**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 #2

Received and published: 10 November 2017

This study provides assessment report of a lumped hydrologic models inter-comparison conducted over sixteen mid-sized Spanish River basins. Streamflow simulations of six hydrologic models are compared against observations using six different statistical metrics. The authors reports that the lumped models show better skill in the humid basins as compared to those achieved in the sub-humid and semi-arid basins. They also provide some qualitative guidance on selection of statistical metrics used for model verification. The overall impression I have after reading this work is that it falls quite far way from qualifying as a research article, and as such it appears to me written like a (technical) report.

I have also a hard time to identify novelty or new insights that can be gained from this study – considering that the main finding (models do better in humid catchments) is

well demonstrated in prior works. Another thing that the manuscript seriously lacks is that there is virtually no discussion on why the different models show different behavior – and as such it goes in the direction of showing analysis results to "what" sort of questions and not discussing about "why". How does the selected six hydrologic models differ in terms of other main hydrological fluxes and states (e.g., soil moisture storage and evapotranspiration – besides the aggregated streamflow behavior)?

What is the main motivation behind this model inter-comparison study? I do not have much issue with usage of lumped models, but the authors show clearly motivate their research work - also taking into account why/how did they come up with six models (justification for selected models). As presented now, it appears to me that the authors just came up with six models and then perform the inter-comparison study, without a clear motivation and goal.

I would also suggest authors to tone down the part discussing about lumped and distributed model advantages/limitations. Specifically distributed models have their own advantages, which goes beyond just reproducing the aggregated streamflow dynamics at a catchment outlet. I also do not agree with the author sentence about unavailability of spatial datasets to establish distributed models (line 27, page 2). These days we have now access to lot of spatial datasets (either from remote sensing or ground based) and now it is becoming routinely to apply (fully or semi) distributed models. Please modify your argument.

Page 4, line 1: You mean PET? ET is estimated by hydrologic models. Right?

Section 3.1: Please provide list of calibration parameters and their ranges and optimal values. How did you consider the varying number of calibration parameters among models in your overall ranking process?

Section 3.2: What is the objective function used in the model calibration? How sensitive the model (ranking) results are to the selection of particular objective function?

Section 4: There are only results presented in this section and virtually no discussion, so please modify the section title. How do your results fits to the previous inter-comparison study? Discuss your results in context of previous study – there is not a single citation in this section linking back your results to previous findings.

Page 6, line 25: To which period (calibration or verification) these results belong? Check also elsewhere. Could you present the calibration and validation period results separately? Is the model ranking similar between two periods?

Page 8, line 1: What do you mean by "irregular"?

I find caption of nearly all figures not very informative. For example in Figure 4 – is it for the calibration or verification period? Unit of flow should follow the SI system. Figure 5 and 6 – again to which period, does these results correspond and to which river basins? Figure 7 - please provide indication to calibration and validation periods.

Figure 2: What information is gained by having this figure in the paper - once we have already Fig 1. The quality of Figures 2 and 7 is also very poor. Hard to visualize them and also distinct features could not be identified as presented.

Overall I find there are lots of tables with lots of information inside - that are not thoroughly discussed in the text. Try reducing them to only couple of informative tables and discuss them in detail. Tables like 2 and 3 could be easily dropped out do not add much to the overall message of this paper.