Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/hess-2022-320-RC3, 2022

Comment on hess-2022-320

Anonymous Referee #3

Referee comment on "Utility of Deep Learning and Data-rich Regions in Predicting Monthly Basin-scale Runoff in Ungauged Regions" by Manh-Hung Le et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-320-RC3, 2022

I found this manuscript is very confusing. I am not sure about their numerical experiments. Before I have a good understanding of their experiments I cannot give a good review on the results. Below are my comments for now. I am happy to give more detailed review after I have a better understanding of their numerical experiments from their revised manuscirpt.

- The title mentioned "quantifying uncertainties in geographic extrapolation". I am wondering how the authors quantified the uncertainties. This uncertainty quantification is one of the objectives of this study if I understand the authors correctly, but I did not see any related discussion in the introduction till the results analysis.
- The conclusion in the abstract said "This study provides insight into the selection of input datasets and ML algorithms with different sets of hyperparameters for a geographic streamflow extrapolation." I am wondering what the insights are specifically.
- The effectiveness of transfer learning depends on the similarity of the source and the targe. I am wondering whether the authors performed a similarity analysis which I think it is important to analyze the effectiveness of the extrapolation. And it might explain that adding more sample data from the souces did not improve the performance in predicting the targets.
- Line 107, what "hypothesis"?
- Why specifically chose these three ML methods? How about the more recently widely used LSTM network? It is known that these three chosen ML methods cannot learn the temporal dependence and the memory effects of the dynamic inputs on streamflow outputs.
- Did the authors consider the influence of lagged P and T on current streamflow when they designed the numerical simulations?
- Please be specific about the input and output data. Both spatial and temporal data were considered, how the authors split the data for training-validation-testing in terms of both space (i.e., catchments) and time period. The description of 25%-25%-50% of the total number of data is very vague. I do not know what the total number of data represent?
- I am confused about the local-based models. It said "using target catchments to train

the ML algorithms", did it also include the source catchments or just target catchments?

- Figure 2. I am confused about the total data, i.e., training is about 25% of total. Did this total data include all five regions (source +target) or just source/target?
- Table 3 and the 7 experiments need more explanation. I am not sure what these 7 experiments are.
- Line 241, for each of these 100 simulations, the hyperparameter tuning was performed and the best results were presented? Please clarify.