Hydrology and
Earth System
Sciences

Discussions

EGU

Open Access

Interactive comment on "Behind the scenes of streamflow model performance" *by* Laurène J. E. Bouaziz et al.

Anonymous Referee #3

Received and published: 1 July 2020

This manuscript proposes a multi-objective model evaluation to compare a number of different hydrological catchment models. While this is certainly a valuable task, I honestly have very split feelings about this study. The general idea of multi-objective testing is not new, but very important, and the comparison of several models is an interesting novel aspect. However, I have a number of fundamental concerns, which would require new data and computations to be addressed.

1) The study is based on only three catchments. Several studies have shown how variable results between catchments can be, and these days with more and more data sets being available, the use of just three catchments seems a bit surprising for this type of study.

2) The study addresses different storages, including snow storage. However, the im-

portance of snow in the test catchments is minor. I could not find any information on the relative importance of snow (the info of about one month of snow cover is incomplete as this does not say anything about the amount of water stored as snow). Still, my general understanding is that snow does not play any major role in these catchments. This is probably also the reason why the authors can get away with not using any elevation zones for modelling snow processes.

3) Each of the storage estimation used for model testing is associated with significant observation uncertainties. There is also a scale-mismatch which results in additional uncertainties. These issues have to be considered!

4) Another point that seems to be missing is that each of the models of course also is affected by parameter uncertainties (which will influence the simulated storages). Perhaps I am missing something, but as I understand, single parameter sets are used for each model. This is not sufficient; we know that the same model can result in very different internal simulations because of parameter uncertainty. This leaves me wondering how much of the differences presented here are due to parameter uncertainty rather than due to model differences.