Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-368-RC1, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Benchmarking a Catchment-Aware Long Short-Term Memory Network (LSTM) for Large-Scale Hydrological Modeling" *by* Frederik Kratzert et al.

Anonymous Referee #1

Received and published: 28 August 2019

This very interesting paper of Kratzert, et al. compares the quality of the predictions of various hydrological models with three variants of the Long Short-Term Memory (LSTM) deep learning network. One of these variants, the novel EA-LSTM, is trained using meteorological data and catchment similarities as an additional input and is analysed in detail highlighting the superiority of such a network. In general the paper is very well written and it is worth to be published after some minor changes. Some comments:

1. Maybe you could explain better the differences of the analysis of the single model and the ensemble mean approach. On page 13, lines 317-320 you write: "To assess statistical significance for single models, the mean basin performance (e.g. mean NSE

per basin and across all seeds) between two different model settings was compared between different model configurations." What's the difference between model settings and configuration? If I understood it correctly the difference in the verification of the single models and the ensemble mean is: Single model: From 8 ensemble model runs, you get 8 different predictions and you calculate the verification measures (e.g. NS values) for each of it and take the average (+/- Std? in Table 2); whereas in the Ensemble mean approach, for example this measure is calculated taking the mean of the 8 predictions? 2. For clarity reasons I would not include the single model outcome in Figure 3, because this a random outcome and would look different for each ensemble run.

3. Nice to have the significance reported, which is most often not shown. Although the precision of these p-values is extremely high and the differences are probably rather neglectable caused by noise.

4. Regarding the modified NSE. Wouldn't it be easier to normalize the streamflow data (e.g. using the BoxCox transformation)? So you don't have to event a new measure and adding a constant in order to achieve stable results.

5. Looking at the results, I would conclude that the EA-LSTM is very interesting for this analysis, but for practical applications the LSTM with the coupled meteo data and catchment attributes is even more efficient and is less complex. That's why I would like to see the results of this model also in Figure 4, 5 and Table 3.

6. Are the catchment attributes kept static for all days of the year? For example the monthly mean of leaf area index could be easily varied depending on the month of the year?

7. I would suggest to delete the UMAP analysis, since the method is not explained and the results are a bit confusing.

I'm looking forward to see the results of your planned work with a higher number of

ensemble members and the inclusion of the dynamic and the static input in the EA-LSTM input layer!

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-368, 2019.