An argument for parsimony in differentiable hydrologic models

Sandeep Poudel and Scott Steinschneider

Corresponding author: Sandeep Poudel, sp2596@cornell.edu

Key

Black font: Reviewer comments

Bold blue font: Author responses

Acronyms: OOS-time = out-of-sample in time; OOS-space = out-of-sample in space

---

This article is within the scope of the journal. The manuscript is very well organized and easy to follow. The experiments are of general interest to the community that is developing differentiable models, including myself. The novelty is fairly low, and the main motivation of static parameters from MLP demonstrates a lack of sufficient literature review. The results are all mostly expected based on my familiarity with the literature. This suggests to me that the experiments were all done correctly. I was able to access all the open-source code, and have reviewed it, and it is very nicely presented. My main concern is with the novelty of result 1 and the interpretation of result 2. Result 3 seems to be the strongest contribution of this work.

We thank the reviewer for the careful review of both the manuscript and the code repo, and for the constructive feedback. Below we provide a point-by-point response to each comment.

Line 80: There have been differentiable model studies in the past that use MLP for static parameters, See:

- Bindas, T., Tsai, W.-P., Liu, J., Rahmani, F., Feng, D., Bian, Y., et al. (2024). Improving river routing using a differentiable Muskingum-Cunge model and physics-informed machine learning. Water Resources Research, 60, e2023WR035337. https://doi.org/10.1029/2023WR035337.
- Song, Y., Bindas, T., Shen, C., Ji, H., Knoben, W. J. M., Lonzarich, L., et al. (2025). High-resolution national-scale water modeling is enhanced by multiscale differentiable

physics-informed machine learning. Water Resources Research, 61, e2024WR038928. https://doi.org/10.1029/2024WR038928

The reviewer is correct that MLP-based parameterization has appeared in prior differentiable modeling work. We do not claim novelty in its use for parameter learning. Rather, our novelty lies in providing a systematic comparison of a truly static parameterization approach via MLP against a dynamic LSTM-based approach for catchment-scale rainfall-runoff modeling on a large-scale benchmark dataset, and in demonstrating that a single static parameter set obtained with an MLP-based differentiable model matches the performance of its dynamic LSTM-based counterpart. To our knowledge, we are the first to make this claim about the parsimony of differentiable rainfall-runoff models in this context.

Regarding the two specific references pointed out by reviewer: Bindas et al. (2024) use MLP-based parameterization for river routing, not for the rainfall-runoff process, which is the focus of our work. Song et al. (2025) use both MLP and LSTM in a national-scale high-resolution modeling framework, but their motivation for adopting MLP is primarily computational. At high-resolution national scale, LSTM introduces a large computational burden and using MLP in combination with LSTM to handle static and dynamic parameterization separately substantially reduces this cost. Our motivation is fundamentally different. We propose MLP because it produces truly static parameters that are directly analogous to those of the process-based model, are more parsimonious and interpretable, and most importantly our results show they match the performance (both OOS-Time and OOS-Space) of dynamic LSTM-based parameterization. We will revise the manuscript to more clearly distinguish our contribution from these prior studies.

Line 270: it seems you are describing your attribution methodology in the main text, but then you don’t present the results until Appendix D. This is only an editorial/stylistic comment, but perhaps it would be better to save the methodology of the appendix experiments/results for those appendix sections?

Thank you for this suggestion, in our revised manuscript we will move attribution methodology to the Appendix D.

Result 1 and description on Line 330: this interpretation of the results seems to be a little conflicting. On the one hand you claim the attributes aren't used for learning physical relationships because you can get the same results with only lat/lon (spatial interpolation), on the other hand, you claim lat/lon degrades the temporal time split because the network can't memorize. What about the interpretation that static attributes aren't used to "memorize", but rather to inform a learnt physically meaningful relationship? Would this not also be supported by the nonsensical feature sets failing to spatially generalize?

We note that Reviewer #1 had a similar comment, and we agree that the manuscript did not discuss this result with sufficient clarity. However, our results taken together do not support the interpretation that static attributes inform physically meaningful relationships. We elaborate below, and note that our response below is the same as the one provided to Reviewer #1.

We agree that the true static attributes contain meaningful information the network can use. However, we do not think the current results clearly show that the network is learning *physically* meaningful process relationships in an interpretable sense (e.g., that high clay content leads to faster runoff response). Rather, our results suggest that the primary role of the static attributes is to help the network infer each basin's geographic context, which in turn supports OOS-space generalization. The strongest evidence for this is that lat/lon alone yields OOS-space performance comparable to the full set of 30 physical attributes, whereas permuted attributes substantially reduce OOS-space performance (0.42 vs. 0.62). This implies that the network is mainly using the static attributes to determine where a basin sits geographically, where many attributes are similar across relatively homogeneous regions, rather than to learn physically meaningful hydrologic process relationships. A secondary role of the static attributes is to provide a high-dimensional input space that supports memorization of basin-specific behavior, as shown by the fact that OOS-time performance is retained even when the 30 static attributes are nonsensical.

There are therefore two possible interpretations for how a neural network learns from static attributes. Under Interpretation A (Physical relationship learning), the network learns meaningful physical mappings between static basin attributes and hydrological processes. Under this interpretation, a network using 30 physical attributes should outperform a network using 2 geographic coordinates in OOS-space generalization. Under Interpretation B (Geographic coordinate learning), the network uses static attributes mainly to infer a basins geographic context, learning that certain combinations of catchment properties are associated with particular regions and their characteristic rainfall–runoff behavior. Under this interpretation, latitude and longitude should perform comparably to the 30 physical attributes for OOS-space generalization.

Our results are more consistent with Interpretation B, as summarized in the table below:

Static attributes used as NN input	**OOS Time generalization**	**OOS Space generalization**
30 physical attributes (no lat and lon)	**Yes, because the high-dimensional input supports basin memorization**	**Yes, because the attributes are used to learn geographic positioning and the model can locate unseen basins geographically**
Lat and lon only	**No, because only 2 inputs are insufficient for basin memorization**	**Yes, because geographic positioning is directly encoded with lat and lon and spatial generalization is retained**
30 nonsensical physical attributes	**Yes, because the high-dimensional input still supports memorization despite carrying no physical meaning**	**No, because nonsensical inputs prevent geographic learning and spatial generalization is lost**

We acknowledge that the original manuscript did not explain these points with sufficient clarity, and we will carefully revise the relevant sections to reflect this discussion.

Result 2: "Ensemble models outperform single-parameter configurations", this is a well-known phenomena, and the reason why most of the impactful results from DL and differentiable modeling papers use ensembles.

We agree with the reviewer. And our full sentence for this result reads as follows: 'Ensemble models outperform single-parameter configurations, and a simpler MLP-based static parameterization matches the performance of a dynamic LSTM-based parametrization.'

The finding that ensemble models outperform single-parameter configurations is well established in the literature, and we do not claim novelty there. Our novelty lies specifically in the latter part of the sentence: once we move to ensemble-based models, MLP-based single static parameter matches the performance of a dynamic LSTM-based parameterization. We will revise the manuscript throughout to ensure this is the focal claim.

Result 3: this is an interesting result.

We are glad the reviewer finds this result interesting.