Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2021-113-RC1, 2021

Comment on hess-2021-113

Anonymous Referee #1

Referee comment on "Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling" by Sam Anderson and Valentina Radic, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-113-RC1, 2021

Manuscript Number: HESS-2021-113

Title: Evaluation and interpretation of convolutional-recurrent networks for regional hydrological modelling

This manuscript presents an interesting application of deep learning approach for modelling streamflow responses across 226 streamflow gauges in southwestern Canada. This paper is well written and mostly easy to follow. I find the application of DL approach for streamflow simulation to be quite innovative and worthy addition to the growing body of literature in this field. The paper will be of widespread interest to the community. Overall, the paper offers plenty of interesting work, however, some effort is needed to communicate the results more effectively and highlighting key findings and novelty in view of the journal audience, e.g., better describe temperature and spatial perturbations results by linking with previous studies. Additional discussion is also needed on what the DL method brings to the table in comparison to the traditional process based models. Furthermore, while the application of DL methods for streamflow simulation is interesting, it is not entirely clear how this approach could be used for real world applications. There are also questionable choices on some of the methods and data used.

I also find the most figure captions lacking in details and could be expanded to provide more details. This will avoid readers having to scroll up and down the paper to understand the details in figures. I hope these comments are helpful and I look forward to reading the revised manuscript. My detailed comments are given below.

Major comments

- It is not clear what the application of DL method bring to the table in comparison to the traditional process based hydrologic models. This is an important question as the application of DL methods may be limited to predicting within the range of training datasets. Additionally, while the authors outlined the development and evaluation of DL method for streamflow simulation as their objectives, it is not stated how the DL method could be used for real world applications beyond the proof of concept type approach presented in this paper.

- The clustering method divided the study region into six clusters based on seasonal streamflow, latitude and longitude variables in order to fine-tune the model training. However, there are a number of studies in the region which describe the spatial heterogeneity of the region. For instance, streamflow responses in the lee- and windward side of coast and rocky mountains, as well as mountainous and interior plains are know to be quite different (e.g., Moore 1991; Shrestha et al. 2012). Therefore, I would think including variables like slope and aspect will be able to better characterize the spatial heterogeneity and provide clusters that better capture the variability in the streamflow response. Better clustering can potentially improve the model fine-tuning and model performance in several regions, especially in the Eastern slopes of the Rocky Mountains where the model performed relatively poorly.
- It is surprising to see that the study used 0.75° x 0.75° resolution ERA5 reanalysis data, especially given that the authors stated finer resolution climate data may improve model performance (L637). I wonder why the authors did not used to the finer resolution data readily available for the region (e.g., Werner et al. 2019)?
- The authors described the DL methods as if the study is on image/video processing. While the methods may be same as image/video processing, there is a need to rephrase section 4 in terms of hydro-climatic modelling.

Specific comments

- L95-110: The objective and novelty need to be revised by clearly describing what the DL method bring to the table compared to the process-based models, and how the DL method could be used for real-world application.
- L135: What are the range of basin areas for the selected stations?
- L138: Naturalized flow generally means regulated flow adjusted with regulation/abstraction removed. Correct term is natural flow.
- L140-145: 40% missing data can lead to challenges in model setup. Wondering if the model performance was inferior for basins with missing data than basins with complete data sets?
- L158-170: As stated earlier including slope and aspect may improve the cluster selection and model performance.
- Figure 2: State in figure caption how the discharge values are normalized. Similarly, the authors need to provide more details in all Figure captions.
- L189: It is not clear how the gridded weather data is mapped to the streamflow stations, are the nearest grid cells or mean values from several grid cells taken?
- L315-316: Since previous 365 days of data are required, is Jan. 1 1980 is the first day used for streamflow training?
- L364: The spatial perturbation section is hard to follow, how was the amplitude of 1 used in perturbation of climate fields?
- L411: Also say temperature perturbations are constant throughout the time period.
- L425: on what basis/reference was the criteria for freshet timing defined?
- The results in Figure 4b have not been adequately described in the text.
- Figure 6: name the basins for which these example results are presented.
- L492: Given that streamflow at a hydrometric station is response to precipitation and temperature over the entire drainage basin, it is to be expected that there are higher sensitivity in response by including areas near and within the station. This need to be clarified, in the context of how big the drainage basins are, and whether the inclusion of precipitation and temperature variables from a wider region improved the model performance.
- L562: How is the intensity of freshet calculated?
- L562-580: The results in Figures 9 and 10 seem to be consistent with previous climate

change impact studies in the region. This is quite promising and is perhaps one of results the authors can highlight further. I suggest expanding the discussion in this section by linking with previous climate impacts studies.

- L589: Rephrase the sentence, it appears as if previous studies also used deep learning.
- L626-627: While it is true that the non-contributing areas may have played a part in DL results in parts of eastern cluster, the cited studies are outside of the study region and not directly comparable. Also non-contributing areas may not be a factor for the entire region. There are maps available which outline the extent of non-contributing areas.
- Table 2 heading: State the period of test set used. Also it should be clarified in the heading that various validation periods were used in reference models.
- Discussion and Conclusions: the changes suggested above also applies to the results and discussion section.

References

Moore, R. D., 1991: Hydrology and water supply in the Fraser River basin. Water in Sustainable Development: Exploring Our Common Future in the Fraser River Basin, 21–40.

Shrestha, R. R., M. A. Schnorbus, A. T. Werner, and A. J. Berland, 2012: Modelling spatial and temporal variability of hydrologic impacts of climate change in the Fraser River basin, British Columbia, Canada. Hydrological Processes, **26**, 1840–1860, https://doi.org/10.1002/hyp.9283.

Werner, A. T., M. A. Schnorbus, R. R. Shrestha, A. J. Cannon, F. W. Zwiers, G. Dayon, and F. Anslow, 2019: A long-term, temporally consistent, gridded daily meteorological dataset for northwestern North America. Scientific Data, **6**, 180299, https://doi.org/10.1038/sdata.2018.299.