

Hydrol. Earth Syst. Sci. Discuss., referee comment RC4
<https://doi.org/10.5194/hess-2021-211-RC4>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-211

Derek Karssenberg (Referee)

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-RC4>, 2021

The manuscript proposes machine learning (with extensive observational data as input) to identify spatially and temporally varying values of parameters in an extremely simple forward simulation model of hydrology. A small number of parameters represents how water fluxes are partitioned over various storages (however they vary in space and time!). Results are evaluated 1) by comparing prediction error with existing global hydrological simulation models, and 2) by exploring spatio-temporal patterns in parameters values and providing possible explanations of these patterns in terms of processes.

The work presented is innovative. The authors have recently published an article on the same model and data sets (Kraft et al, 2020 – referenced in the manuscript), but the current manuscript extends on this by providing a more extensive evaluation of the results (and it seems minor adjustments in the methodology). Because of the innovation I am in the opinion this is a promising manuscript. However, it needs considerable revision in particular regarding the presentation of the material: figures are often very unclear (it is sometimes even unclear what attributes are shown), the text is rather long and could be condensed providing at the same time more focus. Regarding the latter: this seems to be a proof-of-concept paper. It is thus not essential to provide a complete evaluation of the model. Instead, I believe it is more important to properly explain the methodology and key outcomes. In revising the paper I suggest the authors to possibly leave some of the results out (e.g. less figures, simpler figures) – it wouldn't harm the paper but may make it more accessible.

My main comments are:

Figures (and tables)

Most of the figures are hard to understand. Legends are often missing, panels are included without any explanation of what is shown, variables are plotted for which it has not been explained how they are calculated. This needs considerable improvement. Please enlarge the figures as well. In my detailed comments I pin-point a few things that are not clear,

please consider this as examples (not a complete list).

Objective function

The objective function contains four different observational data (terrestrial water storage, evapotranspiration, runoff, snow water equivalent). It is completely unclear how each of these are weighted in the objective function. Do you 'calibrate' against standardized values of these attributes? If so, how are these standardized? Please explain. Note that it is to be expected that this weighting has a strong influence on the results, e.g. if more 'weight' is assigned in the objective function to runoff, the model will perform better in runoff prediction. I do not expect you to explore different objective functions but at least the objective function needs to be given and it needs to be explained that this is quite arbitrarily chosen. Note that for instance in Bayesian data assimilation 'weights' of observations will depend on the uncertainty associated with the observations (high uncertainty -> low weight). This is not the case in your approach.

Training, validation, testing data sets

In machine learning, one uses training and validation sets in the procedural step of model building. Model evaluation then is done using a data set not used for model building (this is often referred to as the test data set in the machine learning world). It seems you are not separating validation and testing data sets. This is an important issue – evaluation of the model (all the performance metrics provided, almost all plots provided with model outcomes) needs to be done on data that are not at all involved in the model building phase. If you are evaluating on the same data as used for building the models, this should be clearly indicated in the manuscript and implications of this extensively discussed.

Context, aim of modelling

I could argue that in your comparison of your approach to modelling (hybrid modelling) and existing global hydrological models you are comparing apples and oranges. The main aim of hybrid modelling (as defined in your manuscript) is prediction (in the statistical sense): estimating variables at time steps for which observational data are not available. Existing global hydrological models however aim at scenario analysis (and oftentimes prediction as well). Scenario analysis may involve evaluating effects of climate change, effects of future changes in water allocation (e.g., irrigation, domestic water use), effects of future changes in land use, etc. Hybrid modelling is not suitable at all for scenario analysis as it almost completely relies on observational data on the system. If the system changes, it won't work anymore. I may exaggerate somewhat (to make my point clear), and you may disagree in which case I challenge you to convince me otherwise. In any case I suggest you 1) mention this difference in the introduction of the manuscript and 2) discuss this issue in the Discussion section. I consider this important in particular because this is a 'proof-of-concept' paper and it is thus important to position this work in the broader context.

Detailed comments (please note my comments on the figures are not complete)

p. 3, end of introduction

The introduction gives a good overview of past work in the domain. However, on line

68-72) it remains somewhat unclear what the contribution of this paper is. I suggest stating this more explicitly and also to state more explicitly what this paper adds compared to your previous publication (Kraft et al., 2020).

Code sharing

I strongly suggest sharing the code of your model on a public repository (e.g., GitHub). It will make your work more credible and will enable other researchers to build on your research.

p. 5, line 113

Clearly state what Q refers to. It is the amount of runoff generated in a pixel (or area of land), not the discharge from the pixel (streamflow). The latter can only be calculated in a spatial model that does channel routing.

p. 5, line 123

Adjust the numbering, is this a nested numbered list?

p. 6, model description

Is this a spatial model, i.e. does it include spatial interactions in the time transition functions? I don't think so, it is a local (point model). Please state this clearly.

p. 6, figure 1

The caption is too long. Reduce it and explain concepts in the main text.

p. 7, neural network

I am unsure you are building the machine learning model for all cells at once (single model) or for each cell separately (number of models equals number of cells). If the former, the method you propose is, I believe, fully non-spatial (point model for hydrology, identified separately for each cell with observations for that cell). Please explain this clearly.

p. 9, line 196

Why DELTA T instead of T?

p. 9, model training

I am wondering how you train the model. Machine learning models typically do not run forward in time. However, in this application, they are fed by temporally changing data, in a forward timestep approach. How is this done? Please explain. Providing the code would help as well.

Figure 2
Too small.

Section 2.5
Which runs of the global hydrological models were used?

Caption Table 3
What is 'median-cell level'? Please explain.

Figure 3
Why are you not including results for ET and Q as well?

Figure 4
Explain root phase and variance error in Methods. I am in the opinion however this plot could be left out.

Figure 5.
What are the small panels on the left and right side? What is plotted on the x-axis? The figure is hard to understand. It needs to be simplified (and possibly include more detailed information in a digital supplement or appendix).

Figure 7.
What is represented by each line in the figure? A single location? A single year? 'quantile of the spatio-temporal distribution' is not easy to understand.

Figure 8.
What is represented by the colours?

Figure 10.
Too small. Consider selecting runs and plotting these.

Figure 11.
Too small. What is along the x-axis? Months? It runs up to 13.

Discussion section

The discussion is interesting but it is somewhat long. Consider reducing it somewhat focusing on the main things (that are relevant to the research objective and questions).

Derek Karssenberq (reviewer)