Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-221-RC2, 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Assessing water supply capacity in a complex river basin under climate change using the logistic eco-engineering decision scaling framework" *by* Daeha Kim et al.

Anonymous Referee #2

Received and published: 31 July 2018

General Comments:

This manuscript describes a method of extending a bottom-up climate risk assessment by using logistic regression to estimate the probability that a water system will meet minimum performance criteria over a planning horizon based on the values of climate variables. The method is demonstrated through a case study of water management in the Geum River Basin in South Korea. The Geum River is host to two dams which are managed for water supply, flood control, and environmental flows. The case study analyzes two alternative operating policies' ability to meet both water supply goals and instream flow requirements under a broad range of potential changes in average temperature, average precipitation, and precipitation variability. It is interesting to see the framework applied for multiple sub-basins within a larger system, and important to acknowledge uncertainty that an operating policy will meet a performance goal within specific climate scenarios.

The text is poorly written and organized, with many strangely used words that inhibit understanding. Key examples include “successive”, “sub-component”, and “risk of system failure,” which are applied in ways that are not standard in the literature and never clearly defined. Many crucial details related to the methods and motivation do not become clear until carefully examining the results section. For example, I believed the logistic model was simply modeling the water supply/environmental flow reliability as a function of climate variables rather than the risk of falling short of the reliability threshold until carefully examining the figures and results. This was the main point of the manuscript, so it is critically important that it is immediately apparent upon reading the abstract and within every part of the manuscript. The text requires substantial rewording and re-organization to clearly summarize the methodological contribution and motivation earlier in the text, better define scientific notation, and ensure new words and concepts are defined clearly the first time they are introduced.

While the goal of the logistic model is a worthy one, it is not clear that the framework has been well executed in the case study or that the novel technical contribution bears sufficient relationship to the EEDS framework to be named for it. This lack of clarity may be a symptom of the confused text. However, based on my understanding of the case study, the methods used to execute the case study are flawed in several important ways. Further, the interpretation of results relies on questionable assumptions related to the fitness of GCM projections for water system risk assessment. Both the manuscript and analysis require major revisions.

Specific Comments:

Logistic regression model: (1) Limited calibration set: It is my understanding that the logistic model was calibrated from 434 binary values that correspond to either water supply reliability or environmental flow reliability meeting a threshold under 434 unique combinations of three climate variables. If my understanding is correct, this would mean that there is one response (binary performance metric) per climate scenario (this should be clarified in the manuscript if that is incorrect). This is a very limited data set for analyzing risk of failure resulting from internal climate variability, especially given that each scenario-specific stochastic trace was (a) only 20 years long, and (b) initially identical to every other weather sequence in the analysis that had then been perturbed from the original trace to match a unique combination of average precipitation, average temperature, and precipitation coefficient of variation using quantile mapping. To characterize the effects of internal variability on risk of failure over a planning period, it would be preferable to use the binary reliability outcomes from many more stochastic realizations of weather sequences within each combination of climate variables. With a single stochastic trace perturbed into many climate scenarios, the modelled risk of failure is likely to be driven entirely by the climate scenario rather than the actual risk of missing a performance target under internal climate variability, and furthermore heavily biased across the climate response function by the single stochastic realization used to generate all climate scenarios. This seems to be the opposite of the intentions described in the introduction. (2) I do not see any part of the manuscript that assesses the performance of the logistic regression model using out-of-sample data. This is critical to the manuscript's success because it would provide evidence that the loss of information from modelling the risk of failing a satisficing criterion rather than evaluating the risk of failure through many simulations at each combination of climate variables could be worth the savings in computation time. (3) It is not clear whether there are separate logistic regression models for each sub-basin, performance metric, etc. How many logistic regression models are there in this case study? One per sub-basin, to model simultaneously meeting water supply reliability and environmental flow requirements? Two per sub-basin, each modelling risk of failing one of the objectives' minimum performance criterion? One, with sub-basins represented through dummy variables? If the model is used to predict risk of failing mutual satisficing rather than risk of failing one performance threshold, would the model structure work if the two objectives were in tension (as in the Poff et al. 2015 case study) rather than aligned (as they are in this case study)? This section needs to clearly list the explanatory variables and document the dependent variables much more clearly.

Water system modelling framework: (1) Synthetic weather generator and streamflow temporal resolution: A daily weather generator is used to generate perturbed weather sequences and run them through a runoff model to generate streamflow. After simulating climate-changed streamflow using the runoff model, daily streamflows are aggregated to monthly flow. Why aggregate ex post rather than using a computationally cheaper weather generator and/or runoff model that is designed to operate at the monthly temporal resolution? (2) Temporal aggregation and precipitation coefficient of variation (cv): Perhaps the monthly streamflow resolution is the reason precipitation coefficient of variation was not a strong predictor of performance metrics? The authors should consider this possibility and potentially discard precipitation cv from their analysis, which might be better served by more stochastic realizations in each climate scenario rather than more climate variables. (3) Climate response surface: The sampling of average precipitation and precipitation coefficient of variation (cv) is coarse (20% increments). I suggest sampling these factors at tighter increments. (4) The computational expense of conducting bottom-up climate risk assessment is mentioned several times in the text. How computationally intense is the Geum water system model to evaluate?

Role of GCM projections in the case study: (1) GCMs are limited in their ability to simulate land/ocean/atmospheric mechanisms, especially those that take place at sub-grid scale resolution. This limits the information that can be credibly derived from projections for water resources planning. Precipitation coefficient of variation (CV), one of the climate variables used in the case study, is not well represented in GCMs so it is questionable to infer precipitation CV from GCM projections. This is why GCM projections are not shown on some of the response surfaces in Poff et al. 2015 (in response to page 3, Line 18-19 of this manuscript). (2) This manuscript repeatedly mentions GCM counts as though GCM count in the feasible region on the climate response surface could be a decision criterion (e.g. page 3 line 19), and perhaps to some stakeholders it would be. However, this could also imply an attempt to quantify risk across the entire sampled climate space. Uncertainty quantification via ensembles of GCM projections is a challenging research question in its own right and would not be well treated by simply counting GCM projections from an arbitrary ensemble. Indeed, the point of bottom-up decision frameworks for climate risk management is avoiding this type of reliance on GCM projections with little scientific basis. Since this manuscript is designed to build on a bottom-up risk assessment framework, it is strange that so much emphasis is put on understanding performance under GCM projections in the text and figures.

Titling the framework: As mentioned above, it is not clear whether the logistic model is designed to model the risk of failing to mutually satisfice the eco-engineering performance thresholds or the risk of failing to meet one performance threshold. If the latter, the main technical contribution seems as appropriate for any single-objective climate response surface type risk assessment as for multi-objective climate response surface analyses, though it is applied here in a multi-objective climate response surface analysis. I would suggest the authors re-frame the analysis and revise the title to put the focus on the manuscript's main technical contribution, which is analyzing and communicating probabilistic information through a climate response surface (with an eco-engineering case study) rather than presenting a novel decision framework.

Technical corrections (typing errors, etc.)

Word choice: The meaning of the terms "successive", "risk of system failure", and "sub-components" in the context of this analysis is not clear from the text.

Page 2, line 31: Whatley et al. 2014 should be Whateley et al. 2014

Page 3, section 5: "However, all assessments using the response surfaces have focused on the “expected performance” rather than risk of system failure” Is this true? I thought many decision scaling papers evaluated reliability, which is risk of failure. . . Figure 9: Labels on X axis would be clearer in words. Also, isolating the results of the analysis to GCM projections is totally counter-intuitive here. The point of bottom-up climate response surface analyses is to avoid relying on GCMs in climate risk management. Figure 2: It is not clear where and how the logistic model comes into this framework based on Figure 2. Figures: None of the response surface figures include precipitation CV as one of the axes, though this is one of the sampled climate variables. The reasoning behind this should be clarified in the text.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-221, 2018.