Review of

Continuous streamflow prediction in ungauged basins:
Long Short-Term Memory Neural Networks clearly outperform
traditional hydrological models

by Richard Arsenault, Jean-Luc Martel, Frédéric Brunet, François Brissette, and Juliane Mai
[Paper #2022-295]

General evaluation

The authors provided satisfactory responses to most of my questions. They clarified the utilization of LOOCV, and the training of the LSTM with the given dataset. The upgraded literature review is well organized, yet they could have surveyed so many other relevant references (from North America) addressing the same topic and not solely focusing on those studies conducted by one specific research group. Thus, to further improve the quality of the manuscript, I believe, there are still a few issues that need be dealt with before the paper is deemed acceptable. Thus, for the time being, I would say moderate revisions are necessary and required.

Comments/suggestions/recommendations

P4 On this page, the authors attempted to complete the literature review, the work done by Feng D, Lawson K, Shen C. (Prediction in ungauged regions with sparse flow duration curves and input-selection ensemble modeling. arXiv preprint arXiv:2011.13380. 2020 Nov 26) should be cited since it is related to the concept brought in this study and has been applied in North America as well.

P11 Line 335, according to this sentence written by authors: « *Streamflow was not scaled itself using this approach, since the model outputs and target values are not part of the model training computations and thus have no impact on the numerical convergence efficiency*. ». This sentence raises two questions: (i) If the authors have not used the target values during the training, how do they train the model? Indeed, one of the essential elements to train and calibrate ML models is the target since the model cannot decipher the physics. (ii) I believe, the rationale behind the authors' work; that is the normalization of the target is not necessary, is not quite correct. Indeed, a target variable with a large spread of values, in return, may result in a large error gradient values; causing the weight values to change significantly, leading to an unstable learning

process and the occurrence of a rather slow learning process. Please correct the text accordingly to avoid any confusion.

P12 The response to the question about using Leaky ReLU was satisfactory. Yet, there is a need for additional clarification. If the LSTM structure used in this study is like the one proposed by Kratzert *et al*. (2018) (It appears that the authors just used two LSTM units to make a more complex one), the authors must mention the work done by Kratzert *et al*. when they are discussing the model structure. This way, more details will be provided to the readers to extract the information regarding model structure (e.g., hyperparameters).

P16 According to the following sentences, « *For example, in this study one catchment has a much larger area than almost all the others. For a hydrological model-based regionalization approach, this might skew the regressions between catchment descriptors and model parameters. LSTM, on the other hand, are strongly non-linear and are thus not bound to these limitations. They could also use these data to better predict streamflow processes at scales between the small and large catchments.* ». Is this where the authors explain the use of large catchments in LSTM modeling? Since I am not convinced with the explanations provided by the authors on how they rationalize including large catchments in their modeling. If that is because of the complexity of the model, it should be clarified thoroughly. Please provide more reasoning.

P17 In the part highlighted in blue, can the authors verify what do they want to explain to the readers? Since it is not quite clear which point (within the comments) is addressed here.

P28. To avoid any confusion, please correct the information provided in the caption of Figure 3. It is mentioned that *N*-1 catchments are used during training which is not correct. These catchments are used during training and validating using an 80-to-20 split, respectively.

Figures and Tables

None, all the tables and figures are well organized.

Editorial comments

None, this is a well-written paper.