Reviewer 1

RC: I enjoyed reading the paper, it is well written and it addresses both a methodological (i.e., integration of UA and SA into decision-robustness frameworks) and practical (i.e., the management of Umbeluzi river basin) issues. I support the effort in integrating UA and SA into established decision-support frameworks, I do however have the following issues:

AR: We thank the reviewer for this outstanding review. His constructive comments helped us improve the manuscript and strengthen our analysis.

General comments

RC: one of these frameworks proposed in the literature (perhaps the first one), i.e., Robust-Decison Making (RDM), in fact finds robust solutions heavily relying on a technique therein called Scenario Discovery, which to me can rightfully be assumed as a SA technique (i.e., factor mapping as also pointed out by Pianosi et al. (2016) http://dx.doi.org/10.1016/j.envsoft.2016.02.008). RDM has also been extensively adopted to address water planning problems https://link.springer.com/chapter/10.1007/978-3-030-05252-2_7. My suggestion would be to integrate such literature in the intro where it is stated that only few studies dealt with this problem before.

AR: We thank the reviewer for the comment. Sensitivity Analysis has indeed been previously applied in robust decision making, and the suggested literature will be added into the manuscript. However, it might also be worth to point out few key aspects:

1. The reviewer is right: factor mapping (FM) and scenario discovery (SD) are key to determine which uncertain input combination might cause robust policies to perform poorly. Those techniques usually classify the (uncertain) input samples as 'behavioural' or 'non-behavioural', depending on whether the response variable (in this application: the objective function) exhibits a certain pattern or not. Even though both FM and SD can both be assumed as SA techniques, in our study the classification between "behavioral" and "non-behavioral" perturbations is performed through the GLUE uncertainty analysis (Beven and Freer, 2001). Nonetheless, as mentioned by Saltelli et al., (2008) and Pianosi et al., (2016), UA and SA are closely related; to the point that their main distinction can be considered that '*UA focuses on quantifying the uncertainty in the output of the model, while GSA focuses on apportioning output uncertainty to the different sources of uncertainty'*. For instance, the GLUE UA has been developed starting from some of the basic ideas of Regional SA and factor mapping. Following the above, SA is not used here for scenario discovery, but builds upon it by providing a sensitivity index which measures quantitatively the main sources of vulnerability for reservoir operation.

We will clarify these aspects in the revised manuscript. Specifically, we will complement our uncertainty analysis section (3.4.1) by explicitly mentioning the strong interconnection between scenario discovery-UA and SA, but also the underpinning assumptions which makes such techniques inherently different among each other.

2. RDM in water resources has indeed been applied to address water planning problems. However, fewer are the studies where quantitative SA (i.e. where each input factor is associated with a quantitative and reproducible evaluation of its relative influence) has been employed for reservoir operation in a multi-objective context.

In order to account for the important application of RDM to pure planning problems, the literature proposed by the reviewer will be integrated in the manuscript.

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of hydrology, 249(1-4), 11-29.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software, 79, 214-232.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... & Tarantola, S. (2008). Global sensitivity analysis: the primer. John Wiley & Sons.

RC: perhaps related with the previous point: I think the UA and SA steps should more strongly be integrated into the robustness search. Afterall, robustness broadly defines the performance of the system under uncertainty and the UA and SA steps should be *part* of it. In the proposed framework instead (Fig. 3), the search for robustness and the UA and SA steps are reported as two different steps. How can the intial robust policies be updated/ameliorated based on SA analysis results? I think one could change the current framework either bringing the UA/SA step into robustness, or establishing some sort of feedback loop between the two. This would also affect the results, where it would be great to comment upon how, in light of the GLUE and PAWN results, one could increase the robustness of the previosuly found strategies. This could perhaps come in the form of a table where e.g. a description of how the best policies for each stakeholder change going through the three steps of the frameworks, i.e., from optimal to robust to robust+SA.

AR: We thank the reviewer for the comment. However, robustness, uncertainty and sensitivity analysis are already deeply interconnected into the proposed framework.

Such interconnection can be observed from three different perspectives:

1. RA, UA, and SA are all based upon similar mathematical techniques, since they all rely upon simulating iteratively the system following a perturbation on the input set.
2. RA, UA, and SA are all performed using same set of input perturbations, i.e., the very own states of the world generated during robustness analysis are also those employed during both UA and SA. The idea behind this choice is that, given a set of scenarios, robustness analysis identifies the most robust policies, UA determines those scenarios leading to acceptable system performances and SA provide a quantitative measure of the vulnerability of each sector to changes in each uncertain input factor. This consideration leads to:
3. During robustness, the states of the world are used to force the model in order to discover robust policies. However, robustness analysis per se does not ensure any knowledge of the scenarios which might yield to acceptable system performance (behavioral perturbations) or any insights upon a quantitative measure of the specific sensitivity (i.e., sensitivity index) of each dimension of the objective function to the various input realizations. Conversely, while UA and SA provide knowledge on each of the aforementioned aspects, they are not performed on the whole set of alternatives designed via optimization, but only on those who are robust (discovered via robustness analysis) against the system perturbations embedded in the states of the world. Considering this, they offer a complementary perspective about how a certain policy behave (i.e.: is it robust? What is the range of variability of the objective function? Which factor is most likely to yield to a failure in meeting a specific objective?) in response to the perturbations themselves.

In addition to the above, the quantitative measure of vulnerability offered by the sensitivity index is specifically tailored upon the policy for which it is computed for. Therefore, it is to be intended as valid only for the policy itself, with little or no possible extrapolation in terms of its robustness. For instance, let's suppose to have two policies (A and B). Let A to be characterized by marginal variations in the objective function value for all the system perturbations (i.e., robust), while the opposite is true for B. If, among all the input factors, streamflow is the responsible for most of the output variability for both A and B (no matter the magnitude of the variability itself), then SA will classify them as both highly sensitive to streamflow, with a sensitivity index close to one, regardless their robustness.

Consequently, the idea of conditioning robustness on SA results is unfortunately not applicable in the proposed framework. However, considering the value of the reviewer suggestion, we will develop the 'limitation and future research direction' section accordingly. Specifically, we will explicitly mention that future research could be developed towards implementing SA methods that establish feedback loops with RA (while the opposite is true and already implemented) in order to enhance the robustness of operating policies against uncertain exogenous factors.

Specific comments

RC: the model description may not be immediate to a reader not used to such models. It would be good to make clear what are the policies, the uncertainties etc. etc. - a good possible framework to follow is the XLRM (actually linked to RDM)

- better description of the GLUE method

- when introducing RID and RUD (at about line 380) it isn't immediately clear that the considered least robust alternative is not the asbolute least robust, but rather the least robust among the Pareto set (i.e., performing best for at least one stakeholder). Also, it would be ideal trying to condense the info from figures 7 to 9 into one figure - or at least plotting them together as they are "the same".

- better captions

- some remaining typos/language issues

AR: we would like to thank the reviewer for the thorough review of the manuscript, we appreciate how addressing all the specific comments provided above will undoubtfully improve the overall quality and readability of the paper. Therefore, we took them in consideration while revising the document.