

Interactive comment on “Resampling and ensemble techniques for improving ANN-based high streamflow forecast accuracy” by Everett Snieder et al.

Anonymous Referee #1

Received and published: 4 November 2020

This manuscript compares different resampling methods and different ensemble-building techniques to improve ANN-based flood forecasts (high streamflow). Those resampling methods and ensemble techniques are also combined, resulting in a total of 16 variants, that are compared to a base model. The base model is a classic Multilayer Perceptron with 10 hidden neurons and 25 input variables, trained using a stop training approach and the Levenberg-Marquardt algorithm. All the 16 variants and the base model are applied to simulate (I think) the streamflow for two rivers in Canada.

The manuscript is very well written, well organized and very clear. However, I am sorry to say that I find the originality and contribution to be very low, too low in my opinion

C1

for a publication in HESS. The research was certainly conducted with great care, but ensemble techniques have already been used for quite a while in hydrology (both in the hydro-informatics community and beyond). While the authors mention that most of the resampling and ensemble techniques they used are not common in hydrology, this is still just an additional application of existing techniques, some of which have already been compared. Further, there is very little discussion and analysis of the results. The author conclude that boosting methods provide only marginal improvements, they offer very little explanation.

In addition, I am not convinced that improving only specifically high flows, sometimes at the expense of what the authors call "typical flows", is the way to go. I also fail to see it as a major contribution to the hydrological sciences, worthy of publication in an international journal.

I will still detail some major and minor comments below, but my main concern regarding this paper is the level of the contribution, that I unfortunately find too low.

Detailed comments (apart from the contribution/originality issue):

- Abstract, lines 1-2: The affirmation "(...) are increasingly used for operational flood warning systems" should be supported by references to operational flood forecasting systems (not journal articles but documentation from company websites or government websites). At the moment, the affirmation is not only unsupported, but also opposite to my experience. To the best of my knowledge, there are very, very few operational agencies who use ANNs for flood forecasting, despite their use in research for more than 25 years. Also, this affirmation in the abstract somewhat contradicts page 10 lines 157-158 ("Consequently, such models may not be suitable for flood related applications such as flood warning systems").
- Page 2 lines 24-30: One of the causes of errors when simulating high flows in

C2

northern locations such as Canada is the occurrence of ice jams. Ice jams are very common and not accounted for in any way by typical hydrological models. Maybe it is possible to account for them using ANNs, but I'm not sure. I think this is one major aspect that should have been present in the manuscript, both on page 2 but also when presenting the Bow River and Don River (it would be important that those rivers be ice jams free, otherwise you have to account for that). Another issue regarding high streamflow that is not discussed by the authors is the fact that those readings are extrapolations from the rating curve. The rating curve of a gauging station is typically constructed with very few (if any) observations of very high streamflow. Therefore, this part of the rating curve is very uncertain, and this is what we use to obtain "observations".

- Page 6, Table 2 and also Page 7 line 128: Why did you use the Levenberg-Marquardt algorithm? Although it is a popular algorithm, it was shown to have oscillation problems around local minima. I think the use of this algorithm should be better justified. See for instance Kwak et al. (2011)
- Page 6 line 116: What is the forecasting horizon? You mention the word "predict" here, but from reading the manuscript it seems like you perform simulations. If they are really forecasts, I think the forecasting horizon should be specified.
- Page 10 line 170: I think there is a mistake here "(...) studies that featuring each (...)"
- Page 11: the distinction between RUS and ROS seems extremely thin to me.
- Page 13 lines 256-257: The definition of ESP that you provide here corresponds to *Extended* Streamflow Prediction, as per Day (1985), not *Ensemble* streamflow predictions. Ensemble streamflow forecasts (or predictions) can be obtained by a variety of manners, including feeding a hydrological model with dynamical meteorological ensemble forecasts.

C3

- Page 14: Why 20 members?
- Results: Why do you aggregate the ensemble into a deterministic simulation and therefore evacuate the information about the uncertainty? Why not use ensemble-based performance assessment criterion such as the Continuous Ranked Probability Score, logarithmic score, etc.?
- Page 19 lines 457-459: How can overtraining happen if you have used stop training? I think this has to be explained more.
- Page 21, Figure 7: From a general perspective, I don't see how decreasing the quality of simulation for typical flow values could be positive, even if the simulation of high flows is improved. Typical and especially very low flows are also important.
- Page 26 lines 531-536: The analysis and discussion are very thin.
- Page 26, section 4.2: I disagree with the idea of fine tuning the hyper parameters differently and specifically for each model, as it would violate the ceteris paribus principle, making it difficult to isolate and compare the influence of ensemble techniques and resampling techniques.
- Page 26-27 lines 562-563: Ensembles are already quite common in ANN-based flow forecasting model, so this is not a very useful recommendation.

References:

Day, G. (1985). "Extended Streamflow Forecasting Using NWSRFS." J. Wat. Res. Plan. Mgmt., 10.1061/(ASCE)0733-9496(1985)111:2(157), 157-170

Kwak Y., Hwang J. and Yoo C-J. (2011). A new damping strategy of Levenberg-Marquardt algorithm for Multilayer Perceptrons, Neural Network World, 21, 327-340

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2020-430>, 2020.

C4