Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-424-RC1, 2017

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Assessment of lumped hydrological balance models in peninsular Spain" *by* Julio Pérez-Sánchez et al.

Anonymous Referee #1

Received and published: 18 September 2017

General comments:

The present paper aims to conduct a comparative analysis among six lumped hydrological models applied to streamflow simulation in 16 watersheds in Spain. The watersheds have different climatic regimes with more than 30 years of data. Models are used to generate monthly streamflow, and are compared with respect to six quality metrics. The Bressiani classification scheme is used to assess models performance trying to investigate which model is satisfactory for a given watershed. Even though the paper is well-written, the main problem is not stated well, and the rationale behind using lumped hydrological models (rather than semi-distributed models) is not convincing enough (see specific comments). I believe that this study, rather than a research article, is an exploratory analysis of a number of models, which can be presented in

terms of a technical paper. I also have a number of comments that could improve the quality of this work if authors decide to resubmit the manuscript in future.

Specific comments:

- My main issue with the current study is the fact that authors point out only to a couple of troubles in distributed and semi-distributed modeling, and then decide to use lumped models for streamflow predictions. I argue that there has been a great deal of advances in modeling in the past decades, which has resulted in ease-of-use and high identifiability of semi-distributed models for water management. Lumped models are no longer a justifiable option for water resources modeling. I agree that they can be used for regional analysis, but for watersheds like those studied in this paper, I totally recommend parsimonious semi-distributed models. Using lumped models, even though streamflow can be reproduced satisfactorily at coarse spati-temporal scales (such as monthly, that is considered in this paper), no management decisions could be made regarding water resources. For example, one cannot evaluate the impact of land use change or urbanization on water dynamics. Overall, using lumped models in this study needs to be justified in a better way.

- Regarding the use of lumped models for streamflow predictions, authors say that "The suitability of a model depends on the basin and specific regional characteristics." However, they never justify why lumped models are suitable for the watersheds under study. My impression is that the modeling scheme has been selected just because of simplicity, rather than basin and regional characteristics. Moreover, even though authors list more models in the introduction section, they only use six models. Why?

- It is claimed that this work takes into account the stochastic behaviour of the natural streamflow and the climatic variables. However, nowhere in the paper are any probabilistic analyses employing probability density functions. Just using time series of forcings and streamflow does not mean that the stochastic behavior is taken into account. Moreover, they also say that "We intend to calibrate parsimonious models by

considering the stochastic behaviour of the natural streamflow and climatic variables." However, no details on model calibration (e.g., algorithmic and computational settings) are provided in the manuscript.

- All quality metrics used in this study are aggregate measures that quantify simulations-observations match in an average sense. I believe that, since the ultimate goal in this study is to use model predictions for water management, authors need to utilize more specific quality metrics including those that target low-flows, high-flow, timing, etc. I suggest looking at Gupta and Kling (2009), Yilmaz et al. (2008), and Shafii and Tolson (2015), some examples in a large literature on diagnostic model evaluation.

- Model verification (i.e., in the time frame 1995–2010) has not been demonstrated in the paper. Only Figure 7 graphically shows time series in that time frame, but no quality metrics are calculated and no comparison are made either.

Technical corrections:

- Tables and figures captions need to be longer providing more details. Also, consider merging Figure 1 and 2 in one figure. Fonts in Figure 4 need to be larger. Figure 7 does not demonstrate how well models perform, because it is extremely busy. I recommend using Flow Duration Curve instead of streamflow time series. It makes comparisons conclusions easy to follow.

- On page 7, it is mentioned that "ensemble of models is performed". But there is no information on how it is done. Is it the average of all models, or what?

- Consider rewording the sentence "The range of missing values moves from 2% to 8% in the stations considered" to "Missing values ranges from 2% to 8% in the stations considered"

- Model parameters need to be provided in a table, along with the prior ranges used for model calibration, as well as the optimal values obtained by calibration.

- On page 9, reword "models were proved to perform well" to "models performed well"

- Define "best fit" in Tables 5-7. What conditions need to be satisfied so the model becomes a "best fit" in a watershed?

References: Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22(18), 3802-3813. Shafii, M., and B. A. Tolson (2015), Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives, Water Resources Research, 51(5), 3796-3814. Yilmaz, K. K., H. V. Gupta, and T. Wagener (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resources Research, 44, W09417.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-424, 2017.