Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-375-RC4, 2018

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "A comprehensive sensitivity and uncertainty analysis for discharge and nitrate-nitrogen loads involving multiple discrete model inputs under future changing conditions" *by* Christoph Schürz et al.

F. Pianosi (Referee)

francesca.pianosi@bristol.ac.uk

Received and published: 8 October 2018

The manuscript presents an interesting application of uncertainty and sensitivity analysis to the SWAT model. The aim is to assess the dominant controls of long-term discharge and nitrate-nitrogen load predictions under climate and land use change, while also taking into account the intrinsic uncertainty in the model, i.e. parameter and set-up uncertainty. The analysis is solid and provides interesting insights about the model behaviour. Although the specific findings are only relevant to the investigated model and case studies, their discussion is interesting for the wider community of SWAT users

Printer-friendly version

Discussion paper

and in general users of environmental impacts assessment models, as it demonstrates the type of findings yielded by GSA and their implications for the refinement and use of the model. The visual analysis introduced in Figure 4-8 is a simple and yet effective complement to quantitative GSA approaches.
Overall the paper is well structured and well written, and I think it should be accepted for publication.

Below are some points that could be addressed to improve the manuscript clarity before publication.

[1] Language is at times unclear - some examples are given below as Minor points. I also have a general comment about the use of the term "sensitive". The authors use it as interchangeable with "influential" however I find this confusing, because "sensitivity" is an attribute of the output, not of the inputs. I would say that "input x1 is influential on the output" or "the output is sensitive to input x1" but I would not say that "input x1 is sensitive" - this is confusing. Some examples of these unclear occurrences are also given below under Minor points, however if the authors accept my remark they should check the entire manuscript.

[2] The definition and use of the behavioural parameter sets is slightly unclear. I think the confusion started on P. 10 L. 6-7 with the sentence
"For all SWAT model setups of the Schwechat and the Raab catchments we identified non-unique parameter sets that adequately simulated daily observation of discharge and NO3-N loads".
Does it mean that you identified one behavioural parameter set for **each** model setup, or that you identified one behavioural parameter set to be applied **in all** the set-ups? If the former, then how is the dependency between parameterisations and model setups accounted for in the GSA? If the latter, then the underlying assumption is that the same parameter values can effectively represent processes at different aggregation scales (ie for different definitions of the subbasins and HRUs)? This should be clarified.
On a parallel note, I find it interesting that out of 100,000 sampled parameterisations

only 43 and 52 where found behavioural. This is not uncommon in calibration of complex hydrological models but still worth highlighting. It would also be interesting to see whether these behavioural parameterisations are clustered in specific regions of the parameter space or if they are scattered across the sampled ranges, which would indicate a certain amount of interactions between the parameters. This could be illustrated for example through a parallel coordinate plot.

[3] GSA was applied using 7000 samples of the input factors. How was this number chosen? Did the authors checked the adequacy of this sample size? The fact that the ranking based on the sensitivity indices in Figure 2 is confirmed by the visual analysis of Figure 4-8 is reassuring, yet formal methods exist to assess the robustness of the GSA results to the chosen sample size (for example, using bootstrapping confidence intervals as in Sarrazin et al. 2016 or a dummy parameter as in Zadeh et al 2017, both cited in the manuscript). It would be good to include more discussion of this point in the manuscript.

[4] The PAWN method was applied using a sampling scheme different from the one originally presented in Pianosi and Wagener (2015), in order to handle discrete-valued input factors. I understand the idea is to consider as fixed points x_i^j all the possible values that the discrete input factor x_i can take. Hence, for each input factor, the number of fixed points coincides with the number of possible values (n_i) that the input can take. If my interpretation is correct, then the text is misleading when it says (P. 13 L. 28) that "a generic random sample of the size N was drawn and subsetted with N/n_i subsets for all x_i^j"
as the generic sample is divided into n_i (and not N/n_i) subsets. Is this right?
Also, if I understand the strategy correctly, then the inputs with small number of possible values (for instance the land use scenario) are associated with conditional distributions based on a very large number of samples (around N/n_i=7,500/2 in the case of land use scenarios), while the inputs with large number of possible values (for instance the parameterisation) are associated with conditional distributions based on much smaller

Printer-friendly version

Discussion paper

number of samples (around 7,000/43). Do you think using such different sample sizes could have had an impact on the estimation of the KS values and hence of the PAWN sensitivity indices?
Finally, a new sampling strategy was recently proposed for PAWN (Pianosi and Wagener, 2018). While this new strategy is still designed for continuous inputs, and hence could not be used here, it would be good to mention its existence for readers who may want to apply PAWN in the future (as for the case of continuous inputs this would be recommended over the strategy in the 2015 paper).

[5] I think the discussion in Section 4.2 is interesting but potentially slightly misleading. The authors clarify that "several assumptions were made in the development of scenarios that are highly subjective". I understand the importance of highlighting the subjectivity inherent in the scenario definition *if the goal of this study was to make projections* of the future evolution of the two catchments. However, this is not the objective when doing GSA. GSA answers the question: "how much output variation do we get if we vary the inputs *within certain ranges*?" The answer yielded by GSA (i.e. the sensitivity indices, the input ranking, etc.) is certainly conditioned upon the chosen ranges, however this is "intrinsic" to the question asked, regardless of how the choice is made - be it an "objective" calibration exercise (as done for the parameterisations) or a "storylines" approach. In other words, I think the point is to justify why certain scenarios are considered for the impacts assessment study; once they have been selected for that purpose, it follows that they would be used in the GSA too if one wants to know their relative influence with respect to other input factors of the model. So, I do not agree with the sentence (P. 26 L. 13-14) "For the SA of the simulated variables the diversity of the developed scenarios is essential.": diversity may be important for the impacts assessment (is it?) but not necessarily for the GSA. If a limited set of scenarios were selected for the impacts assessment, I would use that set for the GSA even if it is not diverse.

MINOR POINTS

Printer-friendly version

Discussion paper

P. 1 L. 15: "scenario inputs" should be "input scenarios"
P. 2 L. 5: "the precipitation of the climate scenarios" sounds a bit odd, maybe "precipitation projections"
P. 3 L. 3-4: "An assessment is only as good as the dominant contributors of uncertainty in such a modeling chain." Unclear. Something seems to be missing in this sentence: an environmental impacts assessment is only good if dominant contributors of uncertainty are... what? identified? removed? ...?
P. 3 L. 11-12: "model computations" should be "model evaluations" (or "runs" or "executions")
P. 3 L. 19: "Most applications utilize GSA to identify and rank continuous model parameters". Unclear: GSA does not "identify parameters" at most "identify influential parameters"
P. 3 L. 21: "Although," Comma should be removed
P. 3 L. 26-27: "An OAT analysis however presumes linear models and non-correlated inputs". Not sure OAT requires a linear model, for instance the Morris method uses a OAT approach and yet is typically applied to nonlinear models. More generally, why should GSA be applied to a linear model at all? If the model is linear than the effect of each input on the model output is simply proportional to the input variation, no need to do GSA to know that.
P. 4 L. 4: "complex". Unclear. What is the definition of a "complex" input?
P. 4 L. 5: "No study is known to us that takes advantage of GSA in the scope of environmental impact studies." What is the definition of "environmental impact studies" here? I would say that GSA has been applied to such studies before, e.g. Anderson et al (2014); Butler et al (2014); Le Cozannet et al (2015)
P. 4 L. 13-16: Very long sentence, consider splitting into two.
P. 8 L. 3: "Although," Comma should be removed.
P. 8 L. 14-15 "The SWAT model setups for the Raab and the Schwechat involved decisions for the selected number of subbasins of a model setup and the definition of the HRUs." Convoluted sentence.

P. 8 L. 15: "Both modifications": which modifications? Unclear
P. 9 L. 2: "involving". Unclear. Maybe "which requires"?
P. 9 L. 11: "to define of the thresholds". Remove "of"? In general, the entire sentence is a bit unclear. How is the "aggregation error" defined? Error in which variable, and with respect to what "correct" value?
P. 9 L. 14: "In a pre-analysis step," In the GSA literature, this kind of "pre-analysis step" is often called a "screening" analysis, as it aims at screening out the non-influential parameters. Maybe worth mentioning the term as it would be familiar to many readers.
P. 9 L. 14: "relevant parameters". Relevant to what? Maybe better "influential"
P. 9 L. 21: "FDCs". Explain the acronym
P. 13 L. 5: "To identify the impact of" maybe better "To measure the relative importance of"
P. 13 L. 7: "PAWN involves". Unclear what "involves mean. Maybe better "PAWN uses"
P. 13 L. 11: "the sensitivity of a model input x for a target variable y". Sensitivity is an attribute of the output, not the input. I would rephrase as "the sensitivity of a target variable y to a model input x".
P. 13 Eq. (1) and (2). The mathematical notation could be made clearer. I find it odd that in Eq. (1) KS takes as subscript the index of the fixed point (j) while its argument remains the generic input x_i. This choice also makes it more difficult to understand how maximisation occurs in Eq. (2). I think using the notation $KS(x_i^j)$ in both equations would make things much clearer.
P. 13 L. 23: "possible states". Why "states"? The term was never used with this meaning before. I would rather say "possible values".
P. 13 L. 24: "a lower sensitivity of the input x_i on the target variable y". Again, rephrase as either "a lower sensitivity of the target variable y to the input x_i" or "a lower influence of the input x_i on the target variable y".
P. 13 L. 28: "subsetted with" Not sure "subset" can be used as a verb. Maybe better "divided into"
P. 13 L. 29. "... were used for the sensitivity assessment". I would link this to the

mathematical notation just introduced above and say: "... were used as target variable y".
P. 14 L. 10-12: "In this study, we consider all execute model setups to be plausible..." I do not understand this clarification. What other approach would have been possible? To discard some simulations because deemed not plausible? And how would you define then what is plausible and what is not? Please clarify.
P. 14 L. 17: "low number of each input". Unclear. Do the authors mean "low number of inputs" (i.e. 5 inputs) or "low number of possible values taken by each input"?
P. 14 L. 24-28. This sounds like a repetition of what just said in the methodology section, I do not think is needed. I would rather use this opportunity to explain how to read Figure 2 (what is the difference between the panels and how to read each circle plot).
P. 15 L. 17: "highly sensitive" replace by "highly influential"
P. 15 L. 25-26: "their overall sensitivities follow the general trend of the climate scenarios to a large extent". Unclear, please rephrase.
P. 15 L. 33: "difference that is visible for the two". Unclear, do you mean "difference that is visible between the two"?
P. 15 L. 33-34: "how the reference period relates to the uncertainty bands in amplitude". Unclear what this means.
P. 16, caption of Fig. 2: "Model input sensitivities for signature measures". Replace by "sensitivities of signature measures to model inputs". And later on "sensitivities of" should be replaced by "sensitivities to"
P. 17, text and Figure 3: what does "specific discharge" mean? Why "specific"?
P. 17 L. 6: "show a difference". Does this mean "show an increase"? If so, I would use "increase", it makes it easier for the reader to follow.
P. 19 L. 12: "While a grouping...". Remove "While".
P. 24 L. 20: caused future land use change" Maybe "caused **by** future..." ?
P. 26 L. 31-32: "The application of sampling strategies for SA usually do not account for the circumstances that one model input constrains any other model input". I do not fully agree. There is an increasing literature on GSA methods applicable to the case of

dependent inputs, see for instance Mara and Tarantola (2012; 2017).
P. 27 L. 17-18: "by a factor of up to 5... up to 8". Do these numbers come out of a comparison of PAWN indices? If so, I am not sure I would draw such quantitative comparison. PAWN indices are (maximum) KS values: what is the practical interpretation of "a factor of 5" between KS values? I find it difficult to imagine.
P. 28 L. 3: "the lack of tool that allow the practitioners access to such methods". Not sure I understand what the authors mean here. Several GSA software tools are available (some are reviewed for example in Pianosi et al 2015). So what is the problem here? That they are not "friendly" enough for practitioners to use them? Or that they are not sufficiently tailored to SWAT applications? Pls clarify.

REFERENCES

Anderson, B., Borgonovo, E., Galeotti, M., Roson, R., 2014. Uncertainty in climate change modeling: can global sensitivity analysis be of help? Risk Anal. 34 (2).

Butler, M.P., Reed, P.M., Fisher-Vanden, K., Keller, K., Wagener, T., 2014. Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. Environ. Model. Softw. 59.

Le Cozannet et al (2015) Evaluating uncertainties of future marine flooding occurrence as sea-level rises, Environmental Modelling Software, 73.

Mara and Tarantola, 2012, Variance-based sensitivity indices for models with dependent inputs, Reliability Engineering and System Safety

Tarantola and Mara, 2017, VARIANCE-BASED SENSITIVITY INDICES OF COMPUTER MODELS WITH DEPENDENT INPUTS: THE FOURIER AMPLITUDE SENSITIVITY TEST, International Journal for Uncertainty Quantification

Pianosi and Wagener (2018), Distribution-based sensitivity analysis from a generic input-output sample, Environmental Modelling Software

Pianosi, F., Sarrazin, F., Wagener, T. (2015), A Matlab toolbox for Global Sensitivity

Analysis, Environmental Modelling Software, 70.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-375, 2018.

Printer-friendly version

Discussion paper