CC1

I read this paper with interest, as I agree that such metamorphic tests on ML hydrologic models are needed to assess their appropriateness for certain hydrologic modeling applications like projections under climate change.

My main comment is that I think the authors could contextualize their study with past work that has conducted a similar exploration. The first paper that I am aware of which attempted a metamorphic test on an LSTM was Razavi (2021) (see their Figure 11). They only considered an LSTM fit to one site, and so there are limitations to that work, but I think it is important to recognize it. Afterwards, Wi and Steinschneider (2022) conducted a similar metamorphic test as conducted in the present study, using both 1) an LSTM and physics-informed LSTMs fit to 15 sites across California, as well as an LSTM fit across the entire CAMELS dataset. They found related challenges with LSTM projections under warming as found in this work.

Therefore, I recommend that the authors adjust their Introduction to recognize these past studies, and then to articulate how their work provides a contribution over these past studies. I believe this is very straightforward, as the present study 1) considers changes in precipitation as well; 2) explores responses separately by basin elevation and temperature; and 3) explore sensitivity to calibration choices (this later one was particularly helpful to see). In addition, I might adjust the Summary and Conclusion to discuss the results of the present study in comparison to the metamorphic results seen in Wi and Steinschneider (2022), in order to help synthesize related results in the literature.

References

Razavi, S. (2021). Deep learning, explained: Fundamentals, explainability, and bridgeability to process-based modelling, Environmental Modelling and Software, 105159, https://doi.org/10.1016/j.envsoft.2021.105159.

Wi, S., & Steinschneider, S. (2022). Assessing the physical realism of deep learning hydrologic model projections under climate change. Water Resources Research, 58, e2022WR032123. https://doi.org/10.1029/2022WR032123

We thank the reviewer for pointing our attention to these papers. We fully agree with adapting the introduction and conclusions in the sense suggested by the reviewer.

These papers were also useful for identifying newer papers that cite these papers and are useful for additional references regarding hybrid models in the discussion:

Ng K.W., Huang Y.F., Koo C.H., Chong K.L., El-Shafie A., Ahmed, A.N. A review of hybrid deep learning applications for streamflow forecasting. Journal of Hydrology 625 (2023) 130141.

Razavi, S., Hannah, D.M., Elshorbagy, A., Kumar, S. Marshall, L., Solomatine, D.P., Dezfuli, A., Sadegh, M. and Famiglietti, J. (2022) Coevolution of machine learning and process-based modelling to revolutionize Earth and environmental sciences: A perspective. Hydrological Processes. 2022;36:e14596.

Zhong, L., Lei, H., & Gao, B. (2023). Developing a physics-informed deep learning model to simulate runoff response to climate change in Alpine catchments. Water Resources Research, 59, e2022WR034118.