Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-88-RC2, 2016

Hydrology and Earth System Sciences Discussions

Interactive comment on "Searching for an optimized single-objective function matching multiple objectives with automatic calibration of hydrological models" *by* F. Tian et al.

Anonymous Referee #2

Received and published: 22 May 2016

1 General comments

The paper by Tian et al. presents a single-objective calibration method producing similar flow simulations compared to a multi-objective calibration approach. The paper provides a synthetic review of the relevant litterature and deals with an important topic, namely the choice of an objective function while calibrating a rainfall-runoff model. The paper introduces an innovative objective function with a variable exponent applied to the model residuals (named OEV), whereas most of the litterature published so far has focused on data transform applied to streamflow data (e.g. square root or log transform). Finally, the authors have covered several important issues related to model

calibration including an attempt to apply the method over a large number of catchments and the impact of the transition from a calibration to a validation period.

However, we believe that the paper requires significant improvements before it can be accepted for publication. Four points require the author's attention:

- The authors need to clarify their approach in dealing with model uncertainty. Whatever objective function is used, a single-objective calibration method always produces a single parameter set. Consequently, it cannot quantify model uncertainty unless it is placed within an uncertainty framework (e.g. Bayesian methods) where residual error modelling is part of the model calibration. The authors haven't mentioned such a framework in their paper, as a result their single-objective calibration method has no ability to quantify model uncertainty and cannot be compared with a multi-objective calibration method from the aspect of estimating uncertainty bounds. We suggest removing all reference to model uncertainty in the paper, especially in Section 4.4 and Figure 7.
- The conclusions of the paper rely heavily on Figure 2, 5 and 6. However, the figures are hard to read and aggregate a lot of information, which makes it difficult to judge on the validity of the associated comments. We suggest adding several tables or figures to clarify the findings. For example:
  - Figure 2: The impact of the two metrics ROCE and SFDCE is difficult to distangle from the two other metrics (NSE and TRMSE). The use of the size and color of individual dots is not recommended for such a dense plot. We suggest using standard 2d scatter plots showing the relationships between two metrics only.
  - Figure 5 and 6: Hydrograph plots are generally noisy and hard to comment on, unless they are plotted over a very short time scale (e.g. a flood event). We suggest redrawing the plots using flow duration curves (similar to figure

7) and 2 or 3 flood events. In addition, the analysis of hydrographs on 8 sites is not sufficient to judge on the quality of model calibration method. If the intend of the authors is to show the similarity between the OEV calibration and an alternative method, we suggest computing a similarity metric between simulations produced from the two methods. For example the four objective functions NSE, TRMSE, ROCE and SDFCE can be used where $Q_{o,t}$ is replaced with the simulated value produced with the OEV calibration. Low values of these metrics would suggest that both simulations are similar. This approach would offer a quantitative approach in the comparison of simulations.

- We believe that one of the reasons behind the similarity between the single and multi-objective calibration results reported by the authors comes from the lack of diversity in the objective functions selected for the multi-objective calibration exercise. More specifically, the NSE and $TRMSE$ are both metrics that compute the sum of squared residuals of flow simulations (see detailed comment #1). $TRMSE$ uses an Box-Cox transform with exponent 0.3, which is not putting a very strong emphasis on low flows. As suggested by Pushpalatha et al. (2012), an exponent between -1 and 0 ($\log$ function) would be more appropriate. Such an exponent would clearly distinguish a calibration based on NSE, which focuses on high flows as indicated by the authors, from a calibration based on TRMSE.
- In their conclusion, the authors claim that “the methodology was applied to 196 (...) watersheds” (page 25, line 387). Such a large scale testing provides a strong support for the authors’ conclusions. However, we noted two important shortcomings in the way the authors used the MOPEX catchment dataset:
  - First, the authors reduced the catchment data set from an initial list of 438 catchments to 196. The authors indicate that the catchments were selected “because of the applicability of the Xinanjiang model”. This point is a ma-

jor problem because we believe that a calibration method should not be tested on good behaving catchments only. Trade-off in calibration between different objective functions appear when the model cannot reproduce the full extent of the flow regime, which is generally a synonym of poor model performance. As a result, we suggest expanding the dataset used in this paper to catchments where the model does not perform well, and check if the paper conclusions still hold in less favourable modelling conditions.
  - Second, the authors tested their calibration method on 196 catchments, but only reported validation statistics on 8 “representative catchments” (see page 17, line 305) that were selected “arbitrarily” within each OEV group (see Page 16, Line 287). The authors do not provide additional details on the rationale behind this selection process. We believe that this point constitutes a major issue in the paper, where important conclusions are drawn from a very small sub-sample of the initial dataset. We strongly recommend reporting the validation results on the full set of 196 catchments, or if possible, on the complete MOPEX dataset.

Detailed comments are provided in the following section.

2 Specific comments

1. Page 11, Line 194, Equation 3: Please reformulate TRMSE to match the formulation of NSE. We suggest using the following formulation:

$$TRMSE = 1 - \frac{\sum_t \left(Q'_{s,t} - Q'_{o,t}\right)^2}{\sum_t \left(\overline{Q'_o} - Q'_{o,t}\right)^2} \tag{1}$$

This reformulation would convert TRMSE to a dimensionless metric and greatly facilitate the comparison between the two metrics. Alternatively, the formulation of both metrics can be combined using the following notation

$$NSE(\lambda) = 1 - \frac{\sum_t \left[(Q_{s,t}+1)^\lambda - (Q_{o,t}+1)^\lambda\right]^2}{\sum_t \left[\overline{(Q_o+1)^\lambda} - (Q_{o,t}+1)^\lambda\right]^2} \tag{2}$$

The analysis could proceed based on $NSE(1)$ (equivalent to the current NSE) and $NSE(0.3)$ (equivalent to the current $TRMSE$).

2. Page 13, Line 232: "We selected one representative from each category". Please precise the procedure of catchment selection. As indicated in the previous section, we strongly recommend using this catchment sub-sample for illustration purpose only, and avoid drawing conclusions from such a limited catchment dataset.

References

Pushpalatha, R., Perrin, C., Le Moine, N., and Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *J. Hydrol.*, 420:171–182.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-88, 2016.