Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2017-28-AC2, 2017
© Author(s) 2017. CC-BY 3.0 License.

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 presents an approach to quantify the strength of the bijective relationship between model parameters to performance measures. The proposed method explores the model parameter-performance space using regression trees with the goal to detect performance measures that can uniquely identify a parameter. The regression trees are first developed by casting model parameter as explanatory variables and performance measure as prediction variables, and then by exchanging the explanatory and prediction variables. These trees are developed for two catchments in Germany. The main idea presented in the study is interesting and results contribute valuable insights towards model diagnostics.

Reply: We thank the reviewer for summarising our methodology and for emphasising

the value of our study for the hydrological community.

C: However, there are a few issues that should be addressed. First, the introduction requires revision so as to remove repetition of ideas (see for example, lines 11 and 32 on Page 2), and to provide more background.

R: In the revised version of the manuscript, we will rework the introduction, as also raised by comments from the other referees and the Editor. In this way, we will reduce repetitions in the introduction. As also raised by the other reviewers, we will add a discussion of how our study is related to sensitivity analysis to improve the overall presentation of our study.

C: The need for multiple performance measures to identify unique aspects of the hydrograph is well motivated, but why this has remained a challenge is not discussed.

R: We certainly agree that the use of multiple performance criteria was already emphasised. It is known that each performance criterion is related to different aspects of the hydrograph. However, it is still unclear a) which performance criteria are appropriate for parameter value selection and b) which performance criteria are related to which (type of) parameters. We will address this issue also by including additional references to this point when revising the introduction and provide a more focused introduction in the revised version of the manuscript.

C: Some well-known issues are parameter interaction, limited information content in hydrologic time series data that allows identification of only a handful of parameters, and uncertainties in input as well as streamflow data (Beven, 2011).

R: Certainly, parameter interaction and limited information content complicate the interpretation of model results. These points also complicate the understanding which and how model parameters control the hydrological behaviour in models. We see our approach as a contribution to a more detailed use of available information with the aim to achieve a more precise identification of all relevant model parameters. We have shown how the use of different performance criteria help in identifying model parameters. Thus, based on our approach the parameter identification can be improved systematically. The connective strength shows how precisely a model parameter can be identified based on the existing information content. Since we selected a two-step approach by using at first the performance criteria and secondly the model parameters as explaining variables, we can provide some statements to the aspect of parameter interaction. In RTpar with the model parameters as explaining variables, each model parameter is individually assessed. There is an indirect impact since the model parameter values are different for each model simulation. Thus, by comparing RTpar between the different model parameters runs and thus different model parameter sets, we can see whether the performance criteria which are influenced by changes in this parameter are the same. Moreover, in comparing the ten RTperf applications, we can see whether the same model parameters affect different performance criteria. This is also an insight for parameter interaction. Thus, our study contributes to better parameter identification also under consideration of parameter interactions. We will consider this point while revising the mauscript.

C: Complicating this further is the time varying nature of parameter sensitivity (Herman et al. 2013).

R: We agree that we have to add the current state-of-the-art in time-varying sensitivity analyses to the introduction. Moreover, we will relate our approach in the discussion to the recent approaches in sensitivity analyses. Since we have presented studies on sensitivity analyses during the last years (Guse et al., 2014, 2016a,b), in the revised version of the manuscript we will refer to these studies by also considering other major recent studies on this topic such as Herman et al. (2013a,b) or Van Werkhoven et al. (2008).

C: Another important issue is the comparison of the proposed method to already existing sensitivity analysis methods, which also attempt to identify relationship between model performance and parameters. In my understanding, it is the partitioning of the

model performance space by parameter values that is unique about regression trees, but the order of importance of parameters should ideally be the same as that derived using sensitivity analysis methods.

R: We agree with the reviewer that in general similar results in parameter ranking are expected in sensitivity analysis. This can be proved by comparing this manuscript with our recent work (Guse et al., 2016a,b) on sensitivity analyses on model results (and not on performance criteria). The parameter ranking could be different when using different performance criteria. This was shown in our study, since we focused more on the relationship between performance criteria and model parameters. We examine which performance criteria are appropriate to identify at best the parameter values. The novelty is that we assessed this relationship bijectively from both sides. At first, as usual we used performance criteria as target variable (such as typically in sensitivity analysis) and secondly we used model parameters as target variable. This cannot be realized in a sensitivity analysis and is a clear benefit of our approach compared to sensitivity analysis. However, as described in the previous comment, we will relate our approach to sensitivity analyses in the revised version of the manuscript and discuss the advantage of our approach in the discussion chapter.

Specific Comments:

C1. Method description: Further information on implementation of regression trees is warranted. For example, the metric: ‘percentage contribution of each explaining variable’ is used throughout the manuscript without an explanation to how it is actually estimated by regression trees. It is expected that any method will have some error or uncertainty associated with its results, so what levels of ‘percentage contribution’ are significant? If any cross validation analysis during tree construction was used, it should be explained.

R1: We will improve the description of percentage contribution in the revised version of the manuscript. In this context, we are not aware of a threshold value stating that a percentage contribution is relevant above this value. Within the construction of the regression tree, a cross-validation analysis was automatically realised in the r-package rpart (Therneau and Atkinson, 2010). Hereby, it was investigated how an additional split leads to a better explanation.

C2. Methodological choices: It is mentioned in the manuscript that it is likely some performance measures are correlated (Lines 6-9, Page 7), these correlations are also presented in Figure 2. However, all performance measures are used for generating trees for RTpar. Could this potentially be the reason behind poor connective strength between performance measures and parameters for high flows? It should be discussed whether the regression tree algorithm can deal with correlated input. If not, correlated performance measures should ideally be reduced to an uncorrelated set. In fact, the same holds true for the use of model parameters as independent variables in RTperf, the presence of parameter interaction will affect the results to some extent.

R2: The selection of the performance criteria impacts the results of RTpar. Using similar performance criteria, the percent contribution for single performance criterion may decrease. However, as the comparison between both catchments shows, there are large differences in the results of the two catchments. The high similarities in five performance criteria in the Treene catchment are not observed in the Saale catchment. The idea is to cover different aspects of hydrological behaviour by using these ten performance criteria which consider different aspects of hydrological behaviour. Depending on catchment characteristics and the way the model can produce the hydrological behaviour in these catchments, the impact of performance criteria might be different. Thus, it is not fully clear in advance which performance criterion is best suited to represent a certain model parameter. Due to that, we are convinced that it is required to use all these performance criteria. In addition, we tried to cover with our selection the most common performance criteria as well as the different aspects of hydrographs. However, the issue of correlated performance criteria needs to be certainly considered in the interpretation of connective strength. Concerning model parameters, we do not

agree to remove model parameters because of parameter interactions in RTperf. The idea is to detect which model parameters impact a certain performance measure. In this context, it would not be helpful to remove interacting model parameters. We are not aware of a similar approach in model calibration. In the case of two relevant, but correlated model parameters, we have nevertheless to identify appropriate values for both model parameters and not only for the most relevant one.

C3. Background data: The time period of analysis, values of catchment average precipitation, temperature, etc. should be provided. An appendix with some details on the model structure and implementation of SWAT can be considered to make the study independent of prior applications of the model to these catchments. As the main focus is model diagnostics, it is essential that readers are aware of the model structure and the details of its implementation.

R3: We agree that more information of the SWAT model would be helpful to understand our model diagnostic approach. Thus, we will provide more detailed information of the SWAT model in the revised version of the manuscript.

C4a. Issue of CN2 (Line 16, Page 11): It is surprising that no appropriate performance criteria is found to relate to CN2, which is generally a sensitive parameter in SWAT. One reason can be the low variation assigned to it (only within +/- 10 of base value, see Table 1). On the other hand, some other parameters are allowed to vary within much larger ranges (GW_DELAYfsh between 1-50, RCHRGssh between 0.2-0.8, etc.). It is later found that these parameter display high connectivity to performance measures.

R4a: We have selected parameter ranges based on former studies (Guse et al., 2014, 2016a, Pfannerstill et al., 2014, 2015) and we think that they are well justified. Even for other model parameters such as RCHRGssh which can range from 0 to 1, we have reduced the parameter ranges to minimize the number of inappropriate model simulations. The parameter ranges were selected so that the processes are adequately represented and unrealistic parameter combinations (values) are intended to be avoided.

An increase of parameter ranges can always result in a higher risk of unrealistic high or low relevances of a certain hydrological component. In our experience, an increase in the parameter range can increase the impact of a model parameter in investigating its influence on a performance criterion. However, we are not aware of an example of a SWAT model in which a parameter with a low relevance becomes strongly relevant after increasing a well-justified parameter range. In the case of CN2, we have to add that the initial parameters were already carefully checked. Thus, a value for a certain land use type of 50 means that the CN2 was varied in these HRUs from 40 to 60 which is assumed to be sufficient to maintain the landscape heterogeneity. Moreover, we like to highlight that surface runoff is only of low relevance in the Treene catchment and of medium relevance in the Saale catchment. A higher relevance of CN2 would be expected for catchments with higher relevance of surface runoff. Thus, the selected ranges are seen as representative to reproduce the process accurately and to avoid unrealistic high or low contributions of surface runoff.

C4b: Please also mention the units of parameters in Table 1.

R4b: Units will be added.

C5. Threshold of performance: Figure 1 shows that negative NSE and KGE values were also allowed in the tree construction. The issue of using parameter sets related to highly degraded performance has been raised and addressed by earlier studies (Kelleher et al. 2013). Should a threshold of performance be fixed and only those parameter sets that perform above it considered for further analysis?

R5: We are aware that the performance of model runs is different and that a reduction of a certain performance criteria shows a lower quality of the model run. However, we are not aware of a consistent approach to identify thresholds for "good“ and "poor“ model runs. All selections of thresholds are somehow arbitrary. Thus, we prefer to use the whole data set of 2000 model simulations. However, we can add to the manuscript that the values of the performance criteria can have a high range and that the ranges

are shown in the correlation plots.

C6. Convergence of results with number of LHS samples: 2000 parameter sets are used in the analysis but no discussion on the stability of results w.r.t number of LHS samples is provided. One way to test this is to look at the agreement between current results with those from a subset of 500, and 1000 sets. Typically, the number of sets after which little fluctuation in results is seen is used.

R6: We see 2000 parameter sets as a good number to represent the parameter space accurately for our purpose. According to our experiences with the SWAT model and the Latin-Hypercube sampling a reduction to 500 and 1000 does not lead to a good coverage of the parameter space. In this case, the information content might be too low for a realistic construction of regression trees, since the number of model runs within a subset reduces from node to node and in the case of 500 models finally the number of model runs is too low in a subset to provide reasonable results.

C7. Equation 3, Page 7: Please elaborate how the RSR calculation is implemented. Say there are only 10 flow values for 0-5 percentile range for observed flow but 100 such values are present for simulated flow, how is RSR then calculated?

R7: Both, observed and simulated discharge time series are available for the same modeling period, i.e. both have the same length. In constructing the FDCs, both time series are separately considered. However, the number of days for 5% of the total time series is identical. Thus, we have the same number of flow values for example in the 0-5 percentile range. Based on this, the equation for the RSR can be applied.

Technical Corrections

C1. Line 1, Page 1: Consider replacing ‘parameters are used to adapt the model to the conditions of the catchment’ with ‘parameters are used to represent the time-invarying characteristics of the catchments’.

R1: Changed.

C2. Line 1, Page 2: Consider replacing ‘In models’ with ‘In rainfall runoff models’.

R2: Changed.

C3. Lines 7-9, Page 2: It is now generally accepted that parameters may or may not be identifiable (Beven, 2011).

R3: We are aware of studies on parameter identifiability. However, we still think that it might be worth to investigate the parameter identifiability using different performance criteria and to understand why a certain model parameter cannot be identified and whether this is related to the selection of the performance criteria.

C4. Line 21, Page 4: Explain ‘high drainage activities’.

R4: High drainage activities mean that a high percentage of agricultural areas is covered by drainages. We will describe in the revised version of the manuscript that large parts of agricultural areas are drained.

C5. Line 28, Page 4: Replace ‘temporally’ with ‘temporal’.

R5: changed

C6. Line 22, Page 11: Remove ‘extremely’.

R6: changed

C7. The text size in Figure 3 (lower panel) should be increased for visibility.

R7: changed

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.

Therneau, T.M. and Atkinson, B. (2010): Rpart: Recursive partitioning. R package, http://CRAN.R-project.org/package=rpart.

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

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