Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2021-605-RC1, 2022

Comment on hess-2021-605

Anonymous Referee #1

Referee comment on "Large-sample assessment of varying spatial resolution on the streamflow estimates of the wflow_sbm hydrological model" by Jerom P.M. Aerts et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-605-RC1, 2022

GENERAL COMMENTS

This study is focused on the spatial resolution effect of the wflow_sbm model on the streamflow performance over the CONUS domain. The streamflow performance is evaluated through the KGE score at three spatial resolutions: 3km, 1km and 200m. To this end, the authors follow a benchmark approach in which they compare their results against a statistical benchmark in order to select or reject their simulations. The main conclusion of the study is that, besides some strong locality in the scaling behavior, finer resolutions do not implicate a better streamflow performance.

Although I find this work interesting and appropriate for the scope of this journal, there is still room for improvement before I can recommend its publication. In its current form, the manuscript should be reconsidered after major revision. I hope the comments below will help the authors improve their manuscript.

MAJOR COMMENTS

- The streamflow performance of the different model instances is evaluated through the KGE score. Although the authors state in L150-L152 that they assessed the KGE score for both a calibration and a evaluation period, it seems that the results are mainly focused on the evaluation period: the CDFs of Figure 7 correspond to the evaluation period, and at least the map in Figure 8d also corresponds to the evaluation period according to the figure caption. It is not clear if Figures 8a, b and c also correspond to the evaluation period. The calibration results briefly appear in Figure 3 for an example basin, but I consider this insufficient. Therefore, my recommendation is to include the CDFs for the calibration period in Figure 7 (see also next two comments), and clearly distinguish between calibration and evaluation scores in the figure captions.

- Similarly to the NSE score, KGE can be decomposed into three parts: the coefficient of correlation, the ratio of the mean values and the ratio of the standard deviations (Gupta et al., 2007; Knoben et al., 2019). All these CDFs should be present in the manuscript, as they will help understand why the KGE values are as they are. Apart from the CDFs for KGE, Figure 7 should collect the CDFs for these three component (not necessarily for the MARRMoT ensemble, although it would be more than welcome). These new results should be discussed as well.

- The two-fold statistical benchmark (one for the mean and one for the median) produces a poor performance (Figure 6d) that wflow_sbm can easily beat for most of the basins (Figure 6b). Although this is not a problem, I feel curious about why the KGE values are so low for the statistical benchmark. Then, the decomposition of the KGE score mentioned above should also be done for the statistical benchmark and should be incorporated into Figure 7 (a multi-panel figure where the plotted lines can be differentiated from each other may be the best way to show all this). This will help understand why the "mean statistical benchmark" outperforms the "median statistical benchmark" (Figure 6c). In particular, the ratio of the mean values will provide an interesting insight: is the ratio of the mean values closer to one for the "mean statistical benchmark"?

- The Discussion section is not structured and is written as a single block. It can be clearly divided into two parts: one part discussing the benchmark selection and one part discussing the spatial scaling effect. For sure, the new CDFs will strengthen the results and will enrich the discussion.

- I also miss in the discussion some recent and important references for the CONUS domain: for example, Mizukami et al. (2017) (already cited in the Introduction) and Rakovec et al. (2019) also carried out a large-domain calibration exercise and followed a benchmark approach to evaluate their results for the CONUS basins. Are the results of this study similar to their results?

References

Gupta et al. (2009): http://dx.doi.org/10.1016/j.jhydrol.2009.08.003

Knoben et al. (2019): https://doi.org/10.5194/hess-23-4323-2019

Mizukami et al. (2017): https://doi.org/10.1002/2017WR020401

Rakovec et al. (2019): https://doi.org/10.1029/2019JD030767

MINOR COMMENTS

Title

- The title is extremely long and sounds like a sentence extracted from the abstract or the conclusions. I would suggest a more concise title, something like "Large-sample assessment of spatial scaling effects on the streamflow estimations of a distributed hydrological model". The reader will find that "finer spatial resolution does not necessarily lead to better streamflow estimates" in the abstract. In any case, I will leave this open to the authors.

Section 2.1.1 The CAMELS data set

- The authors point out three reasons behind failed runs: errors during parameter derivation, errors during run time and missing streamflow observations. While the last one is clear, the other two are not properly described. What do the authors mean by "errors during parameter derivation"? Is this related to the parameter estimations from external sources prior to calibration? Or is it related to the calibration procedure? On the other hand, what do the authors mean by "errors during run time"? I suggest a more detailed description.

Section 2.2.3 Model Runs & Calibration

- The parameter KsatHorFrac is the only parameter subject to calibration, and the rest of the parameters are derived from external sources. Firstly, the parameter range for KsatHorFrac should be indicated here and not in L198 when the results are presented. Secondly, it is not clear if the selection of this parameter is based on prior studies, on calibration recommendations for wflow_sbm, or on a sensitivity analysis carried out by the authors. Some information is provided in L60-L62, but I find confusing to read this in the introduction. I suggest mentioning this information in section 2.2.3 as I feel it belongs here.

- How is the model calibrated? Do the authors use a calibration algorithm? Is it based on a Montecarlo experiment? No details are given on the calibration procedure, only L153-L154 state that “the calibration procedure finds an optimal parameter value based on the KGE objective function of streamflow estimates at the basin outlet”. The calibration procedure should be properly described.

Section 2.3.2 Comparison of Streamflow Estimates

- The last sentence in L187-L188 seems incomplete, or at least has no cohesion with the previous sentence.

Section 3.4 Benchmark selection

- Instances of “Figure 7” throughout the paragraph seem to refer to Figure 6.

Section 3.5 Streamflow estimates of model instances

- “Figure 5” in L249 seems to refer to Figure 7.

- The colorbar in Figure 8c should indicate “KGE difference” or “â□□KGE”. “KGE value” is not correct.

Section 4 Discussion

- Should “their” in L303 be “there”?