Dear Dr. Ding,

thanks a lot for your feedback. We will reply below to your comments. Your comments are *italic*; our replies are highlighted **bold**. The line numbers in **red** are referring to the manuscript version you commented on.

Best regards,
Juliane Mai and co-authors

Comparing a LSTM model and a one-step-ahead river forecast model

The study team reaches a most profound conclusion that the Machine Learning LSTM-lumped model outperforms 12 other physically based models for Great Lakes - Ottawa River region (Lines 14-16, Abstract, Main Result (1)).

I'm curious how the LSTM model (Sect. 2.4.1, S.2.1) compare with a simple one-step-ahead forecast model, AR(2), a second-order autoregressive process of the streamflow (only). The latter is constructed as follows (e.g., [Ding, 2018]):

$$Q[t+1] = 0 + O[t] + (O[t] - O[t-1]) = 2O[t] - O[t-1] \tag{1}$$

in which:

$O[t]$ and $Q[t]$ are the observed and simulated discharge, respectively, at current timestep t. This pre-defined AR(2) has a constant of zero, and lag 1 and 2 coefficients of 2 and -1, thus having a fixed variance for an observed hydrograph.

A comparison between the two on one of their study watersheds will help demonstrate their performances. One candidate could be Gauge ID 02GA047 - Speed River at Cambridge which has been used for validation purposes (Table S15). This is located on the east side of the Grand River, Ontario, opposite to the University of Waterloo campus, and has a drainage area of 782 sq. km.

On the Speed River, does a LSTM model that has been calibrated globally for the Great Lakes-Ottawa River region outperform an AR(2) too?

This is an out of scope analysis. All models were not allowed to use streamflow on day t-1 as an input variable to predict streamflow on day t. Perhaps this reviewer mistakenly thought the LSTM uses the previous day streamflows to predict streamflow for the current day. This is explicitly stated as not being the case in Section 2.4.1:

line 225: Streamflow is not part of the input variables.

The study design called for all models to utilize the same set of geospatial and forcing inputs to build the models, hence adding a new model with a new input variable not used by other models is completely inconsistent with our inter-comparison study design. A one day autoregressive model is incapable of simulation beyond the one day time horizon; these models are being assessed in their skill in simulating streamflow throughout a 7-year validation period. Future work by others can assess such a time series model against our suite of GRIP-GL models.

References

John Ding. Interactive comment on "On the choice of calibration metrics for "high flow" estimation using hydrologic models" by Naoki Mizukami et al. Hydrol. Earth Syst. Sci. Discuss., 2018. URL https://doi.org/10.5194/hess-2018-391-SC1.