

Reviewer #1

Summary and Recommendation

Nearing et al. test how near-real time streamflow observations can effectively be used in Long Short-Term Memory (LSTM) rainfall runoff models. They compare an autoregression (AR) approach with a data assimilation (DA) approach and test, additionally, how sensitive AR is to random gaps in the data. The manuscript (MS) is easy to follow, well-structured and suits the scope of HESS. I particularly liked the comprehensive appendix in combination with a short MS. I think that this MS can be published after some minor revisions and provide only some smaller comments and questions below.

Sincerely

Ralf Loritz

Questions and comments:

- Reading the MS I would have like to see a couple of detailed results from three or four catchments where the AR or DA worked particular good or bad and what the (hydrological and ML) reason for this might be (to underpin the discussion of Appendix F and G). For instance, what could be the reason that DA and AR reduces the predictive performance of a few of your models (Fig. F2)? You state that: "*We are unsure of the reason for this, but it warrants further exploration.*" (Line 352) maybe zooming into one of the catchments could help to give a better explanation.

Thank you for the suggestion. We do agree with the notion of this comment. Indeed, We tried this extensively before submitting the manuscript and did not find a cohesive story to tell. We do not have a strong issue with this either way, and could be convinced to add this type of analysis, however we prefer not to because we didn't find any value in doing so. It's just extra text for very little payoff. Since this was a negative result, we only included the statistical analysis in appendix F. Again, happy to take this advice if the reviewer or editor thinks it's critical, however we did make this decision consciously, and we prefer not to.

- I find it a bit unrealistic how you added the missing data. I would assume that a broken gauging station is not working for a couple of days or maybe weeks in a row and wonder how this would alter your results (e.g. all streamflow data available for training but then two weeks or more only simulated data during testing with a closer focus on particular that period and not the entire testing period).

This is a good point. We will include this type of analysis in the revision (withhold time-continuous periods of data). Also, we will include an “ungauged basin” analysis, where we test both methods (DA, AR) on catchments where there is no gauge data at all during training and/or inference (we will use gauged catchments and withhold the data, so as to have data for evaluation).

- Showing how the variance or the Shannon entropy changes of your simulations in addition to the median would be interesting (Fig.1 and Fig.3). If it remains constant, I would mention that the spread of the predictions is not affected by the data availability.

Yes, this is a strange oversight on our part. We will include CDF plots in the revised manuscript.

Personal comment: Three of the seven Co-Authors have presumably not contributed to this “technical note” as they are not mentioned in the author contribution section.

Thank you, we will add contributions by all authors in the revision.