Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., author comment AC4
https://doi.org/10.5194/hess-2021-211-AC4, 2021

Reply on RC2

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-AC4, 2021

Reply to comment https://doi.org/10.5194/hess-2021-211-RC2

General comments:

This is an interesting paper demonstrating the feasibility of reproducing the simulations of global hydrological models (GHMs) using a hybrid approach (H2M). The latter is based on a toy model consisting of a series of bulk reservoirs, coupled to a statistical model based on machine learning. Results are encouraging. I am not sure that the comparison of H2M with GHMs is completely fair because the precipitation dataset used to force H2M (GPCP) is based on observations, while the one used to force GHMs is derived from the ERA-Interim reanalysis. Another reason why the comparison may be unfair is that spatial resolution of GHMs is degraded from 0.5 degree to 1 degree to be compared to the H2M simulations. Also, it should be emphasized that some GHMs are uncalibrated models. I was not able to do a complete review of this work because some Figures are not readable.

Recommendation: major revisions.

Response:

We are grateful for the comments and suggestions, which helps us to improve the manuscript. Once again, we apologize for the issue with the figures and we understand that you could not provide a full review under the given circumstances. We also thank you for the second comment after the correction of the figures. Here, we reply to both comments (https://doi.org/10.5194/hess-2021-211-RC2, https://doi.org/10.5194/hess-2021-211-RC3).

Regarding the general comments, we have now added an analysis of the same WFDEI forcings as used in eartH2Observe to make the comparison as fair as possible. In fact, this was also suggested by reviewer 1. The analysis reveals that the H2M performance is similar when using WFDEI. This additional result will be added to the final version in the appendix (see Figure A in the supplement to this response).

On the "fairness" of the comparison, we agree that the H2M modeling framework is driven by the observation data, and by design, it should be expected to be closer to the observation. One of the main motivations for hybrid modeling is that it can make use of large amounts of data, and learn from data (including data-specific biases).

As you mentioned, most of the GHMs, not only those in eartH2Observe, are often either not calibrated at all or calibrated against river discharge and are not able/allowed to learn from the observed data. Within the GHMs, only LISFLOOD uses catchment-based runoff for calibration, as clearly mentioned in L 289.

We would also like to emphasize here that the evaluation of the performance of H2M against the GHMs is actually the benchmarking that is more essential to the qualitative validation of the H2M rather than devaluating the state-of-the-art GHMs, as both approaches have their strengths and weaknesses. We tried to make this clear by stressing that the model performance comparison is not the core component of this study (L 290), but shows the strengths (the local adaptivity) of the approach.

Lastly, regarding the spatial aggregation, we agree that the information in the half degree simulation is potentially altered due to aggregation to one degree. We assume that the differences within the 2 half degree grid cells are much smaller than the spatial variability across the globe. In fact, such aggregation and disaggregations are quite common in CMIP model intercomparisons, where the models with different spatial resolutions are evaluated against each other as well as combined together to generate model ensembles. Nonetheless, we will add a clarification of the assumption in the revised text.

In summary, we think that you raised very essential questions that would benefit the manuscript and guide the readers to evaluate the results in a way we have intended to. We will gladly address the issues in the revised manuscript.

Below, we indicate the answers to your particular comments, which we consider very useful as well.

Particular comments:

L. 5 (Abstract): Is H2M a new model developed in this study? What is the added value of this approach with respect to more traditional modeling approaches? What is the meaning of H2M acronym?

Response: The model was first introduced here: https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1537-2020 and is now further developed and evaluated in this study. The acronym (hybrid hydrological model, H2M) was used without introduction in the abstract and the benefit of using a hybrid approach could be stated more clearly, which we will fix in the revised version.

L. 95 (22 static variables): unclear because 4 lines correspond to static variables in Table 1, not 22.

Response: We used four data products, some of them representing multiple variables (e.g., land cover fractions of water bodies, wetlands, artificial surfaces, tundra, permanent snow and ice, etc.), 20 variables (used to be 22 in the text but actually it is 20) in total. We will improve this aspect in the next version.

L. 113: procuct?

Response: Fixed.

L. 166, 174 (soltmax, softplus): all readers may not be familiar with these machine learning technical terms. They should be defined.

Response: We will add the definitions to the manuscript.

L. 199 (model training): more details should be given on the used machine learning approach. Is a local training (one statistical model for each model grid cell) performed or a global training (all model grid cells together represented by the same statistical model)?

Response: It is a global model processing each grid-cell individually, i.e., one model learns the dynamics of all pixels. We will make this clearer in the revision.

L. 243 (Table 2): CWD and SStor are written here for the first time and were not defined before. A clear definition should be given. The definition of CWD given in the next paragraph is not clear.

Response: We will improve this in the revision.

L. 242 (selection of models): How was model selection made? In Schellekens et al., 10 models are considered.

Response: We only selected the models for which groundwater storage was available (mentioned in L 243).

L. 284 (Table 3): the period of time for which the comparison was made should be indicated.

Response: We will add the time period (2009 to 2014) to the table caption.

L. 286 (model intercomparison): Could be completed with a water balance Table similar to Tables 7 and 8 in Schellekens et al.

Response: We will add a table showing ET, Q, Precip, and delta Storage to the revised version (see Table A in the supplement to this response).

L. 293: Mean or median scores are little informative in case of non-Gaussian score value statistical distribution. Could you plot score histograms instead?

Response: This could be a misunderstanding: we report the performance of the global signal (`spatially aggregated mean', terminology to be improved on request of reviewer 1), and the cell level median. The cell level distribution is also shown, at least for the NSE in Figure 3. Showing boxplots as in Fig. 3 for all the metrics would be too extensive in our opinion, as the manuscript is already loaded with figures. However, we can add the respective figures to your or the Editor's request.

L. 340 (Amazon basin): the Amazon area was affected by droughts (2005, 2010, 2015). Are these drought events visible in the simulations performed in this study?

Response: Figure 6 shows the Amazon region (T1 S-AM tropical) in detail. From the years you mentioned, only 2005 and 2010 are covered by our simulations. In both cases, the H2M models reproduce the GRACE patterns quite well.

Reply to comment https://doi.org/10.5194/hess-2021-211-RC3.

Thank you for the additional comments!

In Figure 10, CWD is indicated as one of the 3 considered variables, while the Figure itself shows SM.

Response: This will be fixed in the revision.

In the whole paper, there is a confusion between CWD and SM.

Response: This issue was also brought up by reviewer 1 and we agree that this aspect needs to be improved. We have to use both terms as our model simulates CWD, while the GHMs simulate SM. Whenever we compare the models, we use negative CWD dynamics as SM dynamics. We try to make this clearer by using better figure labels and by providing interpretations in the text (e.g., "higher CWD (dryer soil)" etc.).

L. 249 ("We consider the dynamics of CWD to correspond to SM and thus, the terms are used interchangeably when talking about soil moisture dynamics"): has to be clarified.

Response: We agree that this aspect needs to be improved.

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2021-211/hess-2021-211-AC4-supplement.pdf