Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-216-RC1, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Implications of Model Selection: A Comparison of Publicly Available, CONUS-Extent Hydrologic Component Estimates" *by* Samuel Saxe et al.

Janneke Remmers (Referee)

janneke.remmers@wur.nl

Received and published: 2 July 2020

Summary

This study evaluates the effects of model disagreement for publicly available estimates of the hydrological cycle. These estimates were created by hydrologic modelling, re-analysis and remote sensing. Multiple statistical methods were employed to perform this comparison, e.g. the coefficient of variation, the Mann-Kendall trend test, Sen's slope and the Spearman's Rho test. The datasets were compared based on precipitation, actual evapotranspiration, root zone soil moisture and discharge. The study focuses strongly on the Land Surface Models in the results, discussion and conclu-

sion. The main conclusions are that the disagreement between models varies spatially and the authors recommend the use of model ensembles. In my opinion, this study suits HESS well.

Major comments:

I think this study supplements the previous studies concerning the evaluation of publicly available datasets. The use of literature is extensive, which supports why this study supplements the existing science. In the introduction this is done thoroughly and in detail. I have understood that there are two reasons: discharge has not been compared that often and the comparison is often made with in-situ observations, not between the models. In the conclusion, the reasons why this study is supplementary is revisited, though mainly the latter reason. Within the results and discussion, I missed this focus. In both sections, it seems to me that the focus is put on P and AET. While R has received less attention in the past, it was not discussed as much as I expected it to be. For example, Fig. 7 does not contain a subplot with R. It seems that a subplot can be easily added, since one subplot is empty.

The study has been set up well, in generally. The use of different statistical methods complement each other well. There are three aspects that can either be improved or be stated more clearly.

1) I think the readability would be improved if a more clear structure would be implemented in the statistical part of the methods, as is already done in the results. I would suggest to use subsections and state more clearly how each subsection contributes to the main goal and relates to the other subsections.

2) I was confused by the different time periods applied for the analysis. By implementing a clearer structure, I think this will be partly remedied. On top of that, I would suggest to clarify the purpose of each time period more clearly.

3) The relative imbalance is calculated for 8 out of 10 ecoregions in the cases studies.

Two regions, “Tropical west forests” and “Southern semi-arid highlands”, are excluded due to their relatively limited extent, respectively 0.27% and 0.82% of the total CONUS land surface. Yet, three other ecoregions (“Temperate sierras”, “Marine west coast forest” and “Mediterranean California”) have a relatively limited extent too (respectively 1.11%, 1.35% and 2.06%). Thus, I wondered why the two former ecoregions are excluded, but the latter three are not. Concerning the imbalances, the terminology of ‘imbalances’ is proper, as opposed to using ‘residuals’. The implicit assumption here is that the residuals, i.e. model uncertainty, are considered constant for every case study. Clark et al. (2008) show that model uncertainty differs between climates. Especially, more arid climate have higher residuals. Thus, the analysis based on the relative imbalance can only be semi-reasonably compared within 1 ecoregion. I do think imbalances is an interesting concept to use.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., ... & Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research, 44(12).

As mentioned above, the originality of the study can be shown more clearly in the results/discussion. Aside from that, I would suggest to critically examine both sections as well regarding the text and the figures. Some aspects of the analysis seem odd to me. For example, line 464 states “Runoff datasets show the most consistent spatial distribution of $u > 0$ across the study ecoregions.” In the methods, it is stated that u has a range between 0 and 1 (line 294). Thus, to me the runoff datasets do not show a great consistency, since basically the whole range is still possible except 0 (based on the text). Another example is the description of figure 3c (line 355). I did not see 3 distinct clusters initially. However, maybe it is more easily visible if it is more visible how the different hydrologic models are clustered. Then, I have some suggestions for different figures to increase their clarity:

1) Fig. 3: continuing on the just made comment. I think it would be beneficial for your

analysis to indicate what type of model each hydrologic model is. Maybe a sub- or superscript can be added, e.g. 1 = LSM, 2 = CM and 3 = WBM.

2) Fig. 7: I have already mentioned this. One subplot is missing. Here a subplot concerning R can be added.

3) Fig. 8: The visibility of the 10th and 90th percentile lines is minimal. I would enhance this more.

4) Appendix 2.5 – 2.9: The dataset names overlap each other in multiple graphs. Also, these graphs are an addition to figure 6. However, these figures in the appendix are referred to more than figure 6 itself. To me, this seems odd.

5) Appendix 2.10: Here several bins are used. The first two are clear to me. I am not sure if I understood the last two completely. The bin with $p < 0.50$ does this mean 0 – 0.50 or 0.05 to 0.50? And the final bin with $p > 0.05$, should include most of the points of the previous bins. But these are not shown in the graph. That is why I though this bin is maybe $p < 0.05$. Have I understood this correctly?

6) General remark: There are quite a few figures in the whole study and most figures have quite a high information density. Because of this, I thought it was difficult to see where the research was heading. If possible, I would suggest to consider what the main results of this study are and focus on those results.

So to summarise, I would suggest to shorten the introduction; to structure the statistical methods more clearly; to adapt the case studies; and to thoroughly revise the analysis and use of figures.

Minor comments:

Besides these major comments, I have noticed some smaller issues. Most are quite easily changed.

1) The abstract is too long.

2) The datasets employed in this study were selected based on their public availability, which is I think a proper criterion to base the selection on. I did notice that the three types of datasets are not equally represented. Does this bias the results? Is it possible to reasonably compare the different types for 1 variable, e.g. for AET?

3) Throughout the main body of the manuscript, I noticed the abbreviation RZSM (root zone soil moisture) was used. Only in the introduction, the abbreviation SM (soil moisture) is used. I was initially confused because of this, since to me it seems to describe the same object. So I would recommend to be chose either one of the abbreviation for the whole manuscript, unless I have perhaps interpreted this wrongly.

4) Finally, the numbering of the sections and one table is incorrect. The methods section is numbered '3', which I think should be '2'. I also noticed that Table 2 has in its header number '1'. I would recommend to check the full manuscript for its numbering, both the sections and the figures/tables.

I look forward to reading the revised manuscript.

Kind regards, Janneke Remmers Wageningen University

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-216, 2020.