Reply to CC2 of hess-2024-2147 (John Ding)

Dear John Ding,

Thank you for your comment on our manuscript hess-2024-2147. Please find your comment below in blue, and our reply in black.

AR2 second-order autoregressive process of the streamflow

Besides the LSTM, HBV and a hybrid of the two, the authors may wish to revisit an autoregressive baseline model called AR(2) or AR2. This, an acceleration-based metric, is expressed by:

Qar2[t+1]=2Qobs[t]-Qobs[t-1], see Azmi et al. (2021, SC1, Eq. 1).

The subject was previously discussed between me and Uwe Ehret, the current closing author, on a storm event scale in a different but related context (ibid., AC1, Table 1). To summarize my take of our discussion, below are two main points:

1) a third-order AR model, AR-3 (Model-07, therein) when rounding off the time lag coefficients, is identical to AR2, and

2) it outperforms an ANN model (Model-08) by an NSE value of 0.99 to 0.12.

For the 531 CAMELS-US basins (Lines 125-130, and Figure 2), can we infer from point 2 above that an AR2 will be a better performing model? Let's consider this a hypothesis for falsification in another open discussion forum. In theory, an AR2 projection hydrograph over/under shoots the observed peak/trough flows - just visualize a USDA-SCS triangular unit hydrograph having an upslope and a downslope projection. This is in contrast to the authors' finding that 'all [three of their] models underestimated extreme flow scenarios,' (Line 243).

References

Azmi, E., Ehret, U., Weijs, S. V., Ruddell, B. L., and Perdigão, R. A. P.: Technical note: "Bit by bit": a practical and general approach for evaluating model computational complexity vs. model performance, Hydrol. Earth Syst. Sci., 25, 1103–1115, https://doi.org/10.5194/hess-25-1103-2021, 2021.

From the comment, we assume the John Ding suggests we should add an AR2 model to the set of models applied and compared in the manuscript. Although AR-models are known to perform well for streamflow time series due to their high temporal autocorrelation, we do not see merit in adding this model to the study, as it does not contribute to the goal of the paper, which is to investigate how well hybrid combinations of conceptual and LSTM-based hydrological models do in extrapolation towards very high streamflow, and to compare this to the two natural candidates for benchmark models, i.e. conceptual-only and LSTM-only. We therefore prefer to not include an AR-model into our work.

John Ding also raises the question " can we infer from point 2 above that an AR2 will be a better performing model?", with reference to Azmi et al. (2021). Respectfully we think this question is unrelated to the manuscript under review, and therefore invite the author to contact us outside the discussion forum of the manuscript to discuss this matter.

Sincerely,

Eduardo Acuna Espinoza on behalf of all co-authors