Identifying the connective strength between model parameters and performance criteria By Guse et al.

General Comments

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

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.

Specific Comments

1. 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.

2. 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.

3. 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.

4. 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. Please also mention the units of parameters in Table 1.

5. 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?

6. 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.

7. 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?

Technical Corrections

1. 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’.
2. Line 1, Page 2: Consider replacing ‘In models’ with ‘In rainfall runoff models’.
3. Lines 7-9, Page 2: It is now generally accepted that parameters may or may not be identifiable (Beven, 2011).
4. Line 21, Page 4: Explain ‘high drainage activities’.
5. Line 28, Page 4: Replace ‘temporally’ with ‘temporal’.
6. Line 22, Page 11: Remove ‘extremely’.

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

References

Beven, K.J., 2011. Rainfall-runoff modelling: the primer. John Wiley & Sons.

Herman, J.D., Reed, P.M. and Wagener, T., 2013. Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. Water Resources Research, 49(3), pp.1400-1414.

Kelleher, C., Wagener, T., McGlynn, B., Ward, A.S., Gooseff, M.N. and Payn, R.A., 2013. Identifiability of transient storage model parameters along a mountain stream. Water Resources Research, 49(9), pp.5290-5306.