Risk Analysis, Vol. 36, No. 12, 2016 DOI: 10.1111/risa.12582

Scenario Discovery with Multiple Criteria: An Evaluation of the Robust Decision-Making Framework for Climate Change Adaptation

Julie E. Shortridge[1,*] and Seth D. Guikema[2]

There is increasing concern over deep uncertainty in the risk analysis field as probabilistic models of uncertainty cannot always be confidently determined or agreed upon for many of our most pressing contemporary risk challenges. This is particularly true in the climate change adaptation field, and has prompted the development of a number of frameworks aiming to characterize system vulnerabilities and identify robust alternatives. One such methodology is robust decision making (RDM), which uses simulation models to assess how strategies perform over many plausible conditions and then identifies and characterizes those where the strategy fails in a process termed scenario discovery. While many of the problems to which RDM has been applied are characterized by multiple objectives, research to date has provided little insight into how treatment of multiple criteria impacts the failure scenarios identified. In this research, we compare different methods for incorporating multiple objectives into the scenario discovery process to evaluate how they impact the resulting failure scenarios. We use the Lake Tana basin in Ethiopia as a case study, where climatic and environmental uncertainties could impact multiple planned water infrastructure projects, and find that failure scenarios may vary depending on the method used to aggregate multiple criteria. Common methods used to convert multiple attributes into a single utility score 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.

KEY WORDS: Climate change; deep uncertainty; robust decision making

1. INTRODUCTION

In recent years, there has been increasing concern and discussion over deep uncertainty in the risk analysis field.(1) The term "deep uncertainty" is commonly used to refer to situations where probabilistic models of uncertainty cannot be confidently determined or agreed upon(1) or where frequentist probabilities based on repeatable events cannot be developed.(2) 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.(3) This has led to interest in robust decision frameworks,(4) which include methods such as robust decision making (RDM),(5)

[1]Department of Geography and Environmental Engineering, The Johns Hopkins University, Baltimore, MD, USA.
[2]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA.
[*]Address correspondence to Julie Shortridge, Department of Geography and Environmental Engineering, The Johns Hopkins University, 3400 N. Charles Street, Ames Hall 317, Baltimore, MD 21218, USA; jshortridge@jhu.edu.

decision scaling,[6] and info-gap decision theory.[7] 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.[4] These frameworks can be particularly useful in situations characterized by poorly understood nonlinear or threshold responses[8] or many stakeholders with conflicting values and beliefs about the future.[9]

RDM is one such framework that has been applied to a number of climate adaptation problems.[10–13] It is a multistep, iterative approach that includes both analytical and deliberative components.[5] 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.[5] 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.[14] Scenario discovery uses the patient rule induction method (PRIM)[15] 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.[10,14] 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 decisionmakers 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,[10,16] system reliability,[17] expected utility,[18] or a single aggregated performance score.[5] A number of evaluations that do consider multiple criteria apply scenario discovery over each criterion separately.[12,17,19,20] 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.[11,21–23] 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 decisionmakers.

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 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 that 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.

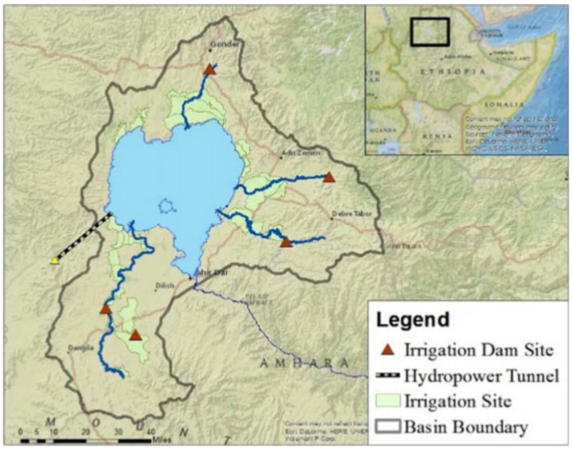

Fig. 1. Map of Lake Tana, the surrounding watershed, and planned infrastructure developments.

2. METHODS

2.1. Study Area

Lake Tana is the source of the Blue Nile River, located in the highlands of northwest Ethiopia at an elevation of approximately 1,790 m. The lake has a surface area of approximately 3,000 square kilometers, and the catchment draining to the lake encompasses approximately 12,000 square kilometers (Fig. 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.(24) 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 1,000 mm/year to over 1,800 mm/year.(25) 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. Population growth and expansion of agricultural and pastoral land use in the region have resulted in substantial deforestation and land degradation.(26–28)

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-m 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 7,000 hectares. There are five other reservoirs being planned for construction in the basin, ranging in volume from 80 to 220 MCM.(24) 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. A summary of existing and planned water infrastructure is shown in Table I.

While these projects have the potential to generate important economic growth for the region, there are a number of uncertainties that could impact their performance in the future. Climate change could dramatically impact the amount of water available in the basin, particularly considering the long life-span of the proposed infrastructure. However, projected changes in climate for Ethiopia are highly uncertain, with climate models disagreeing on even the direction of precipitation change.(29) Land cover in the basin has been dramatically altered over the past few decades, and could change further due to increasing agricultural development or expansion of conservation efforts. This impacts the amount and timing of runoff in the basin's rivers, as well as amount of sediment that will be introduced to them. Finally, there are few data available in the basin with which to estimate certain operational parameters of this infrastructure even under current conditions. These data limitations make it difficult to predict future rates of reservoir sedimentation and evaporative losses with any degree of confidence.

2.2. Simulation Models

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 data sets(30) and agricultural land cover as reported by data taken from

Table I. Existing and Proposed Water Resources Infrastructure in the Lake Tana Basin

Project Name	Type	Annual Demand (MCM) Min	Annual Demand (MCM) Max	Reservoir Capacity (MCM)	Average Annual Flow into Dam Site (MCM)	Catchment Area (km^2)	Irrigable Area (km^2)
Existing Projects							
Koga	Irrigation reservoir	62	86	83	114	185	70
Tana-Beles	Hydropower tunnel	2681	2681	NA	NA	NA	NA
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
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

Rientjes *et al.*,[26] Gebrehiwot *et al.*,[27] and Garede and Minale.[28] 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,[31] artificial neural network,[32] generalized additive model,[33] and random forest model.[34] Each basin's model was compared to a climatology model that 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 are discussed by Shortridge *et al.*[35]

The second component of the simulation model was a water evaluation and planning (WEAP[36]) water allocation model developed for the basin by Alemayehu *et al.*[24] 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.[36] 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.*[24]

2.3. 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 II). 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.

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 multimodel ensemble projections for the East Africa region for the period 2081–2100 under all representative concentration pathways.[29] Additionally, there is concern that climate change could result in an intensification of precipitation, even when overall amounts of precipitation decrease.[37–39] 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

Table II. Uncertain Parameters

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

sampling results from various rivers in the basin,[40,41] while future agricultural land cover was assumed to range from 50% to 90% based on values experienced over the past 50 years.[26–28] 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, 5,000 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[42] 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,[12,43] 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 5,000 samples, the change in precipitation was randomly selected to be either positive or negative with an equal probability. For the nth sample, a parameter $\overline{\Delta P_n}$ was calculated as in Equation (1) and a parameter $\overline{\Delta \mathrm{Int}_n}$was calculated as in Equation (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 $\overline{\Delta \mathrm{Int}_n}$, and a coefficient of variation equal to 0.5.

$$\overline{\Delta P_n} = \begin{cases} \frac{(\Delta T_n - \Delta T_{\min})}{(\Delta T_{\max} - \Delta T_{\min})} \times \Delta P_{\max} & \text{if} \quad \Delta P > 0 \\ \frac{(\Delta T_n - \Delta T_{\min})}{(\Delta T_{\max} - \Delta T_{\min})} \times \Delta P_{\min} & \text{if} \quad \Delta P < 0 \end{cases} \tag{1}$$

$$\overline{\Delta \mathrm{Int}_n} = \frac{(\Delta T_n - \Delta T_{\min})}{(\Delta T_{\max} - \Delta T_{\min})} \times \Delta \mathrm{Int}_{\max} \tag{2}$$

Each of the 5,000 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 5,000 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.[44] 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.[45] 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.*[46] 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.

Table III. Performance Metrics

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%

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 III). Previous studies have identified 1,784.75 m as the minimum elevation that Lake Tana can reach before negative impacts to the navigation and fishing industries begin to occur.[47] Alemayhu *et al.*[24] 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 III shows baseline results for each metric, assuming that the infrastructure was operated under historic climate conditions, an annual specific sediment yield of 1000 tons/km^2, 50% agricultural land cover, and an evaporation coefficient of 1.0. 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.

2.4. Scenario Discovery

The RDM framework is a multistep, 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.[5] The second analytical component, termed "scenario discovery," aims to identify the conditions that cause unsatisfactory performance in a preferred alternative. In this work, we use the scenario discovery process to identify the conditions that cause unsatisfactory performance for the proposed infrastructure described in Table I.

The scenario discovery process uses the results of the 5,000 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.[15] The objective of the PRIM algorithm is to find a region of the input variable space X that results in particularly low values of the output variable $Y = f(X)$. This region is made up of one or more "boxes" B that can be defined by simple logical conditions involving the value of individual input variables. To identify these boxes, the algorithm uses a process of top-down successive refinement, referred to as "peeling," followed by bottom-up successive expansion ("pasting"). The peeling phase begins with a box B containing all of the data. At each iteration, a small subbox b^* is removed, resulting in a smaller box equal to B-b^*. The subbox b^* chosen for removal is selected from a set of n candidate subboxes $C(b)$, each of which is defined by a single input variable x_j, to minimize the mean value of y within the resulting box.[15] This process is continued until the size of the box falls below a prespecified 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 subboxes to minimize the mean in the new larger box $B+b^*$. This process continues until the mean of the larger boxes starts to increase. This 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 $f(x)$ assumes low values.[15]

The PRIM algorithm was implemented using the SD toolkit package in R.[48] This package provides an interactive implementation of the PRIM algorithm on a binary output variable, thus identifying scenarios that result in performance y below some user-defined threshold. The package generates a trade-off curve showing the sequence of boxes identified during the peeling 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.[49] 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 trade-off 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.

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 IV. 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 multiattribute 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.*,[23] Herman *et al.*,[21,22] and Lempert *et al.*[11] This method is demonstrated in Equation (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 5,000 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 i, and w_i is the weight assigned to attribute i.

$$Y_n = \prod_{i=1}^{I} y_{i,n} \tag{3}$$

$$Y_n = \sum_{i=1}^{I} w_i u_{i,n} \tag{4}$$

Normalized weights were calculated using the rank sum weighting procedure based on four different possible rankings of attribute importance.[50] A summary of the four weighting schemes evaluated is presented in Table IV. For the additive aggregation schemes, an aggregated minimum acceptable performance threshold is calculated using Equation (4) and the performance thresholds presented in Table III.

3. RESULTS

A summary of the simulation results for each individual performance metric is presented in Table V. It is apparent that accounting for uncertainty in the parameters listed in Table II has 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 36–66% 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

Table IV. Weighting Schemes Used to Calculate Aggregate Performance Scores

	Multiplicative	Additive—Equal Weighting	Additive—Agricultural Priority	Additive—Hydropower Priority	Additive—Environmental Priority
Irrigation water reliability	NA	0.2	0.5	0.33	0.05
Hydropower water delivery	NA	0.2	0.33	0.5	0.05
Lake levels	NA	0.2	0.06	0.06	0.4
Tis Issat falls coverage	NA	0.2	0.06	0.06	0.2
Environmental flow coverage	NA	0.2	0.06	0.06	0.3
Acceptable performance threshold	1.0	0.75	0.87	0.88	0.69

Table V. Simulation Results

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%

information about the relative sensitivity of the different metrics to the uncertain parameters listed in Table II. 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.

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 VI 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 that 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 Fig. 2.

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

Table VI. Failure Scenarios for Individual Performance Metrics

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		

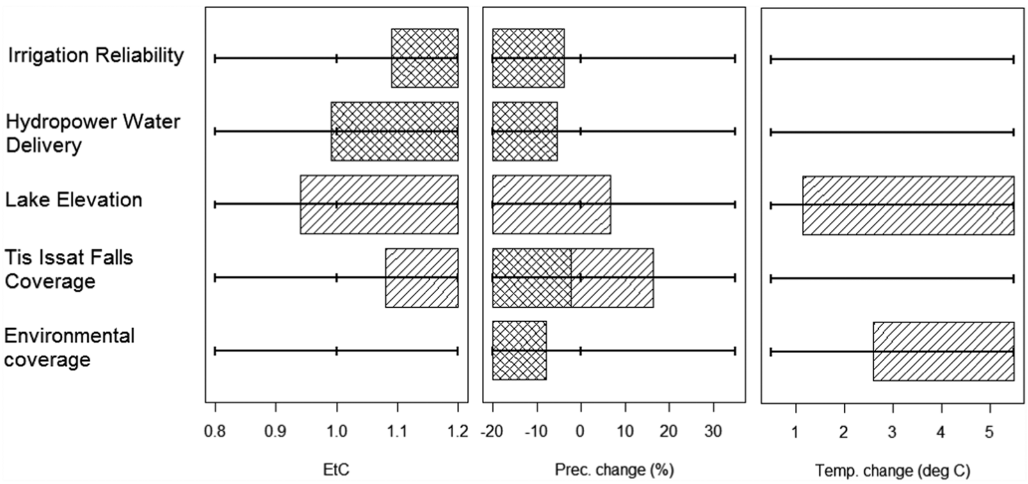

Fig. 2. Failure scenarios for individual performance metrics. Diagonal lines indicate a condition that has to occur in conjunction with specific conditions regarding the other parameters identified by diagonal lines. Boxes with hash marks indicate conditions that are sufficient to cause failure on their own, regardless of the values taken on by other parameters.

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.

Table VII and Fig. 3 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 multicriteria 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

Table VII. Failure Scenarios for Aggregated Performance Scores

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		

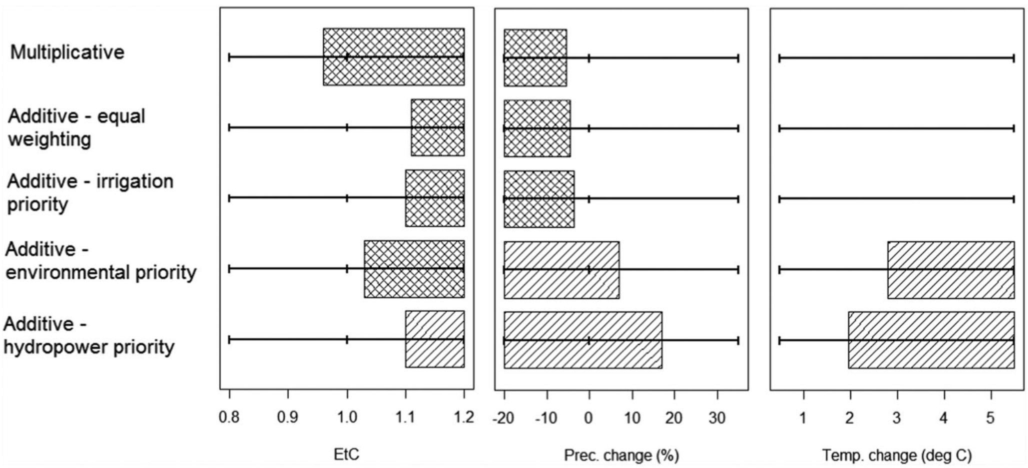

Fig. 3. Failure scenarios for aggregated performance scores. Diagonal lines and hash marks are as in Fig. 2.

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 Fig. 4 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[51]), 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.

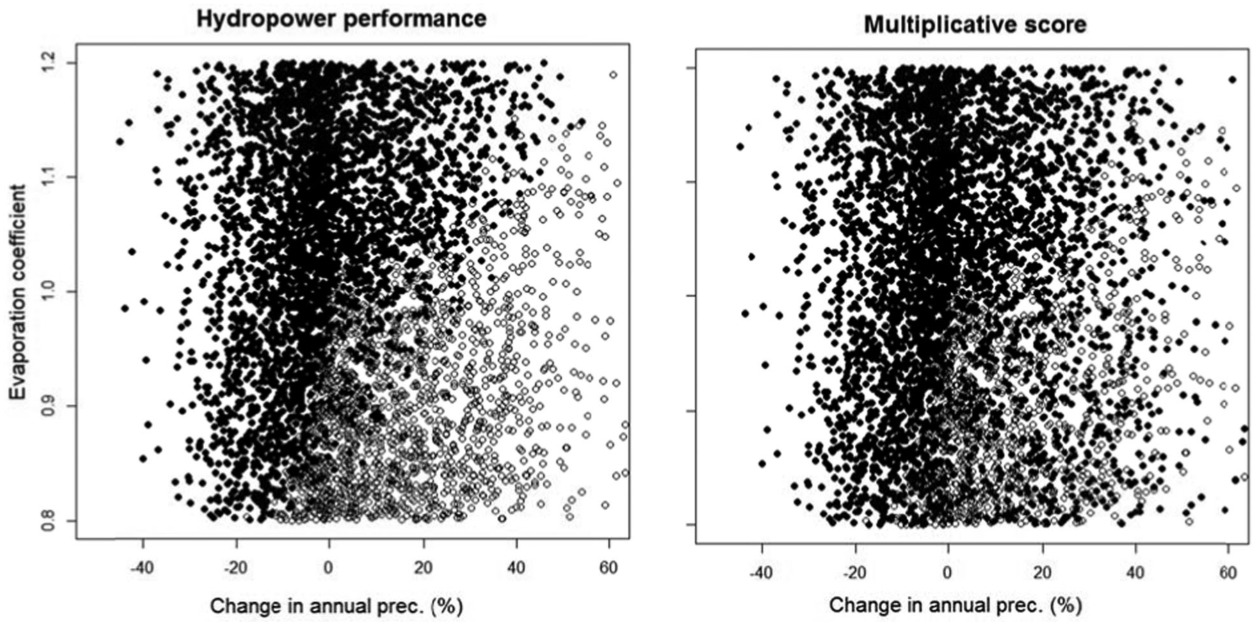

Fig. 4. Scatterplots showing simulations with hydropower performance and multiplicative aggregated performance above their respective thresholds. Filled-in circles represent simulations where the threshold was not met, and empty circles indicate simulations where it was.

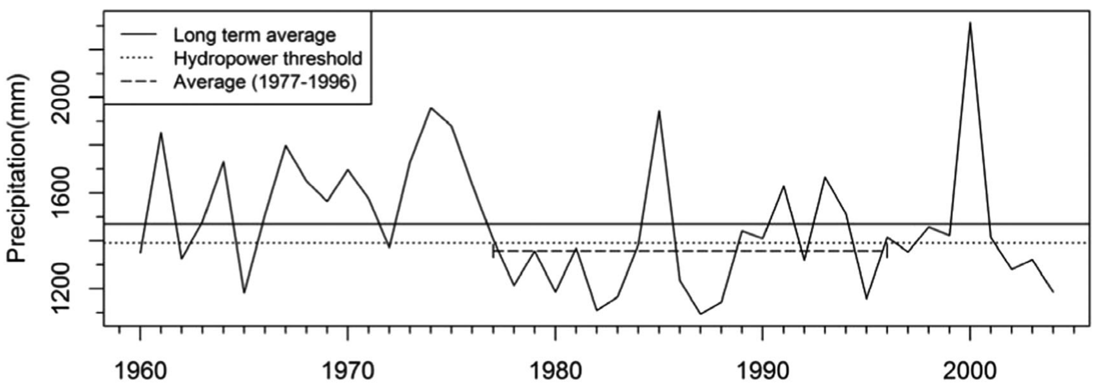

Fig. 5. Total annual precipitation (Gilgel Abbay). Horizontal lines indicate the thresholds for annual average precipitation identified by the PRIM algorithm.

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.[9] 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,[3] 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.[22] 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 (Fig. 5), 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 1,360 mm, which is approximately 8% less than the long-term average of 1,470 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.*[16] 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 decisionmakers 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[10] 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 of 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, decisionmakers 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.

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 was 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 decisionmakers 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 failures 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.

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 like 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.

ACKNOWLEDGMENTS

This research was supported by a National Defense Science and Engineering Graduate Fellowship and by National Science Foundation Grants 1149460 (CMMI), 1069213 (IGERT), and 1331399 (SEES). This research was conducted while Dr. Guikema was affiliated with the Department of Geography and Environmental Engineering at Johns Hopkins University. This support is gratefully acknowledged. The authors also thank the International Water Management Institute and the Tana Sub-Basin Organization for sharing data, models, and documentation related to the proposed infrastructure in Lake Tana. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding sources.

REFERENCES

1. Cox LAT. Confronting deep uncertainties in risk analysis. Risk Analysis, 2012; 32(10):1607–1629.
2. Aven T. On how to deal with deep uncertainties in a risk assessment and management context. Risk Analysis, 2013; 33(12):2082–2091.
3. Kunreuther H, Heal G, Allen M, Edenhofer O, Field CB, Yohe G. Risk management and climate change. Nature Climate Change, 2013 24; 3(5):447–450.
4. Weaver CP, Lempert RJ, Brown C, Hall JA, Revell D, 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, 2013; 4(1):39–60.
5. Lempert RJ, Groves DG, Popper SW, Bankes SC. A general, analytic method for generating robust strategies and narrative scenarios. Management Science, 2006; 52(4):514–528.
6. Brown C, Ghile Y, Laverty M, Li K. Decision scaling: Linking bottom-up vulnerability analysis with climate projections in the water sector. Water Resources Research, 2012; 48(9):W09537.
7. Ben-Haim Y. Robust rationality and decisions under severe uncertainty. Journal of the Franklin Institute, 2000; 337(2):171–199.
8. Lempert RJ, Collins MT. Managing the risk of uncertain threshold responses: Comparison of robust, optimum, and precautionary approaches. Risk Analysis, 2007; 27(4):1009–1026.
9. Hallegatte S, Rentschler J. Risk management for development-assessing obstacles and prioritizing action. Risk Analysis, 2015; 35(2):193–210.
10. Lempert RJ, Groves DG. Identifying and evaluating robust adaptive policy responses to climate change for water

management agencies in the American West. Technological Forecasting and Social Change, 2010; 77(6):960–974.
11. Lempert R, Kalra N, Peyraud S, Mao Z, Tan SB, Cira D, Lotsch A. Ensuring robust flood risk management in Ho Chi Minh City. World Bank Policy Research Working Paper, 2013. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2271955. Accessed August 10, 2015.
12. Groves DG, Fischbach JR, Bloom E, Knopman DS, Keefe R. Adapting to a Changing Colorado River: Making Future Water Deliveries More Reliable Through Robust Management Strategies. Santa Monica, CA: RAND, 2013.
13. Groves DG, Bloom E. Robust Water-Management Strategies for the California Water Plan Update, 2013. RR-182-DWR. Santa Monica, CA: RAND Corporation, 2013. Available at: http://www.waterplan.water.ca.gov/docs/swan/2013/RR182_FNLCompiled.pdf, Accessed April 16, 2015.
14. Lempert R. Scenarios that illuminate vulnerabilities and robust responses. Climatic Change, 2013; 1–20.
15. Friedman JH, Fisher NI. Bump hunting in high-dimensional data. Statistics and Computing, 1999; 9(2):123–143.
16. Lempert R, Sriver RL, Keller K. Characterizing Uncertain Sea Level Rise Projections to Support Investment Decisions. California Energy Commission, 2012.
17. Groves DG, Water Research Foundation, New York State Energy Research and Development Authority, WSAA (Association). Developing Robust Strategies for Climate Change and Other Risks: A Water Utility Framework, 2014. Available at: http://www.jstor.org/stable/10.7249/j.ctt14bs4mg, Accessed May 12, 2015.
18. Hall JW, Lempert RJ, Keller K, Hackbarth A, Mijere C, McInerney DJ. Robust climate policies under uncertainty: A comparison of robust decision making and info-gap methods. Risk Analysis, 2012; 32(10):1657–1672.
19. Popper SW, editor. Natural Gas and Israel's Energy Future: Near-Term Decisions from a Strategic Perspective. Santa Monica, CA: RAND, 2009.
20. Groves DG, Bloom E, Johnson DR, Yates D, 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.
21. Herman JD, Zeff HB, Reed PM, Characklis GW. Beyond optimality: Multistakeholder robustness tradeoffs for regional water portfolio planning under deep uncertainty. Water Resources Research, 2014; 50(10):7692–7713.
22. Herman JD, Zeff HB, Characklis GW. How should robustness be defined for water systems planning under change? Journal of Water Resources Planning and Management, 2015; 0(0):04015012.
23. Kasprzyk JR, Nataraj S, Reed PM, Lempert RJ. Many objective robust decision making for complex environmental systems undergoing change. Environmental Modelling & Software, 2013; 42:55–71.
24. Alemayehu T, McCartney M, Kebede S. The water resource implications of planned development in the Lake Tana catchment, Ethiopia. Ecohydrology & Hydrobiology, 2010; 10(2-4):211–221.
25. Achenef H, Tilahun A, Molla B. Tana Sub Basin Initial Scenarios and Indicators Development Report. Bahir Dar, Ethiopia: Tana Sub Basin Organization, 2013.
26. Rientjes THM, Haile AT, Kebede E, Mannaerts CMM, Habib E, Steenhuis TS. Changes in land cover, rainfall and stream flow in Upper Gilgel Abbay catchment, Blue Nile basin—Ethiopia. Hydrology and Earth System Sciences, 2011; 15(6):1979–1989.
27. Gebrehiwot SG, Taye A, Bishop K. Forest cover and stream flow in a headwater of the Blue Nile: Complementing observational data analysis with community perception. AMBIO, 2010; 39(4):284–294.
28. Garede NM, Minale AS. Land use/cover dynamics in Ribb Watershed, North Western Ethiopia. Journal of Natural Sciences Research, 2014; 4(16):9–16.
29. Van Oldenborgh G., Collins M, Arblaster J, Christensen JH, Marotzke S., Power SB, *et al*. Annex I: Atlas of global and regional climate projections. Pp. 1311–1394 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. Cambridge, UK and New York: Cambridge University Press, 2013.
30. Harris I, Jones PD, Osborn TJ, Lister DH. Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. International Journal of Climatology. 2014; 34(3):623–642.
31. Quinlan JR. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence. Singapore: World Scientific, 1992.
32. Ripley BD. Pattern Recognition and Neural Networks. Cambridge, UK: Cambridge University Press, 1996.
33. Hastie T, Tibshirani R. Generalized Additive Models. London: Chapman, Hall, 1990.
34. Breiman L. Random forests. Machine Learning. 2001; 45(1):5–32.
35. Shortridge JE, Guikema SD, and Zaitchik BF. Empirical streamflow simulation for water resource management in data-scarce seasonal watersheds. Hydrology and Earth Systems Sciences. Discussion paper available at: http://www.hydrol-earth-syst-sci-discuss.net/hess-2015-413/, Accessed February 7, 2016.
36. Sieber J, Purkey D. WEAP (Water Evaluation and Planning System) User Guide. Somerville, MA: Stockholm Environment Institute, 2015. Available at: http://www.weap21.org/, Accessed May 1, 2015.
37. Barnett DN, Brown SJ, Murphy JM, Sexton DMH, Webb MJ. Quantifying uncertainty in changes in extreme event frequency in response to doubled CO_2 using a large ensemble of GCM simulations. Climate Dynamics, 2006; 26(5):489–511.
38. Kharin VV, Zwiers FW. Estimating extremes in transient climate change simulations. Journal of Climate, 2005; 18(8):1156–1173.
39. Easterling DR, Meehl GA, Parmesan C, Changnon SA, Karl TR, Mearns LO. Climate extremes: Observations, modeling, and impacts. Science, 2000; 289(5487):2068–2074.
40. Water Works Design & Supervision Enterprise (WWDSE), Tahal Group. Irrigation and Drainage Projects in Lake Tana Sub-Basin: Gilgel Abbay Final Feasibility and Detailed Design Report, 2009.
41. Water Works Design & Supervision Enterprise (WWDSE). Gumara Irrigation Project Feasibility Study Report, 2008.
42. Bankes S. Exploratory modeling for policy analysis. Operations Research, 1993; 41(3):435–449.
43. McJeon HC, Clarke L, Kyle P, Wise M, Hackbarth A, Bryant BP, Lempert RJ. Technology interactions among low-carbon energy technologies: What can we learn from a large number of scenarios? Energy Economics, 2011; 33(4):619–631.
44. Gleick PH. Methods for evaluating the regional hydrologic impacts of global climatic changes. Journal of Hydrology, 1986; 88(1–2):97–116.
45. Penman HL. Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1948; 193(1032):120–145.
46. Kebede S, Travi Y, Alemayehu T, Marc V. Water balance of Lake Tana and its sensitivity to fluctuations in rainfall, Blue Nile basin, Ethiopia. Journal of Hydrology, 2006; 316(1-4):233–247.
47. SMEC International. Hydrological Study of the Tana-Beles Sub-Basins, 2008.

48. Bryant BJ. sdtoolkit: Scenario discovery tools to support robust decision making. R package version 2.33-1, 2014. Available at: http://CRAN.R-project.org/package=sdtoolkit, Accessed January 1, 2015.
49. Bryant BP, Lempert RJ. Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. Technological Forecasting and Social Change, 2010; 77(1):34–49.
50. Stillwell WG, Edwards W. Rank Weighting in Multiattribute Utility Decision Making: Avoiding the Pitfalls of Equal Weights. Los Angeles, CA: University of Southern California Social Science Research Institute, 1979.
51. Dalal S, Han B, Lempert R, Jaycocks A, Hackbarth A. Improving scenario discovery using orthogonal rotations. Environmental Modelling & Software, 2013; 48:49–64.