Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-247-RC1, 2018

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Rainfall-Runoff modelling using Long-Short-Term-Memory (LSTM) networks" *by* Frederik Kratzert et al.

Anonymous Referee #1

Received and published: 1 June 2018

Summary: This paper utilizes a data-driven approach, based on recurrent neural networks, to model rainfall-runoff relationships. A novel method is applied to model runoff in catchments in the continental U.S where gage data and meteorological forcings are available, and results are compared with existing process-based model results which are used as a benchmark. The LSTM method presented is tested through various experiments where the network is either trained for individual catchments, large aggregated regional catchments, or a combination approach where models are initialized based on large catchments and then "fine-tuned" to smaller catchments. This study is presented to introduce LSTM as an efficient hydrological modelling approach that is shown to provide similar quality predictions as an existing process-based model.

Novelty: The novelty of this paper is in the LSTM network approach, which is an im-

provement over other types of data driven approaches in its capacity to retain longer time dependencies. The results indicate that this type of model, when adequately trained, provides similar results as a benchmark model and may be useful to estimate runoff in ungauged catchments. The experiments are generally well-described and organized. Overall this is an interesting study that is appropriate for the journal, but I have several comments and suggestions detailed below. They involve the description and advantages of the methodology, linking with existing knowledge of the basins in the study, and suggestions for re-organization.

Comments: In Section 2.1, it is mentioned that the LSTM overcomes the weakness of traditional RNNs to learn long-term dependencies. This seems to be addressed in the additional cell state that stores or "forgets" long-term dependencies. However, it is not clear what the difference would be, for example in a hydrological application, between the two methods. It would be helpful to include a "traditional" or more simple RNN model to the LSTM model on the study dataset to show how this capacity for long-term storage comes into play.

In general, I recommend to expand the description of the methods, particularly the significance of the forget, input, output gates, and hidden states. As it is, readers will have to dig back through 2 cited papers or further on the LSTM method, and I think that a few sentences within this section could go a long way to help interpret what is going on.

Page 6, Line 25: This is not specific and should be more detailed, ". . ..were varied and found to work well in a number of preceding tests" – what values or ranges worked well, and how is "worked well" defined? I think this "initial screening" is also referred to in the conclusion and should be more clearly addressed as to how it was done.

Section 2.1.1: The hydrological interpretation was not very useful until I got to the very end of the paper (Figure 14) where the evolution of a cell state is compared to temperature variables. Since Figure 14 and its associated discussion seem to be an

afterthought in the conclusion section, I would recommend folding this example into section 2.1.1 instead, as they both relate to a “hydrological interpretation” of the data-driven network. Also in Figure 14, some vertical lines through the figure would be useful to better link to the narrative about the thresholds between temperature and cell state.

Section 2.2: The definition of epoch is not quite clear to me – for example, is it the same as the “next iteration” loop in Figure 3 ,or something different? If the same, the idea of the epoch could be illustrated in Figure 3. It makes sense that a higher number of “epochs” in this sense would lead to improvement of the simulation as shown in Figure 4.

Section 2.3: In Figure 5 and discussion throughout the experiments and results sections, it would be useful to refer to the HUC basins (01,03,11,17) by the names of the watersheds or the regions (e.g. Pacific Northwest, Northeast, etc). This may make the results more interpretable for many readers, especially those familiar with climatology in the U.S.

Section 2.4.2: In Line 19, the statement “in our case, the network has to learn the entire hydrological model purely from available data” – should specify that this is true of any data-driven approach, not specific to this case. Also in this section, comment on why fewer epochs were needed for Experiment 2 compared to 1?

Section 2.6: This section breaks the flow of the paper between the description of the experiments and their results. I suggest placing this information earlier in the paper before the experiment descriptions or as an appendix.

Page 12, Line 24: From Figure 6b, this claim is not very apparent to me, that LSTM outperforms the benchmark for more dry catchments (in HUC 11, it seems like it outperforms in the western part but not the eastern part, but the NSE is higher in the eastern part).

Page 12, Line 27: Why is this result surprising, since the LSTM is posed as a method to retain longer-term dependencies? This is a place where it would be advantageous to show how a traditional RNN would not capture these dependencies to prove its capabilities in this area.

Figure 11 and associated discussion in Section 3.2: This may be expected since gages in the Northeast are more closely spaced and homogenous compared to the Central Plains region, where there is a large wet-to-dry gradient between Missouri and Colorado. Some discussion on the characteristics of the regions of interest would be beneficial here (linking back to annual precipitation, other climate characteristics). Also, I don’t think the Basin numbers in Figure 11 are ever defined so there is no way to interpret Figure 11 spatially (e.g. there is no way to look at a certain correlation for a pair of basins and understand why they are very different from each other). Possibly a better way to create this figure would be to order basins by longitude?

Section 4: In the conclusion, it would help to come back to the broad topic in the introduction of hydrological modeling in general, and a discussion of process based models and other types of data-driven models in the context of the results, instead of re-iterating the results. As mentioned previously, Page 20 Lines 18 onward seem to be tacked-on to the end, and would be better placed earlier in the paper and referred back to here.

Finally, a general comment regarding the results: It was found that the regional model performed better for regions with correlated discharge (e.g. the Northeast). However, the basis for the regional model was that more scenarios are present in the dataset (i.e. stated that long dry periods or extreme events may be observed in one catchment in the training, which may help to simulate similar types of events in another catchment). This makes it seem like the regional model should actually benefit for places where discharge is not correlated between stations (i.e. in the Central Plains rather than the Northeast) and spans a wider range of behaviors, whereas the opposite results are found in the study. I think this is linked to the catchment processes, in that in

the Central Plains, rainfall-runoff processes occur differently between basins, so that a set of inputs and outputs for one basin cannot translate to model outputs in another. Meanwhile in the Northeast, climate is very similar between catchments, so while the regional model may not include so many disparate events (input samples are relatively similar), it still serves to improve the overall model of a given catchment. This may be somewhat addressed in the results and discussion, but could be expanded upon and help to discuss the model in a “hydrological process” context.

Minor line by line comments and typos:

Page 7, Line 6: “as well as” Page 7, Line 19: “iteration” Page 10, Line 12: typo in “each the model” Page 10, Line 17: would expand acronym to “deep learning” Page 10, Line 20: “would help to obtain”? Page 10, Line 21: remove “e.g.” Page 10, Line 30: remove comma after “analyze”

Page 12, Line 21: This makes sense that many zero-values would lead to worse predictions, since there are effectively “fewer” data points (in that many samples correspond to zero-flow values) in those training data sets. Could comment here on whether more epochs (greater than 50) would have benefited the model or not for this region?

Figure 7: The acronyms FHV, FMS, FLV should be re-defined in this figure caption. Figure 9 (and Figure 12): tiny text in the insets, should be able to read axis values Page 13, Line 6: “more strongly” Page 13, Line 8: can barely see this from Figure 7a

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-247, 2018.