

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2021-423-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-423

Anonymous Referee #2

Referee comment on "Deep learning rainfall-runoff predictions of extreme events" by
Jonathan M. Frame et al., Hydrol. Earth Syst. Sci. Discuss.,
<https://doi.org/10.5194/hess-2021-423-RC2>, 2021

Recently, deep learning has d more accurate rainfall-runoff predictions than a conceptual and physically-based model; this is to be expected as they are trained to be accurate predictively. It is also interesting that these models can generalise to different catchments and are trying to explain some of the mechanisms leading to runoff.

The paper provides a first-hand comparison of how different models (deep learning vs conceptual vs physical model) fare on unseen extreme events. It is clear that with the comparison, the deep learning-based model outperforms on different accuracy metrics.

The paper also argues that deep learning provides the hydrological sciences community's most accurate rainfall-runoff simulations. While this might be true, but certainly requires comparison with many available different models in different parts of the globe. Moreover, this might be only true when we have a large amount of data available. Nevertheless, I would agree that deep learning provides one of the most accurate rainfall-runoff predictions. In the future, there is much potential for a deep learning-based model for rainfall-runoff prediction.

One of the vital points raised in the paper is about the conceptual flux equations and highlights the potential for improvement with a comparison to MC-LSTM.

Another exciting thing is about the FLV (Bottom 30% low flow bias). Analysing why FLV increases(in magnitude) drastically for MC-LSTM could be an interesting direction to explore. Especially for the low probability years. As theoretically, the machine learning model should have seen such low flow data. The author illustrates that any constraint restricts the space of possible functions that the network can emulate. MC-LSTM is developed primarily to model this type of situation where an entity is conserved. Furthermore, unlike other metrics, which did not deteriorate much, we see a drastic drop (increase in magnitude) in FLV. More analysis in this direction would be interesting for the

readers as well.

The paper highlights the potential of deep learning models to predict extreme events, while the hypothesis is that **the data-driven models lose reliability in extreme events more than models based on process-understanding**. The notion of reliability can be somewhat vague and should be clarified. The paper is only focusing on the predictive reliability here.

Overall, the paper would provide the first comparison on predictive accuracy for unseen extreme events for a deep learning model and a valuable contribution to the hydrological community.