Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2017-28-AC3, 2017

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Identifying the connective strength between model parameters and performance criteria" *by* Björn Guse et al.

Björn Guse et al.

bguse@hydrology.uni-kiel.de

Received and published: 12 May 2017

Comment: This study uses regression trees in a bi-directional framework to estimate the importance of a model parameter in a set of objective functions, as well as the relative importance of each parameter to a given performance measure. The authors conclude that this method permits identifying model parameters with respect to certain performance measures, and I agree with their assessment. I found this paper to be interesting and generally well written. I have a few important comments which I would like to see addressed before considering publication in HESS. I recommend a major revision.

Reply: We thank Richard Arsenault for this very positive statement to our manuscript and for encouraging us to revise the manuscript.

C: First, the author's RT based method should ideally be compared to other proven techniques to estimate parameter importance, such as Sobol' sensitivity analysis (or any global sensitivity analysis). If not integrated into the work directly, differences in expected outcomes should be addressed in the literature review. I recognize that the bi-directional aspect of this work is novel but any advantages of this method should be compared to a proven baseline.

R: We agree that the manuscript would benefit by introducing a discussion on sensitivity analyses also by considering the comments from the other referees and the Editor. Since we have presented studies on sensitivity analyses during the last years (Guse et al., 2014, Pfannerstill et al., 2015, Guse et al. 2016a,b), we will refer to these studies and consider other major recent studies on this topic such as Herman et al. (2013a,b) or Van Werkhoven et al. (2008, 2009).

In the current version of the revised manuscript, the added passage at the end of the discussion to this issue reads as stated below. This text is in an intermediate state and may change during the revision.

"The relevance of model parameters can also be investigated using sensitivity analyses. By comparing the parameter relevances with former studies on temporally resolved parameter sensitivity analyses, it becomes apparent that the overall ranking is similar (Guse et al., 2014, 2016b). While we used here ten different performance criteria in regression trees, the parameter sensitivities were separately derived for the five segments of the FDC. In both cases, the differences in parameter relevances between different hydrological conditions were shown and are consistent.

As a further development, the bijective analysis of the relationship between model parameters and performance criteria was introduced here. The interpretation of this relationship from sites is an advantage compared to classical analysis of the impact of model parameters on performance criteria such as realised in sensitivity analyses (van Werkhoven et al., 2008; Herman et al., 2013a,b). With our approach, it is possible to investigate which performance criteria are appropriate for a certain model parameters.

A core advantage of RT is that subsets of the simulation runs are constructed in a structured way. By subdividing the simulation set based on the major influencing variables at each branch, two distinct subsets occur which differs in the values of model parameters as well as of the performance criteria. As shown in the example of a RT, good model simulations are separated from poor simulations. And also the parameter values within each subset are different. With this subset construction, it can be detected whether a model parameter has the highest explanatory power in a certain branch."

C: Page 4, lines 4-6: This sentence is very confusing, please rewrite differently.

R: We are aware that we have to explain the RTpar method in a clearer way. In the revised version of the manuscript, this sentence will be rewritten.

C: Page 4 lines 7-8: This means that the model parameter has no relevant impact on other performance measures. Perhaps give a clear example of how this can be achieved in the case of a hydrological model with highly interacting parameter sets. My previous work in parameter identifiability suggests that a large part of the relative importance of a parameter on a performance measure comes from its interactions with other parameters.

R: One aspect of introducing the connective strength is to show whether a model parameter can be clearly identified by one specific performance criterion. This means that this performance criterion is precisely related to the process which is related to this parameter. One example is the strong bijective relationship between the baseflow recession coefficient ALPHA_BFssh and the RSR for the very low segment. Despite of parameter interaction, it becomes apparent that this parameter is best identified by the RSR for the very low segment which makes sense, since ALPHA_BFssh is the major low flow parameter. Moreover, we like to mention another point: Our core aim is to improve parameter identification. Thus, in the best we want to consider a model pa-

rameter in isolation. Thus, a model parameter which is always insensitive or only due to parameter interaction is in both cases a parameter with a low parameter identifiability.

C: Page 7, lines 4-5: By using 2000 samples with hypercube sampling, are the authors not effectively working in spaces where parameter combinations might not make physical sense? Usually the model parameters, during calibration, will self-regulate to attain sensible parameter values. With a LHS approach, perhaps some combinations are tested here which are out of the bounds that the model can work with appropriately. More information regarding this aspect would be interesting.

R: We agree with the reviewer that within a parameter sampling, different qualities of parameter combinations are considered. Certainly, some of them make more sense than others. This aspect is considered at first by constraining the parameter ranges according to our experiences to reduce unrealistic parameter combinations, which lead to non-physically process description. We will more clearly state in the revised version of the manuscript that we constrain the parameter space to reduce unrealistic parameter combinations. Second, by calculating the performance criteria, we detect the quality of the model runs. However, since the selection of thresholds on performance criteria is somehow arbitrary, we think that it is a more consistent approach to do not restrict the data set in a second step due to the performance criteria. Otherwise we would have to justify ten thresholds (for each performance criterion). In this context, we like to emphasise that the problem of analysing parameter sensitivities with model simulations of different performance appears also in sensitivity analyses. In particular, studies with multiple performance criteria show that it is very difficult to subdivide a data set into "good" and "poor" performing model runs.

C: Furthermore, the parameters do not seem to be normalized in their ranges, therefore allowing some parameters more leverage over the performance measures. If I interpreted this correctly, then some of the results would be trivial since the larger boundaries will naturally have more effect on the performance measure and thus the parameter will be more "important". The use of a LHS methodology in an uneven search space will bias the results (as an extreme example, if ESCO bounds were set between 0.995 and 1.005, then the parameter would definitely not be considered important). The choice of boundaries, then, induces a methodological bias in the results. I am not sure how to solve this problem, perhaps by performing multiple calibrations and taking the envelope of the parameter sets, but this also has its drawbacks.

R: We are aware of this problem. As it is also typical in sensitivity analyses, the selection of the lower and upper bounds of the parameter values is critical. However, we have selected the parameter ranges based on several former studies with the SWAT model and think that we have selected reasonable ranges (Guse et al., 2014, 2016a, Pfannerstill et al., 2014, 2015). Our experience with increasing ranges (within reasonable areas) is that the general pattern of the results is not largely affected.

C: Also, the parameters seem to be evaluated on the entire time series. In a snowmelt-dominated catchment, the parameters are highly time-variant. How could this affect the method's robustness?

R: Certainly, the results would be different if we would apply it only to a sub-period or only to winter month. However, we think that this issue is still a big challenge in hydrological modelling. For example, we are not aware of a consistent approach how to identify or calibrate snow parameters using daily resolution and a reasonable sub-period. It is still challenging to consider sub-periods for model calibration. We agree that this issue is relevant, but in our opinion out of the scope for this study. However, we could mention this point at the end of the discussion as an outlook for future research.

C: I think Figure 1 can be omitted completely without any loss of information in the paper. It is fairly well described in the text.

R: Based on the positive feedback during the presentation of this work at the General Assembly of the European Geosciences Union (EGU) in April 2017 in Vienna, we intend to keep this figure in the revised manuscript.

C: Page 12, lines _20-25: I have the feeling that some of these strong connections are trivial. If I had had to guess in advance, I would have guessed that Evapotranspiration (ESCO) is probably strongly linked to bias (KGE_beta), and that mid flows and lower were also affected by baseflow recessions and to some extent evaporation due to the relative scale of a fixed evaporation rate on total available volumes. Once interactions are important, then the method seems to "get lost" in a sense, as there is no clear path to identifiability (as demonstrated in the discussion). I think sensitivity analyses would provide the same information while also informing on the different order sensitivities.

R: We agree that some results could be expected. To give an idea of this issue, we have selected two contrasting catchments. Our results clearly show that there are lots of differences even in these two catchments. Moreover, we agree that some relationships such as between ESCO and KGE_beta are not surprising. On the other side, other relationships such as between the curve number and the RSR for high flows do not appear to be strong, even if this could be expected as well. Thus, we think that it worth to check the relationships whether they are really as relevant as expected. Most importantly, a sensitivity analysis would only provide one-directional results, while our approach is focused on the bijective relationship between model parameters and performance criteria.

References:

Guse, B.; Reusser, D. E.; Fohrer, N. (2014): How to improve the representation of hydrological processes in SWAT for a lowland catchment - Temporal analysis of parameter sensitivity and model performance, Hydrol. Process., 28: 2651–2670. doi: 10.1002/hyp.977.

Guse, B.; Pfannerstill, M.; Strauch, M.; Reusser, D.; Lüdtke, S.; Volk, M.; Gupta, H.; Fohrer, N. (2016a): On characterizing the temporal dominance patterns of model parameters and processes, Hydrol. Process., 30(13), 2255-2270, doi:10.1002/hyp.10764.

Guse, B.; Pfannerstill, M.; Gafurov, A.; Fohrer, N.; Gupta, H. (2016b): Demasking the integrated information of discharge: Advancing sensitivity analysis to consider different hydrological components and their rates of change, Water Resour. Res., 52, 8724-8743, doi:10.1002/2016WR018894.

Herman, J.D.; Kollat, J.B.; Reed, P.M.; Wagener, T. (2013a): From maps to movies: high resolution time-varying sensitivity analysis for spatially distributed watershed models. Hydrology and Earth System Sciences, 17, 5109–5125.

Herman, J.D.; Reed, P.M.; Wagener, T. (2013b): Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. Water Resources Research, 49, doi:10.1002/wrcr.20124.

Pfannerstill, M.; Guse, B.; Fohrer, N. (2014): Smart low flow signature metrics for an improved overall performance evaluation of hydrological models, J. Hydrol, 510, 447-458, doi:10.1016/j.jhydrol.2013.12.044.

Pfannerstill, M.; Guse, B.; Reusser, D.; Fohrer, N. (2015): Process verification of a hydrological model using a temporal parameter sensitivity analysis, Hydrol. Earth Syst. Sci., 19, 4365-4376, doi:10.5194/hess-19-4365-2015.

van Werkhoven, K.; Wagener, T.; Reed, P.; Tang, Y. (2008): Characterization of watershed model behavior across a hydroclimatic gradient. Water Resources Research 44: W01429. doi: 10.1029/2007WR006271

van Werkhoven, K.; Wagener, T.; Reed, P.; Tang, Y. (2009): Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources 32(8): 1154–1169.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2017-28, 2017.