Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/hess-2022-53-RC3, 2022

Comment on hess-2022-53

Anonymous Referee #3

Referee comment on "Using a long short-term memory (LSTM) neural network to boost river streamflow forecasts over the western United States" by Kieran M. R. Hunt et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-53-RC3, 2022

General Comments

This paper is an exciting work moving machine learning models from a research lab to an operational setting. The manuscript is quite comprehensive and compares extensively to existing operational methods, highlighting advantages and challenges associated with the machine learning model (LSTM). Overall LSTM performs better than the other methods, which is expected given the nature of the machine learning model (if we have enough data).

Specific comments

The manuscript claims that it has been the first time that LSTM has been used in a hybrid system to create a medium-range weather forecast. In sync with the comment - RC1 Clear distinction should be made with hybrid models. This can be achieved by infographics (pictorial representation) of different models and variables used with a bounding box illustrating what is called a hybrid system and how it varies for different kinds of models. This would also help the readers to understand the entire workflow.

Concerning line 53 "In this regard, studies fall into two categories – either seeking to create a model capable of replicating existing streamflow observations or seeking to create a model capable of forecasting streamflow at some future time. Several highly illustrative studies approach the former topic….." Though the manuscript proposes two different categories but essentially, from a machine learning perspective, they might not be very different. Replication is also a form of prediction for a machine learning model. Distinction based on this might not be appropriate here. A rather significant difference is the use of streamflow at previous timesteps. Or in general, the first approach could be using only the drivers (precipitation, temperature radiation, wind) where we have almost no explicit information about the inherent state of the catchment (things like how moist is the soil, how much snow we have in the catchment, which might melt) to make predictions of streamflow (series of papers published by researchers at Johannes Kepler University Linz is in this direction). While the second category could be where we explicitly include the inherent state information (flow at previous time steps) about the catchment, which will have more potential for better predictions.

For the LSTM model, 23 variables have been chosen for predictions; while not doing any extensive hyperparameter search (optimum number of variables), some rationale needs to be provided on why those variables were chosen (if any kind of qualitative selection was made).

Figure 2 of the paper is really interesting, and we can see that for some catchments, through different epochs NSE (and RMSE) changes a lot for models initialised with different weights. Is it normal for all machine learning models to vary a lot after hyperparameter tuning? Secondly, we also see that the variability in NSE (and RMSE) is also high when models are trained for 100 epochs, and they perform worse than the best models trained with 10 epochs. This could also result from overfitting. While the figure provides a nice way to represent the uncertainties associated with the model, it might also make them look more uncertain than they actually are. As there are methods in machine learning to decrease this variability, it would be interesting to see how models perform if trained for 100 epochs with early stopping criteria.

The discussion section is focused on the use of the convolution layer, a big challenge would be handling the different sizes of catchments. Generalised 1D representation of 2-D values might be a research direction. The other could be the use of graph neural network. A good example could be in the direction of the paper "Spatial and Temporal Aware Graph Convolutional Network for Flood Forecasting"

Feng, Z. Wang, Y. Wu and Y. Xi, "Spatial and Temporal Aware Graph Convolutional Network for Flood Forecasting," 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1-8, doi: 10.1109/IJCNN52387.2021.9533694.

Minor technical comment :

Figure 2 A suggestion would be to either create a plot with a colour gradient for the density of points, else making the marker size smaller can help in illustrating where most of the points are (reducing the overlap).