

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1  
<https://doi.org/10.5194/hess-2021-211-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on hess-2021-211

Anonymous Referee #1

---

Referee comment on "Towards hybrid modeling of the global hydrological cycle" by Basil Kraft et al., Hydrol. Earth Syst. Sci. Discuss.,  
<https://doi.org/10.5194/hess-2021-211-RC1>, 2021

---

Kraft et al present a convincing example of the potential of hybrid modelling. The H2M framework combines a neural network with hydrological model constraints. I definitely see the value of this research line in the context of hydrological modelling, demonstrated in this manuscript by the comparison with several established GHMs. Unfortunately, the manuscript is quite hard to read, especially because in the figures several letters dropped of the axes, which made it a puzzle to find out what was shown where (I did not manage to solve this puzzle for Fig 10), this made it hard to estimate if all conclusions are robust / valid. Besides, some sections and choices are hard to follow for an average HESS reader with average ML knowledge (as I consider myself that way..). Below I indicate this in more detail, hopefully the authors will be able to improve and clarify this in a next version.

Please solve the axes issues for all figures. In Fig 4, for instance the N is missing on the y-axis, and the 40 and 60 have dropped off, and the legend is unclear (my guess is it should be H2M and GHMs). In Fig 5/6, first letters of the month dropped off x-axis, and it took a while before I realize the two middle panels show variation over the years (not only because the numbers dropped off, it would be helpful to add a label 'years', in the same way it would be helpful to add a y-axis label TWS or SWE). In Figure 10 I don't know which variable is on which corner of the pyramid. Perhaps updated figures can be uploaded in response to this review, so that other reviewers can use these figures.

Besides the axes-issues, I think the figures themselves are anyways challenging read. There is a very high information density in each figure, but the figures are often not directly showing what is most interesting. For Fig 5/6 for instance, one could consider to show the difference between the models and the observations in a barplot, rather than their temporal dynamics. Figure 2 is only very very briefly introduced, even though the climate regions are extensively used for all figures.

It is unclear why the authors have decided to use the forcing from three different data sources, which makes the study more sensitive to inconsistencies / non-closure of balance, etc. Besides, it remains undiscussed how these data compare to the data used in

the earth2Observe project, because it might explain some of the differences with the GHMs.

Grid cells with large withdrawals have been removed but it is unclear which data source was used to identify cells with groundwater withdrawals.

The procedure with the static input layers is unclear to me. First, they are compressed to 30 (l.95-100 on p4) and then from 30 to 12? (l.140 p7).

In the validation, large negative NSE values were rescaled, but in Table 3 the spatial mean NSE is given. This makes the numbers provided here not comparable to NSE values obtained in other studies. Question is if it should still be called NSE, this can be misleading. In general, section 3.1 is hard to follow, because it is not directly clear what the spatially averaged signal is - is that averaged globally?.

Overall, the manuscript has a very high information density, and a combination of unclear figure axes and sometimes unclear terminology ("spatially averaged signal" as an example, but for instance also expressing soil water as a deficit requires the reader to pay a lot of attention) makes it difficult to completely understand what happens where. As written above, I see the potential of hybrid modelling and the potential of the approach of this study (comparing it to GHMs, exploring NN identified patterns), it would therefore be a waste if readers give up the reading because it is so challenging. Besides, it makes it difficult to estimate whether the conclusions are robust/valid, so I hope the authors can help me, average HESS reader, by increasing the clarity and readability.

Lieke Melsen  
Wageningen University