

Hydrol. Earth Syst. Sci. Discuss., author comment AC1
<https://doi.org/10.5194/hess-2022-56-AC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on CC1

Marvin Höge et al.

Author comment on "Improving hydrologic models for predictions and process understanding using neural ODEs" by Marvin Höge et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-56-AC1>, 2022

We thank the participant in the community discussion for the contribution and for providing constructive feedback. Please find our replies in italic.

This is an interesting article. There are minor issues that I think would increase the value of it (in no particular order):

- $M(t)$ (the flow between the two buckets in M_0 , which I guess is a snow melting flow) is not formally defined in the model description.

We introduce melting M in l.134. and in l.148, and we reference appendix A1 where M is defined for the original conceptual model in Equation A4. We are open for further specification w.r.t. what the formal definition might be lacking but we are sure that it exists.

- There is no explanation of why different errors were applied to different modules of the hybrid models. There is a comment in the discussions regarding the known failures of models to capture discharge peaks, this performance is strongly linked to the type of error (or the noise distribution in a stochastic modeling context) that was used to train the model. Did the chosen errors for training improve this? why? what were the criteria to choose the different errors?

There was no error function assigned. The models are deterministic and were trained using the available data points. We applied the same procedure as was done in the reference study by Jiang et al., 2020, in order to ensure comparability.

- I could not find a link to the software, and "be made available in the near future" is too ambiguous. The software should be part of the publication work, citing:

*"An article about computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result."
-- Buckheit and Donoho*

We fully agree, it was always our intention to provide the code. For this purpose, we have set up a Github repository at the time of our submission. However, we wanted to have a discussion about the approach at first in a review process, and only want to release the code of a method that is considered interesting by the scientific community. Now, after

receiving encouraging feedback, we will release the code in the coming revision.

- The data-driven relations learned by M50/100 are clearly tuned to the data. Assuming the proposed mechanism is causal and universal, wouldn't it then make more sense to train these modules in the totality of the data, not per catchment? On the one hand, it is well established that non-causal data-driven models can easily outperform causal models (e.g. a causal structure $X \rightarrow Y \rightarrow Z$ with noise in X larger than in Z will cause data-driven models to choose Z to be the best predictor of Y , alas non-causal). On the other hand, it is unlikely that NN models will use relations that go beyond the scope of the data, hence the optimal relations found per catchment might be reflecting circumstantial relations, but the mechanisms proposed are supposed to be principled mechanisms, not circumstantial.

We agree. This is where our current research is headed. We work on training the models on multiple basins. One of the points we want to investigate is exactly whether the learned relations go beyond the scope of the data. The physical structure of the model enforces a sort of regularization that we think might help in this respect.