is Dead: Whither Water Management?, Science, 319(5863), 573–574, 2008.

Mintesinot, B., Verplancke, H., Van Ranst, E. and Mitiku, H.: Examining traditional irrigation methods, irrigation scheduling and alternate furrows irrigation on vertisols in northern Ethiopia, Agricultural Water Management, 64(1), 17–27, doi:10.1016/S0378-3774(03)00194-X, 2004.

Mitchell, T. D. and Hulme, M.: Predicting regional climate change: living with uncertainty, Progress in Physical Geography, 23(1), 57–78, 1999.

Montgomery, D. C., Peck, E. A. and Vining, G. G.: Introduction to Linear Regression Analysis, 5th Edition, John Wiley & Sons. [online] Available from: http://proquestcombo.safaribooksonline.com.proxy1.library.jhu.edu/book/math-and-science/9780470542811 (Accessed 17 August 2015), 2012.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. and Veith, T. L.: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Trans. Asabe, 50(3), 885–900, 2007.

Mosleh, A. and Bier, V. M.: Uncertainty about probability: a reconciliation with the subjectivist viewpoint, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 26(3), 303–310, doi:10.1109/3468.487956, 1996.

Nassopoulos, H., Dumas, P. and Hallegatte, S.: Adaptation to an uncertain climate change: cost benefit analysis and robust decision making for dam dimensioning, Climatic Change, 114(3-4), 497–508, 2012.

National Research Council: Informing Decisions in a Changing Climate, The National Academies Press, Washington, D.C., 2009.

New Zealand Climate Change Centre: Climate change adaptation in New Zealand: Future scenarios and some sectoral perspectives. Nottage, R.A.C., Wratt, D.S., Bornman, J.F., Jones, K. (eds)., Wellington, New Zealand., 2010.

Parker, A. M., Srinivasan, S. V., Lempert, R. J. and Berry, S. H.: Evaluating simulation-derived scenarios for effective decision support, Technological Forecasting and Social Change, 91, 64–77, doi:10.1016/j.techfore.2014.01.010, 2015.

Parson, E. A.: Useful global-change scenarios: current issues and challenges, Environmental Research Letters, 3(4), 045016, 2008.

Paté-Cornell, M. E.: Uncertainties in risk analysis: Six levels of treatment, Reliability Engineering & System Safety, 54(2), 95–111, 1996.

Penman, H. L.: Natural Evaporation from Open Water, Bare Soil and Grass, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 193(1032), 120–145, doi:10.1098/rspa.1948.0037, 1948.

Poccard, I., Janicot, S. and Camberlin, P.: Comparison of rainfall structures between NCEP/NCAR reanalyses and observed data over tropical Africa, Climate Dynamics, 16(12), 897–915, 2000.

Popper, S. W., Ed.: Natural gas and Israel's energy future: near-term decisions from a strategic perspective, RAND, Santa Monica, CA., 2009.

Pushpalatha, R., Perrin, C., Moine, N. L. and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420–421, 171–182, doi:10.1016/j.jhydrol.2011.11.055, 2012.

Quinlan, J. R.: Learning with Continuous Classes, in Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, World Scientific, Singapore., 1992.

Ranger, N., Reeder, T. and Lowe, J.: Addressing "deep" uncertainty over long-term climate in major infrastructure projects: four innovations of the Thames Estuary 2100 Project, EURO J Decis Process, 1(3-4), 233–262, doi:10.1007/s40070-013-0014-5, 2013.

R Development Core Team: R: A language and environment for statistical computing., R Foundation for Statistical Computing, Vienna, Austria. [online] Available from: http://www.R-project.org, 2014.

Reid, S. G.: Acceptable risk, Engineering safety, 138–166, 1992.

Rientjes, T. H. M., Haile, A. T., Kebede, E., Mannaerts, C. M. M., Habib, E. and Steenhuis, T. S.: Changes in land cover, rainfall and stream flow in Upper Gilgel Abbay catchment, Blue Nile basin – Ethiopia, Hydrology and Earth System Sciences, 15(6), 1979–1989, doi:10.5194/hess-15-1979-2011, 2011.

Ripley, B. D.: Pattern Recognition and Neural Networks, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 1996.

Schaefli, B. and Gupta, H. V.: Do Nash values have value?, Hydrol. Process., 21(15), 2075–2080, doi:10.1002/hyp.6825, 2007.

Schuol, J. and Abbaspour, K. C.: Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa, Advances in Geosciences, 9, 137–143, 2006.

See, L., Solomatine, D. P., Abrahart, R. and Toth, E.: Hydroinformatics: computational intelligence and technological developments in water science applications—Editorial, Hydrological Sciences Journal, 52(3), 391–396, doi:10.1623/hysj.52.3.391, 2007.

Setegn, S. G., Srinivasan, R., Melesse, A. M. and Dargahi, B.: SWAT model application and prediction uncertainty analysis in the Lake Tana Basin, Ethiopia, Hydrol. Process., 24(3), 357–367, doi:10.1002/hyp.7457, 2010.

Setegn, S. G., Rayner, D., Melesse, A. M., Dargahi, B. and Srinivasan, R.: Impact of climate change on the hydroclimatology of Lake Tana Basin, Ethiopia, Water Resources Research, 47(4), n/a–n/a, doi:10.1029/2010WR009248, 2011.

Sheffield, J., Goteti, G. and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, J. Climate, 19(13), 3088–3111, doi:10.1175/JCLI3790.1, 2006.

Shortridge, J., Guikema, S. D. and Zaitchik, B. F.: Data-driven streamflow prediction for climate change impact analysis, Hydrology and Earth System Sciences Discussions, 2016.

Shortridge, J. E. and Guikema, S. D.: Scenario discovery with multiple criteria: an evaluation of the robust decision making framework for climate adaptation, Risk Analysis, in press, 2016.

Shortridge, J. E., Falconi, S. M., Zaitchik, B. F. and Guikema, S. D.: Climate, agriculture, and hunger: statistical prediction of undernourishment using nonlinear regression and data-mining techniques, Journal of Applied Statistics, 0(0), 1–24, doi:10.1080/02664763.2015.1032216, 2015.

Sieber, J. and Purkey, D.: WEAP (Water Evaluation and Planning System) User Guide, Stockholm Environment Institute, Somerville, MA, USA. [online] Available from: http://www.weap21.org/ (Accessed 1 May 2015), 2015.

Simane, B., Zaitchik, B. F. and Mesfin, D.: Building Climate Resilience in the Blue Nile/Abay Highlands: A Framework for Action, International Journal of Environmental Research and Public Health, 9(12), 610–631, doi:10.3390/ijerph9020610, 2012.

SMEC International: Hydrological Study of the Tana-Beles Sub-Basins., 2008.

Solomatine, D. P. and Ostfeld, A.: Data-driven modelling: some past experiences and new approaches, Journal of Hydroinformatics, 10(1), 3, doi:10.2166/hydro.2008.015, 2008.

Sorooshian, S. and Dracup, J. A.: Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases, Water Resour. Res., 16(2), 430–442, doi:10.1029/WR016i002p00430, 1980.

Soundappan, P., Nikolaidis, E., Haftka, R. T., Grandhi, R. and Canfield, R.: Comparison of evidence theory and Bayesian theory for uncertainty modeling, Reliability engineering & System safety, 85(1), 295–311, 2004.

Stakhiv, E. Z.: Pragmatic Approaches for Water Management Under Climate Change Uncertainty1, JAWRA Journal of the American Water Resources Association, 47(6), 1183–1196, 2011.

Steenhuis, T. S., Collick, A. S., Easton, Z. M., Leggesse, E. S., Bayabil, H. K., White, E. D., Awulachew, S. B., Adgo, E. and Ahmed, A. A.: Predicting discharge and sediment for the Abay (Blue Nile) with a simple model, Hydrological Processes, n/a–n/a, doi:10.1002/hyp.7513, 2009.

Stephenson, D. B., Collins, M., Rougier, J. C. and Chandler, R. E.: Statistical problems in the probabilistic prediction of climate change, Environmetrics, 23(5), 364–372, 2012.

Stillwell, W. G. and Edwards, W.: Rank Weighting in Multiattribute Utility Decision Making: Avoiding the Pitfalls of Equal Weights., University of Southern California Social Science Research Institute, Los Angeles, CA, USA., 1979.

Sudheer, K. P. and Jain, A.: Explaining the internal behaviour of artificial neural network river flow models, Hydrological Processes, 18(4), 833–844, doi:10.1002/hyp.5517, 2004.

Tamene, L. and Vlek, P. L. G.: Assessing the potential of changing land use for reducing soil erosion and sediment yield of catchments: a case study in the highlands of northern

Ethiopia, Soil Use and Management, 23(1), 82–91, doi:10.1111/j.1475-2743.2006.00066.x, 2007.

Tang, S. and Dessai, S.: Usable Science? The U.K. Climate Projections 2009 and Decision Support for Adaptation Planning, Weather, Climate, and Society, 4(4), 300–313, doi:10.1175/WCAS-D-12-00028.1, 2012.

Tebaldi, C. and Knutti, R.: The use of the multi-model ensemble in probabilistic climate projections, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1857), 2053–2075, 2007.

Tebaldi, C., Smith, R. L., Nychka, D. and Mearns, L. O.: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, Journal of Climate, 18(10), 1524–1540, 2005.

Tucker, W. T. and Ferson, S.: Probability bounds analysis in environmental risk assessment, Applied Biomathematics, Setauket, New York., 2003.

Van Griensven, A., Ndomba, P., Yalew, S. and Kilonzo, F.: Critical review of SWAT applications in the upper Nile basin countries, Hydrology and Earth System Sciences, 16(9), 3371–3381, doi:10.5194/hess-16-3371-2012, 2012.

Van Oldenborgh, G. ., Collins, M., Arblaster, J., Christensen, J. H., Marotzke, S. ., Power, S. B., Rummukainen, M. and Zhou, T.: Annex I: Atlas of Global and Regional Climate Projections, in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 1311–1394, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S-PLUS, Springer Science & Business Media., 2013.

Walley, P.: Statistical reasoning with imprecise probabilities, Chapman and Hall London., 1991.

Wang, P., Dong, S. and Lassoie, J. P.: The Large Dam Dilemma: An Exploration of the Impacts of Hydro Projects on People and the Environment in China, Springer Science & Business Media., 2013.

Washington, R., Harrison, M., Conway, D., Black, E., Challinor, A., Grimes, D., Jones, R., Morse, A., Kay, G. and Todd, M.: African Climate Change: Taking the Shorter Route, Bulletin of the American Meteorological Society, 87(10), 1355–1366, doi:10.1175/BAMS-87-10-1355, 2006.

Water Works Design & Supervision Enterprise (WWDSE): Gumara Irrigation Project Feasibility Study Report., 2008.

Water Works Design & Supervision Enterprise (WWDSE) and Tahal Group: Irrigation and Drainage Projects in Lake Tana Sub-Basin: Gilgel Abbay Final Feasibility and Detailed Design Report., 2009a.

Water Works Design & Supervision Enterprise (WWDSE) and Tahal Group: Irrigation and Drainage Projects in Lake Tana Sub-Basin: Jema Final Feasibility and Detailed Design Report., 2009b.

Water Works Design & Supervision Enterprise (WWDSE) and Tahal Group: Irrigation and Drainage Projects in Lake Tana Sub-Basin: Megech Final Feasibility and Detailed Design Report., 2009c.

Water Works Design & Supervision Enterprise (WWDSE) and Tahal Group: Irrigation and Drainage Projects in Lake Tana Sub-Basin: Ribb Final Feasibility and Detailed Design Report., 2009d.

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

Whateley, S., Steinschneider, S. and Brown, C.: A climate change range-based method for estimating robustness for water resources supply, Water Resources Research, 50(11), 8944–8961, doi:10.1002/2014WR015956, 2014.

Wilby, R. L.: Uncertainty in water resource model parameters used for climate change impact assessment, Hydrol. Process., 19(16), 3201–3219, doi:10.1002/hyp.5819, 2005.

Wilby, R. L. and Dessai, S.: Robust adaptation to climate change, Weather, 65(7), 180–185, doi:10.1002/wea.543, 2010.

Wilby, R. L., Abrahart, R. J. and Dawson, C. W.: Detection of conceptual model rainfall—runoff processes inside an artificial neural network, Hydrological Sciences Journal, 48(2), 163–181, doi:10.1623/hysj.48.2.163.44699, 2003.

Williamson, R. C. and Downs, T.: Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds, International Journal of Approximate Reasoning, 4(2), 89–158, 1990.

Wood, S. N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(1), 3–36, doi:10.1111/j.1467-9868.2010.00749.x, 2011.

Wood, S. N.: On p-values for smooth components of an extended generalized additive model, Biometrika, ass048, doi:10.1093/biomet/ass048, 2012.

Woodward, M., Kapelan, Z. and Gouldby, B.: Adaptive Flood Risk Management Under Climate Change Uncertainty Using Real Options and Optimization, Risk Analysis: An International Journal, 34(1), 75–92, 2014.

World Commission on Dams: Dams and Development: A new framework for decision-making, Earthscan Publications Ltd, London and Sterling, Virginia., 2000.

Yates, D., Sieber, J., Purkey, D. and Huber-Lee, A.: WEAP21—A Demand-, Priority-, and Preference-Driven Water Planning Model: Part 1: Model Characteristics, Water International, 30(4), 487–500, doi:10.1080/02508060508691893, 2005.

Yu, J., Qin, X., Larsen, O. and Chua, L. H. C.: Comparison between response surface models and artificial neural networks in hydrologic forecasting, Journal of Hydrologic Engineering, 19(3), 473–481, 2013.

VITA

Julie Shortridge was born in Cincinnati, Ohio on March 10, 1983. She grew up in Terrace Park, OH, Mt. Pleasant, SC and Tualatin, OR, graduating from Tualatin High School in 2001. After receiving a B.S. in Environmental Engineering Science from the University of California – Berkeley in 2005, she went on to work at ARCADIS U.S. as an environmental engineer/project manager and for the United Nations Environment Programme as a consultant on community-level disaster preparedness. She joined the Department of Geography and Environmental Engineering at Johns Hopkins University in August, 2011, receiving her M.S.E. in 2012 and completing her Ph.D. in 2016. Her research focuses on adaptation to climate change risks and the accompanying issue of deep uncertainty, with a particular focus on water infrastructure and other coupled human-natural systems. She is interested using systems engineering methodologies, including risk and decision analysis, optimization, simulation, and statistical modeling, to address these issues. Following completion of her doctorate, Julie will be joining the Biological Systems Engineering department at Virginia Tech as an assistant professor.