Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., author comment AC3
https://doi.org/10.5194/hess-2021-211-AC3, 2021

Reply on RC1

Basil Kraft et al.

Author 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-AC3, 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.

Response:

Thank you very much for the comments and suggestions and for helping us to improve the manuscript. First of all, we want to apologize for the issues with the figures. We understand that this made the review difficult and we appreciate your efforts to "solve the puzzle".

We agree that some things need to be simplified. We plan to reduce the complexity Figures 5 and 6 (only show global signal, move current Figures to Appendix), and remove Figure 4. We will also try to better explain the machine learning aspects. The NSE transformation, to deal with large negative values, could be avoided by just truncating the figure axes, which is easier to understand and we will change accordingly. Furthermore, we will improve the sometimes confusing terminology of CWD and SM, although we still need both terms as the former is used in the H2M, while the other is used when comparing the results to the GHMs.

Below, we provide answers to your comments.

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).

Response: We will fix the issues in the next version.

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).

Response: We will add labels `Month', `Year', `TWS', and `SWE' to the respective axes of Fig. 5 and 6.

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.

Response: Labels for Fig. 10 will be fixed.

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.

Response: We think that showing the dynamics is important for our storyline and would prefer to keep the plots. However, we agree that the Figures are very complex and hard to grasp. We, therefore, decided to only include the global signals of TWS and SWE (Figures B and C in the supplement to this response) and their MSC and IAV components. As the global signal does not tell the full story, we want to move the original Figures 5 and 6 (with improved axis labels) to the Appendix.

Figure 2 is only very very briefly introduced, even though the climate regions are extensively used for all figures.

Response: We will provide more details on the climatic regions (Figure 2) as you suggest.

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.

Response: We agree that the uncertainties in the forcing data have a big impact on the model simulations. In the choice of our forcing data, we wanted to stick as close as possible to observations, and they are unfortunately from different sources. Now, for consistency, we trained our model with WFDEI (as used in eartH2Observe). As the results are very similar, we think that the findings are robust (see Figure A in the supplement to this response). We plan to add the figure to the appendix.

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

Response: We used an ad-hoc solution based on Rodell (2019; https://www.nature.com/articles/s41586-018-0123-1). We removed all regions with only groundwater depletion (#7, #12, and #14). We will add this explanation to the manuscript.

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).

Response: The static data was compressed in a pre-processing step: For each 1-degree

grid-cell, we had 30 latitude x 30 longitude high-resolution cells for 20 variables = 18'000. Instead of feeding this high-dimensional data into the model, we compressed the data using an autoencoder and ended up with a vector of 30 values (non-interpretable) for each grid cell. This data was used in the model, but before adding the 30 values as an input to the LSTM, the data was further compressed to 12 values. In this step, the compression is 'learned' as part of the model tuning. We will better explain this aspect in revision.

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?.

Response: Agreed, we need to be more precise on the terminology. The spatial mean is the global signal, i.e., averaging each time-step separately across space, leading to a single, global time-series. As all values for the spatial mean in the table are positive, the transformation, that deals with large negative values, has no effect. For the cell-level median, a single negative value was reported, it would be ~-0.8 instead of -0.65 (as reported). We believe that negative values of NSE just indicate that the predicted values are worse than the observed mean, and the magnitude of -0.65 or -0.8 does not change the interpretation. However, it seems indeed that the transformation of negative NSE causes confusion and we will use the original NSE without transformation in the revision. The only impact this has is the one value in Table 3 and Figure 3, where large negative values occur in some boxplots. Instead of transforming the negative NSE, we will truncate the y axis limits in the figure. Your suggestion is much cleaner than transforming the values, which did not add any additional information.

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.

Response: We will revisit every figure, unify the terminology, and clarify the machine learning methods. With these changes, the information density could be trimmed. However, this study constitutes a first assessment of hybrid modeling on global scales and we want to be fully transparent and comprehensive about the strengths and weaknesses of the approach. Thus, we provide as much insights as possible, allowing the readers to grasp the potential but also the limitations of the approach. The presented results were selected to cover two major aspects of the study: model performance and different hydrological responses. We are of the opinion that covering both these aspects are critical to maintain the value of the study. We hope that changes proposed before would make the manuscript more readable. In addition, we will make additional effort to avoid repetition and improve conciseness in the revision. Specifically, we plan to simplify Figures 5 and 6, and whenever SM or CWD appears in a Figure, we change the label to SM (-CWD), and indicate `wet -> dry' label to facilitate the interpretation (Fig. 7, 8, and 9). Further, the text will be improved where SM and CWD are mentioned.

Please also note the supplement to this comment:

https://hess.copernicus.org/preprints/hess-2021-211/hess-2021-211-AC3-supplement.pdf