Response to Reviewer #1's comments

[0] Greetings! This study presents PLSTM-Reg v1.0, a regional physics-encoded LSTM framework that effectively simulates reservoir operations by integrating mass-balance constraints with deep learning. The methodology demonstrates strong spatial and temporal generalizability, particularly through its synergy with remote sensing data. Overall, I think this is a very high-quality, well-organized study that addresses a critical challenge in hydrological modeling, which fully meets the requirements for publication in *Geoscientific Model Development* (GMD).

Response: Thank you for your positive and encouraging feedback. We have carefully addressed your comments and suggestions to further improve the clarity of the manuscript, as detailed below.

[1] I have two minor suggestions. First, I recommend that the authors move the descriptions of static attributes to the main text, as they are essential parts of the model structure.

Response: Thank you for your suggestion. We have revised the description of the 21 static attributes (previously Table S1 in the Supplement) and moved it to **Table A1 in Appendix A** of the manuscript. This improves their visibility and accessibility while preserving the flow of the Data and Methods section.

[2] Second, I would like to see a little discussion on which of these attributes have the most important impacts on the model results and which are less so.

Response: Following your suggestion, we conducted a feature attribution analysis using the Integrated Gradients (IG) method. Details of the IG methodology and the feature-importance results are provided in Supplementary **Text S3** and **Fig. S6**, respectively. In the main text (**lines 335-340**, Section 3.1), we now state:

"*Furthermore, we conducted a feature attribution analysis using the Integrated Gradients (Sundararajan et al., 2017) method (see Supplementary Text S3). The analysis shows that the model relies primarily on operational parameters (e.g., fill fraction, reservoir functions, and storage size ratio) together with regional topographic and hydro-climatic factors (e.g., elevation, precipitation, and temperature), whereas geometric variables (e.g., dam length, depth, and surface area)* *exert a comparatively minor influence (see Fig. S6).*"

The new Fig. S6 is also provided below:

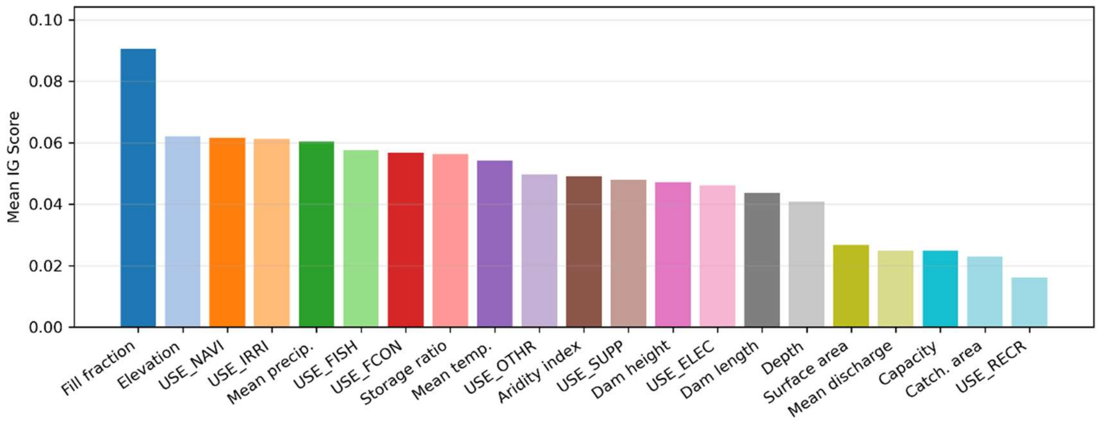

*Figure S6. Mean Integrated Gradients (IG) importance scores of the 21 static attributes (see **Table A1** in Appendix A) for reservoir release predictions. IG scores were normalized within each reservoir-day and then averaged across reservoirs, dates, and ensemble repeats. Higher values indicate greater relative contribution to model predictions.*

[3] I also have three technical questions that I hope the authors could clarify to improve reproducibility. 1) What is the specific sequence length used for the training and testing of the sequence-to-sequence regional model?

Response: Thank you for this question. Different sequence lengths were used for short-term prediction and long-term simulation experiments. To clarify this and ensure reproducibility, we have added corresponding details in Section 2.3.

In **lines 219-221**, the original sentence was revised as: "*In short-term prediction (Experiment I), PLSTM models were trained in a sequence-to-one manner* <u>*with a sequence length of 180 days*</u> *to predict next-day release and storage.*"

In **lines 226-229**, we added: "*The sequence length for both model training and testing was set to 1460 days (4 years). A sensitivity analysis using 2-, 4-, and 6-year sequence lengths yielded similar performance across all tested settings (Supplementary Fig. S1), indicating that the results are not strongly sensitive to this choice.*"

[4] 2) Why did the authors choose a recursive sequence-to-one model instead of a direct sequence-to-sequence approach (e.g., outputting a 7-day vector) for the 7-day

lead time forecast?

Response: Thank you for this insightful question. We adopted the recursive sequence-to-one architecture for two reasons. First, it is the standard approach widely used in deep learning hydrology (e.g., Feng et al., 2020; Jiang et al., 2022; Kratzert et al., 2018). Second, and more importantly, it is required by the model structure. As shown in Table 1, reservoir storage is used as a dynamic input feature in Experiments I and III. Therefore, predicting the state at day *t*+1 requires the storage at day *t*, making autoregressive recursive prediction necessary for multi-day prediction.

To clarify this design choice, we added the following text in **lines 222-224**: "*This structure follows standard hydrological prediction practice (Dong et al., 2023a; Kratzert et al., 2019) and is required because the model uses storage at day t to predict release and storage at day t+1.*"

[5] 3) Did the authors encounter gradient vanishing or truncation issues, given that the hard physical constraints in the PLSTM cell might truncate the gradient flow during backpropagation?

Response: Thank you for this insightful question. Preliminary experiments show that PyTorch's native clamp() function remained stable during backpropagation and did not hinder model optimization, despite the possibility of gradient truncation when constraints are exceeded. Similar implementations are also adopted in recent differentiable hydrology frameworks such as hydrodl2 (mhpi, 2026). We therefore used clamp() for its simplicity and exact enforcement of physical bounds.

To clarify this point, we added the following text in **lines 177-181** of Section 2.2: "*It is worth noting that although hard physical bounds implemented via the torch.clamp function may truncate gradients when constraints are exceeded, preliminary experiments showed stable optimization without requiring smoothing techniques (e.g., tanh smoothing). Similar implementations are also adopted in recent differentiable hydrology frameworks such as hydrodl2 (mhpi, 2026)*"

References

Dong, N., Guan, W., Cao, J., Zou, Y., Yang, M., Wei, J., Chen, L., and Wang, H.: A hybrid hydrologic modelling framework with data-driven and conceptual reservoir operation schemes for reservoir impact assessment and predictions, Journal of Hydrology, 619, 129246, https://doi.org/10.1016/j.jhydrol.2023.129246, 2023a.

Feng, D., Fang, K., and Shen, C.: Enhancing Streamflow Forecast and Extracting Insights Using Long-Short Term Memory Networks With Data Integration at Continental Scales, Water Resources Research, 56, e2019WR026793, https://doi.org/10.1029/2019WR026793, 2020.

Jiang, S., Zheng, Y., Wang, C., and Babovic, V.: Uncovering Flooding Mechanisms Across the Contiguous United States Through Interpretive Deep Learning on Representative Catchments, Water Resources Research, 58, https://doi.org/10.1029/2021WR030185, 2022.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning, Water Resour. Res., 55, 11344–11354, https://doi.org/10.1029/2019WR026065, 2019.

mhpi: mhpi/hydrodl2, 2026.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic Attribution for Deep Networks, arXiv:1703.01365 [cs], 2017.