Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-320-RC1, 2022

Comment on hess-2022-320

Anonymous Referee #1

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-RC1, 2022

Title: Streamflow Estimation in Ungauged Regions using Machine Learning: Quantifying Uncertainties in Geographic Extrapolation

General:

This paper attempts to make predictions of monthly averaged streamflow in data scarce regions with machine learning models that were trained in data rich regions. The test their predictions with different permutations of training regions. As expected, the models perform better with different climates and catchments attributes in the training set. Interestingly, however, the results suggest that models trained in North and South America are more reliable than models trained in Europe. They also find, as expected, that extreme gradient boost outperforms support vector mache and random forest. The paper is written fairly well, with exceptions noted below, and provides additional support for the well established conclusion that machine learning models trained on diverse data sets can be useful outside the basins which they are trained. This paper expands that conclusion by transferring the learned models to entirely new regions, in particular to data sparse regions, which is important, as the authors point out.

It was not clear to me if these models were forward looking or backward. I am not entirely sure how useful a monthly average streamflow prediction is in practice, especially if the forcings which drive the prediction are aggregated over that particular month, which would have the prediction a backward estimate. If, however, the forcings are aggregates from the previous month, then this is valuable to water resources management. I ask the authors to make this clarification in their data and methodology sections.

This paper omits non-machine learning models from the study because they are harder to set up. And unfortunately there is no benchmark model/s presented. I believe that this could potentially draw criticism. I do fully understand the need for easy-to-use models in

some situations. I would encourage the authors to rethink their framing of the model selection in the introduction and conclusion. Perhaps it would be good to make a case for the benefits of easy-to-use models, and make a case that these shallow learning models are suitable for monthly averaged streamflow over the state-of-the-art LSTM mode, which has been shown again and again to outperform other streamflow models, even when trained out of sample.

Abstract:

Line 21 has double periods.

Introduction:

Lines 38 and 39 claim about stream gauges being the most accurate way to measure streamflow is vague and trivial. Are you making a distinction between remote sensing and in situ measurements? There are many methods of gauging a stream, some more accurate than others. I'm not sure what is the purpose of the sentence, remove or clarify.

Lines 78 and 79: If there is a good argument that ML is not **the** most promising approach, I'd like to see a citation. Otherwise just state it directly as "machine learning models are arguably one of the most promising approaches"

Line 86: I'm not sure it is obvious what at "traditional" hydrological model is.

Line 107-108: I assumed your hypothesis was about ML model's ability to transfer learning from one region to another, but here you claim that you use ML models because they are easier to set up?

Lines 108-109: I think the last sentence of this paragraph is fragmented. What kind of water resources prediction? In what context are the water resources secure or insecure?

Data:

Line 127: What is the rational for removing values greater than 2,000 cms?

Line 140: Can you make it clear if your model is making a forward or backward prediction? Is your monthly forcing aggregates from the same month in which your monthly averaged streamflow comes from?

Figure 1: What unit is catchment density?

Methodology:

Lines 207-209: This wording is a little confusing. Can you rephrase to make it clear that the validation set was used to tune the hyper-parameters? Meaning, your training set is used to get the model weights, and then you check the quality of those weights by calculating an error on the validation set, then modify a hyper-parameter and train again, then check the quality of the new weights on the validation set. And to be clear, you do not calculate any error on the test set until the hyper-parameters have been chosen, right?

Table 3: Consider moving the regions into the table header, instead of as a note.

Results and Discussion:

Line 236: "The local-based models also served as benchmark models" This should be moved to the methods section.

Limitations and further studies:

Line 326: In parentheses you have "daily or monthly", but I think you meant "daily or hourly"

Conclusions

Line 334-335: "ML algorithms to quickly test our hypothesis since ML algorithms could be easier to set up than traditional hydrological models." I think this is a bad reason to us ML. There is no use doing a study with one tool instead of another simply because it is easier.

Line 351: double periods.