Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2021-566-RC1, 2021

Comment on hess-2021-566

Lukas Gudmundsson (Referee)

Referee comment on "Hydrological concept formation inside long short-term memory (LSTM) networks" by Thomas Lees et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-566-RC1, 2021

The paper submitted by Lees et al. aims at advancing the interpretability of neural networks used for rainfall-runoff modelling, focussing in particular on Long Short Term Memory (LSTM) architectures. LSTMs (a special type of neural network) have in recent years been popularized for rainfall-runoff modelling. The resulting models perform very well but lack a stringent physical interpretation. An interesting property of LSTMs is that they contain internal states – similar to internal states (storages) in classical hydrological models. Lees et al use statistical "probes" to link these LSTM states to independent estimates of soil moisture and snow storage. The analysis shows that LSTM states mimic soil moisture and snow storage dynamics although these quantities were not used for model calibration.

This paper is to my knowledge the first systematic attempt to interpret the internals of an LSTM used for rainfall-runoff modelling in a systematic manner, although un-systematic examples have been emerging in the literature. The application of LSTMs for rainfall-runoff modelling is state of the art and the use of "probes" based on linear regression makes intuitively sense. Beside this, I personally value the authors effort to assess the robustness of their results by (i) applying the "probes" to both re-analysis based and remote-sensing based soil moisture estimates and (ii) by randomization based statistical testing.

The objective of the study is clearly stated: I.e., to test if internal states of an LSTM used for rainfall-runoff modelling at many catchments at once can be linked to independent estimates of soil moisture and snow. The paper is well structured to meet this objective and the data and methods are chosen accordingly.

Nonetheless the study leaves some open questions which the authors may want to further explore:

- An obvious limitation of the proposed approach is that it relies on independent estimates of soil-moisture and snow (or other variables). Therefore, it does not allow for a self-contained interpretation of LSTM states.
- It would be interesting to know how much of the variance of the internal state vectors is captured in the resulting soil moisture estimates.
- How many independent signals are present in the 64 states? (e.g. estimated as the number of dominant principal components)

Minor issues:

- Inconsistency: Equation 1-2 use i_t as input. Equations 3-4 use x_t
- The paper somehow requires that the reader is familiar with how LSTMs work. I acknowledge the authors choice not to repeat the LSTM definition, also since it is available in many other publications. Nonetheless, this made it a bit more difficult to fully understand the paper.
- Elastic net: I assume that the description of the elastic net regularisation might be quite cryptic to readers who are not familiar with this tool (I am). Also: Given the large number of samples (# catchments x # time steps) and the relatively low number of predictors I wondered if a linear regression would perform equally well.