https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

PLSTM-Reg v1.0: A regional physics-encoded LSTM model for simulating reservoir operations under data scarcity

Bin Yu[1], Yanan Chen[1], Yi Zheng[123*]

[1] School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China
[2] Shenzhen Municipal Engineering Lab of Environmental IoT Technologies, Southern University of Science and Technology, Shenzhen, China
[3] State Environmental Protection Key Laboratory of Integrated Surface Water - Groundwater Pollution Control, Southern University of Science and Technology, Shenzhen, China

* *Correspondence to*: Yi Zheng (zhengy@sustech.edu.cn)

Abstract. Representing reservoir operations in large-scale hydrological models remains difficult due to complex release decisions and scarce operational records. Here, we develop PLSTM-Reg v1.0, a regional deep learning framework with physics encoded to simulate reservoir operations across diverse systems. The framework is evaluated using 256 representative reservoirs across the Contiguous United States, focusing on three core capabilities: temporal generalization to unseen periods, spatial transfer to unseen reservoirs, and historical data reconstruction. Under temporal testing, the regional model improves 1-day-ahead release forecasts from a median Kling–Gupta Efficiency (KGE) of 0.83 to 0.96 relative to local counterparts, and reduces poorly simulated cases (KGE < 0.8) from 41.8% to 2.3%. For long-term simulation, storage performance reaches a median KGE of 0.79, a modest gain over local models (0.76) but with notable robustness for reservoirs with large capacity. When transferred to unseen reservoirs, the model substantially outperforms widely used rule-based schemes: median KGE rises from 0.55 (best benchmark) to 0.73 for release and from 0.22 to 0.59 for storage, and the proportion of usable simulations (KGE > 0.5) increases from 56.6% to 89.8% for release and 14.4% to 61.7% for storage. In historical storage reconstruction, incorporating monthly satellite-derived surface area strengthens storage estimates and enables reconstruction accuracy comparable to models trained with local records. These results demonstrate that cross-reservoir deep learning combined with physical knowledge provides a scalable scheme for representing human water management within large-scale hydrological and land surface models under widespread data scarcity.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

1 Introduction

Approximately 2.8 million dams exist globally (Grill et al., 2019), including over 90,000 in the United States (DeNeale et al., 2019). Reservoirs sustain societies through flood control, water supply, and hydropower (Haemmerli et al., 2024; Zhao et al., 2020), while profoundly altering water cycles (Chalise et al., 2021; Fan et al., 2024). However, hydrological and land surface models struggle to represent their operations due to complex human decision-making (Longyang and Zeng, 2023; Turner et al., 2020) and scarce publicly available operational records (Steyaert and Condon, 2024), creating substantial biases in large-scale water resource assessments (Li et al., 2025; Turner et al., 2020, 2021). Although large-scale models such as PCR-GLOBWB (Sutanudjaja et al., 2018) and the NOAA National Water Model (Cosgrove et al., 2024) incorporate simplified regulation schemes (Hanasaki et al., 2008a; Shen et al., 2025; Wisser et al., 2010; Zajac et al., 2017), they remain unable to realistically reproduce observed flows (Steyaert and Condon, 2024; Vora et al., 2024). Scalable approaches for data-scarce reservoirs are thus urgently needed to improve the representation of human-water interactions in these large-scale models.

Historically, reservoir operations have been simulated using two main classes of methods. The first comprises generic rule-based models (Hanasaki et al., 2008a, b; Wisser et al., 2010; Zajac et al., 2017), widely applied in large-scale hydrological studies. These models employ simplified empirical rules based on reservoir purpose, size, and storage state. While interpretable, they cannot fully capture the complex, nonlinear decision making of real-world operators (Steyaert and Condon, 2024; Yang et al., 2021) and often require water-demand data for parameterization (e.g., Dong et al., 2022). Recent studies increasingly focus on the second class – data-driven approaches. Machine learning models such as decision trees, Long Short-Term Memory (LSTM) networks, and their variants can effectively learn realistic operation patterns from historical records (Chen et al., 2022; Dong et al., 2023a; Ford and Sankarasubramanian, 2023; Yu et al., 2025). However, these models are typically developed for individual reservoirs which requires site-specific calibration, limiting transferability across systems. Improving generalizability of machine learning-based reservoir operation models remains an open research challenge.

Regionalization is commonly used to transfer parameters of hydrological models from gauged to ungauged basins (Guo et al., 2021). In reservoir modeling, analogous work has regionalized parameters for simplified rule-based schemes. Turner et al. (2021) assigned

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

parameters to data-scarce reservoirs using nearby "donor" systems with similar operational purposes, while Steyaert et al. (2025) predicted rule parameters via Random Forest using static reservoir attributes. Yet these methods are still limited by the simplified rules themselves. Meanwhile, regionalization through deep learning (DL) has advanced rainfall–runoff modeling by training regional models on large and diverse watershed datasets to learn general links between static attributes and hydrological responses (Jiang et al., 2020; Kratzert et al., 2019; Yu et al., 2024), often outperforming local models. Li et al. (2024) further examined 452 data-rich U.S. reservoirs and identified recurring operational patterns tied to reservoir attributes (e.g., regions, sizes, functions), implying learnable structure. Extending DL-based regionalization to reservoir simulation is therefore appealing, but its feasibility remains unclear because reservoir operations reflect human decisions rather than purely physical processes, and it is not yet known whether regional patterns can generalize these anthropogenic behaviors.

This study aims to leverage knowledge learned across many reservoirs to address the longstanding challenge of simulating reservoir operations under data scarcity. We develop PLSTM-Reg v1.0, a regional physics-encoded LSTM framework designed to emulate human decision-making in reservoir operations. A dataset of 256 representative reservoirs across the Contiguous United States (CONUS), each with multi-decadal daily inflow, release, and storage records, provides the basis for evaluation. The framework is examined with respect to three central expectations. The first concerns temporal generalization, testing whether regional training improves predictive skill during unseen periods relative to models trained solely on local records. The second addresses spatial generalization, determining whether the regional model can be reliably transferred to data-scarce reservoirs and outperform widely used rule-based schemes. The third investigates historical reconstruction, evaluating the regional model's ability to recover past reservoir behavior, particularly storage dynamics, using readily available remote-sensing data. Together, these assessments provide a systematic evaluation of whether shared learning across reservoirs can enhance reservoir simulation and offer a scalable pathway for representing human water management in large-scale hydrological modeling.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

2 Data and methods

2.1 Data

This study evaluated 256 major reservoirs across the CONUS (Fig. 1), with daily inflow, release, and storage observations obtained from Chen et al. (2025). Each reservoir has at least 25 years of continuous records spanning 1990–2021 (end dates ranged from 31 December 2014 to 30 April 2021) <10% missing data. To characterize upstream hydrologic conditions, catchment boundaries were delineated using MERIT-Hydro and MERIT-Basins (Lin et al., 2019; Yamazaki et al., 2019) through the "delineator" package (Heberger, 2023). Resulting catchment areas were validated against reported values in the Global Dam Watch (GDW) database (Lehner et al., 2024), and delineations were manually corrected when relative error exceeded 10%. Catchment-scale forcings, including daily mean temperature and total precipitation, were obtained from the Daymet dataset (Thornton et al., 2022) and spatially averaged over each delineated catchment. These forcings complement the reservoir-specific dynamic inputs (daily inflow and storage) used in the modeling framework.

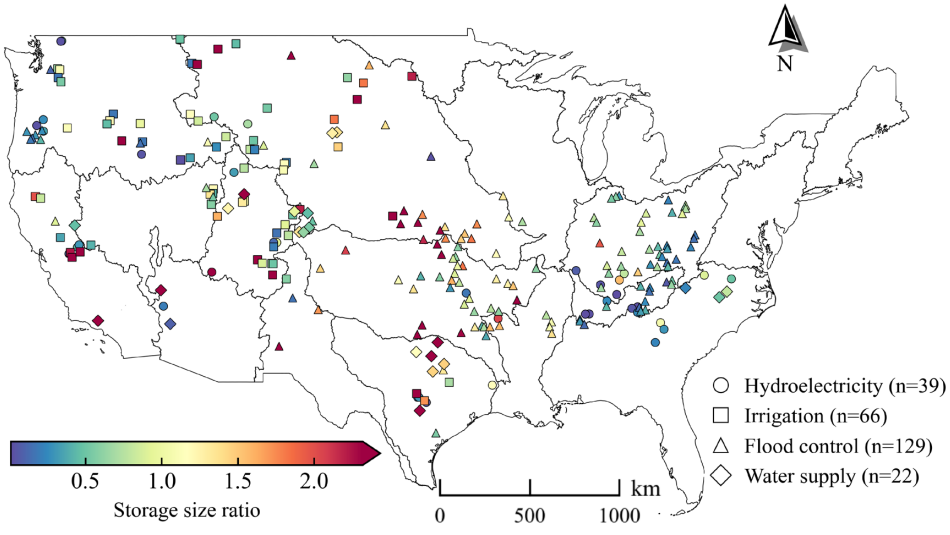

Figure 1. Spatial distribution of the 256 reservoirs across the CONUS. Each marker denotes a reservoir, with marker shape indicating the primary operational purpose and color representing the storage size ratio. This map highlights the geographic breadth and functional diversity of the reservoir network used for model training and evaluation.

For each reservoir, we compiled static attributes from the GDW database, including physical characteristics (e.g., dam height, storage capacity) and primary management purpose. We

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

also derived the catchment-scale static attributes: mean annual precipitation and temperature from Daymet, and the aridity index from the Global Aridity Index dataset (Zomer et al., 2022). In total, 21 static attributes were compiled (Supplementary Table S1). These attributes highlight the diversity of the 256 reservoirs. Storage capacities span 0.1–31.2 km³, and storage size ratios (i.e., storage capacity divided by mean annual inflow) range from 0.01 to 11.41, encompassing from small, near-natural reservoirs to large, heavily regulated ones. Management purposes are distributed across flood control (129), irrigation (66), hydroelectric generation (39), and water supply (22). The reservoirs span 15 of the 18 HUC2 water regions across the CONUS, demonstrating broad geographic and hydroclimatic representativeness. Their spatial distribution and key attributes (primary purpose and storage size ratio) are shown in Fig. 1.

In addition, to evaluate the contribution of remote sensing data to reconstructing operation records for data-scarce reservoirs, we incorporated monthly satellite-derived surface area time series from the Global Reservoir Surface Area Dataset (GRSAD; Zhao & Gao, 2018), which provides data for 6,817 reservoirs listed in the GRanD database from 1984 to 2018. For reservoirs with records extending beyond 2018, we supplemented GRSAD using the SARAH-CONUS dataset (2016–2023) (Yadav et al., 2025). Missing values 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.

2.2 Model architecture

This study introduces PLSTM-Reg v1.0, a regional Physics-Encoded LSTM (PLSTM) framework (Fig. 2a) for reservoir operation modeling. The model processes two categories of inputs I_t: dynamic forcings (precipitation, mean temperature, inflow, storage, and day of the year) and static attributes (Supplementary Table S1). Each input group is first transformed through a dedicated embedding network (a two-layer fully connected neural network, FCNN) to extract dynamic features and static site-specific condition vectors. The resulting embedding streams are concatenated into a combined input vector x_t, which is passed to the core PLSTM layer adopting the physics-encoded architecture of Yu et al. (2025). Unlike a standard LSTM, which maintains an abstract cell state c_t, the PLSTM introduces a physically meaningful internal state s_t that directly tracks reservoir storage. Within this layer (Fig. 2b), the recurrent unit generates hidden states h_t, which are mapped through a fully connected output layer to produce candidate

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

predictions $\tilde{r}_t$. These candidates are then processed by a physical constraint module that enforces mass conservation and valid bounds on release and storage. This architecture ensures physically feasible behavior and prevents long-term mass-balance drift. Full mathematical details are provided in Supplementary Text S1. For comparison, we also implemented local PLSTM models trained separately for each reservoir to assess the value of regional information. Following Kratzert et al. (2024), the local models adopt a simplified architecture to reduce overfitting risks. The network dimensions for both model configurations are summarized in Table S2.

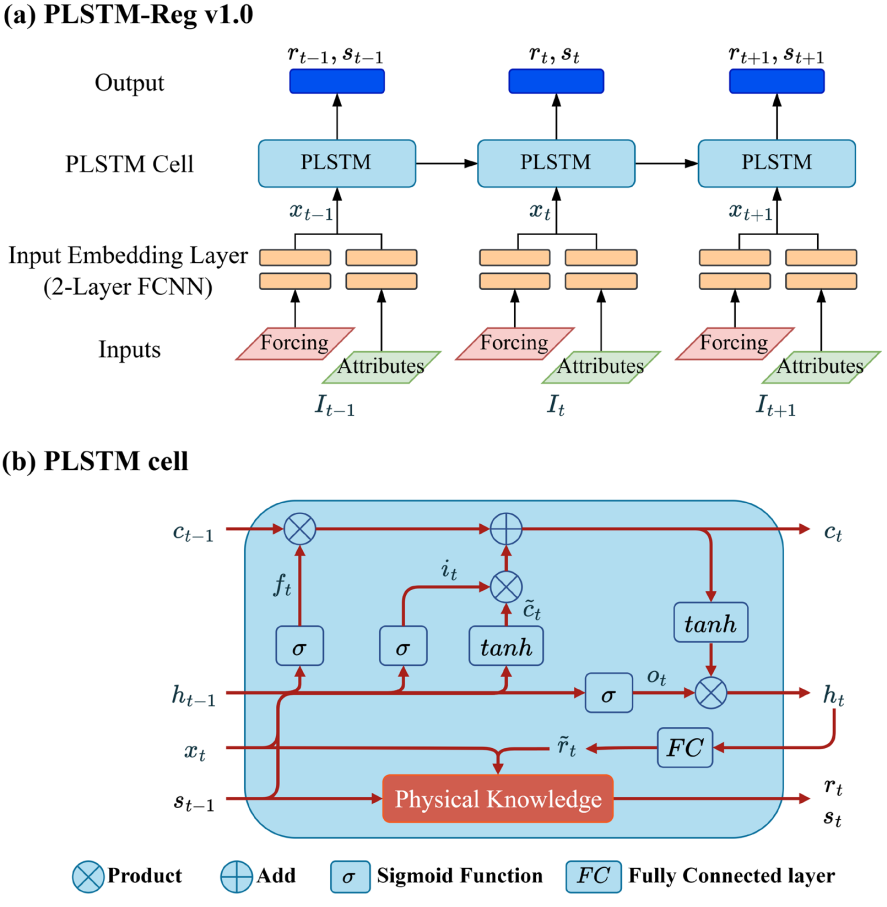

Figure 2. The PLSTM-Reg v1.0 model. (a) Model architecture. Dynamic forcings and static attributes are processed by separate embedding networks (two-layer FCNNs) and fused into the input vector x_t. (b) PLSTM cell. The cell introduces an explicit storage state s_t alongside the standard LSTM states c_t and h_t. The "Physical Knowledge" module (red box) enforces mass balance and physical bounds on the candidate release $\tilde{r}_t$, producing physically consistent release r_t and storage s_t outputs.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

All models were implemented using the NeuralHydrology package (Kratzert et al., 2022), which is built on the PyTorch framework (Paszke et al., 2019). Model optimization was performed with the Adam algorithm (Kingma and Ba, 2015) to minimize the Nash–Sutcliffe Efficiency (NSE; Nash & Sutcliffe, 1970) loss. Regional models were trained for up to 20 epochs with a batch size of 512, while local models were trained for up to 100 epochs with a batch size of 256. To assess modeling uncertainty, each model was trained ten times using different random seeds. For each reservoir, data were split chronologically into training (1990–2004), validation (2005–2009), and testing (2010 onward) periods. The testing period extends to the end of each site's available record, ensuring at least five years of data.

2.3 Experimental design and model evaluation

The proposed regional DL framework was evaluated for temporal and spatial generalization: performance on unseen periods at data-rich reservoirs and on unseen reservoirs in data-scarce contexts. Two real-world application scenarios, short-term forecasting (lead times <= 7 days) and long-term simulation, were tested under both temporal and spatial generalization settings. In addition, we explored the potential of using satellite data to recover historical records for data-scarce regions via our regional models. To support these analyses, we designed five numerical experiments (Experiments I–V), summarized in Table 1 and described in detail later in this section.

Table 1. Five numerical experiments and corresponding model configurations.

ID	Modeling scheme	Generalization type	Dynamic inputs
I	Short-term forecasting	Temporal (Chronological Split)	P, T, DoY, Inflow, Storage
II	Long-term simulation	Temporal (Chronological Split)	P, T, DoY, Inflow
III	Short-term forecasting	Spatial (Cross-Validation)	P, T, DoY, Inflow, Storage
IV	Long-term simulation	Spatial (Cross-Validation)	P, T, DoY, Inflow
V	Long-term simulation	Spatial (Cross-Validation)	P, T, DoY, Inflow, RS

Notes: P (precipitation), T (temperature), DoY (day of year), and RS (remote-sensing-derived monthly surface area).

Experiments I and II evaluate whether regional learning enhances predictive skill for data-rich reservoirs using a chronological split. In short-term forecasting (Experiment I), PLSTM

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

models were trained in a sequence-to-one manner to predict next-day release and storage. For lead times beyond one day, the model is run recursively, using the predicted storage from the previous step as input. In long-term simulation (Experiment II), we tested the models' ability to maintain physical consistency over multi-year reservoir operation using a sequence-to-sequence configuration. In this configuration, only the initial storage is prescribed, and the model simulates storage forward in time based solely on its internal dynamics and the mass-balance and range constraints embedded in the PLSTM cell. In both experiments, we benchmarked our regional models (PLSTM-Reg) against local counterparts (PLSTM-Loc) to quantify the added value of incorporating regional information.

Experiments III and IV evaluate the model's spatial generalizability to data-scarce reservoirs, under both short-term forecasting (Experiment III) and long-term simulation (Experiment IV) settings. We adopted a five-fold spatial cross-validation protocol (Feng et al., 2021; Kratzert et al., 2019) to assess performance on held-out reservoirs. We further benchmarked the regional model against three widely used rule-based schemes: the HANA model (Hanasaki et al., 2008a), the WISS model (Wisser et al., 2010), and the ZAJC model (Zajac et al., 2017). To ensure a fair comparison, each benchmark was implemented with its default, uncalibrated parameters (details in Supplementary Text S2). Furthermore, for all simulations in data-scarce settings (Experiment IV and benchmarks), storage was initialized at 50% of reservoir capacity with a 1-year warm-up period to equilibrate model states.

Experiment V further investigates the value of satellite-derived information for reconstructing historical records in data-scarce reservoirs using PLSTM-Reg. In this configuration (PLSTM-Reg+RS), the regional model takes monthly surface-area time series from GRSAD and SARAH-CONUS as an additional input. Because water surface area reflects underlying storage dynamics, we hypothesize that incorporating widely available remote-sensing observations can guide the learning process and improve daily-scale storage estimates. The added value of remote sensing is quantified by benchmarking this configuration against the standard PLSTM-Reg tested in Experiment IV under long-term simulation settings.

Model performance was evaluated using several widely adopted statistical metrics. Our primary metric is the Kling–Gupta Efficiency (KGE; Gupta et al., 2009), which provides a holistic assessment by decomposing model skill into correlation, bias, and variability components. We also

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

report the NSE and the Pearson Correlation Coefficient (Corr), both widely used for evaluating hydrological simulations. These metrics are defined as:

$$\mathrm{KGE} = 1 - \sqrt{(\mathrm{Corr} - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (1)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{N}(O_t - S_t)^2}{\sum_{t=1}^{N}(O_t - \mu_O)^2} \quad (2)$$

$$\mathrm{Corr} = \frac{\sum_{t=1}^{N}(O_t - \mu_O)(S_t - \mu_S)}{\sqrt{\sum_{t=1}^{N}(O_t - \mu_O)^2}\sqrt{\sum_{t=1}^{N}(S_t - \mu_S)^2}} \quad (3)$$

where O_t and S_t are the observed and simulated values at time t, over N time steps. The components of the KGE are the Pearson Correlation Coefficient (Corr), the variability ratio α (σ_S/σ_O), and the bias ratio β (μ_S/μ_O). μ_O and μ_S are the sample means, and σ_O and σ_S are the sample standard deviations of the observed and simulated series, respectively.

3 Results

3.1 Generalizability to unseen periods

In temporal generalization experiments, the regional model PLSTM-Reg outperforms the local model PLSTM-Loc across all three metrics in short-term forecasting (Experiment I). At the 1-day lead time (Fig. 3 a-c), PLSTM-Reg achieves a median KGE of 0.96 versus 0.83 for PLSTM-Loc, and greatly reduces poorly simulated cases: reservoirs with KGE < 0.8 drop from 41.8% to 2.3%. PLSTM-Reg maintains its advantage across all lead times (1-7 days) (Fig. 3d) with statistically significant improvements (paired Wilcoxon signed-rank test, $p<0.05$), despite the overall performance decline with longer lead times due to error accumulation as expected. To further identify which reservoir types benefit the most, we stratified the performance (shown for the 1-day horizon) by primary operational purpose and regulation capacity (Fig. 3 e-f). PLSTM-Reg delivers significant improvements across all operational categories ($p<0.05$), with the largest gains for irrigation reservoirs. Improvements strongly correlate with regulation capacity: while local model accuracy drops sharply for reservoirs with large storage size ratios, the regional model remains consistently accurate, indicating that cross-reservoir learning helps capture the complex decision-making of heavily regulated systems. As for short-term storage forecasting, both approaches achieve high accuracy due to the autoregressive input structure (Supplementary Fig. S1).

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

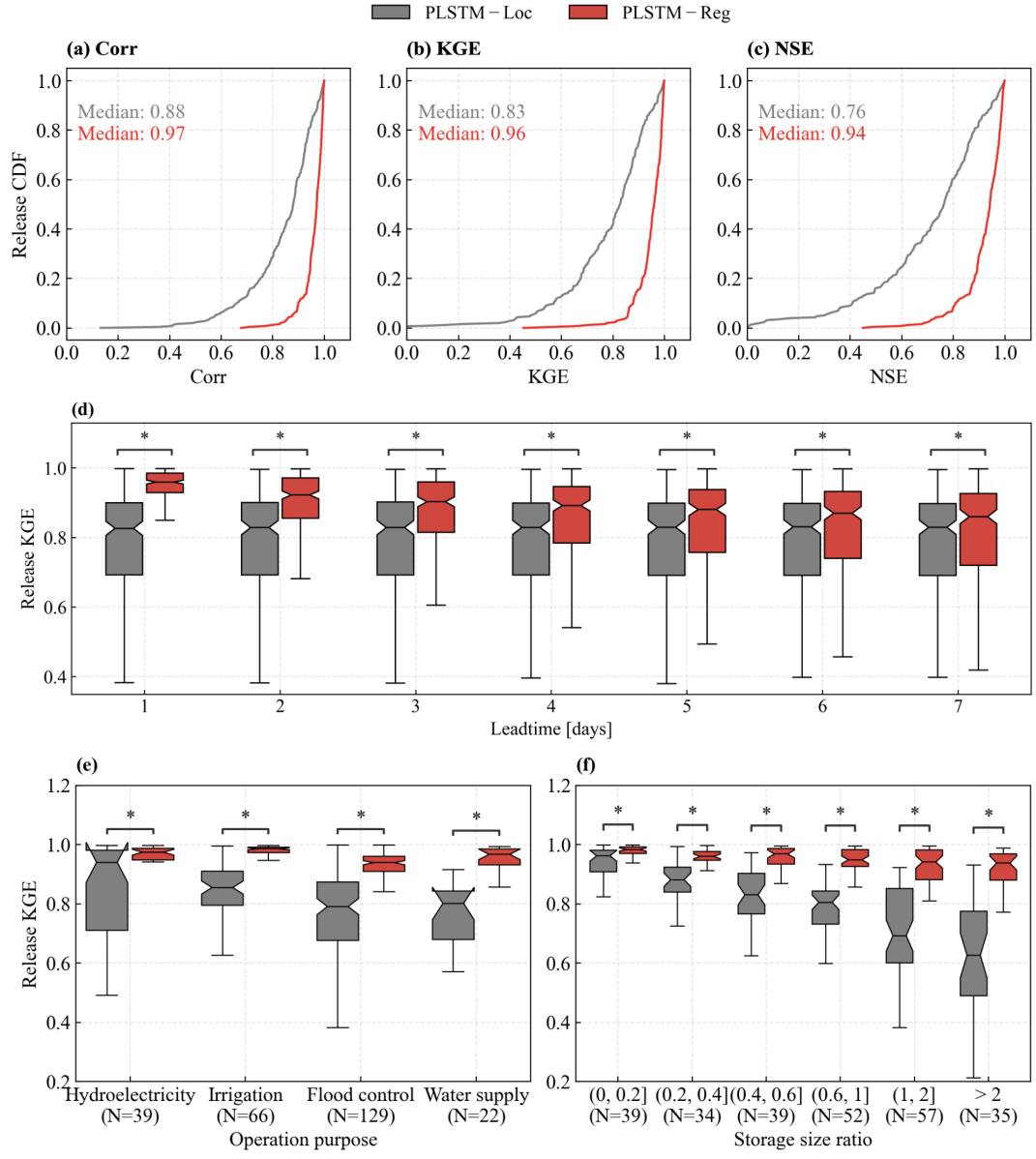

Figure 3. Evaluation of short-term release forecasting performance (Experiment I) across all reservoirs during the test period. (a–c) Cumulative distribution functions (CDFs) of 1-day-ahead release metrics comparing regional and local models. Median scores are annotated. (d) Evolution of release KGE scores across lead times (1–7 days). An asterisk (*) indicates a statistically significant difference (paired Wilcoxon signed-rank test, $p < 0.05$), while "ns" denotes not significant. (e–f) Performance stratified by (e) primary operational purpose and (f) storage size ratio. N denotes the number of reservoirs in each category.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

In long-term simulation (Experiment II), performance differences between the regional and local models are generally smaller than in short-term forecasting (Experiment I). Overall, PLSTM-Reg offers only modest improvements in storage simulation (median storage KGE: 0.79 vs. 0.76; Supplementary Fig. S2), and release accuracy remains similar. However, clear benefits emerge when examining specific reservoir groups. For reservoirs with storage size ratios > 0.2 (i.e., notable regulation capacity), the regional model performs notably better in the lower tail: the 25th percentile of storage KGE improves from 0.37 to 0.44, and the fraction of reservoirs with NSE < 0.5 declines from 34.1% to 29.0%. Together, these results suggest that while overall improvements narrow under long-term simulation, reflecting expected error accumulation, the regional model still provides meaningful robustness by mitigating failures in reservoir groups where the local model underperforms.

The robust performance gains, particularly for highly regulated systems, stem from a space-for-time substitution mechanism that allows the regional model to overcome the limitations of short local records. Complex multi-objective reservoirs (e.g., irrigation and flood control reservoirs) rely on rule structures triggered by specific hydroclimatic states that may never occur within a 15-year training record. For instance, a local model trained exclusively on wet periods cannot learn drought-hedging behavior. By pooling information from hundreds of reservoirs, the regional model absorbs operational strategies from “donor” sites and transfers them to those encountering similar, but locally unobserved conditions. This explains the substantial gains in systems with high regulation capacity (Fig. 3), where limited local data cannot represent the full range of decision-making complexity. Figure 4 further illustrates this advantage using four reservoirs from geographically distinct regions that experience out-of-distribution events during the test period. Here, such events are defined as instances with test inflows, releases, or storage levels exceeding the training envelope by more than 20%. Local models struggle to extrapolate in such regimes, whereas the regional model consistently infers plausible release decisions. By learning generalized operating rules, the regional framework provides more reliable forecasting and simulation, particularly as climate change increases the frequency of hydroclimatic extremes.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

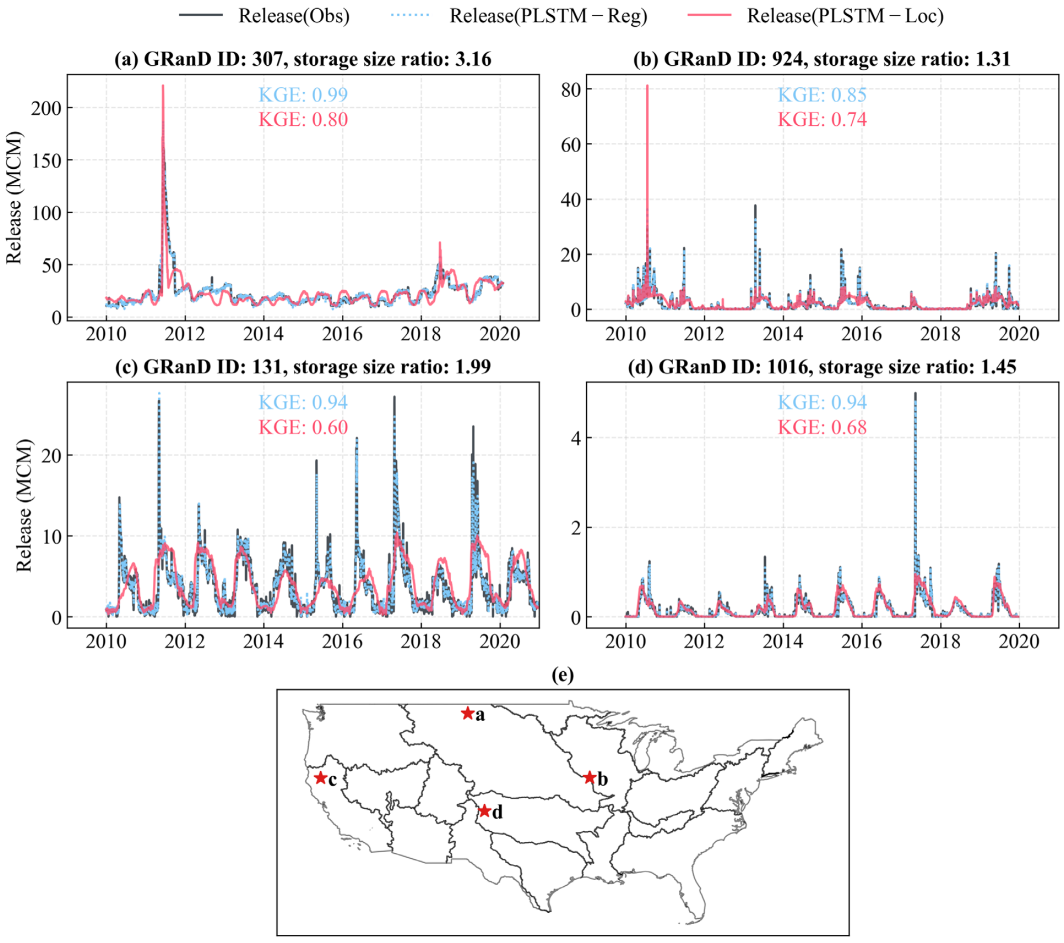

Figure 4. Release time series of four representative reservoirs illustrating the robustness of the regional model during out-of-distribution events at a 1-day lead time. (a)–(d) Release time series for four representative reservoirs; (e) The geographic locations of the selected reservoirs. Release values are expressed as daily total volume in million cubic meters (MCM).

3.2 Generalizability to unseen reservoirs

The regional framework demonstrates strong generalizability to unseen (data-scarce) reservoirs. In short-term forecasting (Experiment III), PLSTM-Reg achieves a median 1-day-ahead KGE of 0.95 (Supplementary Fig. S3b), substantially higher than the site-specific PLSTM-Loc (median KGE 0.83), and this advantage persists across increasing lead times (Supplementary Fig. S3 g–h). This generalization strength also holds in long-term simulation (Experiment IV). PLSTM-Reg markedly outperforms all rule-based benchmarks (Fig. 5), raising median release and

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

storage KGE from 0.55 (ZAJC, best benchmark) to 0.73 and from 0.22 to 0.59, respectively. More importantly, the proportion of usable simulations improves sharply: reservoirs with release KGE > 0.5 increase from 56.6% (ZAJC) to 89.8%, and for storage from 14.4% to 61.7%. These results highlight that the regional model learns adaptive, state-dependent operating strategies directly from data across hundreds of reservoirs, enabling realistic predictions for previously unseen systems—where uncalibrated, generic rule-based operation schemes often fail.

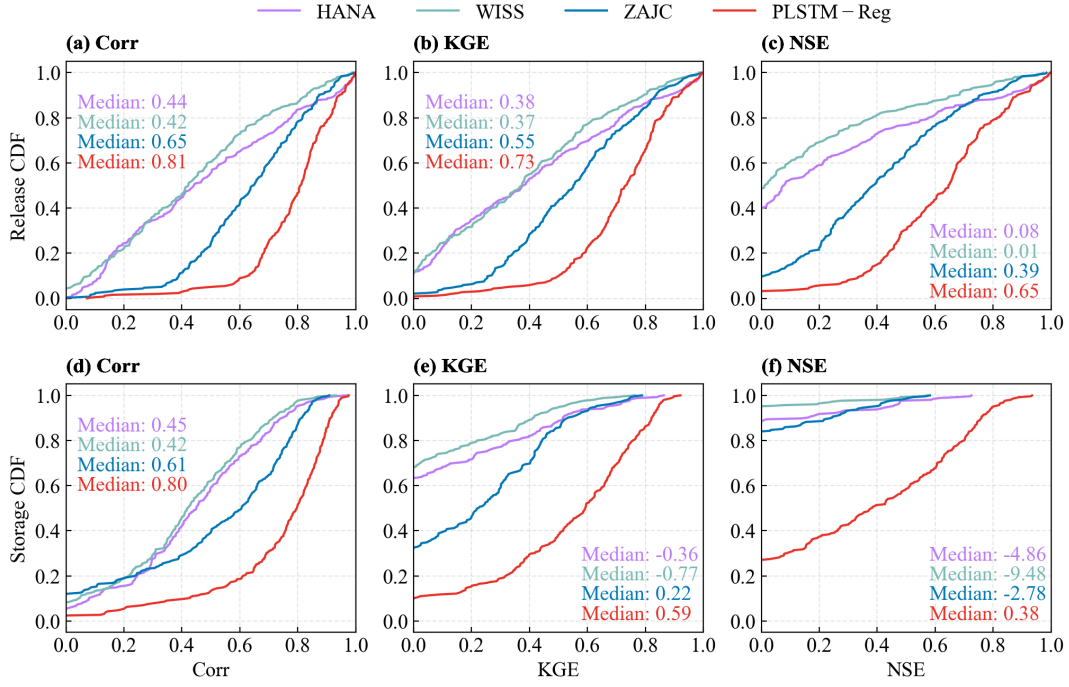

Figure 5. Comparison of release and storage long-term simulation performance between the regional model and three benchmark models. CDFs of three metrics (Corr, KGE, and NSE) are displayed for 256 reservoirs over the full evaluation period. The upper row (a–c) illustrates release performance, and the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

In long-term simulation, performance gains remain statistically significant across reservoir groups differentiated by operational purpose and storage size ratio (Fig. 6). The largest improvements occur for flood control systems, particularly those in the Mississippi Basin (Supplementary Fig. S4), where PLSTM-Reg outperforms ZAJC by a median KGE of 0.24 for release and 0.53 for storage. This reflects the model's superior ability to capture the complex, event-driven operating rules. As exemplified in Supplementary Fig. S5a–b, PLSTM-Reg holds

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

295 releases low and stable during inflow peaks and increases discharge only after the flood peak has passed, closely matching observed flood mitigation strategies. In contrast, ZAJC releases rise synchronously with inflows, diverging from real-world practices and potentially amplifying downstream flood risks. Significant gains also emerge for irrigation and water supply reservoirs (Fig. 6), where PLSTM-Reg reproduces the seasonal release timing and transition patterns
300 characteristic of the western U.S. (exemplified in Supplementary Fig. S5c–d). Performance advantages are also consistent across all storage size ratios (Fig. 6c–d). For release, median KGE improvements exceed 0.1 for all but the smallest reservoirs (storage size ratio < 0.2). Storage gains are even larger, with improvements of >0.2 across all size classes (Fig. 6d). This consistent superiority demonstrates that the regional framework remains robust across a wide range of
305 regulation regimes.

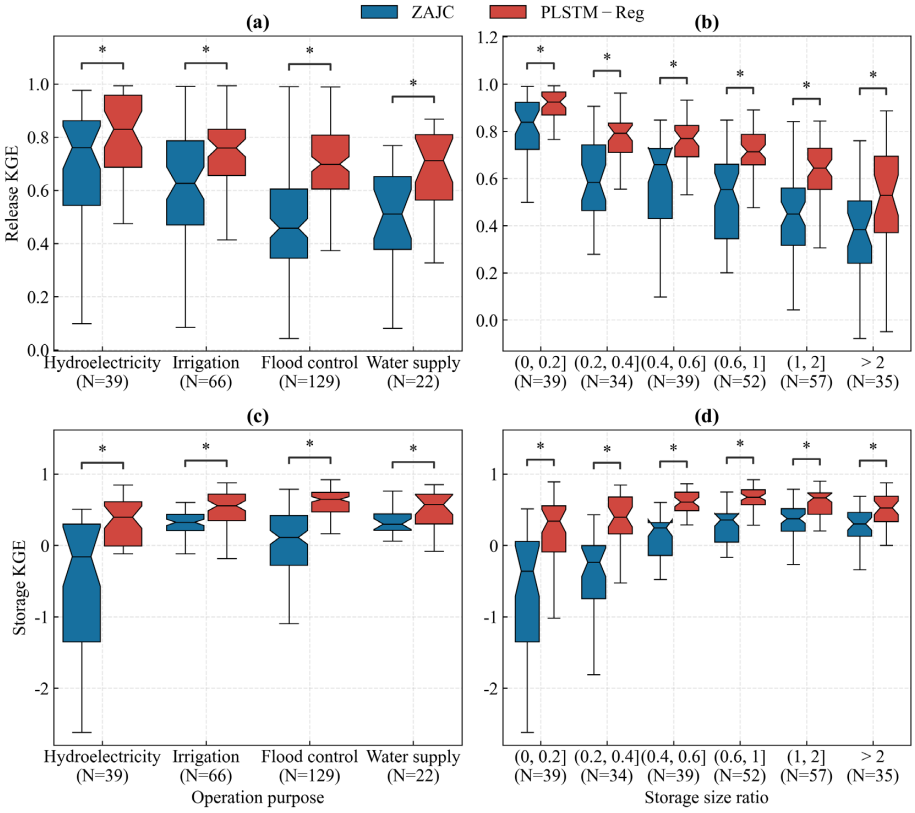

Figure 6. Long-term simulation performance comparison between the regional model (PLSTM-Reg) and the ZAJC benchmark for reservoir release (a, c) and storage (b, d), disaggregated by main operational purpose and storage size ratio. An asterisk (*) indicates

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

a statistically significant difference (paired Wilcoxon signed-rank test, $p < 0.05$), while "ns" denotes no significant difference. N denotes the number of reservoirs in each category.

3.3 Reconstruction of historical records with remote-sensing support

Beyond the CONUS, site-specific reservoir operation records are scarce or inconsistently maintained worldwide. While PLSTM-Reg is able to partially fill this historical data gap through spatial generalization, as demonstrated in Sect. 3.2, its storage simulations remain far less accurate than releases (Fig. 5). Testing on the 256 U.S. reservoirs using cross-validation, Experiment V (PLSTM-Reg+RS) confirms that incorporating widely available monthly remote-sensing observations can guide the learning process and yield measurable gains in storage performance, with median KGE improving from 0.59 to 0.69 and NSE from 0.38 to 0.64 (Fig. 7a). Notably, assimilating satellite data enables the regional model to outperform locally trained models (PLSTM-Loc, Experiment II): the proportion of reservoirs with acceptable storage performance (NSE > 0.5) increases to 64.5%, surpassing the 62.9% achieved by PLSTM-Loc (Supplementary Fig. S2). These results show that publicly available satellite data can effectively correct internal model states, serving as a substitute for missing ground-based records and enabling high-quality historical reconstruction in data-scarce reservoirs.

The spatial robustness of this reconstruction is visualized in Fig. 7b, which compares simulated and observed daily storage trajectories across 18 representative reservoirs spanning diverse hydro-climatic regimes. The model demonstrates high fidelity across a continuum of regulation dynamics, ranging from the humid, inflow-driven systems of the Ohio and Tennessee River basins (characterized by strong seasonal refill cycles) to the arid, storage-driven systems of the Colorado River basin. Crucially, the model captures multi-decadal trends in highly regulated systems, exemplified by the sustained depletion tracks of Lake Mead (GRanD ID 610) and Lake Powell (GRanD ID 597). This capacity to reproduce long-term trends aligns with the spatial distribution of performance gains (Supplementary Fig. S6), which highlights the significant performance improvements in large irrigation and water supply reservoirs of the arid western U.S. These large systems typically operate under "storage-driven" hedging policies (Li et al., 2024), where release decisions depend on storage availability to ensure supply reliability during dry periods. In such regimes, the RS-based surface area series acts as a critical state correction by effectively mitigating the error accumulation in long-term simulation. Consequently, the median storage KGE exceeds 0.7 for all primary reservoir purposes (with the exception of small

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

hydroelectric systems; Supplementary Fig. S7), confirming the framework as a robust strategy for reconstructing historical records across diverse data-scarce reservoirs.

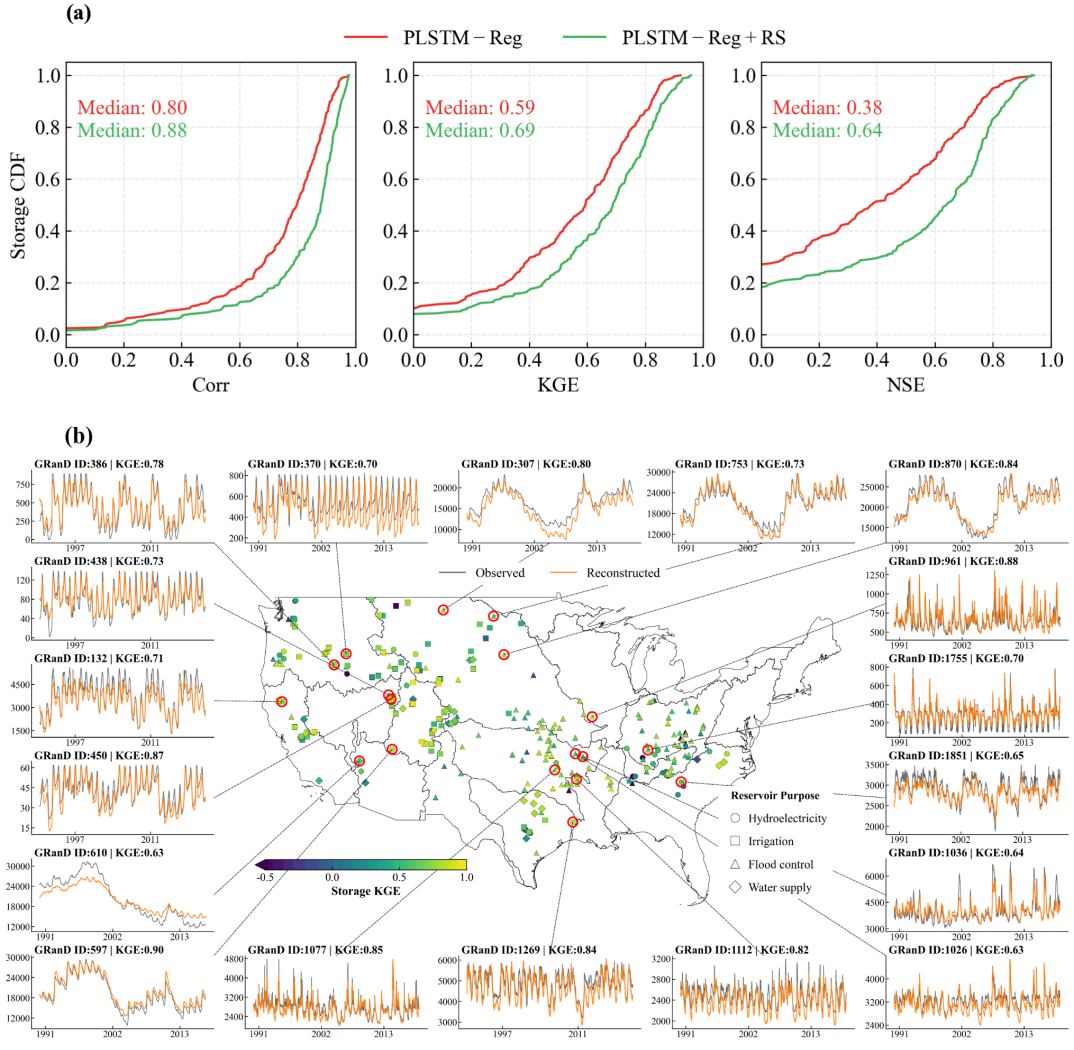

Figure 7. Evaluation of remote-sensing-augmented historical storage reconstruction. (a) Performance comparison between the standard regional model (PLSTM-Reg) and the augmented configuration (+RS). CDFs of Correlation, KGE, and NSE are displayed for 256 U.S. reservoirs over the full data period. (b) Spatiotemporal analysis of the PLSTM-Reg+RS reconstruction. The central map illustrates the spatial distribution of KGE scores (color). Surrounding panels display storage time series for 18 representative reservoirs selected to cover 15 HUC2 regions. Orange and gray lines represent model estimates and in situ measurements, respectively (units: million cubic meters).

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

4 Discussion

4.1 Short-term forecasting vs. long-term simulation

The proposed physics-encoded regional framework demonstrates strong temporal generalizability to unseen periods, while its performance differs between short-term forecasting and long-term simulation tasks. In short-term forecasting, the model is continuously updated with recent storage observations, which implicitly correct prediction trajectories. Conversely, long-term simulation is performed without corrective feedback from observations, making the model highly susceptible to error accumulation. To investigate how these task-specific characteristics influence model behavior, we examine the role of static reservoir attributes by comparing the standard PLSTM-Reg model against a variant without static attributes (PLSTM-Reg-NA). For short-term forecasting (1-7 days), removing static attributes has negligible impact (Supplementary Fig. S8), suggesting that predictions are primarily driven by dynamic inputs. In contrast, long-term simulation performance degrades substantially without static attributes (Fig. 8), with the median storage NSE decreasing from 0.69 to 0.21. The impact is most pronounced for reservoirs with storage size ratio > 0.2, where the proportion of simulation failures (NSE < 0) increases from 9.2% in the standard model to 42.4% in the NA variant.

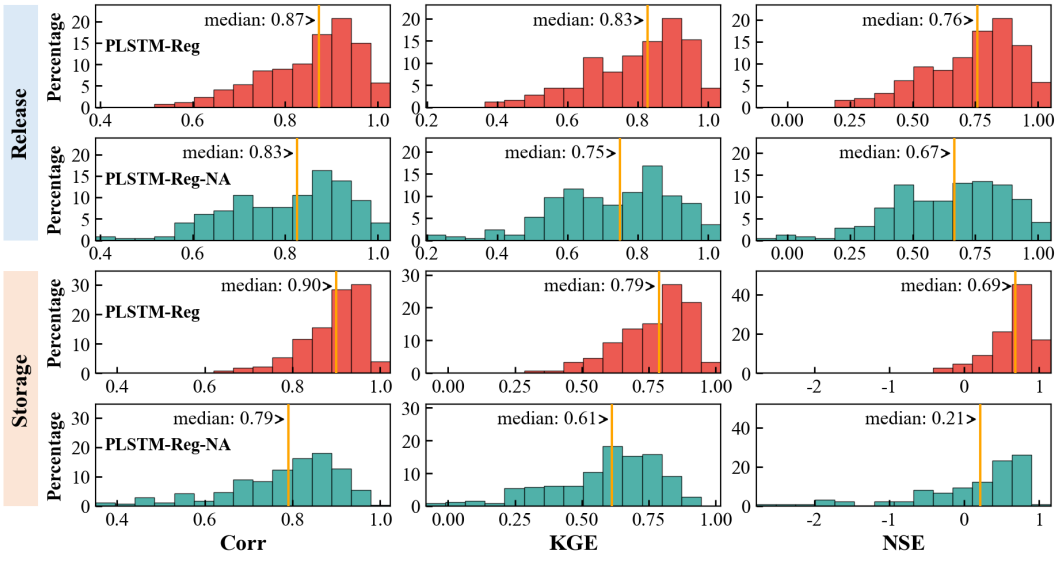

Figure 8. Comparison of long-term simulation performance between the standard regional model (PLSTM-Reg) and the variant without static attributes (PLSTM-Reg-NA). Histograms show the distribution of Correlation, KGE, and NSE for Release (top two rows)

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

and Storage (bottom two rows) during the test period. Orange vertical lines mark the median values. For visualization clarity, statistical outliers (values outside 1.5 × the interquartile range) are excluded.

These results underscore the distinct demands of the two tasks. In short-term forecasting, storage observations provide strong reservoir state information, and the model aims to learn general, state-dependent rules, making site-specific attributes redundant. For example, a storage value near maximum capacity typically requires a high release decision to avoid overtopping, regardless of the reservoir purpose and size. In contrast, long-term simulation lacks corrective state feedback. In this setting, static attributes act as essential "anchors" that define the reservoir's operational identity, enabling the model to adhere to appropriate operational policies for each site and prevent long-term deviation. Specifically, static attributes enable the model to distinguish between different operation strategies, such as multi-year drought management in the arid Upper Colorado region versus seasonal flood control in the humid Ohio region.

4.2 Applications and future directions

Our physics-encoded regional framework has broad potential applications in both real-time water management and long-term water planning. Accurate short-term release forecasts are increasingly important as climate change intensifies hydrological extremes (Gudmundsson et al., 2021; Iles et al., 2024). By capturing realistic decision rules from data, the model can serve as an effective tool for real-time decision support, aiding the assessment of flood and drought risks in complex reservoir systems (García-Feal et al., 2022; Yang et al., 2019). In addition, the regional model has a strong spatial generalization ability (Supplementary Fig. S3) to reservoirs with limited or no historical observations, offering a reliable baseline forecast for newly commissioned or data-scarce systems.

For long-term simulation, the framework offers a physically constrained, operationally realistic alternative to the static, rule-based modules currently used in Global Hydrological Models (GHMs) and Earth System Models (ESMs). This is particularly valuable for future climate change resilience assessments, allowing policymakers to evaluate system vulnerabilities under altered inflow patterns (Longyang and Zeng, 2024). Furthermore, the integration of remote sensing (PLSTM-Reg+RS, Experiment V) is promising for historical data reconstruction. Unlike traditional reconstruction methods constrained by satellite orbital tracks and cloud contamination

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

(Dong et al., 2023b; Wu et al., 2022), our approach delivers continuous, daily-scale records. This is critical for closing the terrestrial water storage (TWS) budget, where anthropogenic reservoir storage change remains a major but poorly constrained component (An et al., 2026; Cooley et al., 2021; Dong et al., 2022). The reconstructed dataset can be further integrated into existing modeling workflows to improve representations of the human-modified water cycle.

A critical next step is expanding the regional framework from continental to global scales. While our results demonstrate the efficacy of regional learning in the U.S., achieving a truly universal model requires compiling global datasets of historical operations. Expanding training diversity is a proven path to performance gains in deep learning hydrology (Kratzert et al., 2024). To handle global heterogeneity in operating rules, future work should explore domain adaptation techniques to improve model transferability across diverse policy paradigms (Shi and Cai, 2025).

In addition, emerging technologies offer opportunities to refine model inputs and architecture. High-resolution elevation data from the Surface Water and Ocean Topography (SWOT) mission (Das and Hossain, 2025) can serve as a critical data assimilation source to correct simulation trajectories in data-scarce reservoirs. Furthermore, the model's static embeddings could be enhanced by geospatial foundation models (e.g., Google's AlphaEarth Foundations; Brown et al., 2025). Finally, integrating Generative AI offers a powerful avenue for enhancing short-term operational robustness. By using probabilistic inflow ensembles from diffusion-based forecasting models (Ou et al., 2025; Yang et al., 2025), the PLSTM-Reg v1.0 framework could be adapted to quantify operational uncertainty, enabling operators to move beyond deterministic predictions and make robust, risk-aware release decisions during extreme events.

5 Conclusions

This study shows that reservoir operations can be simulated robustly using PLSTM-Reg v1.0, a regional physics-encoded deep learning model trained across many reservoirs with high-quality data. Three key insights emerge. First, regional learning delivers clear gains in temporal generalization. Short-term release predictions improve substantially over local models, and low-skill cases are largely eliminated, particularly for reservoirs with high storage capacity or multiple objectives that require diverse operating responses. These benefits demonstrate that information learned from other reservoirs can compensate for limited training data of individual reservoirs.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

Second, strong spatial transfer confirms that reservoir operation rules contain learnable structure that can be generalized. When applied to unseen reservoirs, the model consistently outperforms widely used rule-based schemes for both release and storage, and the fraction of usable simulations increases across operational purposes and regulation strengths. This result indicates that shared behavioral patterns, not explicit operating rules, can provide a reliable basis for simulating systems with sparse or nonexistent operational records. Third, incorporating satellite-derived surface area enables credible reconstruction of storage dynamics in data-scarce settings, producing accuracy comparable to models trained on complete observations. This demonstrates the practical value of integrating remote sensing into regional learning frameworks and highlights opportunities to fill historical gaps in poorly monitored systems.

Together, these findings establish that cross-reservoir deep learning, guided by physical constraints and supported by remote sensing, provides a scalable pathway for simulating human water management at regional and continental scales. The PLSTM-Reg framework offers immediate benefits as a transferable reservoir module in large-scale hydrological modeling and helps to quantify anthropogenic water storage change—particularly as global water systems evolve under climate change. Future work should focus on applying this framework globally, coupling it with existing hydrological and land surface models, and integrating emerging technologies such as geospatial foundation models and generative AI to further improve transferability, robustness, and operational utility.

Code and data availability

The reservoir operation dataset used in this study is available via HydroShare at **https://www.hydroshare.org/resource/092720588e2e4524bf2674235ff69d81/** (Chen et al., 2025). Static reservoir attributes were obtained from the Global Dam Watch (GDW) database, available at **https://www.globaldamwatch.org/database** (Lehner et al., 2024). Meteorological forcings from the Daymet dataset can be accessed at **https://daymet.ornl.gov/** (Thornton et al., 2022). The Global Aridity Index dataset can be downloaded from Figshare at **https://doi.org/10.6084/m9.figshare.7504448** (Zomer et al., 2022). Catchment delineation relied on MERIT-Hydro raster data, available at **https://mghydro.com/watersheds/rasters/** (Yamazaki et al., 2019), and MERIT-Basins vector data, available at

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

https://www.reachhydro.org/home/params/merit-basins (Lin et al., 2019). The Global Reservoir Surface Area Dataset is accessible via the Texas Data Repository at https://doi.org/10.18738/T8/DF80WG (GRSAD; Zhao & Gao, 2018).

The catchment delineation was performed using the delineator package, available at https://doi.org/10.5281/zenodo.7314286 (Heberger, 2023). The model training and evaluation framework was built using the open-source NeuralHydrology Python library, available at https://doi.org/10.5281/zenodo.7063258 (Kratzert et al., 2022). The PLSTM-Reg v1.0 model used in this study is archived at https://doi.org/10.5281/zenodo.18265198 (Yu, 2026).

Supplement

The supplement related to this article is available.

Author contributions

BY and YZ conceptualized the study. All authors contributed to the methodology and formal analysis. BY and YC were responsible for data curation and visualization. BY performed the validation and prepared the original draft. YC and YZ reviewed and edited the manuscript. YZ was responsible for supervision, project administration, and funding acquisition.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Disclaimer

Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

Acknowledgements

We acknowledge the availability of high-resolution satellite imagery from Google Earth, which was essential for the manual verification and correction of dam locations during catchment delineation. The computational resources used in this study were supported by the Center for Computational Science and Engineering at Southern University of Science and Technology. AI tools were used to improve the grammar and fluency of the manuscript.

Financial support

This research has been supported by the National Natural Science Foundation of China (grant nos. 42325702 and 42407072) and the High-level University Special Fund (grant no. G030290001).

References

An, Z., Li, Z., Jin, T., Jiang, W., Yuan, P., Liu, K., Wang, J., and Chen, P.: Water Storage Changes of Lakes and Reservoirs Across Asia (2018–2023) and Their Effects in Flood Control, Geophys. Res. Lett., 53, e2025GL119131, https://doi.org/10.1029/2025GL119131, 2026.

Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., and Kohli, P.: AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data, https://doi.org/10.48550/arXiv.2507.22291, 8 September 2025.

Chalise, D. R., Sankarasubramanian, A., and Ruhi, A.: Dams and Climate Interact to Alter River Flow Regimes Across the United States, Earths Future, 9, e2020EF001816, https://doi.org/10.1029/2020EF001816, 2021.

Chen, Y., Li, D., Zhao, Q., and Cai, X.: Developing a generic data-driven reservoir operation model, Adv. Water Resour., 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, 2025.

Cooley, S. W., Ryan, J. C., and Smith, L. C.: Human alteration of global surface water storage variability, Nature, 591, 78–81, https://doi.org/10.1038/s41586-021-03262-3, 2021.

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., Clark, E., Cabell, R., Casiday, N., Cui, Z., Eicher, K., Fall, G., Feng, X., Fitzgerald, K., Frazier, N., George, C., Gibbs, R., Hernandez, L., Johnson, D., Jones, R., Karsten, L., Kefelegn, H., Kitzmiller, D.,

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

Lee, H., Liu, Y., Mashriqui, H., Mattern, D., McCluskey, A., McCreight, J. L., McDaniel, R., Midekisa, A., Newman, A., Pan, L., Pham, C., RafieeiNasab, A., Rasmussen, R., Read, L., Rezaeianzadeh, M., Salas, F., Sang, D., Sampson, K., Schneider, T., Shi, Q., Sood, G., Wood, A., Wu, W., Yates, D., Yu, W., and Zhang, Y.: NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling, JAWRA J. Am. Water Resour. Assoc., 60, 247–272, https://doi.org/10.1111/1752-1688.13184, 2024.

Das, P. and Hossain, F.: Multi-Satellite Tracking of Surface Water Storage Change in the Era of Surface Water and Ocean Topography (SWOT) Satellite Mission, Earth Space Sci., 12, e2024EA004178, https://doi.org/10.1029/2024EA004178, 2025.

DeNeale, S. T., Baecher, G. B., Stewart, K. M., Smith, E. D., and Watson, D. B.: Current State-of-Practice in Dam Safety Risk Assessment, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), https://doi.org/10.2172/1592163, 2019.

Dong, N., Wei, J., Yang, M., Yan, D., Yang, C., Gao, H., Arnault, J., Laux, P., Zhang, X., Liu, Y., Niu, J., Wang, H., Wang, H., Kunstmann, H., and Yu, Z.: Model Estimates of China's Terrestrial Water Storage Variation Due To Reservoir Operation, Water Resour. Res., 58, e2021WR031787, https://doi.org/10.1029/2021WR031787, 2022.

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, J. Hydrol., 619, 129246, https://doi.org/10.1016/j.jhydrol.2023.129246, 2023a.

Dong, N., Yang, M., Wei, J., Arnault, J., Laux, P., Xu, S., Wang, H., Yu, Z., and Kunstmann, H.: Toward Improved Parameterizations of Reservoir Operation in Ungauged Basins: A Synergistic Framework Coupling Satellite Remote Sensing, Hydrologic Modeling, and Conceptual Operation Schemes, Water Resour. Res., 59, e2022WR033026, https://doi.org/10.1029/2022WR033026, 2023b.

Fan, C., Song, C., Wang, J., Sheng, Y., Lin, Y., Yuan, C., Safat Sikder, M., Crétaux, J.-F., Liu, K., Chen, T., Zeng, F., and Ke, L.: Emerging global reservoirs in the new millennium: Abundance, hotspots, and total water storage, Sci. Bull., 69, 2179–2182, https://doi.org/10.1016/j.scib.2024.04.043, 2024.

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

García-Feal, O., González-Cao, J., Fernández-Nóvoa, D., Astray Dopazo, G., and Gómez-Gesteira, M.: Comparison of machine learning techniques for reservoir outflow forecasting, Nat. Hazards Earth Syst. Sci., 22, 3859–3874, https://doi.org/10.5194/nhess-22-3859-2022, 2022.

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M. E., Meng, J., Mulligan, M., Nilsson, C., Olden, J. D., Opperman, J. J., Petry, P., Reidy Liermann, C., Sáenz, L., Salinas-Rodríguez, S., Schelle, P., Schmitt, R. J. P., Snider, J., Tan, F., Tockner, K., Valdujo, P. H., van Soesbergen, A.,

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

and Zarfl, C.: Mapping the world's free-flowing rivers, Nature, 569, 215–221, https://doi.org/10.1038/s41586-019-1111-9, 2019.

Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., Leonard, M., Liu, J., Müller Schmied, H., Papadimitriou, L., Pokhrel, Y., Seneviratne, S. I., Satoh, Y., Thiery, W., Westra, S., Zhang, X., and Zhao, F.: Globally observed trends in mean and extreme river flow attributed to climate change, Science, 371, 1159–1162, https://doi.org/10.1126/science.aba3996, 2021.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Haemmerli, H. L., Gerlak, A. K., and Swanson, T.: Reimagining hydropower in the United States, WIREs Water, 11, e1735, https://doi.org/10.1002/wat2.1735, 2024.

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, Hydrol. Earth Syst. Sci., 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, Hydrol. Earth Syst. Sci., 12, 1027–1037, https://doi.org/10.5194/hess-12-1027-2008, 2008b.

Heberger, M.: delineator.py: Fast, accurate watershed delineation using hybrid vector- and raster-based methods and data from MERIT-Hydro, , https://doi.org/10.5281/zenodo.10143149, 2023.

Iles, C. E., Samset, B. H., Sandstad, M., Schuhen, N., Wilcox, L. J., and Lund, M. T.: Strong regional trends in extreme weather over the next two decades under high- and low-emissions pathways, Nat. Geosci., 1–6, https://doi.org/10.1038/s41561-024-01511-4, 2024.

Jiang, S., Zheng, Y., and Solomatine, D.: Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning, Geophys. Res. Lett., 47, e2020GL088229, https://doi.org/10.1029/2020GL088229, 2020.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in: ICLR (Poster), arXiv: 1412.6980, 2015.

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.

Kratzert, F., Gauch, M., Nearing, G., and Klotz, D.: NeuralHydrology --- A Python library for Deep Learning research in hydrology, J. Open Source Softw., 7, 4050, https://doi.org/10.21105/joss.04050, 2022.

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

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9, 494–502, https://doi.org/10.1890/100125, 2011.

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.

Li, D., Chen, Y., Lyu, L., and Cai, X.: Uncovering Historical Reservoir Operation Rules and Patterns: Insights From 452 Large Reservoirs in the Contiguous United States, Water Resour. Res., 60, e2023WR036686, https://doi.org/10.1029/2023WR036686, 2024.

Li, D., Vora, A., and Cai, X.: Streamflow simulation improvements enabled by a state-of-the-art algorithm for reservoir routing in the U.S. National Water Model, JAWRA J. Am. Water Resour. Assoc., 61, e13244, https://doi.org/10.1111/1752-1688.13244, 2025.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., and Wood, E. F.: Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches, Water Resour. Res., 55, 6499–6516, https://doi.org/10.1029/2019WR025287, 2019.

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

Longyang, Q. and Zeng, R.: Assessing the effectiveness of reservoir operation and identifying reallocation opportunities under climate change, Environ. Res. Lett., 19, 074010, https://doi.org/10.1088/1748-9326/ad5459, 2024.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Ou, Z., Nai, C., Pan, B., Zheng, Y., Shen, C., Jiang, P., Liu, X., Tang, Q., Li, W., and Pan, M.: Probabilistic Diffusion Models Advance Extreme Flood Forecasting, Geophys. Res. Lett., 52, e2025GL115705, https://doi.org/10.1029/2025GL115705, 2025.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: Advances in Neural Information Processing Systems, 2019.

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

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

Shi, H. and Cai, X.: Extrapolability improvement of machine learning-based evapotranspiration models via domain-adversarial neural networks, Environ. Model. Softw., 187, 106383, https://doi.org/10.1016/j.envsoft.2025.106383, 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, Hydrol. Earth Syst. Sci., 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.

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geosci. Model Dev., 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018, 2018.

Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., and Kao, S.-C.: Daymet: Daily Surface Weather Data on a 1-km Grid for North America (4.1), https://doi.org/10.3334/ORNLDAAC/2129, 2022.

Turner, S. W. D., Doering, K., and Voisin, N.: Data-Driven Reservoir Simulation in a Large-Scale Hydrological and Water Resource Model, Water Resour. Res., 56, e2020WR027902, https://doi.org/10.1029/2020WR027902, 2020.

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, J. Hydrol., 603, 126843, https://doi.org/10.1016/j.jhydrol.2021.126843, 2021.

Vora, A., Cai, X., Chen, Y., and Li, D.: Coupling Reservoir Operation and Rainfall-Runoff Processes for Streamflow Simulation in Watersheds, Water Resour. Res., 60, e2023WR035703, https://doi.org/10.1029/2023WR035703, 2024.

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), Hydrol. Earth Syst. Sci., 14, 1–24, https://doi.org/10.5194/hess-14-1-2010, 2010.

Wu, G., Xiao, X., and Liu, Y.: Satellite-Based Surface Water Storage Estimation: Its history, current status, and future prospects, IEEE Geosci. Remote Sens. Mag., 10, 10–31, https://doi.org/10.1109/MGRS.2022.3175159, 2022.

Yadav, A., Zhang, S., Zhao, B., Allen, G. H., Pearson, C., Huntington, J., Holman, K., McQuillan, K., and Gao, H.: Mapping Reservoir Water Surface Area in the Contiguous United States Using the High-Temporal Harmonized Landsat and Sentinel (HLS) Data at a Sub-Weekly Time Scale, Geophys. Res. Lett., 52, e2024GL114046, https://doi.org/10.1029/2024GL114046, 2025.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resour. Res., 55, 5053–5073, https://doi.org/10.1029/2019WR024873, 2019.

https://doi.org/10.5194/egusphere-2026-1098
Preprint. Discussion started: 17 March 2026

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, J. Hydrol., 579, 124229, https://doi.org/10.1016/j.jhydrol.2019.124229, 2019.

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, J. Hydrol., 602, 126723, https://doi.org/10.1016/j.jhydrol.2021.126723, 2021.

Yang, W., Ji, H., Lonzarich, L., Song, Y., and Shen, C.: Diffusion-Based Probabilistic Modeling for Hourly Streamflow Prediction and Assimilation, https://doi.org/10.48550/arXiv.2510.08488, 9 October 2025.

Yu, B.: PLSTM-Reg v1.0: A regional physics-encoded LSTM model for simulating reservoir operations under data scarcity, , https://doi.org/10.5281/zenodo.18778747, 2026.

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

Yu, Q., Jiang, L., Schneider, R., Zheng, Y., and Liu, J.: Deciphering the Mechanism of Better Predictions of Regional LSTM Models in Ungauged Basins, Water Resour. Res., 60, e2023WR035876, https://doi.org/10.1029/2023WR035876, 2024.

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, J. Hydrol., 548, 552–568, https://doi.org/10.1016/j.jhydrol.2017.03.022, 2017.

Zhao, G. and Gao, H.: Automatic Correction of Contaminated Images for Assessment of Reservoir Surface Area Dynamics, Geophys. Res. Lett., 45, 6092–6099, https://doi.org/10.1029/2018GL078343, 2018.

Zhao, G., Bates, P., and Neal, J.: The Impact of Dams on Design Floods in the Conterminous US, Water Resour. Res., 56, e2019WR025380, https://doi.org/10.1029/2019WR025380, 2020.

Zomer, R. J., Xu, J., and Trabucco, A.: Version 3 of the Global Aridity Index and Potential Evapotranspiration Database, Sci. Data, 9, 409, https://doi.org/10.1038/s41597-022-01493-1, 2022.