

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2  
<https://doi.org/10.5194/hess-2021-566-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Comment on hess-2021-566

Anonymous Referee #2

---

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-RC2>, 2022

---

Lees et al. adapted a novel method used in Natural Language Processing, "probe", to examine the internal function of the Long Short Term Memory (LSTM) model in rainfall-runoff predictions. Their results over 669 catchments in Great Britain show a good correlation between the LSTM internal states with re-analysis and independent soil moisture and snow cover products.

I agree with the authors that this paper could be a stepping stone to a myriad of interesting explorations in the field of hydrology. I also appreciate the authors effort in providing additional analysis in the appendices. However, I have some minor comments about some parts of the manuscript, mostly about the clarity and the tone toward traditional hydrologic models.

- I feel the structure of the Introduction is a bit difficult and redundant for me to follow. I could not get the logical flow here. I found the main objective was stated in both the beginning and the end of the introduction. Why do we need a separate and long paragraph about the interpreting machine learning from other fields? This paragraph disrupts my focus on LSTM interpretability.
- I think the authors don't have to state that LSTM is the best rainfall-runoff model multiple times in the paper (Introduction and Conclusion). While this statement is still debatable, in my opinion, each rainfall-runoff model has its place in the modeling world. LSTM is increasing its popularity because of its robustness, computational efficiency and accuracy. Period. There is no need for bashing one over another.
- Section 2.3 ERA5-Land Data: there is an imbalance between the descriptions of soil moisture and snow depth. I would expect to see more information about snow depth and its accuracy over GB.
- Figure 2: no y label
- Figure 5: no y label
- Line 249: I thought there are only two meteorological drivers (temperature and

precipitation (line 6))?

- Line 268: See the second opinion
- Line 282: See the second opinion
- Line 319: I found a recent paper (Tran et al, Development of a Deep Learning Emulator for a Distributed Groundwater–Surface Water Model: ParFlow-ML. *Water*. 2021) in which the spatial information is included in the LSTM architecture. Do the authors think the probing technique could be use in this architecture? Can the probing technique map between predicted and observed spatially-distributed soil moisture?