Interactive comment on "The new importance measures based on vector projection for multivariate output: application on hydrological model" *by* L. Xu et al.

S. Razavi (Referee)

saman.razavi@usask.ca

This manuscript is on the interesting and important topic of multi-criteria sensitivity analysis. This kind of problem is commonly encountered by hydrologic modellers (and perhaps modelling community in general) that the sensitivity of different model responses (or summary metrics/criteria) to different model inputs can differ, and it may be non-trivial to come up with a unified sensitivity assessment that reflects the collective influence of a model input on multiple outputs. This process, however, should be carefully performed and interpreted as it might result in loss of information (sensitivity of each individual response may tell you an important side of the story in the intended context) or misleading assessment.

Overall, I found the method proposed in the manuscript interesting, but I have major reservations with this manuscript, outlined below. I believe substantial revisions are required. I should note that Razi Sheikholeslami and Shervan Gharari have helped me with this review.

1) First of all, the quality of writing (including English, grammar, organization, etc.) and more importantly equations and notation may be improved significantly. Also, some of the equations are repeated once or twice. Some of the notation might not have been chosen properly.

2) The main motivation of the study, as stated in the abstract, is to analyze the effect of model inputs on "correlated" multivariate output. However, the method proposed assumes that different outputs are orthogonal and their correlation is not accounted for anywhere in the formulations. I might be missing something, but it this is true, then the method doesn't serve for purpose.

3) My understanding is that the proposed method is nothing but a weighting approach that weights the Sobol's indices for each individual model output based on that output variance. More accurately stated, the weight for the indices of an output is the square of the variance of that output. If my understanding is right, then the method might not possess much novelty. In other words, the proposed method is simply a "supposedly objective" weighting approach for different model outputs. In practice, this method results in sensitivity rankings that are the same as (or consistent with) the rankings based on the output with largest weight. The case study results of the manuscript confirms this.

4) Related to the above comment, the weighting approach of the proposed method might not be appropriate. The weighting is overly sensitive to the way that the different model outputs are normalized/became dimensionless (Equation 10 of the manuscript). If you use a different normalization approach, you might get an entirely different assessment. Also, dividing the values by their average (Equation 10) might not be a good

strategy, as the spreads (variances) of different outputs might remain significantly different after the conversion, even of different orders of magnitude. A better strategy might be to standardize the outputs (dividing by standard deviation). But still you would have to deal with the issue raised in the previous comment.

5) Literature review: the manuscript does not provide the reader with the status quo. There is some literature review, but limited and unbalanced. First of all, I suggest the authors have a look at the following paper, which is a fully multi-criteria approach with minimal subjectivity:

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X., & De Gonçalves, L. G. G. (2012). A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis. Journal of Geophysical Research: Space Physics, 117(7), [D07103]. DOI: 10.1029/2011JD016355

Also, at the risk of self-promotion or self-propaganda, I'd suggest the authors have a look at the following papers. The fundamental question is that how variance itself and its decomposition can be meaningful for global sensitivity analysis. Of course, I am not arguing they are not meaningful, but asserting that there are caveats that need to be recognized and taken care of.

Razavi, S., and H. V. Gupta, (2015), What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models, Water Resour. Res., 51, 3070–3092 doi:10.1002/2014WR01652

Razavi, S., and Gupta, H. V., (2016), A new framework for comprehensive, robust, and efficient global sensitivity analysis: II. Application, Water Resources Research, 51 doi:10.1002/2015WR017559 full-text.

Razavi, S., and Gupta, H. V., (2016), A new framework for comprehensive, robust, and efficient global sensitivity analysis: I. Theory, Water Resources Research, 51

doi:10.1002/2015WR017558 full-text.

Also, derivative-based method and elementary effect method are essentially the same but with slightly different numerical implementations for the step size to calculate the numerical derivatives/elementary effects (lines 54-57). Please refer to Razavi and Gupta (2015 WRR).

6) The idea of using Polynomial Chaos Expansion (PCE) is divergent from the core of this manuscript, and may be confusing to readers. Please note that the only reason to use PCE instead of a full Monte-Carlo-based Sobol analysis is computational efficiency, which of course comes at the trade-off of losing accuracy. To me, computational efficiency is a wholly different story and does not go well with the core idea of the manuscript. Note that PCE is a metamodeling approach, and like any other metamodeling approach, may have two major shortcomings. First is the quality of results depends on the quality of fit. So if the response surface is complex, PCE (polynomials) may not fit the response surface well, and as such, the results might be erroneous. Second is metamodels are handicapped for high-dimensional problems, that is why they have typically been used in the literature only for problems with less than $\sim$20 input variables. The authors may find more discussion on these two shortcomings in the following papers:

Razavi, S., B. A. Tolson, and D. H. Burn, (2012), Review of surrogate modelling in water resources, Water Resources Research, 48, W07401 doi:10.1029/2011WR011527. 32 pages.

Razavi, S., B. A. Tolson, and D. H. Burn, (2012), Numerical assessment of metamodelling strategies in computationally intensive optimization, Environmental Modelling and Software, 34(0), 67-86

7) Related to the above comment, if the authors like to include the comparison of PCE and direct Monte-Carlo-based integration of Sobol's indices, it would require a more substantive analysis. Due to stochasticity of these algorithms, any comparison

of this nature would require running multiple replicates of each method to account for sampling variability and randomness, and also, such analysis should be done at different computational budgets (different numbers of function evaluation) to ensure a fair and thorough comparison. The above Razavi et al. (2012 EMS) paper details the elements of such comparisons.

8) This manuscript might be more suitable to be published in an Applied Mathematics journal. The relevance to the HESS community might not have been adequately established. The main (and probably the only) connection is the HBV case study. A little bit more work might be required to strengthen the connection.

Lines 55-69, "it is recommended . . .", readers may wonder who recommends this. The authors?

Section 2 requires a lot of improvement. For example, do you need to have many 3rd level sub-sections such as 2.2.1?

Saman Razavi