Response to Reviewer #3's comments

[0] This manuscript develops PLSTM-Reg, a regional LSTM model trained and evaluated using long-term daily records from 256 reservoirs across the CONUS to simulate reservoir release and storage under temporal and spatial generalization settings. The study is valuable because it uses a high-quality reservoir dataset, considers both short-term and long-term simulation performances, designs multiple experiments to test transferability, and tests the added value of remote-sensing surface area. The manuscript is generally well organized. However, the novelty relative to recent pooled or transferable reservoir-operation models is fully stated, and several aspects of the validation require clarification, including possible information leakage from static attributes, and the fairness of local-model comparisons. Overall, I find the study promising and potentially useful for large-scale hydrologic modeling, but I recommend major revision before publication.

Response: Thank you for the positive assessment and constructive comments. We have carefully revised the manuscript, as detailed below.

Major Comments

[1] Table S1 lists "Average discharge" as an input. This raises two concerns. First, it may partly explain the improvement of the regional model over local models, since local models do not use this static input. Providing an ablation excluding average discharge can be more convincing. Second, in the spatial-transfer experiment, average discharge for held-out reservoirs may create leakage. This would weaken the claim that the model is tested on truly "unseen" or data-scarce reservoirs.

Response:

Thank you for raising this important concern. The "average discharge" attribute is not derived from the reservoir operation records. It is the long-term (1971–2000) average discharge at the dam location, derived from the HydroSHEDS flow-routing scheme combined with downscaled WaterGAP runoff estimates at 15 arc-second resolution (Lehner et al., 2024). Therefore, this variable does not use historical release or storage records from the held-out reservoirs and does not introduce leakage from the target operation data. Moreover, long-term discharge descriptors are commonly used in reservoir operation modeling and parameterization. For example, the storage size ratio used by rule-based benchmark models is calculated from storage capacity and long-term inflow or discharge (Hanasaki et al., 2008a, b; Yassin et al., 2019). Thus, using a

globally available long-term discharge descriptor is consistent with the information typically available for data-scarce reservoir modeling.

To clarify this point, we revised section 2.1 in **lines 133-135** as follows: "*For each reservoir, we compiled static attributes from the GDW database, including physical characteristics (e.g., dam height, storage capacity), primary management purpose, and hydrological context (i.e., long-term mean discharge at the dam location during 1971–2000).*"

We also performed an ablation analysis excluding the GDW mean discharge attribute from the static inputs. As shown in the figure included below, removing this attribute does not substantially change the overall performance of PLSTM-Reg, although a slight performance decrease is observed for some lower-performing reservoirs. These results indicate that the performance advantage of PLSTM-Reg over local models cannot be attributed solely to the inclusion of the mean discharge attribute. Instead, the attribute provides useful hydrological context in some challenging cases while having a limited influence on overall model skill. Because this analysis was conducted specifically in response to the reviewer's concern and does not materially affect the study conclusions, we present the results here for completeness rather than incorporating them into the revised manuscript.

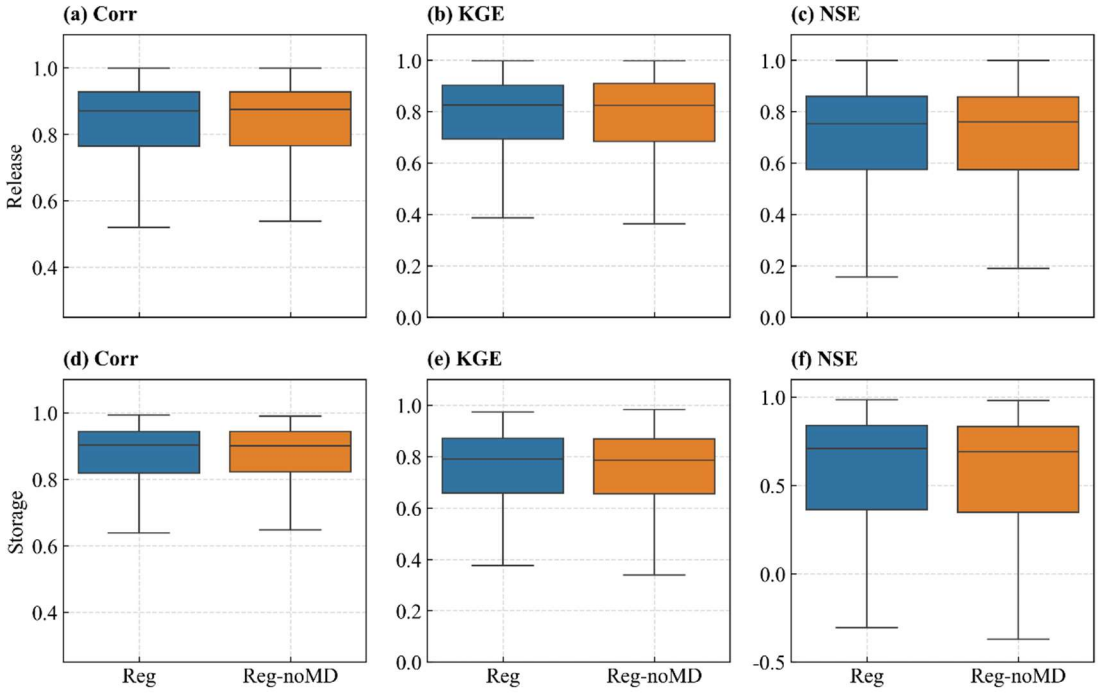

Figure R1. Ablation analysis of the GDW mean discharge attribute in long-term simulation. Boxplots compare release (a–c) and storage (d–f) simulation performance between PLSTM-Reg and PLSTM-Reg-noMD under Experiment II. PLSTM-Reg-noMD denotes the ablated regional model excluding the GDW mean discharge

attribute. Removing this attribute causes no substantial change in overall performance, although some lower-performing reservoirs show slight degradation. This indicates that the GDW mean discharge attribute provides useful hydrological context for difficult cases.

[2] The model seems to assume that reservoir operation depends only on local basin forcings, local inflow/storage, and reservoir attributes. However, many real reservoirs are operated as part of cascade reservoir systems. The authors should clarify whether such cases are included and whether PLSTM-Reg can handle coupled upstream–downstream operation (such as the flexibility to add upstream reservoir information as static input in current framework). If not, the application boundary should be stated more clearly.

Response: Thank you for this insightful comment. The 256-reservoir dataset includes reservoirs located within cascading systems. In the current PLSTM-Reg framework, upstream operational effects are implicitly reflected in the inflow received by downstream reservoirs. Our study focuses on simulating the operation of individual reservoirs rather than explicitly modeling coordinated cascade operations. For applications involving multi-reservoir systems, PLSTM-Reg can be coupled with a river routing module, whereby releases from upstream reservoirs are routed and used as inflows to downstream reservoirs. In this way, the framework can be readily extended to simulate cascade reservoir systems while retaining the same reservoir operation module for each reservoir.

To clarify this, we added the follow discussion in **lines 490-494**: "*Alongside geographical expansion, future development may also extend the framework to cascading reservoir systems. The current PLSTM-Reg framework is designed for single-reservoir simulation, with upstream operational effects represented implicitly through downstream inflows. Extension to multi-reservoir systems could be achieved by coupling PLSTM-Reg with a river routing module that routes upstream releases to downstream reservoirs.*"

[3] The comparison with recent advances is incomplete. Comparing mainly with older rule-based models weakens the significance of the claimed innovation, especially because the key claim is multi-reservoir pooling/regional learning. The following studies should be compared with, or at least discussed:

Tran et al. (2025), "Improving the prediction of daily reservoir releases over the

CONUS using conditioned LSTM." This is a closed predecessor: a pooled/conditioned LSTM across nearly 200 CONUS reservoirs using static reservoir attributes.

The following three are cited in this work, but only as background. They should be discussed more directly against PLSTM-Reg as baselines:

- Ford and Sankarasubramanian (2023), "Generalizing reservoir operations using a piecewise classification and regression approach."
- Turner et al. (2021), "Water storage and release policies for all large reservoirs of conterminous United States."
- Chen et al. (2022), "Developing a generic data-driven reservoir operation model."

Response: Thank you for this suggestion. We revised the manuscript to compare PLSTM-Reg more directly with recent regional or transferable reservoir-operation models, including Tran et al. (2025), Ford & Sankarasubramanian (2023), Chen et al. (2022), and Turner et al. (2021).

We agree that Tran et al. (2025) is closely related to the present study. However, the two studies address fundamentally different objectives. Tran et al. (2025) focuses on next-day release prediction and relies on observed storage inputs without enforcing mass balance, whereas PLSTM-Reg is designed for both short-term prediction and long-term release–storage simulation through an explicit storage state and mass-balance constraint. Therefore, Tran et al. (2025) is discussed in the Introduction but is not used as a direct benchmark for the long-term simulation experiments.

Ford and Sankarasubramanian (2023) was not included as a direct benchmark in Experiment IV because its PLRT formulation requires reservoir-specific statistical information for storage and release standardization, which in turn requires historical observations from the target reservoir. This requirement is inconsistent with the held-out, data-scarce spatial-transfer setting considered in Experiment IV.

Similarly, Chen et al. (2022) was not included as a direct benchmark because its module extraction and application-condition classification rely on historical inflow, storage, and release records from the target reservoir. As such, it is not directly applicable to the Experiment IV setting, where no operational records are assumed to be available for the target reservoir.

Following the reviewer's suggestion, we added Turner et al. (2021) as an explicit benchmark in both temporal and spatial long-term simulation experiments. For

Experiment II, ISTARF was calibrated using the training period of each reservoir and evaluated on the chronological test period; this comparison is now included in **Supplementary Fig. S4** and mentioned in the main text. For Experiment IV, ISTARF was implemented in a spatial-transfer setting: parameters were fitted only using reservoirs in the training folds, and donor selection for each held-out reservoir was restricted to the training pool to avoid information leakage. The implementation details for both ISTARF settings are provided in **Supplementary Text S2**, and **Fig. 5** has been updated to include ISTARF.

To better contextualize our model and differentiate our approach from existing studies, we substantially revised the Introduction. The revised text (**lines 47-95**) now reads:

"*Rule-based reservoir operation models (Hanasaki et al., 2008a, b; Wisser et al., 2010; Zajac et al., 2017) are widely used in large-scale hydrological studies because of their interpretability and modest data requirements. However, their fixed recommended parameters and simplified operating rules limit their ability to represent site-specific and nonlinear real-world operating behavior (Steyaert and Condon, 2024; Yang et al., 2021). Some progress has been made in recent studies. Yassin et al. (2019) proposed an improved target storage-and-release parameterization that better represents reservoir storage zones and seasonal release behavior, but it requires in-situ storage and release series (or distributions) to determine model parameters. Turner et al. (2021) developed a transferable rule-based framework that learns operating polices from data-rich reservoirs and extrapolate them to neighboring data-scarce reservoirs with similar purposes. However, directly transferring the same policy parameters from source reservoirs to target reservoirs may impose source-specific operating behavior and introduce significant bias.*

Recently, machine learning has been introduced to improve the parameterization and transferability of rule-based reservoir operation schemes. Some studies have used machine learning to estimate parameters for existing operation schemes, such as reservoir-operation parameters inferred from reservoir attributes using Random Forest models (Steyaert et al., 2025) and flood storage capacities estimated using XGBoost to define storage-zone parameters in a target storage scheme (Shen et al., 2025). Researchers have also explored to use machine learning to develop hybrid models for improved reservoir representation. Chen et al. (2022) developed a data-driven framework that extracts interpretable operation modules and associated application conditions from historical operation records. Building on this modular reservoir modeling concept, Li and Villarini (2026) further investigated its generalization using Random Forest models to transfer parameters associated with

typical operation modules, such as constant, linear, and piecewise-linear release relationships, across reservoirs. Although these approaches improve parameter estimation and transferability while retaining interpretability, they generally represent reservoir operations through simplified modules or predefined rules. Moreover, because these modules or parameters are typically identified from individual reservoirs before being transferred, these approaches cannot fully leverage cross-site information during model training, which has been shown to improve the generalization of regional deep learning models (Fang et al., 2022; Kratzert et al., 2024). Consequently, their ability to learn complex operating behavior across diverse reservoirs may remain constrained.

Another line of research uses machine learning directly to reservoir operation modeling in an end-to-end way. Site-specific models based on decision trees (Dong et al., 2023; Yang et al., 2016, 2021), recurrent neural networks (Cheng et al., 2024; Yang et al., 2019), LSTM variants (Longyang and Zeng, 2023; Zhang et al., 2018), and physics-encoded deep learning (Yu et al., 2025; Zheng et al., 2022) have shown strong ability to learn release behavior directly from historical records without prescribing explicit operating rules. However, these models require local operational data for training and therefore have limited applicability to data-scarce reservoirs. To address this limitation, recent studies have begun to explore regional end-to-end models that learn release behavior across multiple reservoirs. Ford and Sankarasubramanian (2023) proposed a regionalized piecewise linear regression tree framework for reservoir release prediction, but its reliance on reservoir-specific release and storage statistics for normalization limits direct transfer to unseen reservoirs without historical operation records. Tran et al. (2025) developed a conditioned LSTM for daily reservoir release prediction across CONUS reservoirs using static reservoir attributes, demonstrating the potential of pooled deep learning for reservoir operations. However, the model focuses on next-day release prediction, relies on observed storage as an input, and does not explicitly enforce mass balance. In contrast, long-term reservoir simulation requires the model to internally evolve storage states while maintaining mass balance, which remains largely unexplored."

In section 2.3, we have added (**lines 233-236**): "*In Experiment II, we also compared PLSTM-Reg with the calibrated ISTARF model (Turner et al., 2021), using parameters fitted from the training period of each reservoir and evaluated on the chronological test period. The implementation details are provided in Supplementary Text S2.*"

In section 3.1, we have added (**lines 303-305**): "*Both PLSTM-Reg and PLSTM-Loc outperform the calibrated ISTARF benchmark, which achieves median KGE values of*

0.64 for release and 0.47 for storage (Supplementary Fig. S4)."

In Supplementary Text S2, we have added: "*The fourth benchmark is the ISTARF model developed by (Turner et al., 2021), which represents reservoir storage and release policies using a 19-parameter formulation. ISTARF employs harmonic regression to define the upper and lower bounds of a seasonally varying Normal Operating Range (NOR) as percentages of storage capacity. During normal operations within this range, target release is computed as a fractional deviation from long-term mean annual inflow and adjusted dynamically according to the current storage state and inflow. When storage falls below or exceeds the NOR, releases are constrained by predefined minimum or maximum bounds, respectively.*

We implemented ISTARF in two ways, corresponding to the temporal and spatial long-term simulation experiments. In Experiment II, ISTARF was calibrated separately for each reservoir using the chronological training period and evaluated on the corresponding test period. This calibrated configuration represents the performance of ISTARF when historical operation records are available for the target reservoir during model fitting. In Experiment IV, ISTARF was implemented in a data-scarce spatial-transfer setting. For each fold, ISTARF parameters were fitted using only reservoirs in the training folds. Donor selection for each held-out reservoir was then restricted to the training pool, following the donor-selection hierarchy based on hydrologic-region similarity, operational-purpose alignment, and geographic proximity. The transferred parameters were used to simulate daily release and storage for the held-out reservoirs."

The revised **Fig. 5** and **Fig. S4** are provided below:

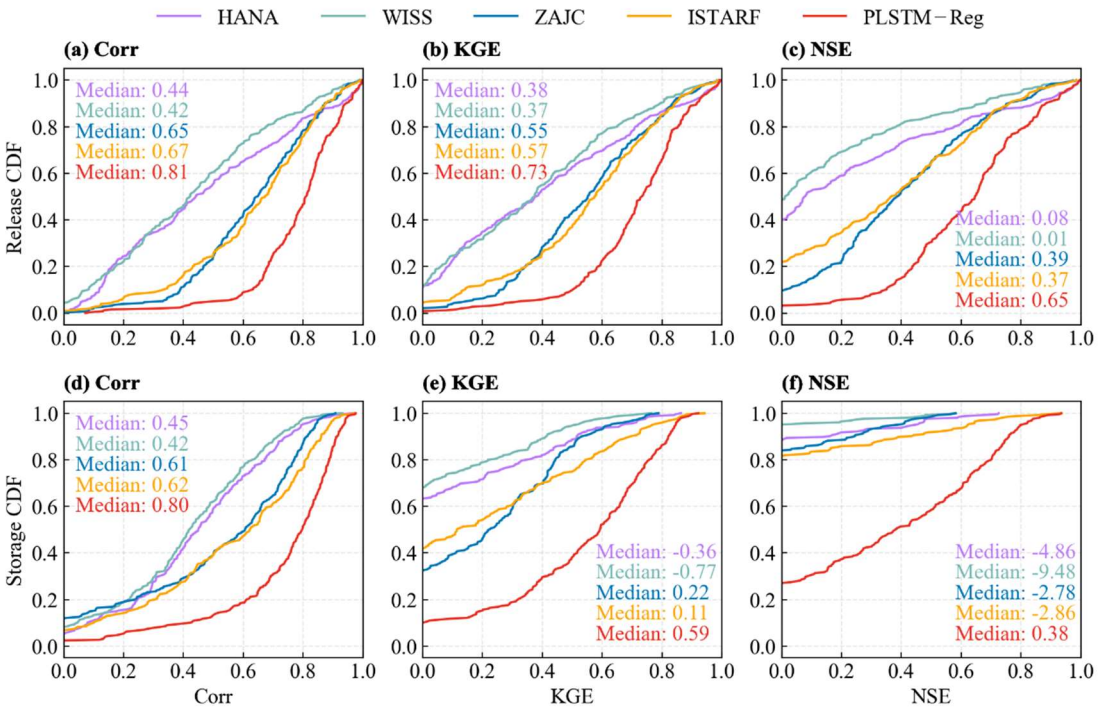

Figure 5. Comparison of release and storage long-term simulation performance (Experiment IV) between the regional model and four benchmark models. CDFs of three metrics (Corr, KGE, and NSE) are displayed for 256 reservoirs over the full evaluation period. The upper row (a–c) illustrates release performance, and the lower row (d–f) shows storage performance. Median metric values for each model are annotated within the respective panels.

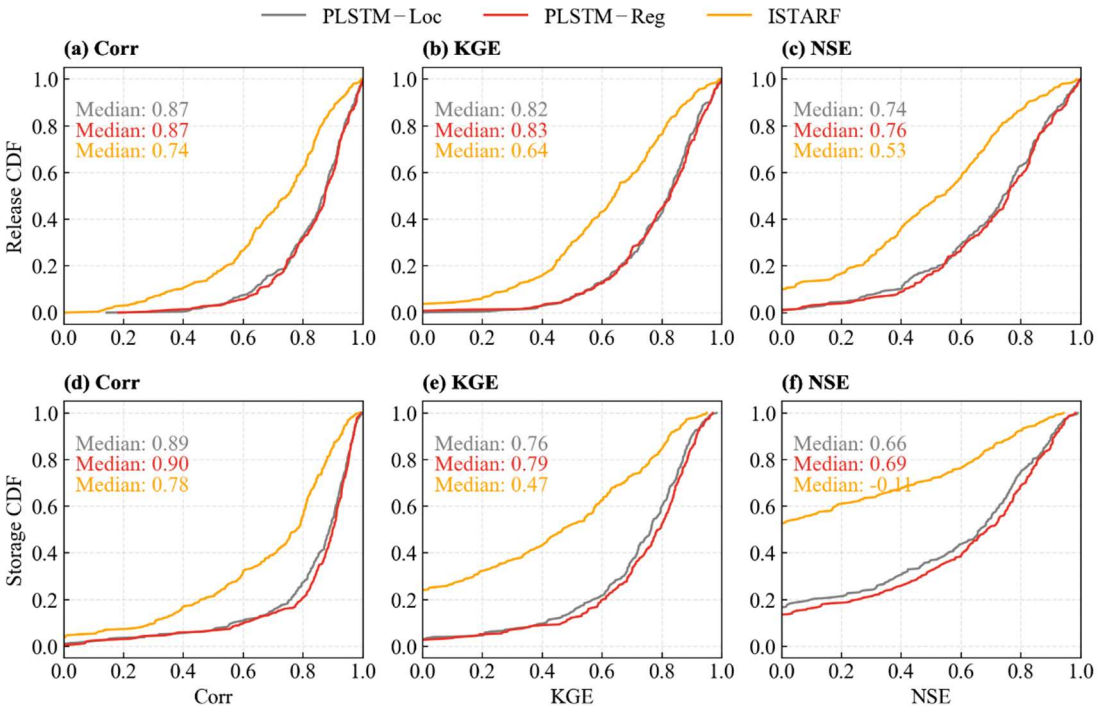

Figure S4. Comparison of release and storage long-term simulation performance

under Experiment II across all reservoirs during the test period. CDFs display release simulation performance in the upper row (a–c) and storage simulation performance in the lower row (d–f). Median metric scores are annotated within the respective panels.

[4] The source of improvement should be interpreted more carefully. The gains over rule-based methods may come from both regional pooling and the flexible neural-network architecture. Simply attributing the improvement to the regional setting is not fully fair. The authors should distinguish the effects of regional training, static attributes, nonlinear LSTM architecture, and physical constraints.

Response: Thank you for pointing this out. We agree that the improvement over the rule-based benchmarks should not be attributed solely to regional training. The contributions of individual components are discussed separately in the manuscript: Section 3.1 evaluates regional training, Section 4.1 evaluates static attributes, and the benefits of the physics-encoded PLSTM architecture have been demonstrated in our previous study (Yu et al., 2025). To clarify this, we revised Section 3.2 and the Conclusions, as follow:

In Section 3.2, we have revised **lines 353-358** to: "*These results highlight the advantages of the proposed framework. The outperformance over generic and uncalibrated rule-based schemes reflects the combined effects of the nonlinear PLSTM architecture, regional data pooling, static reservoir attributes, and embedded physical constraints. Together, these components enable more realistic predictions in data-scarce settings where simplified empirical rules often fail.*"

In the Conclusions section, we have revised **lines 516-520** to: "*This result indicates that the combination of a physics-encoded neural network architecture, regional training, and static reservoir attributes can capture shared behavioral patterns more effectively than simplified empirical rules, providing a more reliable basis for simulating reservoirs with limited or no operational records.*"

Minor Comments

Line 14. "Representative reservoirs" needs clearer definition in the data section to avoid selection bias. Is data length the only selection criteria?

Response: Thank you for pointing this out. We clarified that data length was not the only selection criterion. In Chen et al. (2025), data length and missing-data rate were

the main filtering criteria. The term "representative" refers to the broad geographic, hydroclimatic, and operational diversity of the resulting 256-reservoir dataset, as described in Section 2.1.

In Section 3.2, we have revised **lines 117-119** to: "*The reservoirs were selected based on record availability and data quality, requiring at least 25 years of continuous records spanning 1990–2021 (end dates ranged from 31 December 2014 to 30 April 2021) and <10% missing data.*"

Line 50. Previous models require water-demand data for parameterization, while this model does not use demand data but still performs well. The authors should explain why demand information may be implicitly captured, for example through storage, release history, seasonality, reservoir purpose, and static attributes.

Response: Thank you for this suggestion. We added the following explanation in Section 2.2 (**lines 161-167**): "*Although explicit water-demand data are not used, demand-related operational patterns may be implicitly represented through the model inputs. Seasonal indicators, hydrometeorological forcings, and static attributes such as reservoir purpose provide information related to the timing and magnitude of water demands. Furthermore, historical storage and release records reflect the cumulative effects of reservoir operation decisions driven by water demands and management objectives, thereby providing indirect information on demand-related operating behaviors.*"

Line 124-126. It would be helpful to know the missing ratios of the remote sensing data.

Response: Thank you for this suggestion. We have added the missing ratios of the raw monthly GRSAD surface-area data in Section 2.1 (**lines 150-156**): "*Prior to imputation, the raw monthly GRSAD time series across the evaluated reservoirs had an average missing-data ratio of 3.7%, with reservoir-level values ranging from 0% to 6.6%. For missing values within the GRSAD record, data were imputed according to gap length: gaps of six months or shorter were linearly interpolated, whereas longer gaps were filled with the long-term monthly mean surface area. For periods extending beyond the GRSAD coverage (post-2018), the same climatological monthly mean was used to maintain dataset consistency and avoid cross-product biases.*"

Line 165. The authors use the term “forecasting.” However, if future forcings/inflows are assumed to be known when running the model recursively, this is closer to conditional simulation or hindcast prediction than true operational forecasting. Please clarify.

Response: Thank you for pointing this out. We agree that the setup is not an operational forecast because future meteorological forcings and inflows are taken from observed records. We have updated the manuscript to replace all instances of “short-term forecasting” with “short-term prediction” to avoid ambiguity.

Line 176. Long-term simulation capability depends on the training window length. Longer sequence windows may help the model learn to avoid error accumulation. This does not require a new experiment, but it is worth mentioning in the discussion.

Response: Thank you for this suggestion. We agree that sequence length may affect long-term free-running simulations because storage errors can accumulate over time. We have clarified this in Section 2.3. We have also added a sensitivity analysis using 2-, 4-, and 6-year training sequences. The results (Supplementary Fig. S1) show similar performance across the tested sequence lengths, indicating that the long-term simulation results are not strongly sensitive to this hyperparameter within the tested range.

We revised Section 2.3 (**lines 226–229**) as follows: “*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.*”

Line 197. The authors should clarify how the monthly surface-area time series is used at daily steps. If the current month’s value is used for all days in that month, it may include future information for earlier days.

Response: Thank you for this question. Monthly surface-area data were only used in Experiment V, which is designed for historical reconstruction rather than short-term prediction. Monthly values were assigned to the midpoint of each month and linearly interpolated to daily time steps. The interpolated series may incorporate information from later dates within the same month; however, this is consistent with the retrospective reconstruction objective, as all surface-area observations used in Experiment V are historical data.

We have clarified this point in Section 2.3 (**lines 253–257**): "*To align with the model's daily temporal resolution, monthly surface area values were assigned to the midpoint of each month and linearly interpolated to daily time steps. The interpolated daily series may use within-month information from both earlier and later dates, which is appropriate for the reconstruction setting because the complete historical surface-area record is available during reconstruction.*"

Lines 227–229. It is interesting that release performance differs clearly between PLSTM-Reg and PLSTM-Loc, while storage performance is nearly the same. The authors should explain this.

Response: Thank you for this insightful question. This contrast mainly reflects the strong persistence of reservoir storage and the short 1–7 day prediction horizon. Both models are initialized with observed storage, and storage predictions are updated over only a few days. Because daily inflows and releases are generally small relative to total storage, differences in release prediction have limited impact on storage over such a short period. As a result, storage performance remains similarly high for both PLSTM-Reg and PLSTM-Loc, whereas release predictions more clearly reveal the benefits of regional training.

We revised Section 3.1 (**lines 286–291**) as follows: "*As for short-term storage prediction, both approaches achieve similarly high accuracy because storage is strongly constrained by the observed initial condition and exhibits substantial persistence over the 1–7 day prediction horizon (Supplementary Fig. S3). As daily inflows and releases are generally much smaller than total storage, differences in release prediction accuracy have limited influence on simulated storage over this short period.*"

Figure 3. Since all statistical tests appear significant, the asterisks could be removed and the caption can simply state that all comparisons pass the significance test. Same for Figure 6.

Response: Thank you for this suggestion. We have made changes accordingly.

Lines 273–274. In the "unseen reservoir" experiment, it is unclear how PLSTM-Loc is trained. Does it use local data from the held-out reservoir, or random initial weights? This needs clarification.

Response: Thank you for this question. In Experiment III, PLSTM-Reg was evaluated on reservoirs that were excluded from regional training. The PLSTM-Loc values reported in this section are not models trained for the held-out reservoirs under Experiment III. Instead, they are the site-specific models from Experiment I, trained using each reservoir's local historical records under a chronological split. They are included as a reference benchmark to illustrate that the regional model can achieve performance comparable to, or better than, locally trained models even when transferred to previously unseen reservoirs.

We revised Section 3.2 (**lines 343–349**) as follows: "*In short-term prediction (Experiment III), PLSTM-Reg achieves a median 1-day-ahead KGE of 0.95 (Supplementary Fig. S7b) for reservoirs excluded from regional training. This performance exceeds that of the site-specific PLSTM-Loc models from Experiment I, which achieved a median KGE of 0.83 when trained using local historical records. The advantage of PLSTM-Reg persists across increasing lead times (Supplementary Fig. S7 g–h), indicating that the regional model can successfully transfer learned operating behaviors to previously unseen reservoirs.*"

Line 280. The improvement over rule-based methods may come from both regional pooling and flexible black-box modeling. The manuscript should avoid attributing the gain only to regional learning.

Response: Thank you for pointing this out. We agree that the performance gains over rule-based benchmarks should not be attributed solely to regional learning. This issue is addressed in our response to Major Comment 4, where we clarify that the observed improvements arise from the combined effects of regional data pooling, model architecture, static reservoir attributes, and physical constraints.

Figure 5. Please clarify why the metric values are not consistent with Figure S2. Do they correspond to different experimental settings or evaluation periods?

Response: Thank you for pointing this out. The metric values differ because the two figures correspond to different experiments and evaluation periods. Figure 5 presents Experiment IV, which evaluates spatial generalization to unseen reservoirs over the full available evaluation period. In contrast, Figure S4 (formerly Figure S2) presents Experiment II, which evaluates temporal generalization during the test period only (2010 onward).

We have revised both figure captions to clarify this distinction:

"*Figure 5. Comparison of release and storage long-term simulation performance (Experiment IV) between the regional model and four benchmark models. CDFs of three metrics (Corr, KGE, and NSE) are displayed for 256 reservoirs over the full evaluation period. The upper row (a–c) illustrates release performance, and the lower row (d–f) shows storage performance. Median metric values for each model are annotated within the respective panels.*"

"*Figure S4. Comparison of release and storage long-term simulation performance under Experiment II across all reservoirs during the test period. CDFs display release simulation performance in the upper row (a–c) and storage simulation performance in the lower row (d–f). Median metric scores are annotated within the respective panels.*"

Text S2. Rule-based benchmarks use default parameters, not calibrated ones. It does not necessarily imply that PLSTM-Reg outperforms these schemes.

Response: Thank you for raising this concern. The objective of Experiment IV is to evaluate performance under a spatial-transfer, data-scarce setting in which operational records are unavailable for the target reservoirs. Accordingly, benchmark implementation was designed to reflect the information that would realistically be available under this setting. HANA, WISS, and ZAJC were therefore implemented using their recommended parameters, consistent with their intended use as generic reservoir-operation parameterizations in large-scale hydrological and land-surface models. Calibrating these schemes separately for each held-out reservoir would require target-reservoir operation records and would therefore be inconsistent with the design of Experiment IV.

We also clarified the ISTARF setup. In Experiment IV, ISTARF was implemented using a fold-restricted parameter-transfer procedure, so that held-out reservoir records were not used during parameter fitting or donor selection. As a complementary comparison under data-rich conditions, calibrated ISTARF was evaluated in Experiment II using each reservoir's chronological training period.

To clarify the rationale behind the benchmark implementation and the intended scope of the comparison, we added the following text in Supplement Text S2: "*Because Experiment IV evaluates spatial transfer to held-out reservoirs under data-scarce conditions, benchmark parameterization was designed to reflect information availability in large-scale hydrological modeling. HANA, WISS, and ZAJC were implemented using their recommended parameters, consistent with their typical use as generic reservoir-operation schemes when local operation records are*

unavailable. Calibrating these models separately for the held-out reservoirs would require target-reservoir operation records and would therefore be inconsistent with the spatial-transfer setting. ISTARF was evaluated using the fold-restricted parameter-transfer procedure described above, ensuring that information from held-out reservoirs was not used during parameter fitting or donor selection. As a complementary comparison under data-rich conditions, calibrated ISTARF was also evaluated in Experiment II using each reservoir's chronological training period."

References

Chen, Y., Li, D., Zhao, Q., and Cai, X.: Developing a generic data-driven reservoir operation model, Advances in Water Resources, 167, 104274, https://doi.org/10.1016/j.advwatres.2022.104274, 2022.

Chen, Y., Cai, X., and Li, D.: Historical Operation Data of 256 Reservoirs in Contiguous United States, https://doi.org/10.4211/hs.092720588e2e4524bf2674235ff69d81, 2025.

Cheng, H., Wang, T., and Yang, D.: Quantifying the Regulation Capacity of the Three Gorges Reservoir on Extreme Hydrological Events and Its Impact on Flow Regime in a Changing Climate, Water Resources Research, 60, e2023WR036329, https://doi.org/10.1029/2023WR036329, 2024.

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, 2023.

Fang, K., Kifer, D., Lawson, K., Feng, D., and Shen, C.: The Data Synergy Effects of Time-Series Deep Learning Models in Hydrology, Water Resources Research, 58, e2021WR029583, https://doi.org/10.1029/2021WR029583, 2022.

Ford, L. and Sankarasubramanian, A.: Generalizing Reservoir Operations Using a Piecewise Classification and Regression Approach, Water Resources Research, 59, e2023WR034890, https://doi.org/10.1029/2023WR034890, 2023.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K.: An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing, Hydrology and Earth System Sciences, 12, 1007–1025, https://doi.org/10.5194/hess-12-1007-2008, 2008a.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K.: An integrated model for the assessment of global water resources –

Part 2: Applications and assessments, Hydrology and Earth System Sciences, 12, 1027–1037, https://doi.org/10.5194/hess-12-1027-2008, 2008b.

Kratzert, F., Gauch, M., Klotz, D., and Nearing, G.: HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin, Hydrology and Earth System Sciences, 28, 4187–4201, https://doi.org/10.5194/hess-28-4187-2024, 2024.

Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., van Soesbergen, A., Thieme, M., Garcia de Leaniz, C., Anand, M., Belletti, B., Brauman, K. A., Januchowski-Hartley, S. R., Lyon, K., Mandle, L., Mazany-Wright, N., Messager, M. L., Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Wishart, M., Xing, T., Yang, X., and Higgins, J.: The Global Dam Watch database of river barrier and reservoir information for large-scale applications, Sci Data, 11, 1069, https://doi.org/10.1038/s41597-024-03752-9, 2024.

Longyang, Q. and Zeng, R.: A Hierarchical Temporal Scale Framework for Data-Driven Reservoir Release Modeling, Water Resources Research, 59, e2022WR033922, https://doi.org/10.1029/2022WR033922, 2023.

Shen, Y., Yamazaki, D., Pokhrel, Y., and Zhao, G.: Improving Global Reservoir Parameterizations by Incorporating Flood Storage Capacity Data and Satellite Observations, Water Resources Research, 61, e2024WR037620, https://doi.org/10.1029/2024WR037620, 2025.

Steyaert, J. C. and Condon, L. E.: Synthesis of historical reservoir operations from 1980 to 2020 for the evaluation of reservoir representation in large-scale hydrologic models, Hydrology and Earth System Sciences, 28, 1071–1088, https://doi.org/10.5194/hess-28-1071-2024, 2024.

Steyaert, J. C., Sutanudjaja, E., Bierkens, M., and Wanders, N.: A data derived workflow for reservoir operations for simulating reservoir operations in a global hydrologic model, EGUsphere, 1–38, https://doi.org/10.5194/egusphere-2024-3658, 2025.

Tran, H., Zhou, T., Tan, Z., Fang, Y., and Ruby Leung, L.: Improving the prediction of daily reservoir releases over the CONUS using conditioned LSTM, Journal of Hydrology, 661, 133750, https://doi.org/10.1016/j.jhydrol.2025.133750, 2025.

Turner, S. W. D., Steyaert, J. C., Condon, L., and Voisin, N.: Water storage and release policies for all large reservoirs of conterminous United States, Journal of Hydrology, 603, 126843, https://doi.org/10.1016/j.jhydrol.2021.126843, 2021.

Wisser, D., Fekete, B. M., Vörösmarty, C. J., and Schumann, A. H.: Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network-Hydrology (GTN-H), Hydrology and Earth System Sciences, 14, 1–24, https://doi.org/10.5194/hess-14-1-2010, 2010.

Yang, S., Yang, D., Chen, J., and Zhao, B.: Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model, Journal of Hydrology, 579, 124229, https://doi.org/10.1016/j.jhydrol.2019.124229, 2019.

Yang, T., Gao, X., Sorooshian, S., and Li, X.: Simulating California reservoir operation using the classification and regression-tree algorithm combined with a shuffled cross-validation scheme, Water Resources Research, 52, 1626–1651, https://doi.org/10.1002/2015WR017394, 2016.

Yang, T., Zhang, L., Kim, T., Hong, Y., Zhang, D., and Peng, Q.: A large-scale comparison of Artificial Intelligence and Data Mining (AI&DM) techniques in simulating reservoir releases over the Upper Colorado Region, Journal of Hydrology, 602, 126723, https://doi.org/10.1016/j.jhydrol.2021.126723, 2021.

Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G., and Wheater, H.: Representation and improved parameterization of reservoir operation in hydrological and land-surface models, Hydrology and Earth System Sciences, 23, 3735–3764, https://doi.org/10.5194/hess-23-3735-2019, 2019.

Yu, B., Zheng, Y., He, S., Xiong, R., and Wang, C.: Physics-encoded deep learning for integrated modeling of watershed hydrology and reservoir operations, Journal of Hydrology, 657, 133052, https://doi.org/10.1016/j.jhydrol.2025.133052, 2025.

Zajac, Z., Revilla-Romero, B., Salamon, P., Burek, P., Hirpa, F. A., and Beck, H.: The impact of lake and reservoir parameterization on global streamflow simulation, Journal of Hydrology, 548, 552–568, https://doi.org/10.1016/j.jhydrol.2017.03.022, 2017.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., Liu, X., and Zhuang, J.: Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm, Journal of Hydrology, 565, 720–736, https://doi.org/10.1016/j.jhydrol.2018.08.050, 2018.

Zheng, Y., Liu, P., Cheng, L., Xie, K., Lou, W., Li, X., Luo, X., Cheng, Q., Han, D., and Zhang, W.: Extracting operation behaviors of cascade reservoirs using physics-guided long-short term memory networks, Journal of Hydrology: Regional Studies, 40, 101034, https://doi.org/10.1016/j.ejrh.2022.101034, 2022.