Text S1. Detailed mathematical formulations of the PLSTM cell

The PLSTM cell integrates physical reservoir storage dynamics directly into the recurrent gating mechanism. The mathematical formulation of the PLSTM cell is as follows:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i s_{t-1} + b_i) \quad \text{(S1)}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + V_f s_{t-1} + b_f\right) \quad \text{(S2)}$$

$$\widetilde{c_t} = \tanh\left(W_g x_t + U_g h_{t-1} + V_g s_{t-1} + b_g\right) \quad \text{(S3)}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o s_{t-1} + b_o) \quad \text{(S4)}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad \text{(S5)}$$

$$h_t = o_t \odot \tanh(c_t) \quad \text{(S6)}$$

$$\widetilde{r_t} = FC(h_t) \quad \text{(S7)}$$

$$\widehat{r_t} = \min\left(\max(\widetilde{r_t}, Q_{min}), Q_{max}\right) \quad \text{(S8)}$$

$$\tilde{s}_t = s_{t-1} + x_t^{in} - \widehat{r_t} \quad \text{(S9)}$$

$$s_t = \min\left(\max(\tilde{s}_t, S_{min}), S_{max}\right) \quad \text{(S10)}$$

$$r_t = x_t^{in} + s_{t-1} - s_t \quad \text{(S11)}$$

where i_t, f_t, and o_t represent the input, forget, and output gates, respectively, controlled by logistic sigmoid (σ) and hyperbolic tangent ($tanh$) activation functions. The learnable parameters include bias vectors (b) and weight matrices (W, U, and V) corresponding to the input x_t, hidden state h_{t-1}, and explicitly tracked reservoir storage s_{t-1}. In the state update mechanism, $\odot$ denotes element-wise multiplication, while c_t and $\tilde{c}_t$ denote the cell state and candidate cell state, respectively. Equations (S8–11) define the "Physical Knowledge" module. In Eq. (S8), the candidate release $\widetilde{r_t}$ is constrained within predefined operational limits (Q_{min}, Q_{max}) such as the minimum environmental flow and maximum turbine capacity. In Eq. (S9), this adjusted release $\widehat{r_t}$ is used in the dynamic water balance equation along with the inflow component x_t^{in} to calculate a candidate storage $\widetilde{s_t}$ for the next time step. In Eq. (S10), this candidate storage is further bounded by the reservoir's physical limits (S_{min}, S_{max}) such as dead and maximum storage capacity. In Eq. (S11), the actual release r_t is recalculated via mass balance to account for any storage adjustments, ensuring the system strictly adheres to conservation laws.

Text S2. Implementation details of the three benchmark models

The first benchmark is the HANA model (Hanasaki et al., 2008), which has been widely used in land surface model such as MATSIRO (Pokhrel et al., 2012) and large-scale hydrological model like VIC (Droppers et al., 2020). It controls reservoir release using two operational schemes based on a size ratio, defined as storage capacity divided by mean annual inflow. The model sets the operational year as beginning in the first month when climatological inflow drops below the mean annual inflow, with the initial storage determining the yearly target release. For reservoirs with a size ratio above 0.5,

daily release is linked to mean annual inflow and adjusted according to the deviation of initial storage from a normal storage level, fixed at 85% of capacity. When the size ratio is 0.5 or lower, release is further modified to reflect daily inflow variability. Although an irrigation-specific variant exists, in which releases scale to downstream demand, this study applies the non-irrigation formulation consistently, as irrigation demand data are unavailable.

The second benchmark, referred to as the WISS model, determines reservoir release using a piecewise linear relationship between inflow and outflow, and is a core component of the WBM-plus global hydrological framework (Wisser et al., 2010). Two linear equations are used depending on whether the current inflow exceeds the long-term mean annual inflow. The corresponding slopes are governed by the parameters κ and $\lambda - 1$. Following Wisser et al. (2010), who calibrated the model against operational data from 30 reservoirs worldwide, we adopt the default coefficient values of 0.16 for κ and 0.6 for $\lambda - 1$ in this study.

The final benchmark is the ZAJC model, operates on a target storage-release formulation and serves as the reservoir representation within the LISFLOOD hydrological framework (Zajac et al., 2017). This algorithm partitions reservoir capacity into three operational zones: the dead storage zone (S_{dead}), the conservation zone (S_{cons}), and the flood control zone (S_{fld}). Each zone corresponds to a release target: minimum (Q_{min}), normal (Q_{norm}), and non-damaging flood (Q_{ndf}) releases. Within an active zone, daily discharge varies linearly with storage and inflow. While these boundaries and targets are ideally calibrated using historical operational records, generalized empirical defaults are necessary in data-scarce regions. Following Zajac et al. (2017), we set Q_{min}, Q_{norm}, and Q_{ndf} to the 5th, 30th, and 97th percentiles of natural streamflow, and define the three storage zones at 10%, 30%, and 97% of capacity.

Figure S1.

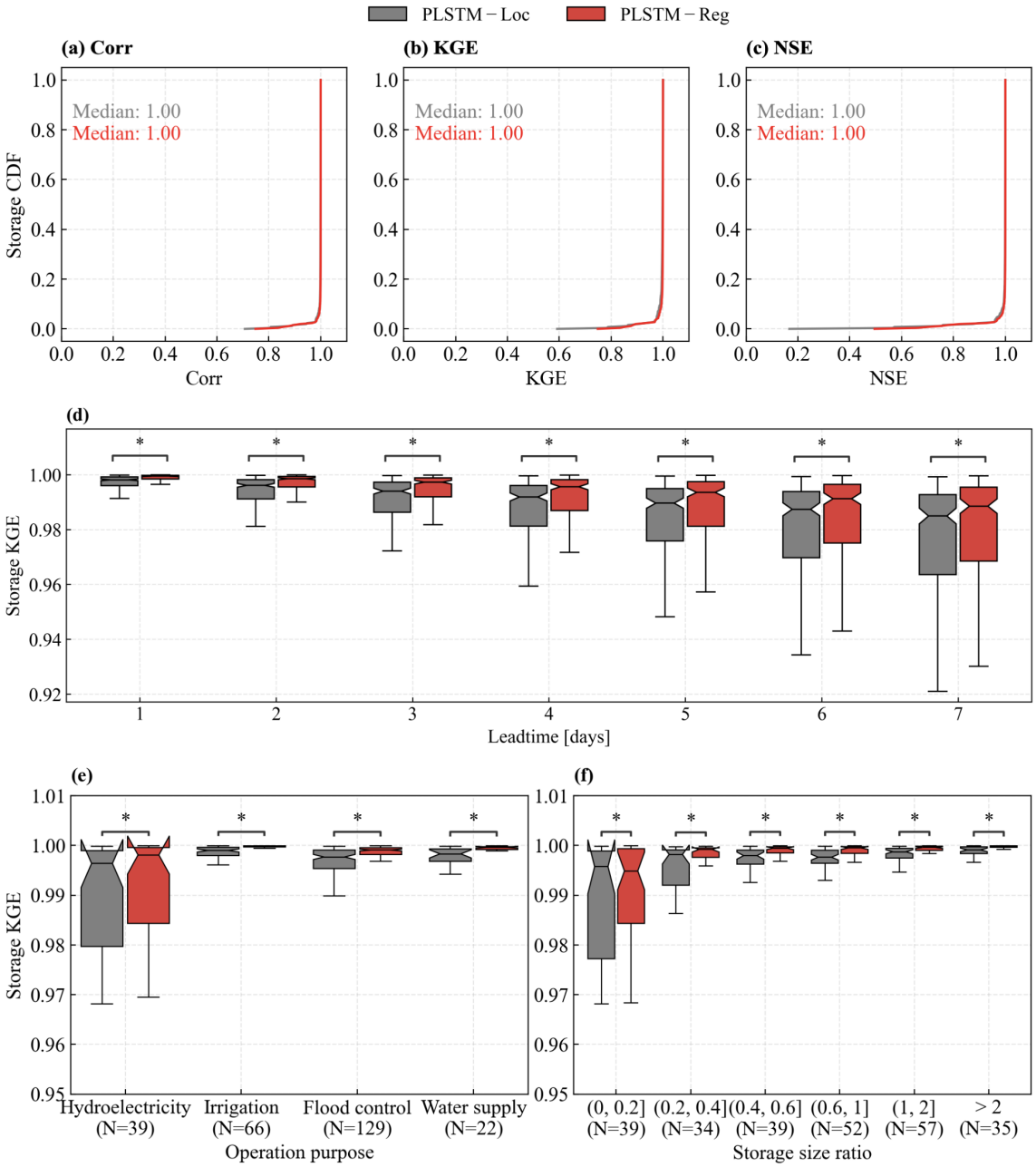

Figure S1. Evaluation of short-term storage forecasting performance (Experiment I) across all reservoirs during the test period. (a–c) Cumulative distribution functions (CDFs) of one-day-ahead storage metrics comparing regional and local models. Median scores are annotated. (d–e) Performance stratified by (d) primary operational purpose and (e) storage size ratio. (f) Evolution of storage 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 no significant. N denotes the number of reservoirs in each category.

Figure S2.

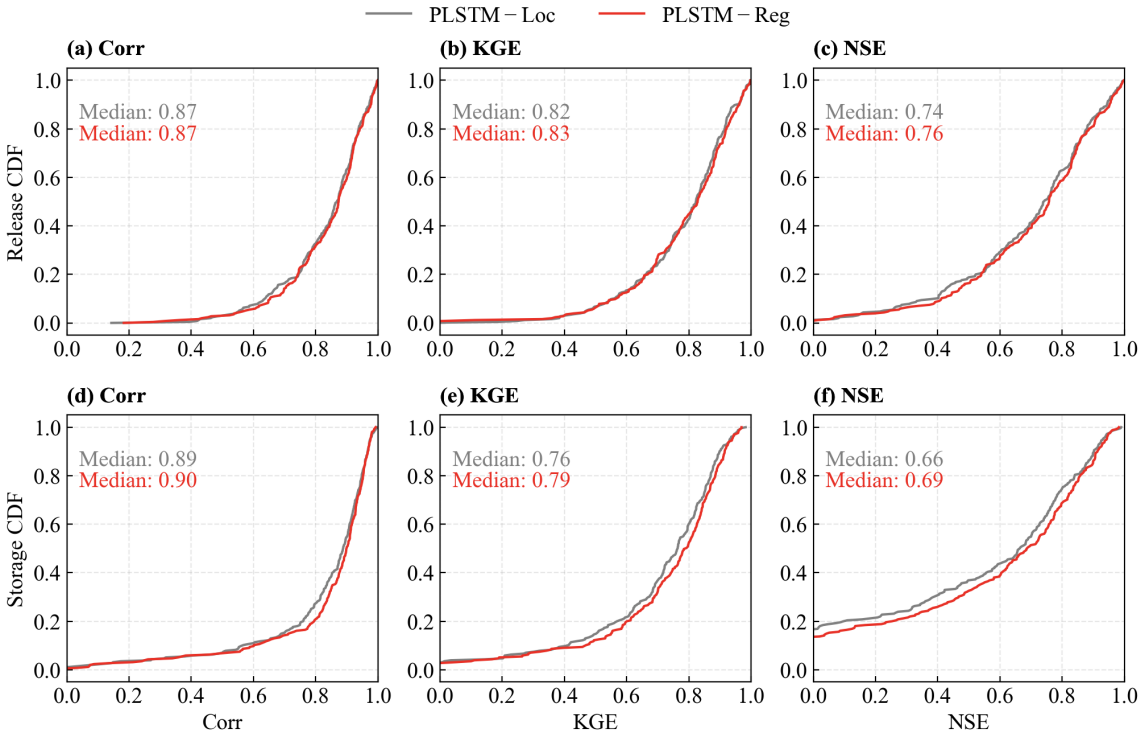

Figure S2. Comparison of release and storage long-term simulation performance between regional and local models across all reservoirs during the test period. CDFs display release simulation performance in the upper row (a–c) and storage simulation performance in the lower row (d–f). Median metric scores are annotated within the respective panels.

Figure S3.

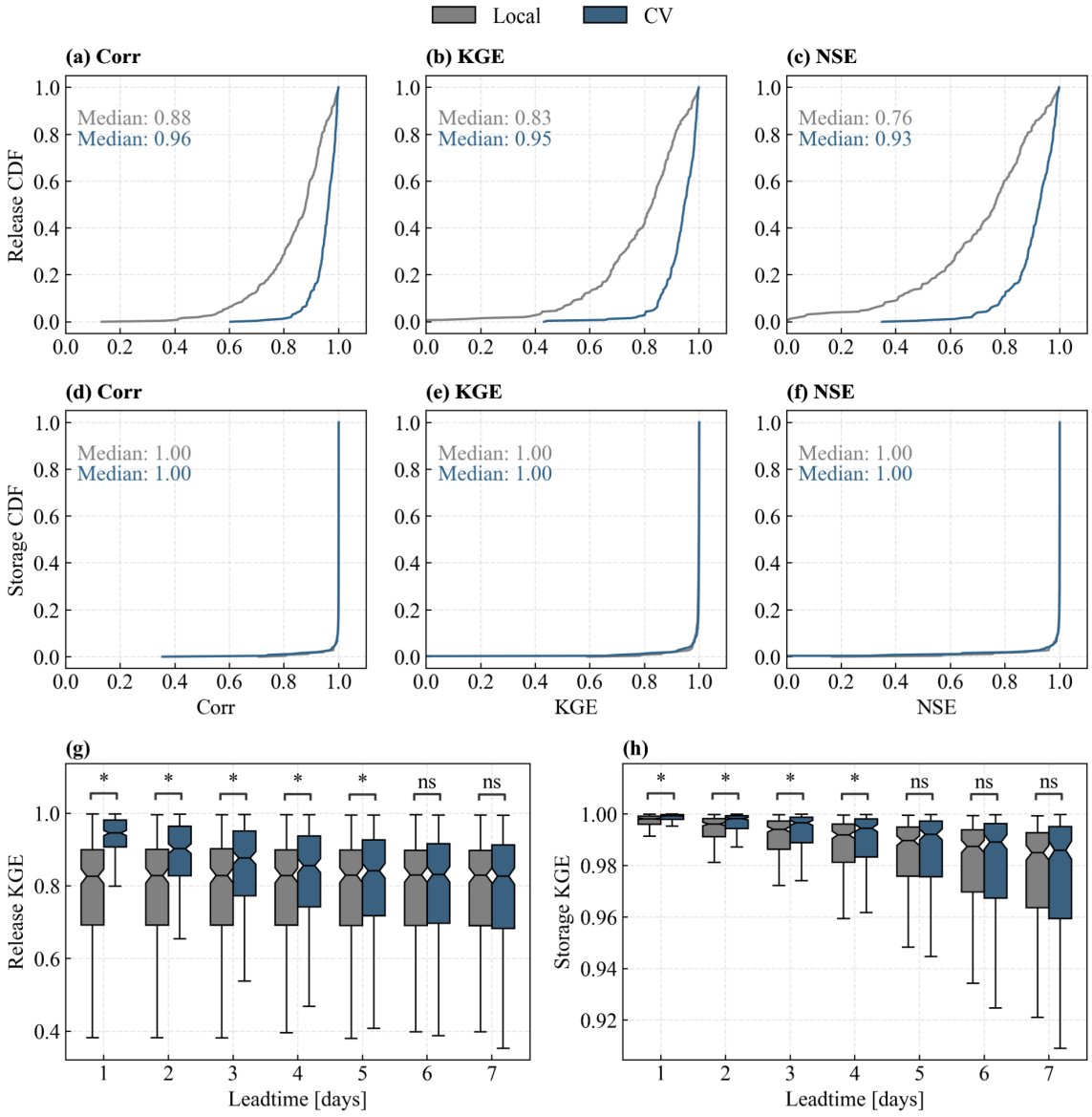

Figure S3. Performance comparison of short-term release and storage prediction among Cross-Validation (CV) and Local models during the test period. The CV model simulates data-scarce conditions via 5-fold cross-validation. CDFs display 1-day ahead performance for release in the upper row (a–c) and storage in the middle row (d–f) over the test period. Median values are annotated within the panels. The final row (g, h) illustrates KGE performance across lead times of 1 to 7 days. Asterisk (*) indicates a statistically significant difference (paired Wilcoxon signed-rank test, $p < 0.05$), while "ns" denotes no significant.

Figure S4.

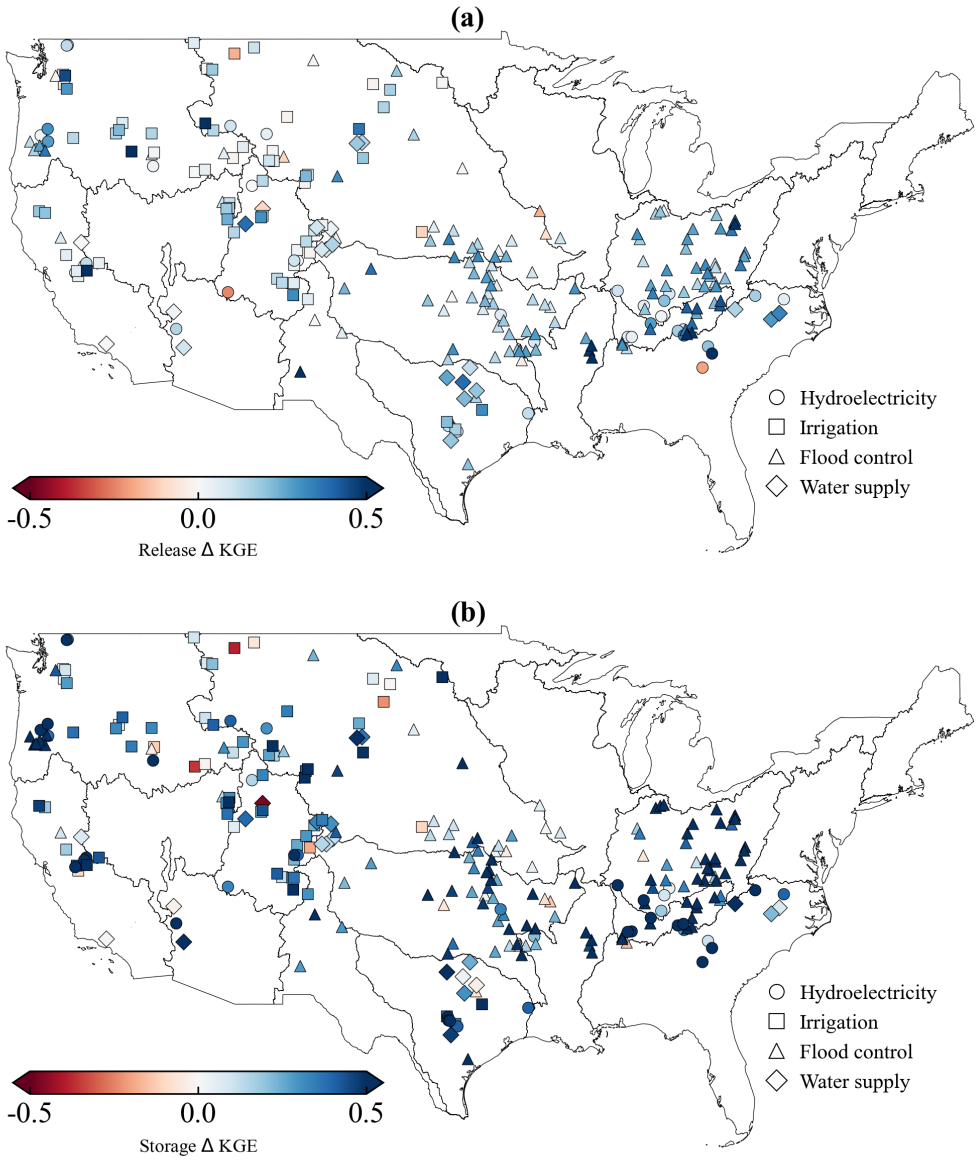

Figure S4. Spatial distribution of release (a) and storage (b) performance differences between the regional (PLSTM-Reg) and ZAJC models. Colors show the change in KGE, and marker shapes indicate primary operational purpose.

Figure S5.

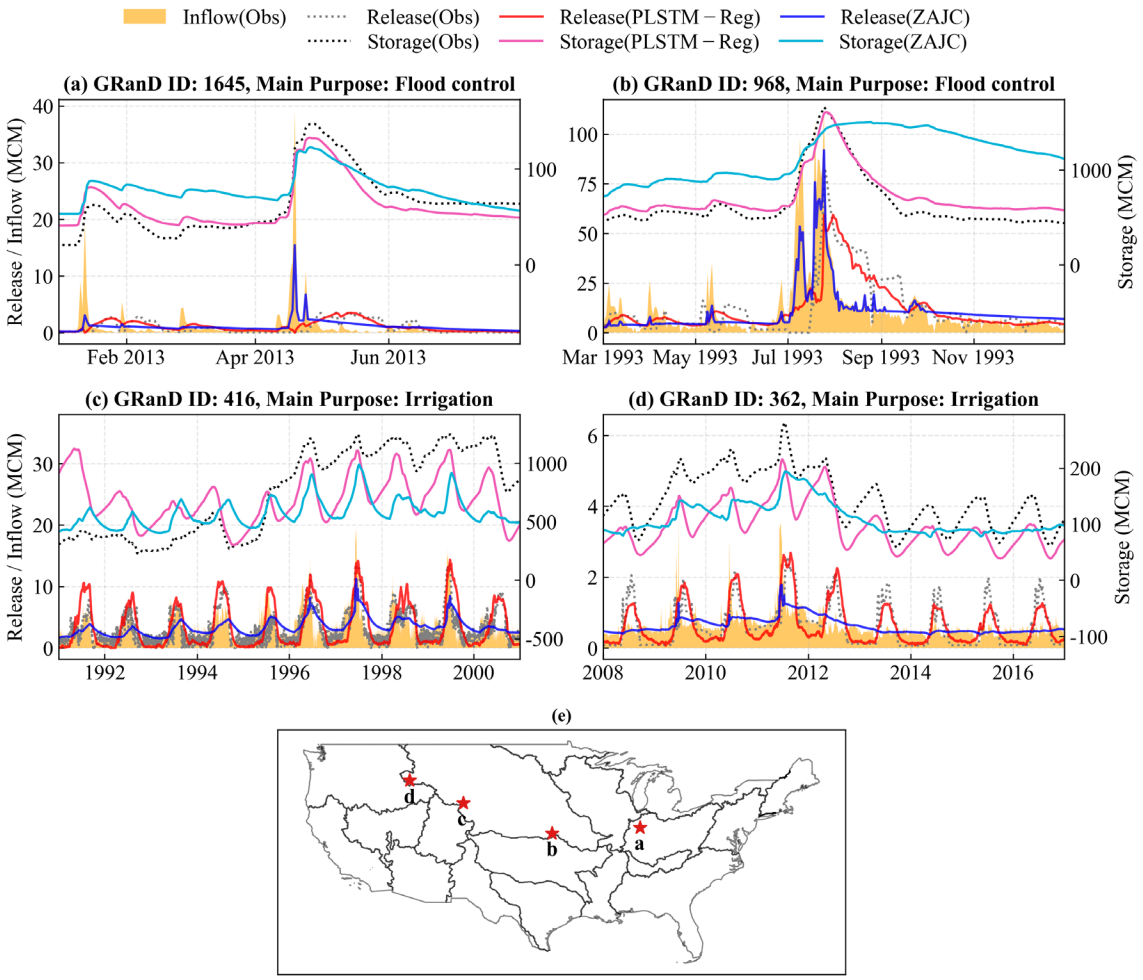

Figure S5. Comparison of operational behaviors between the regional model (PLSTM-Reg) and the generic rule-based benchmark (ZAJC) for selected ungauged reservoirs. (a)–(b) Flood Control Reservoirs, illustrating the regional model's ability to capture peak attenuation and storage buffering compared to the benchmark; (c)–(d) Irrigation Reservoirs, highlighting the regional model's capacity to infer seasonal release pattern; (e) The geographic locations of the selected reservoirs.

Figure S6.

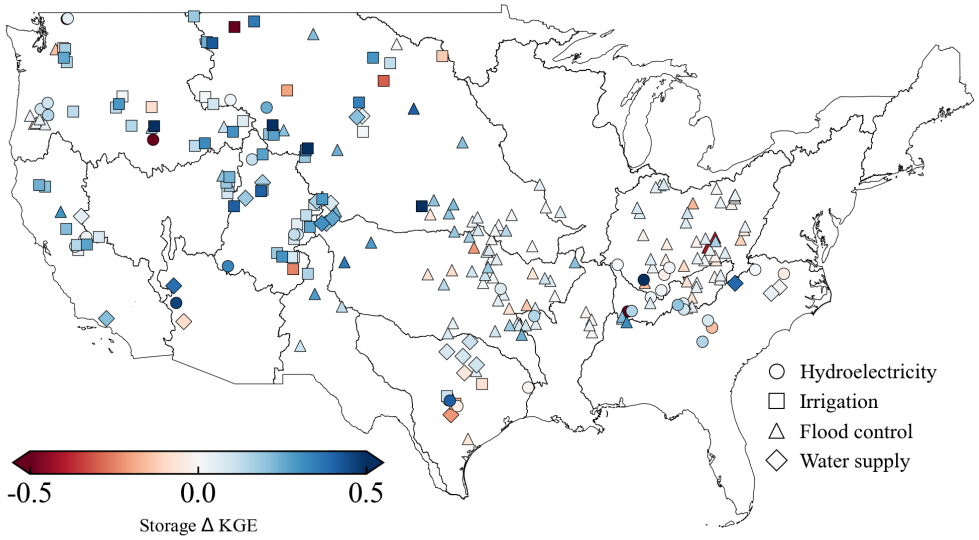

Figure S6. Spatial distribution of storage performance differences between the PLSTM-Reg+RS and PLSTM-Reg models. Colors indicate the change in KGE due to the inclusion of dynamic satellite-derived surface area observations. Marker shape denotes the primary operational purpose of each reservoir.

Figure S7.

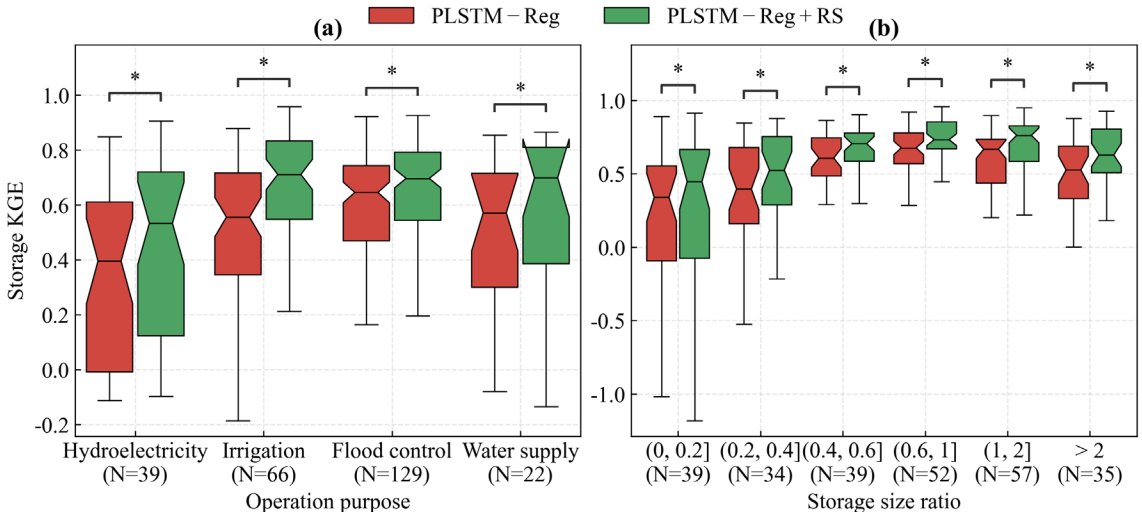

Figure S7. Performance comparison of the standard regional model (PLSTM-Reg) against remote-sensing-augmented configuration (+RS), disaggregated by (a) primary operational purpose and (b) storage size ratio. An asterisk (*) indicates a statistically significant difference (paired Wilcoxon signed-rank test, $p < 0.05$), while "ns" denotes no significant difference. N denotes the number of reservoirs in each category.

Figure S8.

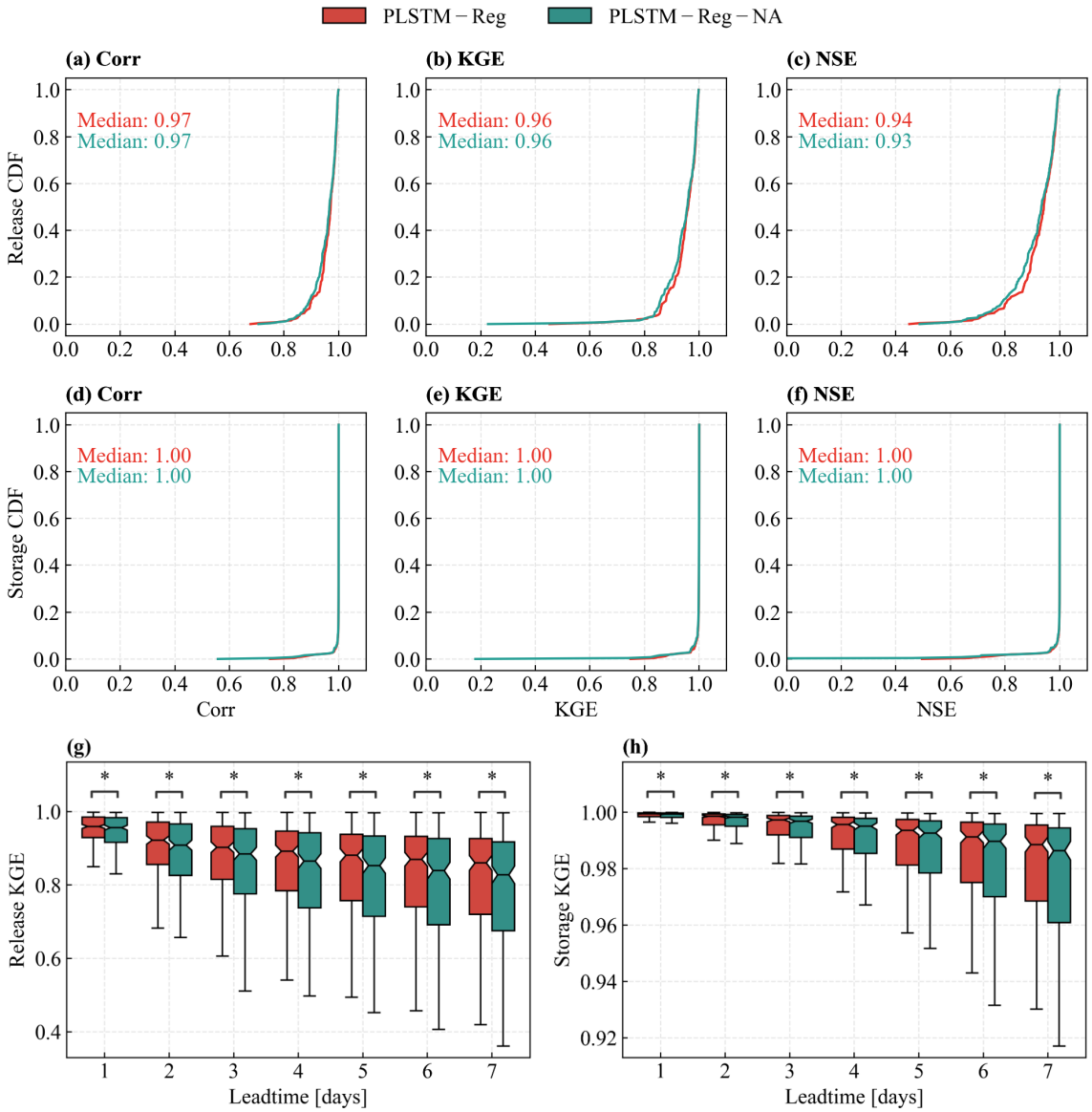

Figure S8. Performance comparison of short-term release and storage prediction among with and without static attributes regional models during the test period. Subplots (a–f) display CDFs of performance metrics for 1-day ahead predictions, where the upper row (a–c) represents release and the middle row (d–f) represents storage. Median performance values are annotated within these panels. The bottom row (g, h) illustrates the evolution of KGE scores across lead times ranging from 1 to 7 days. Asterisk (*) indicates a statistically significant difference (paired Wilcoxon signed-rank test, $p < 0.05$), while "ns" denotes no significant.

Table S1. List of the 21 static reservoir attributes used in this study, including reservoir physical characteristics, operational purpose, and watershed climate descriptors.

Name	Description
Dam height	Height of dam in meters
Dam length	Length of dam in meters
Area	Representative surface area of reservoir in square kilometers
Capacity	Representative maximum storage capacity of reservoir in million cubic meters
Depth	Average depth of reservoir in meters; calculated as ratio between storage capacity and surface area
Elevation	Elevation of reservoir surface in meters above sea level
Catchment area	Area of upstream catchment draining into the reservoir in square kilometers
Average discharge	Long-term average discharge at barrier location in liters per second
Storage size ratio	residence time of water in the reservoir, calculated as ratio between storage capacity and total annual flow
Fill fraction	the long-term mean fraction of reservoir capacity filled, only used in the long-term simulation to anchor the mean storage level.
USE_IRRI	Used for irrigation
USE_ELEC	Used for hydroelectricity production
USE_SUPP	Used for water supply
USE_FCON	Used for flood control
USE_RECR	Used for recreation
USE_NAVI	Used for navigation
USE_FISH	Used for fisheries
USE_OTHR	Used for other purposes
Mean temperature	Annual mean temperature
Mean precipitation	Annual mean precipitation
Aridity index	Aridity index

Table S2. Hyperparameters of network architecture for the PLSTM-Reg and PLSTM-Loc models.

Component	Hyperparameter	Regional	Local
Dynamic Embedding Network	Layer 1 neurons	16	8
	Layer 2 neurons	32	16
Static Embedding Network	Layer 1 neurons	16	N/A
	Layer 2 neurons	32	N/A
PLSTM cell	Hidden neurons	128	32
	Dropout rate	0.2	0.2

References

Droppers, B., Franssen, W. H. P., van Vliet, M. T. H., Nijssen, B., and Ludwig, F.: Simulating human impacts on global water resources using VIC-5, Geosci. Model Dev., 13, 5029–5052, https://doi.org/10.5194/gmd-13-5029-2020, 2020.

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, 2008.

Pokhrel, Y., Hanasaki, N., Koirala, S., Cho, J., Yeh, P. J.-F., Kim, H., Kanae, S., and Oki, T.: Incorporating Anthropogenic Water Regulation Modules into a Land Surface Model, J. Hydrometeorol., 13, 255–269, https://doi.org/10.1175/JHM-D-11-013.1, 2012.

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.

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.