Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/hess-2021-127-RC3, 2021

Comment on hess-2021-127

Anonymous Referee #3

Referee comment on "Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models" by Thomas Lees et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-127-RC3, 2021

This paper describes two versions of a national scale deep learning hydrological model for GB and compares them to 4 conceptual hydrological models from the FUSE framework. The effectiveness of LSTM has been well established in previous studies, and so the novelty of this paper lies in its application to GB catchments. As the code, data and outputs are all freely available, I consider this to be a useful study to hydrologists concerned with modelling GB catchments. I wonder if given the limited scientific insights of this paper may be better placed in the Journal of Hydrology: Regional studies, or Environmental Modelling and Software rather than HESS.

I would like to commend the authors on a very clearly written paper- it was very easy to follow and understand.

My major criticism of the paper is that the authors never demonstrate the model's applicability to a changing climate. Even if the application of LSTM (and all models that rely entirely on calibration) is only for near term flood forecasting, it is likely that we will be modelling events outside of the training data of the model with increasing frequency. I think that an alternative calibration/validation strategy should be examined where extreme events are left out of the calibration of the model, to provide some confidence in its ability to model beyond its training dataset.

My other major criticism is that the authors never discuss the insights gained from the LSTM model. There is no discussion of the sensitivity of the model to the different inputs and how the model ends up being structured. They never provide any evidence to answer their third research question. I think this would add a lot more value to the paper and make it worthy of publication in HESS. In the conclusion the authors state that this will come in a subsequent paper, but I think it would be more valuable here (and some of the detail of the calibration/validation could be moved to the supplementary information).

Some more specific comments follow:

line 19: There are more modern PBSD models than SHE. Reference Parflow, SUMA, SHETRAN, Hydrogeosphere etc.

line 77: there are only 3 research questions

Figure 1: You can format text in python to include superscripts "$mm day^-1$". Reduce point size- they are overlapping and obscuring each other.

Table 1: Nice! Very useful table. Temperature should be referred to with a capital T. Should Xt actually be Xn if it is representing the concatenation of dynamic and static input data for a single catchment?

line 176: Include the link to the prediction and error metrics at the end of the article too.

Table 2: Why these attributes? Was LSTM sensitive to all of these?

line 220: What is an epoch? how does this relate to number of catchments/years of data?

Table 3: How is statistical significance calculated here? Double check that it is the appropriate method.

Figure 3: Nice figure

line 366: I don't think that the catchments with significant snowfall should be included in the comparison if the snow modules of the conceptual models have not been turned on- this does not seem like a fair comparison. Recalculate the statistics leaving these catchments out.

line 367-371: this is a repetition of the previous paragraph.

Figure 5: cut. This is a long paper with a lot of figures. I don't think this figure adds much to the maps.

Figure 6. Label missing on the colorbar

Discussion: Cut all references to the physically based models. The comparisons are not rigorous and so should not be presented.

figure 9: significant correlations are not clear. consider showing this in an alternative way.

line 537: I think this is the most interesting point in the whole paper- I would love to read a lot more about this in the discussion.

Uncertainty: I would like to see some discussion of training models to uncertain flows and uncertain inputs.