Journal of Hydrology: Regional Studies 64 (2026) 103275

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Streamflow prediction in the Danube River Basin using a multi-source graph-integrated GCN-LSTM model

Mingze Sun [a,b], Yu Sun [a,b,*], Xin Yu [a,b], Yuyue Ye [a,b]

[a] Key Lab of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fujian, China
[b] National Engineering Research Center of Geospatial Information Technology, Fuzhou University, Fujian, China

ARTICLE INFO

Keywords:
Streamflow prediction
Graph convolutional network
Long short-term memory
Spatiotemporal modeling
Hydrological simulation

ABSTRACT

Study region: Danube River Basin
Study focus: Streamflow prediction is essential for water resources management and flood-risk mitigation. To better exploit spatial dependencies among gauging stations, we use a Graph Convolutional Network–Long Short-Term Memory (GCN–LSTM) framework for regional streamflow forecasting. The model integrates static and dynamic predictors from ERA5-Land and is evaluated at both monthly and daily temporal resolutions, with benchmarks against MLP, RNN, GRU, LSTM, and biLSTM. We further assess robustness and transferability through station-thinning experiments and cross-station validation at ungauged sites.
New hydrological insights: GCN–LSTM consistently outperforms the conventional baselines and maintains strong predictive skill across station densities, indicating that high accuracy can be achieved without an extremely dense gauging network, although performance drops when fewer than ~30 training stations are available. Daily inputs substantially improve the representation of short-term fluctuations and extreme flows, highlighting the importance of high temporal resolution for capturing rapid hydrological processes. Cross-station tests show a clear accuracy loss at ungauged locations that increases with distance from the training stations, underscoring that spatial generalization remains a key limitation in hydrologically heterogeneous regions.

1. Introduction

Streamflow prediction is a central task in hydrological research and plays a pivotal role in water resource management and decision-making. Accurate forecasting improves monitoring, early warning, and response to hydrological hazards such as floods and droughts, thereby reducing disaster impacts and protecting lives and property (Yang and Liu, 2020). Reliable forecasts are also vital for the rational allocation and efficient management of water use across agricultural, industrial, and domestic sectors (Mumbi et al., 2021; Zhang et al., 2023). Nevertheless, even in regions with relatively well-developed observation systems, such as Europe, the spatial distribution of streamflow gauges remains uneven, with sparse station coverage in some areas and gaps or discontinuities in observation records (Zhang et al., 2025). These limitations constrain the accurate characterization of the spatiotemporal variability of water resources, presenting significant challenges for integrated basin management and disaster risk reduction.

Early hydrological modeling efforts were largely dominated by physically based approaches, including conceptual, distributed, and hydrodynamic models, which simulate watershed responses by representing key processes such as rainfall–runoff generation and flow

* Corresponding author at: Key Lab of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fujian, China.
E-mail address: jade.yusun@outlook.com (Y. Sun).

https://doi.org/10.1016/j.ejrh.2026.103275
Received 28 October 2025; Received in revised form 17 January 2026; Accepted 18 February 2026
Available online 24 February 2026

routing (Kant et al., 2025; Sarker and Leta, 2025). These models offer strong physical interpretability and remain indispensable for process understanding, scenario analysis, and water resources planning when sufficient data and well-calibrated parameters are available. However, their predictive performance is often sensitive to the accuracy of physiographic and meteorological inputs (e.g., topography, land use, and climate forcing), and parameter calibration can be computationally demanding, particularly for large basins or long simulation periods (Herrera et al., 2022; Singh et al., 2025b). In addition, model performance may vary across spatial scales or data availability conditions, which poses practical challenges for large-scale or cross-regional applications without extensive calibration efforts (Razavi et al., 2025; Yan et al., 2025).

With the growing understanding of hydrological processes and advances in computational techniques, streamflow prediction has progressively shifted from physics-driven to data-driven approaches, marking a transition from traditional physical models to intelligent algorithms (Xu et al., 2024b; Yaseen, 2023). Building on this, machine learning methods such as Support Vector Regression (SVR), Random Forests (RF), and Artificial Neural Networks (ANN) have been employed for streamflow prediction (Alquraish and Khadr, 2021; Bargam et al., 2024). However, most conventional machine learning models still depend on manually engineered features and have difficulty capturing spatiotemporal couplings among inputs, which constrains their generalization and representational capacity (Gonzales-Inca et al., 2022; Ho and Goethals, 2022; Jia et al., 2023).Recent advances in deep learning provide a practical pathway to model complex hydrological dynamics from multi-source inputs with reduced reliance on handcrafted features (Dai et al., 2025; Yu et al., 2024a). Among these approaches, Long Short-Term Memory (LSTM) networks have been widely adopted for streamflow forecasting due to their ability to capture long-term dependencies, seasonality, and lagged responses in hydrological time series (Sabzipour et al., 2023; Zhao et al., 2024). Meanwhile, Graph Convolutional Networks (GCN) offer a natural framework for representing inter-station relationships on non-Euclidean structures by propagating information along a graph constructed from spatial proximity or physiographic similarity (Bhatti et al., 2023; Deng et al., 2023; Li et al., 2023). Combining GCN with LSTM therefore provides an integrated means to capture both spatial dependencies across stations and temporal dynamics within each station, which is particularly relevant for large river basins where hydrological responses exhibit strong regional coherence and cross-site coupling.

In this study, we employ a GCN-LSTM framework for streamflow forecasting with the aim of enhancing spatiotemporal modeling capacity and regional generalization. The Danube River Basin is selected as a representative case, with streamflow data from 97 gauging stations provided by the GRDC. Experiments are conducted at both monthly and daily resolutions to assess the effect of input resolution on predictive performance. Building on this, we progressively reduce the number of stations to evaluate model robustness and adaptability in sparse observation conditions. Additionally, leave-station-out experiments are designed to test the model's transferability and spatial generalization to ungauged sites. This study investigates the following research questions: (1) How does temporal resolution affect streamflow prediction accuracy? (2) To what extent does station density influence model stability and robustness? (3) Can the GCN-LSTM model maintain predictive performance and generalization when applied to ungauged locations? By answering these questions, the study provides new insights into the relationships among temporal resolution, spatial density, and generalization. It also offers practical implications for streamflow simulation and water resource management.

2. Study areas

In this study, we selected the Danube River Basin as the case study region, with streamflow observations from 97 gauging stations

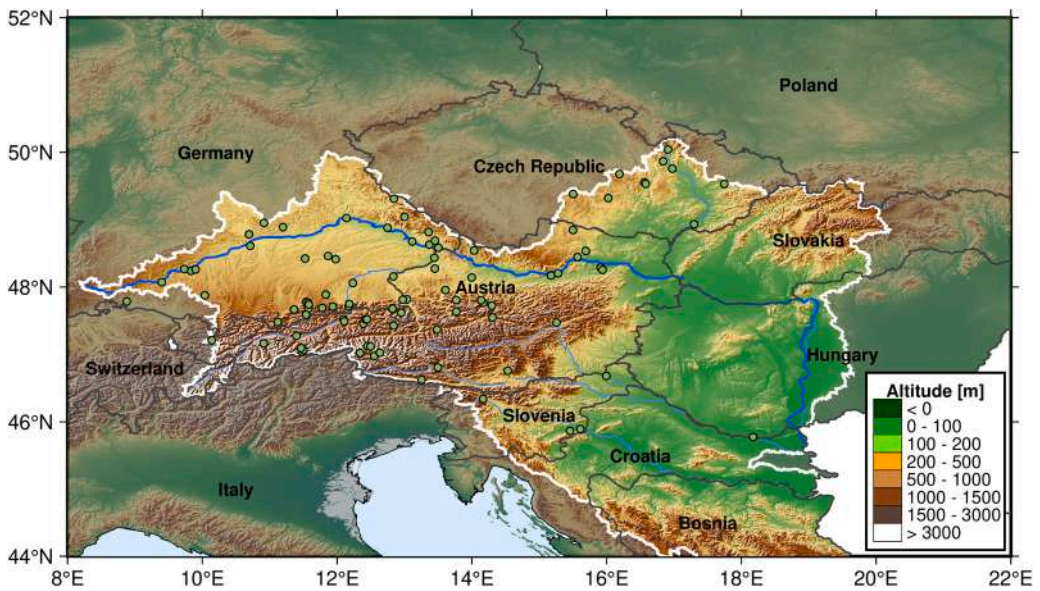

Fig. 1. Study Area and Spatial Distribution of Hydrological Stations.

within the basin used for experiments and comparative analysis (Fig. 1). The Danube is the largest river basin in Europe, spanning multiple countries from the Alpine headwaters in Central Europe to downstream lowland and deltaic environments (Halder et al., 2022). This pronounced longitudinal gradient is accompanied by substantial heterogeneity in both physiography and hydroclimate. The upper basin is dominated by mountainous terrain and Alpine-influenced conditions, where runoff exhibits strong seasonality and is strongly modulated by temperature-dependent processes (Koch et al., 2011). In contrast, the middle and lower reaches encompass broader valleys and plains under temperate to continental climates, where precipitation intermittency, atmospheric demand, and storage–release dynamics jointly govern streamflow variability (Lóczy, 2015; Schiller et al., 2010). This diversity gives rise to complex hydrological processes and pronounced spatial contrasts in streamflow, making the basin a representative area for basin-scale prediction studies (Nichersu et al., 2022). The relatively dense gauging network used here provides broad coverage across major sub-regions, enabling a comprehensive characterization of regional hydrological features and a systematic evaluation of model adaptability and robustness across contrasting hydroclimatic regimes. Accordingly, the Danube River Basin offers an appropriate setting for assessing the predictive performance and generalization capability of the proposed GCN-LSTM framework.

3. Data

A multi-source input dataset integrating static and dynamic features was constructed to support the development and training of the model's graph structure. The details of the static and dynamic features are presented in Table 1 and Table 2. Streamflow observations were obtained from the GRDC, an international repository that provides long-term streamflow records from gauging stations worldwide, while meteorological and land-surface variables were derived from the ERA5-Land reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Mischel et al., 2025; Molteni et al., 1996). ERA5-Land provides a temporally consistent depiction of terrestrial water and energy cycles since 1950. The original ERA5-Land fields are produced at approximately 9 km resolution on a reduced Gaussian grid, and the products distributed through the Climate Data Store are regridded to a regular $0.1^\circ \times 0.1^\circ$ latitude–longitude grid (Clelland et al., 2024; Muñoz-Sabater et al., 2021; Xu et al., 2024a). Because ERA5-Land variables are provided as gridded raster fields whereas gauging stations are point locations, we performed point-to-grid spatial matching for each station. Specifically, for each variable, the value was extracted from the $0.1^\circ \times 0.1^\circ$ cell whose grid-cell center is closest to the station coordinates, using a nearest-neighbour lookup. This mapping procedure was applied consistently to all static and dynamic ERA5-Land variables.

The selected static variables listed in Table 1 were used to describe streamflow generation and storage–release processes across the basin. Specifically, altitude serves as a proxy for temperature-dependent gradients and the propensity for snow accumulation and melt processes (Maier et al., 2022), while soil and vegetation characteristics reflect infiltration capacity, evapotranspiration potential, and subsurface buffering properties (Qiu et al., 2024). Lake coverage represents surface water storage and regulation effects (Cooley et al., 2021). Together, these static descriptors provide physically meaningful information to characterize spatial heterogeneity and catchment similarity among stations, complementing the temporal variability captured by the dynamic forcing variables. The selected dynamic variables listed in Table 2 were used to represent the dominant terms of the basin-scale water and energy balances that govern streamflow variability in the Danube River Basin. Specifically, total precipitation and convective precipitation characterize atmospheric water supply and event-scale inputs (Zhu et al., 2024). Total runoff, surface runoff, and subsurface runoff provide physically consistent land-surface estimates of flow partitioning and delayed contributions, thereby serving as proxies for catchment storage–release behavior (Lucas et al., 2025). Total evapotranspiration and potential evapotranspiration quantify realized water losses and atmospheric demand, respectively, and thus regulate seasonal and interannual runoff efficiency (Gu et al., 2021). Surface latent heat flux, together with net shortwave and net longwave radiation, captures key surface energy constraints that modulate evapotranspiration and indirectly influence snow–rain partitioning through the surface energy balance (Mekonnen et al., 2024). Collectively, these variables provide a coherent hydroclimatic representation of the primary drivers of streamflow. Finally, because several energy- and evaporation-related predictors are physically coupled, we explicitly quantified redundancy and multicollinearity using a predictor correlation matrix. The diagnostic results discussed in Sections 6.3 and 6.4.

For this study, 97 gauging stations within the Danube River Basin were selected from the GRDC dataset. All stations provide

Table 1
Static input variables grouped by hydrological meaning and their data sources.

Data	Data format	Data source
Location & topography		
Longitude	Point	GRDC
Latitude	Point	GRDC
Altitude	Grid	ERA5-Land
Soil properties		
Soil type	Grid	ERA5-Land
Vegetation characteristics		
High vegetation coverage	Grid	ERA5-Land
Low vegetation coverage	Grid	ERA5-Land
High vegetation type	Grid	ERA5-Land
Low vegetation type	Grid	ERA5-Land
Surface water		
Lake coverage	Grid	ERA5-Land

Table 2
Dynamic input variables grouped by hydrological meaning and their data sources.

Data	Data format	Data source
Precipitation forcing		
Total precipitation	Grid	ERA5-Land
Convective precipitation	Grid	ERA5-Land
Runoff components		
Total runoff	Grid	ERA5-Land
Surface runoff	Grid	ERA5-Land
Subsurface runoff	Grid	ERA5-Land
Evapotranspiration		
Total evapotranspiration	Grid	ERA5-Land
Potential evapotranspiration	Grid	ERA5-Land
Radiation & surface energy fluxes		
Surface latent heat flux	Grid	ERA5-Land
Surface net shortwave radiation	Grid	ERA5-Land
Surface net longwave radiation	Grid	ERA5-Land

complete monthly streamflow records from January 1960 to December 2020, enabling analysis of long-term hydrological variability across the basin. For the monthly-scale experiments, the corresponding meteorological and land-surface predictors were obtained from the ERA5-Land monthly product, in which variables are provided as monthly means. Note that the monthly- and daily-scale experiments cover different periods due to the availability and completeness of GRDC streamflow records after quality control. Therefore, cross-scale comparisons are interpreted as complementary evaluations rather than a strict scale-only comparison over identical time

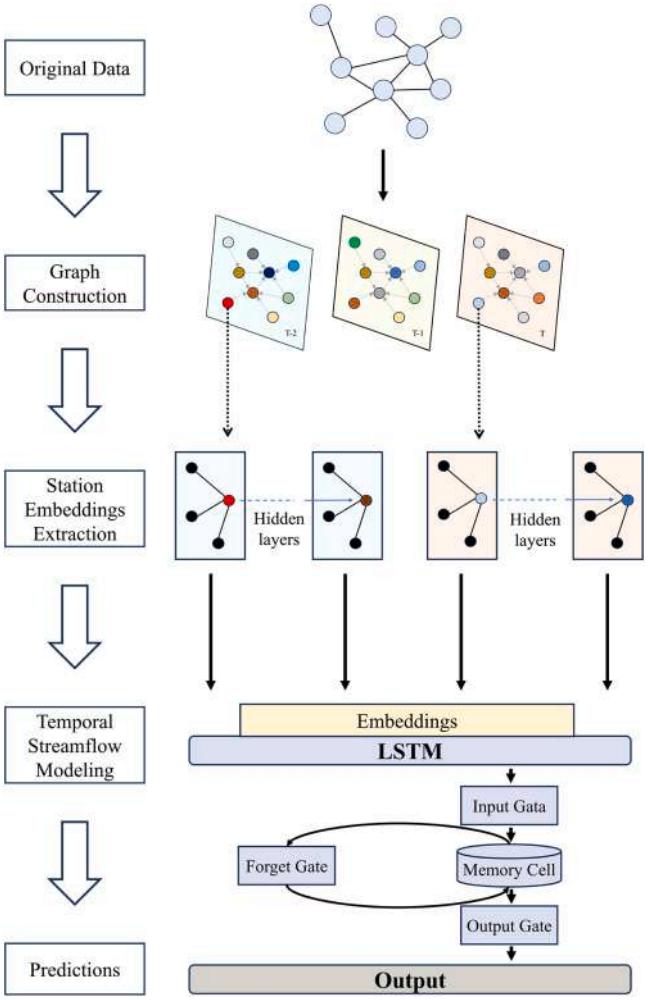

Fig. 2. GCN-LSTM Model.

windows.

4. Methods

4.1. GCN-LSTM

This study applies a deep learning framework that integrates GCN with LSTM networks for streamflow prediction in the Danube River Basin. The model considers both the spatial dependencies among stations and the temporal evolution of streamflow. It further incorporates a multi-head attention mechanism to improve the representation of complex hydrological processes and spatiotemporal modeling performance (Fig. 2). The overall architecture consists of three modules: a graph convolution module, a temporal modeling module, and a fusion prediction module.

4.1.1. GCN module

In river systems, spatial dependencies often exist among hydrological stations due to geographical proximity and dynamic hydrological similarity (Guo et al., 2021; Wang et al., 2021). To effectively represent these spatial characteristics, the stations were first modelled as a graph structure, in which each station corresponds to a node and the inter-station relationships are encoded by an adjacency matrix. Since geographic distance and station static attributes remain constant over time, the resulting graph is time-invariant and is denoted as A. The adjacency matrix was derived based on the geographical distance and the static attributes of the stations.

At each time step t, the input graph is defined as:

$$G_t = (X_t, \quad A) \tag{1}$$

where, $X_t \in \mathbb{R}^{N\times F}$ represents the input feature matrix of all stations at time t; $A \in \mathbb{R}^{N\times F}$ represents the adjacency matrix that characterizes the graph connectivity among stations; and N represents the number of stations.

The adjacency matrix is constructed by jointly considering geographic proximity and physiographic similarity derived from station static attributes. Specifically, for stations i and j, a distance-based affinity and an attribute-based affinity are defined as:

$$S_{ij}^{dist} = \exp\left(-\frac{d_{ij}}{\sigma_d}\right) \tag{2}$$

$$S_{ij}^{attr} = \exp\left(-\frac{\left\|z_i - z_j\right\|_2}{\sigma_z}\right) \tag{3}$$

where d_{ij} denotes the great-circle distance between stations i and j, and $\mathbf{z}_i$ represents the standardized static-attribute vector of station i. The final adjacency matrix is defined as:

$$A_{ij} = \begin{cases} S_{ij}^{dist} \bullet S_{ij}^{attr}, \& i \neq j \\ 0 \quad , \& i = j \end{cases} \tag{4}$$

Unlike traditional Convolutional Neural Networks (CNN), GCN is specifically designed to process graph-structured data and can effectively capture the complex spatial connections among nodes (Bhattacharya et al., 2026; Bhatti et al., 2023). In GCN, the feature update of a node depends on information from its neighbouring nodes, typically involving two key steps: information aggregation and feature transformation. The former collects and integrates information from neighbouring nodes, while the latter encodes the aggregated results through a learnable linear transformation, thereby enhancing the model's nonlinear representation capacity (Feng et al., 2023; Zhang et al., 2022). The GCN model is typically composed of multiple layers, each of which updates the node features. By stacking multiple layers, the model not only captures the local dependencies between a node and its immediate neighbours but also, through multi-hop propagation, indirectly models higher-order spatial correlations among distant nodes, thereby enhancing its ability to represent complex hydrological structures within the basin (Singh et al., 2025a; Zhu et al., 2021). The core of the graph convolution operation lies in the joint computation of the adjacency matrix and the node feature matrix. In this study, we employ three stacked GCN layers, which allow each node to aggregate information from up to 3-hop neighbourhoods. This setting helps capture higher-order spatial correlations within the basin while mitigating the risk of overly deep propagation that may occur in deeper GCN. The mathematical formulation of this operation is expressed as:

$$\widetilde{A} = A + I \tag{5}$$

$$\widetilde{D}_{ii} = \sum_j \widetilde{D}_{ii} \tag{6}$$

$$H_t^{(l+1)} = \sigma\left(\widetilde{D}^{-1/2}\widetilde{A}\widetilde{D}^{-1/2}H_t^{(l)}W^l\right) \tag{7}$$

where, I represents the identity matrix; $\widetilde{A}$ represents the adjacency matrix with self-connections; $\widetilde{D}$ represents the corresponding degree

matrix of $\widetilde{A}$; W^l represents the trainable weight matrix; σ represents the activation function; and H_t represents the node embeddings obtained after graph convolution.

The graph convolution operation allows the output of each node to be influenced not only by its own features but also by those of its neighbouring nodes, thereby achieving feature diffusion and aggregation within the spatial domain and capturing potential hydrological dependencies among stations (Ashraf et al., 2023; Sun et al., 2021). In addition, the output H_t of each node at time t can be regarded as the hydrological state representation of the corresponding station, reflecting its response to the input hydrological variables. This state representation is subsequently propagated through the temporal modeling module, serving as the key feature input for the streamflow forecasting task.

4.1.2. LSTM module

After obtaining the spatial embedding representations of each station, the model introduces an LSTM network to model the streamflow time series. LSTM is an improved type of recurrent neural network capable of effectively retaining long-term dependencies in time series, making it well suited for handling the seasonality and lag characteristics typical of river streamflow (Mehedi et al., 2022; Yu et al., 2024b). For each station, let the graph convolution outputs form a temporal sequence $\{H_{t-T+1}, H_{t-T+2}, \ldots, H_t\}$, which is then fed into the LSTM network. The core of the network is composed of the following gating mechanisms:

Forget gate:

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \tag{8}$$

Input gate:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{9}$$

$$\widetilde{C}_T = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{10}$$

Cell state update:

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C}_t \tag{11}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{12}$$

Output gate:

$$h_t = o_t \odot \tanh(C_t) \tag{13}$$

where, x_t represents the spatial embedding at the current time step; h_t represents the current hidden state; C_t represents the cell state; σ represents the activation function; W_f , W_i , W_o , W_c represent the trainable parameters; and $\odot$ represent the Hadamard product.

To further capture cross-station dependencies beyond fixed graph message passing, we introduce a multi-head self-attention module after the final layer. Unlike temporal attention that reweights different time steps, our attention operates across stations, treating each station as a token.

Given an input historical window, the stacked encoder produces station-wise spatiotemporal representations. We take the final station embeddings from the encoder output and denote them as:

$$H \in \mathbb{R}^{N \times d} \tag{14}$$

where N is the number of stations and d is the embedding dimension. We then apply global self-attention across the station dimension once per sample:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{15}$$

where $Q = HW^Q$, $K = HW^K$, and $V = HW^V$ are linear projections of the station embeddings, and $d_k$is the per-head feature dimension. The resulting attended station representations are finally mapped to streamflow predictions through an MLP.

4.1.3. Fusion prediction module

The model fusion module integrates the spatial and temporal features described above to generate the target streamflow predictions. For each station, the final predicted value is obtained as follows:

$$\widehat{y}_{t+1}^{(i)} = MLP\left(Attn\left(\left\{h_{t-T+1}^{(i)}, \ldots, h_t^{(i)}\right\}\right)\right) \tag{16}$$

where, $h_t^{(i)}$ represents the hidden state of station i at time step t; $Attn(\cdot)$ represents the multi-head attention fusion; and MLP refers to the multi-layer perceptron output layer.

The loss function adopted in this study is the commonly used Mean Square Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(y_{t+1}^{(i)} - \hat{y}_{t+1}^{(i)}\right)^2 \tag{17}$$

where, $y_{t+1}^{(i)}$ represents the observed streamflow, and $\hat{y}_{t+1}^{(i)}$ represents the predicted streamflow.

The model captures spatial dependencies through GCN, temporal evolution through LSTM, and enhances key feature representation with a multi-head attention mechanism, forming an efficient streamflow forecasting framework that jointly learns spatiotemporal dependencies. This framework is well suited for river simulation tasks under complex topological structures and multi-source data scenarios (Wang and Zhu, 2024).

4.2. Data preprocessing

We downloaded all GRDC gauging stations located within the Danube River Basin for both monthly and daily streamflow products and applied a unified preprocessing workflow to ensure record completeness and temporal consistency. Station screening was conducted separately for the monthly- and daily-scale experiments. For the monthly-scale setting, we retained only stations covering January 1960 to December 2020, allowing at most three years of missing months over the full period. For the daily-scale setting, we retained only stations covering 2000–2014, allowing at most six months of missing days in total after quality control. To ensure comparability of station-based analyses across temporal resolutions, the final station set was obtained by taking the intersection of stations that satisfied both criteria.

For stations with small data gaps that remained within the above completeness thresholds, missing streamflow values were filled using time-wise linear interpolation. To avoid introducing artificial variability, interpolation was applied only to short gaps, whereas longer discontinuities were not reconstructed and were effectively controlled by the completeness screening. After gap handling, streamflow and ERA5-Land data were temporally aligned on the same calendar. Monthly experiments used monthly streamflow together with monthly-mean ERA5-Land data, while daily experiments used daily streamflow together with daily ERA5-Land data. This preprocessing ensures that subsequent model training and evaluation are conducted on consistent, quality-controlled time series.

4.3. Model setting

To systematically evaluate the performance of the GCN-LSTM model in streamflow prediction, a series of comparative experiments were designed, covering several representative time-series modeling approaches. The baseline models include Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Bidirectional LSTM (biLSTM), and the proposed GCN-LSTM model. All models were trained with the same hyperparameter settings to ensure a fair comparison (Table 3). Specifically, supervised samples were constructed using a sliding-window strategy with an input sequence length of three time steps and a one-step-ahead prediction horizon. The experimental data were split into training and testing subsets in an 8:2 ratio. During preprocessing, missing values were filled with 0. For each station and each variable channel, Min–Max normalization was applied using pre-computed min/max statistics from the training subset to avoid information leakage; channels with extremely small ranges were set to zero for numerical stability. Model outputs were de-normalized before evaluation. All models were optimized using Adam. To mitigate overfitting, Dropout and L2 weight decay were applied, and early stopping was adopted by monitoring the validation MSE, with the best-performing checkpoint used for final testing.

4.4. Performance assessment

To quantitatively evaluate the predictive accuracy of the model in streamflow simulation, three commonly used error metrics were employed: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). These metrics measure the degree of discrepancy between the model outputs and the observed values. MAE, RMSE, and R^2 were first computed independently for each gauging station using the corresponding observed and simulated time series, and the final values were then summarized by averaging across all stations. Their formulations are given as follows:

Table 3
Experimental configuration and shared hyperparameter settings.

Parameters	Setting
Data split	Train: Test = 8:2
Input sequence length	3
Prediction horizon	1
Optimizer	Adam
Learning rate	1×10^{-2}
Batch size	32
Epochs	60
Regularization	Dropout p = 0.2; L2 weight decay = 1×10^{-4}
Early stopping	Monitor validation MSE; patience = 15

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i^{pred} - y_i\right| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{pred} - y_i\right)^2} \tag{19}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i^{pred}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{20}$$

where, n represents the total length of the time series; y_i represents the observed streamflow at time step i; y_i^{pred} represents the corresponding model-predicted streamflow; and $\bar{y}$ represents the mean of all observed streamflow values. MAE and RMSE measure the average and squared deviations between simulated and observed streamflow, where smaller values indicate higher accuracy (Bringeland and Fotopoulos, 2024; Hodson, 2022; Wu et al., 2024). R^2 quantifies the proportion of variance explained by the model, with values closer to 1 representing better agreement between simulation and observation (Duc and Sawada, 2023; Melsen et al., 2025; Williams, 2025).

To further evaluate the overall capability of the model in capturing streamflow dynamics, the Kling–Gupta Efficiency (KGE) was introduced as a composite performance metric (Gupta et al., 2009). The KGE is calculated as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{21}$$

where, r represents the Pearson correlation coefficient between the predicted and observed values; β represents the ratio of the mean of the predictions to the mean of the observations; and γ is the ratio of the coefficient of variation of the predictions to that of the observations. Here r measures the linear relationship between observed and simulated data, β measures the ratio of the mean of the simulated values to the mean of the observed values, and γ measures the ratio of the standard deviations of the simulated and observed data. As recommended by Knoben et al. (2019), a KGE score of −0.4 is used as a benchmark, which is the score of the observational mean; i.e., KGE > -0.4 represents acceptable performance.

4.5. Station sparsity experiment: K-means-stratified subsampling

We conducted a station-thinning experiment to assess model robustness under different gauging-network densities while maintaining basin-wide spatial representativeness. Let the full station set be $\mathcal{S} = \{1, \ldots, N\}$ with $N = 97$, and each station i has coordinates $\mathbf{p}_i = (\lambda_i, \phi_i)$. We first partitioned the stations into $K = 10$ spatial strata using K-means clustering on $\mathbf{p}_i$, by minimizing the within-cluster sum of squares:

$$\min_{\{\mu_k\}_{k=1}^{K}} \sum_{i=1}^{N} \|\boldsymbol{p}_i - \mu_{c(i)}\|_2^2 \tag{22}$$

where $c(i) \in \{1, \ldots, K\}$ denotes the cluster label of station i, and $\boldsymbol{\mu}_k$ is the centroid of cluster k. Denote the station subset in cluster k as $\mathcal{S}_k = \{i \in \mathcal{S} : c(i) = k\}$ with size $n_k = |\mathcal{S}_k|$.

For a target number of retained stations $M \in \{97, 87, 77, 67, 57, 47, 37, 27, 17\}$, we performed stratified subsampling by allocating a cluster-wise quota proportional to cluster size:

$$m_k = \max\left(1, round\left(M \bullet \frac{n_k}{N}\right)\right), \quad \sum_{k=1}^{K} m_k \approx M \tag{23}$$

To ensure $\sum_{k=1}^{K} m_k = M$ exactly, we adjusted the quotas by adding/removing stations from clusters with the largest rounding residuals until the target M was met. We then randomly selected m_k stations within each cluster to form the retained set:

$$\mathcal{S}^{(M)} = \bigcup_{k=1}^{K} \mathcal{S}_k^{(M)}, \ \mathcal{S}_k^{(M)} \subseteq \mathcal{S}_k, \quad \left|\mathcal{S}_k^{(M)}\right| = m_k \tag{24}$$

This K-means-stratified design avoids overly clustered selections and keeps the retained stations approximately evenly distributed across the basin, so that performance changes primarily reflect reduced station density rather than biased spatial coverage.

5. Results

5.1. Model comparative analysis

To systematically evaluate the performance of different models in streamflow prediction, comparative experiments were conducted

under consistent input–output formats, training epochs, optimizer settings, and data partitioning schemes (Table 4). Overall, the GCN-LSTM model demonstrated superior performance across all three evaluation metrics. Its R^2 value of 0.9525 indicates a strong correlation between predicted and observed streamflow values, suggesting the model effectively captures streamflow dynamics. In terms of error metrics, GCN-LSTM achieved the lowest MAE (114.4235 m^3/s) and RMSE (211.3520 m^3/s) among all tested models.

To make the performance gain more transparent, we further summarize the relative improvements of GCN-LSTM over all baselines in Table S1. Overall, GCN-LSTM achieves consistently better accuracy across metrics, yielding the highest R^2 (0.9525) and reducing MAE and RMSE by 5.9–14.6 % and 3.9–9 %, respectively, compared with the five baseline models. The largest gains are observed against the single-site temporal RNN baseline, where MAE and RMSE decrease by 14.6 % and 9 %, indicating that incorporating spatial dependency learning provides substantial benefit. Improvements are also evident relative to other recurrent baselines, suggesting that the advantage is systematic rather than model-specific. These results support that jointly modeling spatial interactions and temporal dynamics enhances both overall fit and robustness to large errors in streamflow prediction.

To further evaluate the spatial applicability of different models, Fig. 3 shows the distribution of KGE values for different hydrological models at stations in the Danube River Basin. Overall, higher KGE values tend to cluster in the central to middle–lower part of the basin, whereas comparatively lower KGE values appear more frequently toward the far eastern end and several peripheral tributaries. Hydrologically, the middle and middle–lower reaches are typically associated with larger drainage areas and longer routing pathways, so streamflow reflects an integrated response to basin-scale precipitation and evapotranspiration rather than highly localized signals. The combined effects of catchment storage and river network routing tend to damp short-term fluctuations and reduce event-scale noise. In contrast, upstream/headwater areas influenced by Alpine conditions are more strongly affected by snow accumulation and melt, rain–snow transitions, and rapid runoff generation on steep terrain, which introduce nonlinearities and timing shifts that can weaken model–observation agreement at individual stations. Moreover, stations in peripheral or downstream subregions may be additionally affected by reservoir regulation and water withdrawals, which can modify the natural storage–release behavior and locally decouple discharge from meteorological inputs, leading to scattered pockets of reduced KGE.

Among all models, GCN-LSTM (Fig. 3f) exhibits the most spatially consistent performance and achieves higher KGE values at the majority of stations, particularly within the central and middle–lower parts of the basin. Its advantage is reflected in the fact that adjacent stations tend to show similarly high KGE, and the occurrence of isolated low-KGE stations is reduced. This behavior is hydrologically reasonable because streamflow at nearby or hydrologically similar catchments is often driven by comparable large-scale forcing and shares correlated runoff-generation and routing characteristics. By incorporating information from related stations, GCN-LSTM can better represent basin-scale spatial coherence and thus improves robustness across the station network. GRU (Fig. 3c) and MLP (Fig. 3a) also yield relatively high KGE values at many stations, but their spatial patterns are more scattered, indicating stronger sensitivity to site-specific conditions when spatial relationships are not explicitly considered. RNN (Fig. 3b) and LSTM (Fig. 3d) show larger spatial variability and weaker clustering of high KGE, suggesting that sequence-based models alone may not sufficiently utilize the spatial dependence among stations. Finally, BiLSTM (Fig. 3e) yields lower KGE values at most sites, implying that bidirectional temporal modeling does not provide additional benefits under the current input configuration and evaluation setting.

5.2. *Influence of temporal resolution on model prediction performance*

To assess the influence of input data resolution on model prediction accuracy, Fig. 4 presents streamflow simulations and residuals of the GCN-LSTM model, trained with both daily and monthly input data at stations 6246195 and 6243030 in the Danube River Basin. In the figure, the blue solid line represents observed streamflow, the orange dashed line shows model predictions based on daily data, and the green circles indicate predictions derived from monthly input data. The black curve denotes the bias from daily input data, offering an intuitive measure of model fitting accuracy and error fluctuations at different temporal resolutions.

The predictive performance of the GCN-LSTM model exhibits clear differences under varying temporal resolutions. Results from the daily-scale experiment show that the model is significantly more effective at reproducing streamflow dynamics, particularly in capturing the timing and magnitude of flood peaks. The daily input model simulates flood peaks with high accuracy, aligning closely with observed values, which is crucial for hydrological predictions during extreme events. Furthermore, the model maintains stable streamflow during low-flow periods, ensuring reliable predictions across different hydrological conditions. The residuals remain close to zero with minor fluctuations, suggesting that high temporal resolution data—such as daily input—enhances the model's ability to capture short-term dynamics and extreme events like flood peaks. This ability to better represent peak streamflow variations is particularly important in hydrological modeling, where accurately simulating the timing and magnitude of such peaks is critical for flood risk management.

Table 4
The predictive performance comparison of six models.

Models	R^2	MAE (m^3/s)	RMSE (m^3/s)
MLP	0.9471 [0.9431, 0.9515]	125.5113 [69.0150, 190.4830]	219.9383 [122.6216, 332.2067]
RNN	0.9432 [0.9387, 0.9479]	134.0161 [73.4986, 203.2690]	232.4857 [129.7386, 349.3587]
GRU	0.9457 [0.9414, 0.9502]	126.6232 [69.8727, 191.8549]	222.7487 [124.6342, 334.9478]
LSTM	0.9452 [0.9409, 0.9497]	130.6214 [72.9943, 198.6368]	223.4053 [124.5074, 336.8908]
BiLSTM	0.9486 [0.9449, 0.9526]	121.6495 [67.2265, 184.7421]	220.5330 [123.7151, 331.4096]
GCN-LSTM	0.9525 [0.9488, 0.9567]	114.4235 [64.0661, 178.0254]	211.3520 [117.5485, 319.1913]

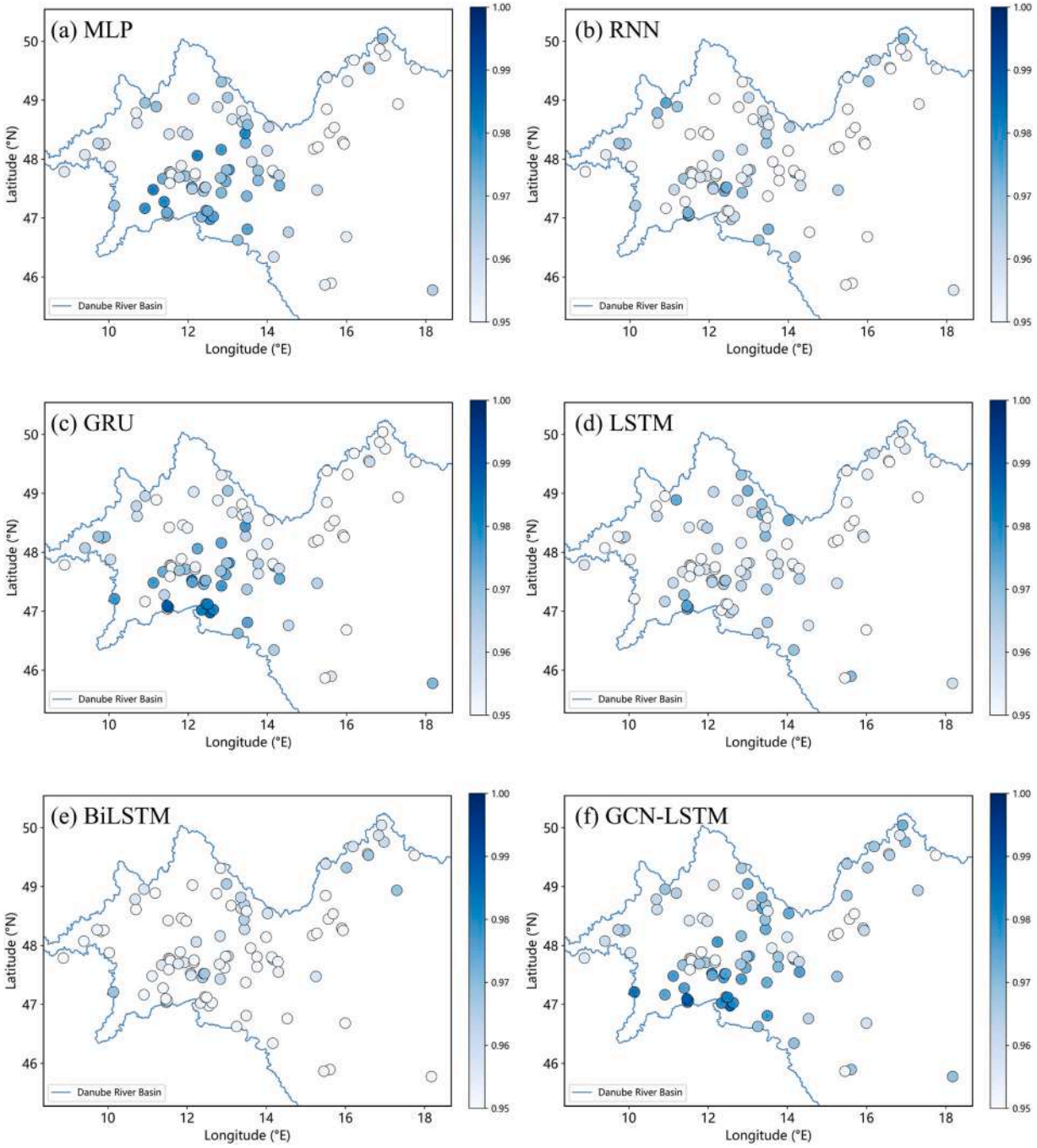

Fig. 3. Spatial distribution of KGE values for different hydrological models (MLP, RNN, GRU, LSTM, BiLSTM, and GCN-LSTM) at stations in the Danube River Basin. The color scale represents the KGE values, with darker shades indicating better performance.

In contrast, the model trained with monthly input data reproduces interannual and seasonal variation trends well but struggles to capture flood peak magnitudes and timings accurately. During flood events, the model tends to underestimate or lag behind the actual peak values, with residual deviations increasing substantially. This delay in peak detection is particularly problematic in scenarios where rapid hydrological responses are necessary for accurate forecasting. As a result, the monthly-scale model is less capable of representing rapid hydrological processes, leading to larger errors and reduced reliability in predictions during extreme events like floods.

To further analyze how temporal resolution affects spatial model performance, Fig. 5 compares the station-wise KGE distribution across the Danube River Basin under two input resolutions. While Table 5 provides a station-wise KGE summary for daily and monthly

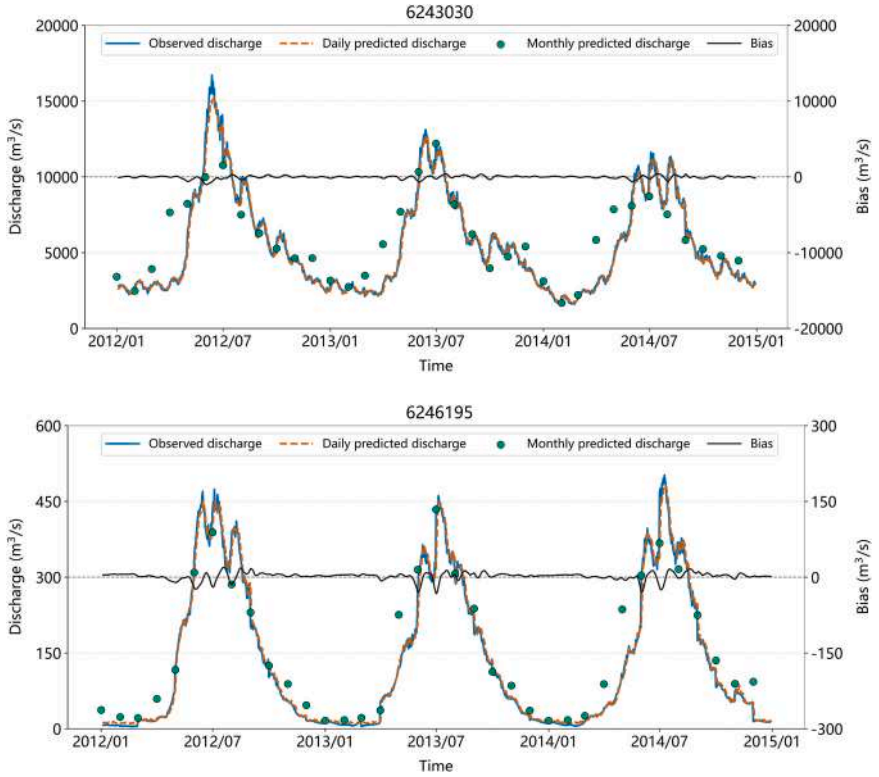

Fig. 4. Flow simulation and bias at station 6246195 in the Danube River Basin using the GCN-LSTM model with daily and monthly input data. Blue and orange lines represent observed and daily predicted streamflow, respectively, while green circles indicate monthly predicted streamflow, and the black line represents the bias based on daily input data.

experiments. Under monthly inputs (Fig. 5a), the model achieves a mean KGE of 0.4426 with a large spread across stations (std = 0.1898). The spatial pattern indicates that the monthly-input setting leads to a highly heterogeneous performance: only a limited number of stations achieve high KGE values, and these are scattered without forming a spatially coherent cluster. In many parts of the basin, KGE values remain low, suggesting that monthly aggregation smooths or removes key hydrological signals required for accurate streamflow prediction. This limitation is particularly relevant for basins such as the Danube, where streamflow variability is shaped not only by slowly varying seasonal water balance but also by sub-monthly processes, including event-driven rainfall–runoff responses, rapid snowmelt pulses in headwater regions, and short-term storage–release dynamics. When inputs are aggregated to monthly means, the timing and magnitude of high-flow events are largely damped, and the model receives weakened information on intra-month variability, which can translate into pronounced spatial inconsistencies and localized areas with low KGE values.

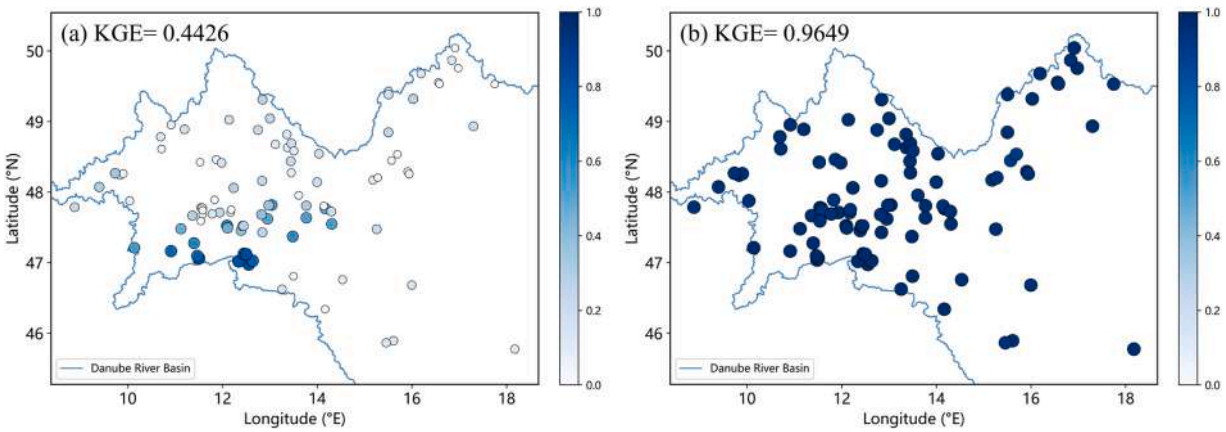

Fig. 5. Spatial distribution of KGE values from GCN-LSTM models trained on different time scales in the Danube River Basin. (a) Monthly input; (b) Daily input. Point size denotes the magnitude of KGE, and color depth (light to dark blue) indicates increasing values.

Table 5
Station-wise KGE summary for daily and monthly experiments, including paired differences and statistical significance.

Experiment	Mean KGE	Std KGE	95 % CI	Wilcoxon W	p-value
Daily	0.9649	0.0136	[0.9621, 0.9675]	—	—
Monthly	0.4426	0.1898	[0.4056, 0.4808]	—	—
Daily − Monthly	0.5223	—	[0.4860, 0.5579]	4753.000	6.090e−18

In contrast, daily inputs lead to a substantial and spatially consistent improvement, with the mean KGE increasing to 0.9649 and the inter-station variability collapsing to a very small range (std = 0.0136). High KGE values are widespread and more uniformly distributed across stations throughout the basin, indicating that the model captures streamflow dynamics consistently across different sub-regions. Importantly, Table 5 confirms that the daily-vs-monthly difference is not only large but also statistically robust: the paired mean gain is ΔKGE = 0.5223 with a bootstrap 95 % CI of [0.4860, 0.5579], and a one-sided Wilcoxon signed-rank test shows a highly significant improvement (p = 6.090e−18). This result suggests that daily inputs provide richer hydrological information, allowing the model to represent both rapid event-scale responses and slower seasonal variations simultaneously. The daily forcing better preserves precipitation intermittency, evapotranspiration demand fluctuations, and runoff generation dynamics, which are essential for reproducing hydrograph variability and improving water-balance consistency at individual stations. Overall, increasing temporal resolution markedly enhances spatial applicability in the Danube River Basin by preserving critical hydrological variability, thereby reducing isolated low-KGE stations and yielding more stable basin-wide performance.

To evaluate model performance under extreme-flow conditions in the Danube River Basin, Fig. 6 provides a consolidated assessment based on the POT framework, combining return-level bias, peak timing error, and peak magnitude error. The Danube Basin covers a wide range of hydro-climatic settings, from Alpine headwaters where snow processes and orographic precipitation are important, to lowland reaches where floodplain storage and human regulation can influence flood wave propagation. This heterogeneity offers a rigorous test of whether the model can reproduce both the magnitude and timing of flood peaks under contrasting hydrological controls.

Fig. 6a summarizes the cross-station distribution of relative biases in POT-based return levels (RL2, RL5, and RL10). Negative

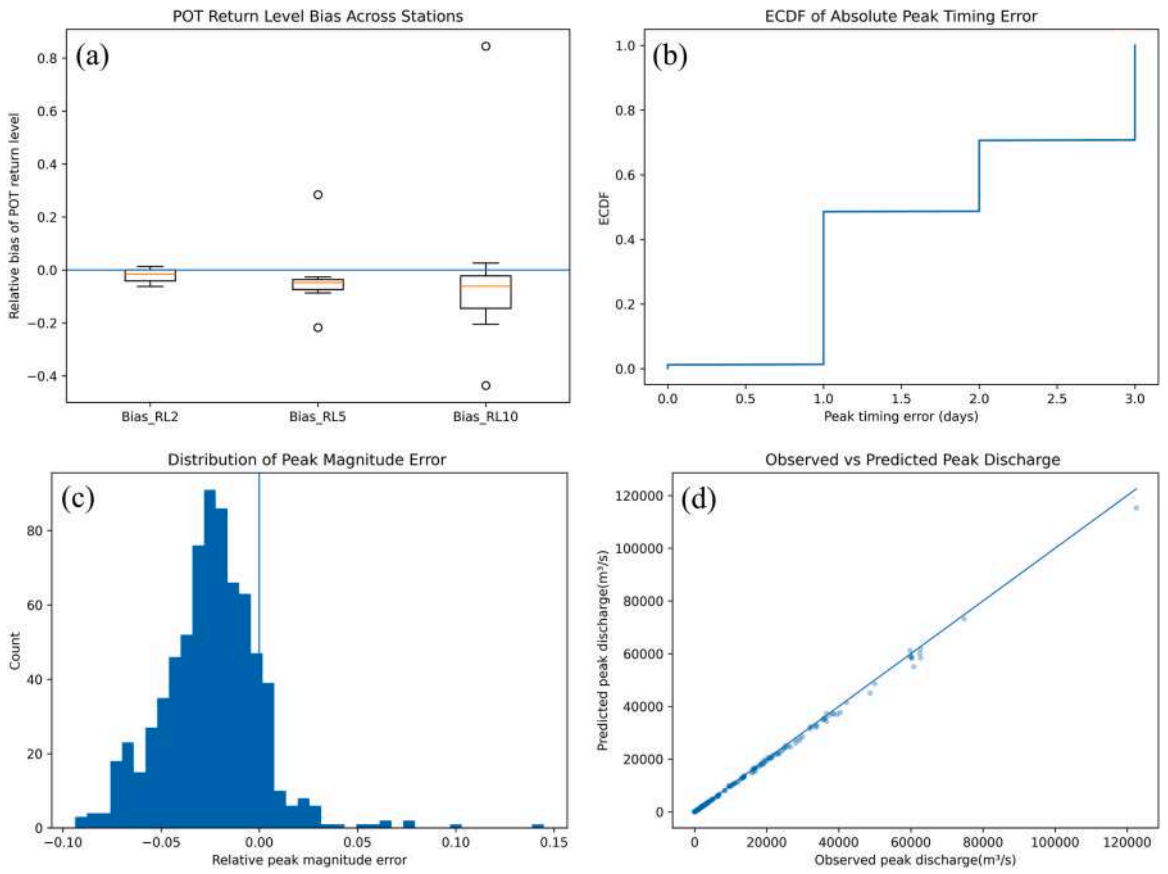

Fig. 6. Peak-oriented evaluation of flood magnitude and timing under the POT framework: (a) cross-station biases of POT-based return levels, (b) ECDF of absolute peak timing error, (c) distribution of relative peak magnitude error, and (d) observed versus predicted peak streamflow.

median biases are found for all return periods, indicating that peak underestimation is a basin-wide feature rather than being confined to a small number of stations. In addition, the bias range increases with return period, with RL10 showing the largest dispersion and more pronounced outliers. This is consistent with hydrological expectations for the Danube: floods with shorter return periods are more strongly governed by synoptic-scale precipitation and seasonal melt signals that are comparatively well represented at the daily scale, whereas rarer floods are more sensitive to event sequencing and state dependence, floodplain storage effects, and reservoir operation. These factors introduce stronger nonlinearity and increase uncertainty, making high-return-period floods more difficult to simulate accurately.

Fig. 6b–d further evaluate event-scale performance in timing and magnitude. The ECDF of absolute peak timing error (Fig. 6b) indicates that many peaks are simulated within a relatively small time offset, suggesting that the model captures the occurrence of flood events reasonably well at the basin scale. However, the tail of larger timing errors implies that a subset of events remains difficult, particularly for floods with rapid rise and short response times that are often associated with intense rainfall in tributaries. The distribution of relative peak magnitude error (Fig. 6c) is centered below zero and shows a clear negative skew, confirming that underestimation dominates peak errors, especially for larger events. This tendency is also evident in the observed–predicted peak comparison (Fig. 6d): while most events lie close to the 1:1 line, underestimation becomes more apparent for the largest peaks. Overall, these results suggest that the daily-scale model reproduces the timing and general scaling of Danube flood peaks, but its performance deteriorates toward the most extreme events, highlighting the continuing difficulty of simulating rare floods in a large basin where both natural storage processes and human regulation play important roles.

5.3. *Influence of station numbers on model prediction performance*

To investigate the impact of station density on model performance, we conducted a station-thinning experiment. Fig. 7(a–i) shows the spatial configuration of gauging stations under progressively sparser networks, where the number of stations is reduced from 97 to 17 in steps of 10 (97, 87, 77, 67, 57, 47, 37, 27, and 17). With increasing thinning, the station network shifts from relatively dense basin-wide coverage to a more discontinuous and uneven distribution. Under mild to moderate thinning (97–47 stations), the

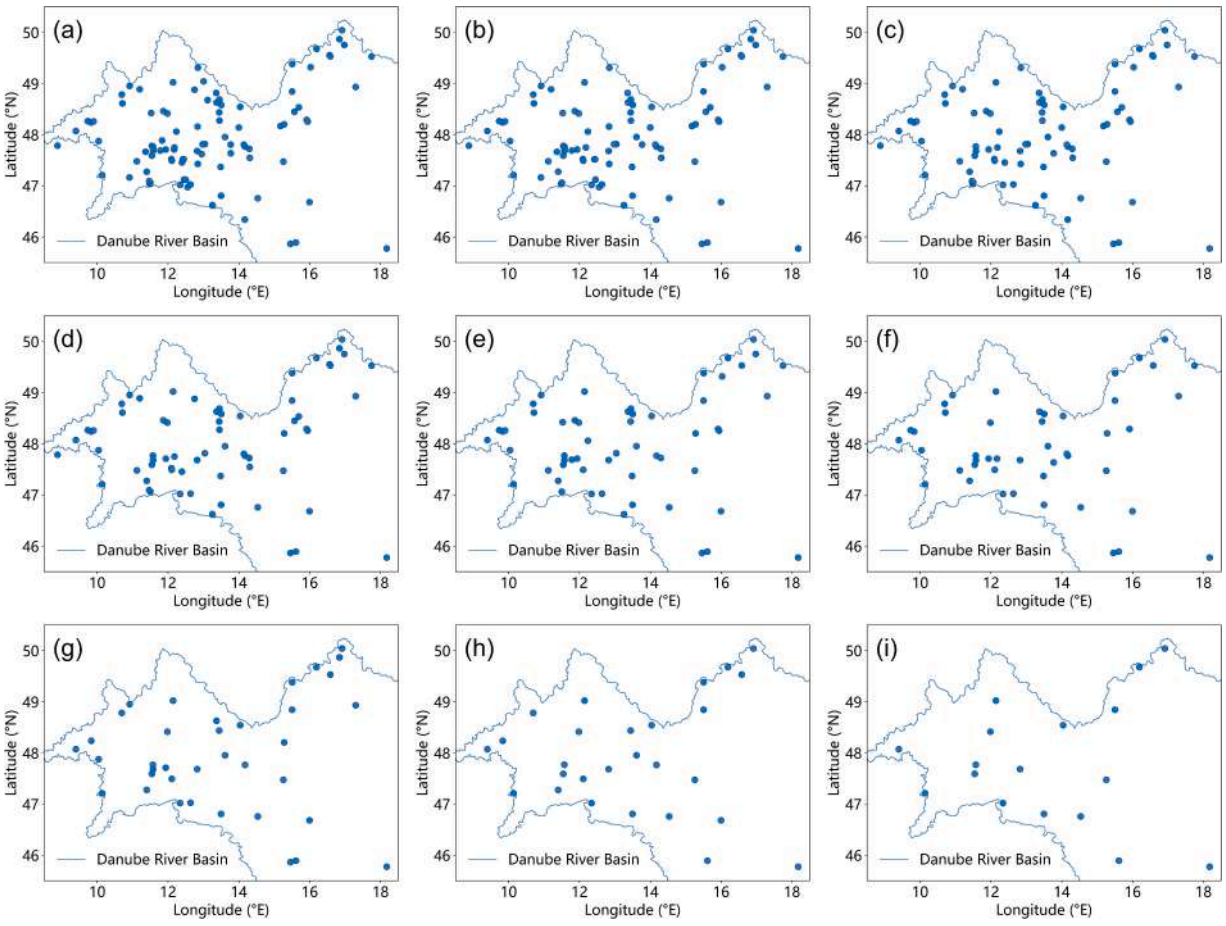

Fig. 7. Spatial Distribution of Hydrological Stations under Progressive Reduction in the Danube River Basin. (a)–(i) illustrate the station distribution after each reduction step of 10 stations.

remaining stations still cover most major parts of the Danube River Basin and retain representation of key upstream, middle, and lower reaches, thereby providing sufficient spatial information to constrain basin-scale variability. Under stronger thinning (≤37 stations), spatial gaps become evident and some sub-regions are represented by only a few stations or none, leading to an increasingly imbalanced sampling of the basin. This is hydrologically important because the Danube exhibits strong regional contrasts in hydroclimate and physiographic controls; insufficient sampling of certain hydroclimatic zones reduces the information available to constrain cross-station dependencies and limits spatial generalization.

The corresponding performance variations are summarized in Fig. 8, where MAE and RMSE are plotted against the number of stations, with the R^2 shown on the secondary axis. Overall, performance remains remarkably stable across the full range of station densities, indicating that the proposed framework does not require an extremely dense gauging network to achieve consistently high predictive skill. In particular, R^2 stays close to unity throughout the thinning experiment and exhibits only minor oscillations as stations are thinned or added, while MAE and RMSE vary gradually rather than abruptly. Although MAE/RMSE show a gentle tendency to decrease with increasing station count, the magnitude of improvement is modest, and no sharp "turning point" is observed—collectively underscoring the robustness of the learned spatiotemporal representations to changes in network density. A shallow, non-monotonic fluctuation in R^2 around intermediate station counts is visible, but its amplitude is small relative to the overall level of fit, suggesting a secondary compositional effect rather than a strong optimum.

This weak sensitivity can be explained by the balance between added information and redundancy in the station set. When the network is relatively dense, many stations are located within the same tributaries or neighboring sub-regions, and their streamflow series are often highly correlated because they share similar meteorological forcing and catchment characteristics. Under such conditions, adding more nearby stations provides limited new information for learning basin-scale patterns, and the marginal benefit for R^2 is therefore small. Moreover, dense networks may include stations with similar hydrological behavior, so the additional data mainly reinforce already well-represented regimes rather than improving coverage of distinct hydroclimatic zones. In contrast, under strong thinning, spatial gaps increase and some regimes may be represented by only a few stations. This reduction in representativeness can lead to a modest increase in MAE and RMSE and a slight decrease in R^2, consistent with weaker constraints on spatial transfer across under-sampled parts of the basin. Overall, the key result is that the performance metrics remain largely stable under station thinning, supporting the applicability of the approach in basins where gauging networks are heterogeneous or relatively sparse.

To assess the spatial generalization ability of the GCN-LSTM model, we conducted a cross-station evaluation in which the model was trained using streamflow observations from 37 gauged stations and subsequently applied to 60 unseen stations within the Danube River Basin. It should be noted that, for each unseen station, the model was required to generate a continuous daily streamflow time series over the full evaluation period (2000–2014) without using any station-specific streamflow observations for calibration, fine-tuning, or adjustment. This experimental setup explicitly tests the model's capability to transfer learned spatiotemporal representations to locations without direct observational support, which is a key requirement for practical applications in ungauged or sparsely gauged regions. The aggregated performance statistics reveal a pronounced discrepancy between training and unseen stations, as summarized in Table 6. At the training stations, the GCN-LSTM achieves an exceptionally high KGE of 0.96, together with relatively small errors (MAE = 114.4235 m^3/s; RMSE = 211.35 m^3/s), indicating that the model can accurately reproduce streamflow variability when local information is explicitly available during training. In contrast, performance degrades substantially when the model is applied to the 60 unseen stations: KGE decreases to 0.24, while MAE and RMSE increase to 663.7236 m^3/s and 888.25 m^3/s, respectively. This consistent decline across all metrics suggests that, although the model exhibits strong fitting capability at trained

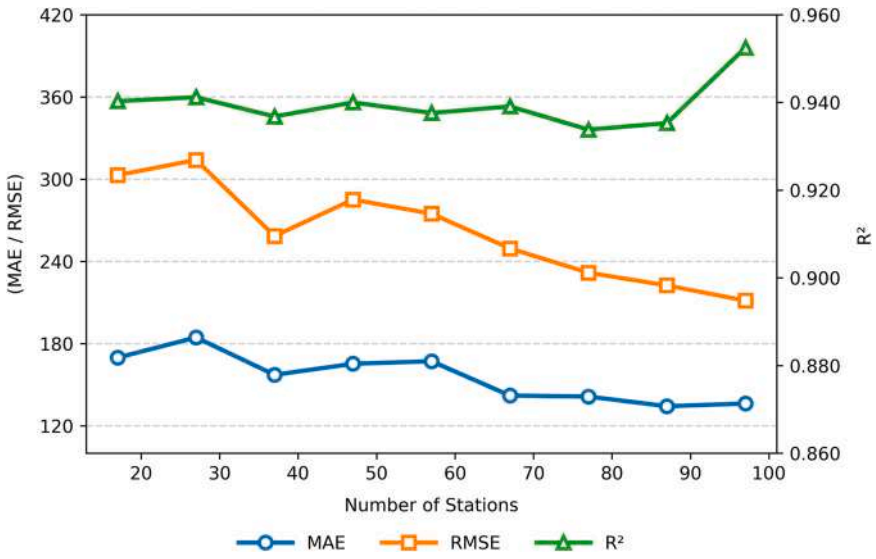

Fig. 8. Variations in Predictive Performance of the GCN-LSTM Model under Different Numbers of Stations. Blue circles represent MAE, orange squares represent RMSE, and blue triangles represent R^2.

locations, its predictive skill is only partially transferable to new stations with distinct hydrological conditions.

To place this performance gap in a spatial context, Fig. 9 shows the distribution of training and testing stations across the Danube River Basin, together with the error distributions at the unseen stations. The spatial layout indicates that both training and testing stations are spread across the basin and cover the main river corridor and major tributary regions, rather than being concentrated within a single sub-region. This reduces the likelihood that the performance degradation is primarily caused by a simple spatial sampling bias. However, the boxplots for unseen stations show a wide range of MAE and RMSE values, even when presented on a logarithmic scale. This dispersion indicates that the loss of performance at unseen sites is not uniform. While a subset of unseen stations maintains reasonable predictive accuracy, others exhibit much larger errors. Such heterogeneity suggests that generalization depends not only on whether a station is excluded from training, but also on how well the hydrological conditions at that station are represented by the training set. In particular, stations associated with distinct hydroclimatic regimes, stronger human influence, or more complex runoff-generation and routing behavior may be harder to predict when similar conditions are under-represented during training.

Motivated by this heterogeneity, we further examined whether the predictive performance at unseen stations is related to their spatial proximity to the training network. Fig. 10 shows the relationships between performance metrics and the distance from each unseen station to its nearest training station. Across the three panels, a consistent tendency is evident: KGE decreases as distance increases, whereas MAE and RMSE increase with distance. This indicates that spatial transferability is partly distance-dependent, and unseen stations located closer to training stations generally achieve better performance, likely because nearby catchments tend to share similar meteorological forcing and hydrological response characteristics, which are more readily transferred by the model. Nevertheless, the considerable scatter around the fitted trends indicates that geographic distance alone cannot fully explain the observed variability in performance, highlighting the influence of additional factors such as basin heterogeneity, anthropogenic regulation, and local catchment characteristics. Overall, Table 6 together with Figs. 9–10 demonstrate that while the GCN-LSTM performs exceptionally well at training stations, its application to unseen stations is associated with a structured and non-uniform degradation in skill.

6. Discussions

6.1. The impact of input data resolution on model performance

Our analysis demonstrates a clear difference in model performance when trained with daily versus monthly input data, highlighting the importance of temporal resolution in streamflow prediction. Specifically, the GCN-LSTM model, when trained with daily input data, shows a marked improvement in capturing both flood peaks and low-flow periods. This is consistent with previous studies that have emphasized the role of high temporal resolution in improving the representation of extreme events and short-term dynamics. The residuals of the daily-scale model remain close to zero with minor fluctuations, indicating its ability to simulate rapid changes in streamflow, which is crucial for flood forecasting (Liu et al., 2024; Lu et al., 2021). In contrast, the model trained with monthly data, while able to capture broader seasonal and interannual trends, exhibits significant limitations in simulating flood peaks. The residuals increase substantially during high-flow events, reflecting a lag in the model's response to rapid hydrological changes. This aligns with findings from previous research, which suggested that models using lower temporal resolution data tend to underestimate peak streamflow and struggle to represent rapid hydrological processes (Gacu et al., 2025; Krajewski et al., 2021). The monthly-scale model performed worse during flood events, underscoring the challenge of simulating peak streamflow accurately when temporal resolution is insufficient (Ren et al., 2023; Yang et al., 2025).

Importantly, several sources of uncertainty and potential confounding factors should be acknowledged when interpreting the magnitude of the daily–monthly performance gap. First, monthly temporal resolution inherently smooths intra-month variability, which can suppress the intensity and timing of short-lived rainfall–runoff responses—processes that are especially relevant in the Danube Basin, where hydroclimatic conditions vary strongly from headwaters to lower reaches (Dunkerley, 2019; Westra et al., 2014). Second, observed streamflow may be influenced by regulation and short-term human interventions (e.g., reservoir operation and flow management), which can alter sub-month dynamics and further amplify the apparent advantage of daily inputs when such variability is not well reflected by monthly forcing (Wu et al., 2019; Yu et al., 2023). These factors do not change the qualitative conclusion that daily inputs provide richer hydrological information for learning streamflow dynamics, but they help contextualize why improvements are most pronounced for peak-flow reproduction. These results underscore the need for high-resolution data, particularly for applications that require precise prediction of extreme events like floods. While daily data significantly enhances model accuracy, it also increases computational demands, particularly for large-scale models (Biazar et al., 2025; Gacu et al., 2025; Marshall et al., 2025). For long-term hydrological analyses focused on seasonal to interannual variability, monthly configurations may offer a more economical baseline, whereas daily configurations are preferable when the priority is flood-relevant dynamics or short-term forecasting. Future work should therefore explore efficient multi-scale strategies that retain long-term consistency while selectively leveraging high-frequency information during hydrologically active periods, thereby balancing accuracy and computational feasibility for

Table 6
Comparison of GCN-LSTM prediction performance between training stations and unseen stations.

Models	KGE	MAE (m^3 /s)	RMSE (m^3 /s)
GCN-LSTM (training stations)	0.96 [0.96, 0.96]	114.42 [64.06, 178.02]	211.35 [117.54, 319.19]
GCN-LSTM (unseen stations)	0.24 [0.19, 0.28]	663.72 [328.83, 1120.23]	888.25 [423.50, 1515.60]

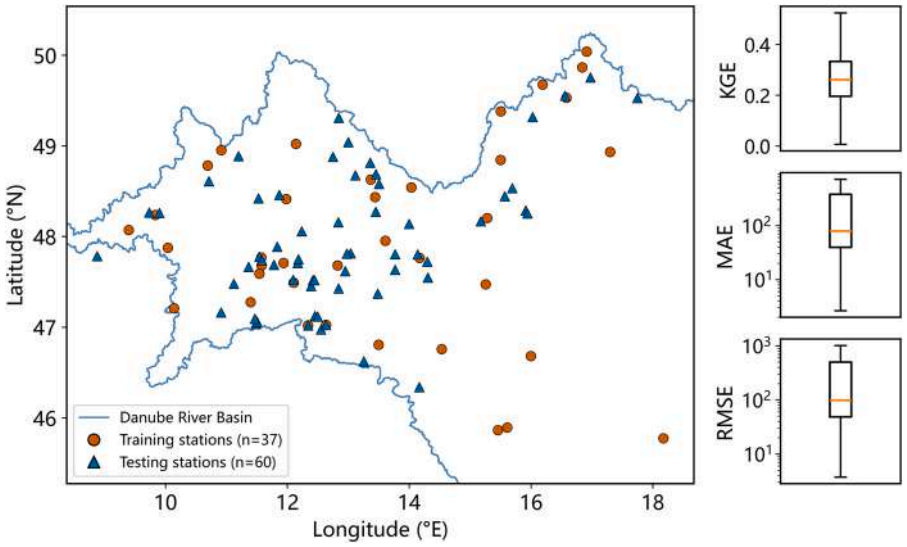

Fig. 9. Spatial distribution of training and testing stations and performance statistics at testing sites in the Danube River Basin.

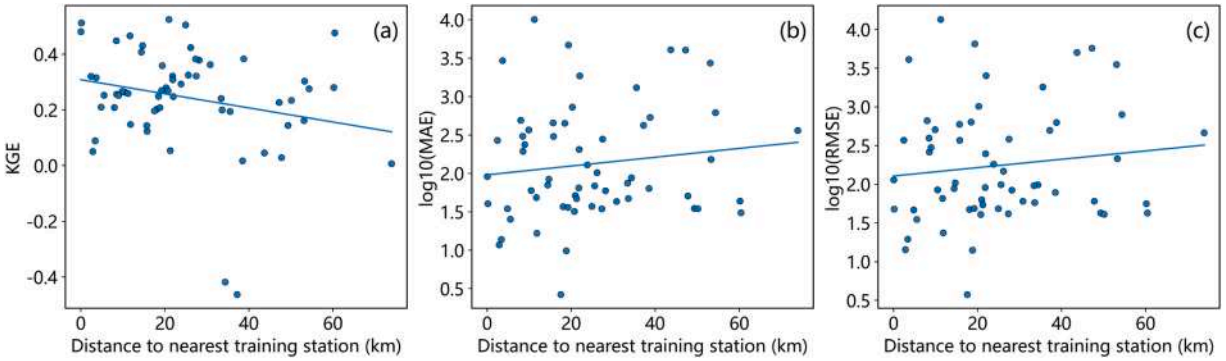

Fig. 10. Relationship between model performance and distance to the nearest training station at testing sites.

decision-making in flood-prone regions (Liu et al., 2025; Wang et al., 2025).

6.2. *Model stability: station reduction experiment and prediction at ungauged stations*

Our station reduction experiment aimed to evaluate the GCN-LSTM model's stability under varying amounts of station data. The results show that, despite reducing the number of stations from 97 to as few as 17, the model maintained stable performance with only minor fluctuations in accuracy. This indicates that the model is stable in handling reductions in the number of stations, demonstrating its resilience across different station configurations. The GCN-LSTM model exhibits a low sensitivity to the density of the input data, which is a desirable property for models intended for real-world applications, where data scarcity is common. This stability, however, comes with a limitation in generalization capability.

Although the model remained robust under the station-reduction experiment, its spatial generalization to ungauged stations still shows room for improvement. In practice, performance declines at distant or weakly represented sub-basins, where streamflow generation and routing behaviours differ from those learned during training (Kumar et al., 2024). A notable symptom is the model's tendency to underestimate peak streamflow at ungauged sites. This can arise because peak events are relatively rare and are therefore under-emphasized by standard regression objectives, leading to conservative predictions that smooth extremes (Bigi et al., 2025). In addition, graph message passing aggregates information from neighbouring stations; when neighbouring nodes are not truly hydrologically analogous, the resulting "spatial averaging" can further dampen sharp peaks (Twaróg, 2025). Moreover, localized extreme precipitation and human regulation can shift or attenuate observed peaks in ways not explicitly represented by the current inputs, thereby amplifying errors at ungauged stations (Potdar et al., 2021).

These findings reinforce a well-known challenge in hydrology: accurate prediction at ungauged locations is often limited by the absence of region-specific information (Gacu et al., 2025). Future work could improve transferability by incorporating richer descriptors of catchment similarity (e.g., additional physiographic and climatic attributes and hydrological signatures such as seasonality

and storage proxies), and by introducing physics-guided constraints or event-sensitive training objectives to better preserve flood peaks. In addition, regional grouping or basin segmentation (e.g., regime-based clustering by tributaries, climate zones, or hydrological signatures) may reduce cross-regime information mixing during message passing and thus enhance generalization. We note that extending the Danube-trained model to entirely different basins or climate zones would generally require re-training or explicit domain-adaptation approaches. While this is beyond the scope of the present study, it represents a meaningful direction for further research (Belvederesi et al., 2022; Gacu et al., 2025; Gonzales-Inca et al., 2022).

6.3. Future directions for model development and improvement

Despite the GCN-LSTM model's strong predictive capability across diverse regions and flow conditions, significant opportunities remain for further refinement and expansion. First, we note a key limitation of this study: although we compared GCN-LSTM with widely used temporal baselines, we did not include additional spatiotemporal architectures such as CNN–LSTM, ST-GCN, or DCRNN. As streamflow prediction is inherently spatiotemporal, broader benchmarking against these models would provide a more comprehensive assessment. We therefore regard identifying and evaluating which spatiotemporal modeling paradigms best represent river-network dynamics as an important direction for future work. At the same time, it should be emphasized that this study does not aim to replace physically based hydrological models. Physically based and conceptual models remain indispensable for process understanding, scenario analysis, and water-resources planning, while the proposed GCN-LSTM framework is intended as a complementary data-driven approach, particularly suited to prediction-oriented applications or settings where extensive physical model calibration is challenging.

Future work should focus on enhancing uncertainty treatment, improving the simulation of extreme events, and strengthening practical applicability across time scales. Current deterministic prediction frameworks often fail to account for uncertainties arising from input errors, model simplifications, and boundary conditions (Khalili et al., 2023; Zhou et al., 2025). To address this, methods such as Bayesian graph neural networks, Monte Carlo dropout, and ensemble learning could be integrated to quantify predictive uncertainty and provide probabilistic forecasts. This would improve reliability for decision-making in high-risk contexts (Fakour et al., 2024; Peng et al., 2025). Furthermore, improving the model's responsiveness to rare extreme events, such as floods and droughts, is crucial. Future research could explore designing loss functions sensitive to extremes or applying reweighting strategies to enhance recognition of rare events, reducing the underestimation of peak streamflow (Laimighofer et al., 2022; Li et al., 2024). Another promising direction is cross-basin transfer learning. By aligning graph structures across basins, enabling embedding matching, and sharing parameters, the model could become more transferable to data-scarce regions. Incorporating basin delineations, climate characteristics, and topographic attributes into graph embeddings could further enhance generalization, contributing to the development of intelligent hydrological models with regional and global applicability (Gacu et al., 2025; Xu et al., 2023).

Finally, while daily inputs have demonstrated clear benefits for short-term dynamics and extreme-event representation, long-horizon applications must also consider computational cost and operational convenience. Practical alternatives include multi-scale strategies that retain a monthly configuration to characterize seasonal–interannual variability while activating daily inference during hydrologically active periods, as well as reducing redundant predictors through aggregation or feature selection. Such designs can preserve much of the event-scale benefit of high-resolution inputs while keeping long-term simulations tractable, thereby strengthening real-world applicability for basin management and forecasting (Gacu et al., 2025; Granata and Di Nunno, 2025; Zhao et al., 2024). Equally important, our redundancy diagnostics indicate substantial collinearity among energy and evaporation-related predictors and among runoff-related fluxes (Fig. S1), suggesting that a more principled predictor reduction could improve interpretability and reduce “double counting” of highly overlapping information. Future work could therefore integrate group-wise feature selection, aggregation of physically equivalent variables, or representation-learning approaches that compress correlated predictors into a compact latent set, thereby improving robustness and facilitating clearer physical interpretation without sacrificing predictive accuracy.

6.4. Physical interpretability and hydrological meaning of the graph structure

Although the proposed approach is data-driven, its spatial design is anchored in hydrologically meaningful information by incorporating physiographic controls and spatial relatedness among stations. The static features represent first-order determinants of streamflow generation and storage–release behavior. These variables provide an interpretable basis for distinguishing catchment similarity and spatial heterogeneity across the Danube Basin (Pastor et al., 2024). We acknowledge that snowmelt and groundwater processes are not explicitly represented through dedicated physical modules. However, their integrated signatures are partially captured through physiographic proxies and meteorological forcing. In particular, elevation-related gradients are closely associated with temperature seasonality and precipitation phase, which control the timing of meltwater contributions (Jahan et al., 2025). Meanwhile, soil and vegetation attributes constrain subsurface storage capacity and delayed release mechanisms that are commonly linked to groundwater buffering and catchment memory (Oswald et al., 2024). By combining these static controls with time-varying atmospheric inputs, the model leverages physically consistent relationships between catchment properties, climate forcing, and observed streamflow dynamics.

The station connectivity defined by geographic proximity and physiographic similarity should be interpreted as an approximation of hydrological relatedness rather than a strict representation of flow routing. Spatial proximity reflects regional coherence in hydroclimate forcing, while physiographic similarity reduces unrealistic connections between nearby but hydrologically distinct catchments (Betterle et al., 2019). This graph design therefore supports basin-scale regionalization and information sharing among

stations under heterogeneous conditions. Nevertheless, it does not explicitly encode upstream–downstream directionality, travel time, or human regulation effects, which may be important in regulated sub-basins or for ungauged locations. Future work could strengthen the hydrological meaning of connectivity by incorporating river-network topology and by explicitly accounting for regulation signals where relevant, thereby improving interpretability and transferability across diverse hydroclimatic settings (Dwivedi et al., 2025). In addition, because multiple ERA5-Land predictors reflect closely related water–energy balance processes, a more compact and physically grounded predictor set may help separate "what drives streamflow" from "how streamflow is transmitted through the network," thereby improving interpretability of both node features and graph message passing.

7. Conclusions

This study employed a multi-source graph-based GCN-LSTM framework for streamflow prediction, integrating the spatial modeling capability of GCN with that of LSTM. The objective was to improve the representation of complex hydrological processes and enhance regional-scale generalization. The main findings are as follows:

(1) Across multiple experiments, the GCN-LSTM model consistently outperformed benchmark models. In experiments using 97 gauging stations in the Danube River Basin, GCN-LSTM achieved the highest R^2 values and the lowest error metrics, demonstrating clear advantages in capturing overall streamflow dynamics and extreme flow responses.

(2) In small-scale experiments within the Danube Basin, replacing monthly inputs with daily inputs and synchronizing both streamflow and dynamic variables to a daily resolution resulted in significantly improved predictive performance. This improvement was especially evident in capturing extreme events and short-term variability, underscoring the importance of high temporal resolution data for modeling rapid hydrological processes.

(3) The station-thinning experiments indicate that model performance remains highly robust across station densities, with R^2 consistently above 0.93 and only minor fluctuations. MAE and RMSE vary gradually rather than sharply as stations are removed or added, implying that station density has only a limited impact on predictive accuracy. A modest degradation is observed when fewer than 30 stations were used, consistent with reduced spatial representativeness.

(4) Cross-validation experiments showed reduced performance at ungauged sites, with further degradation as the distance to the training network increased. These results indicate that generalization remains a key challenge in hydrologically diverse ungauged regions, where limited representativeness of training conditions can degrade performance and lead to underestimation of extreme flood peaks.

CRediT authorship contribution statement

Mingze Sun: Writing – original draft, Software, Figure preparation. **Yu Sun:** Conceptualization, Methodology. **Xin Yu:** Figure preparation. **Yuyue Ye:** Figure preparation.

Ethics, sensitive information, and competing interests

This study does not involve human participants, animals, or sensitive personal data; therefore, ethical approval is not required.

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in the manuscript.

Declaration of Competing Interest

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

Acknowledgment

This research was supported by the National Key Research and Development Program of China (2025YFE0102700) and the National Natural Science Foundation of China (grant 42171426, 42374041 and 42304099).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2026.103275.

Data availability

Data will be made available on request.

References

Alquraish, M.M., Khadr, M., 2021. Remote-sensing-based streamflow forecasting using artificial neural network and support vector machine models. Remote Sens. 13, 4147. https://doi.org/10.3390/rs13204147.

Ashraf, I., Hermes, L., Artelt, A., Hammer, B., 2023. Spatial graph convolution neural networks for water distribution systems. Adv. Intell. Data Anal. XXI Cham 29, 41. https://doi.org/10.1007/978-3-031-30047-9_3.

Bargam, B., Boudhar, A., Kinnard, C., Bouamri, H., Nifa, K., Chehbouni, A., 2024. Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco. Discov. Appl. Sci. 6, 306. https://doi.org/10.1007/s42452-024-05994-z.

Belvederesi, C., Zaghloul, M.S., Achari, G., Gupta, A., Hassan, Q.K., 2022. Modelling river flow in cold and ungauged regions: a review of the purposes, methods, and challenges. Environ. Rev. https://doi.org/10.1139/er-2021-0043.

Betterle, A., Schirmer, M., Botter, G., 2019. Flow dynamics at the continental scale: streamflow correlation and hydrological similarity. Hydrol. Process. 33, 627–646. https://doi.org/10.1002/hyp.13350.

Bhattacharya, R., Nagwani, N.K., Asudani, D.S., Chhabra, G.S., Bhattacharya, S., Kadam, S., 2026. A Comprehensive Overview of Graph Convolutional Network. In: Bhattacharya, R., Rathore, Y.K., Tran, T.A., Swarnkar, S.K. (Eds.), Graph Mining: Practical Uses and Instruments for Exploring Complex Networks. Springer Nature Switzerland, Cham, pp. 1–19. https://doi.org/10.1007/978-3-031-93802-3_1.

Bhatti, U.A., Tang, H., Wu, G., Marjan, S., Hussain, A., 2023. Deep learning with graph convolutional networks: an overview and latest applications in computational intelligence. Int. J. Intell. Syst. 2023, 8342104. https://doi.org/10.1155/2023/8342104.

Biazar, S.M., Golmohammadi, G., Nedhunuri, R.R., Shaghaghi, S., Mohammadi, K., 2025. Artificial intelligence in hydrology: advancements in soil, water resource management, and sustainable development. Sustainability 17, 2250. https://doi.org/10.3390/su17052250.

Bigi, F., Langer, M., and Ceriotti, M.: The dark side of the forces: assessing non-conservative force models for atomistic machine learning, https://doi.org/10.48550/arXiv.2412.11569, 25 October 2025.

Bringeland, S., Fotopoulos, G., 2024. Analysis of gap filling techniques for GRACE/GRACE-FO terrestrial water storage anomalies in Canada. J. Hydrol. 630, 130644. https://doi.org/10.1016/j.jhydrol.2024.130644.

Clelland, A.A., Marshall, G.J., Baxter, R., 2024. Evaluating the performance of key ERA-Interim, ERA5 and ERA5-Land climate variables across Siberia. Int. J. Climatol. 44, 2318–2342. https://doi.org/10.1002/joc.8456.

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

Dai, Z., Zhan, C., Yin, H., Chen, J., Xu, L., Xia, Y., Yang, S., Chen, W., Cao, M., Du, Z., Zhang, X., Yan, B., Ma, Y., Wang, H., Moeini, F., Soltanian, M.R., Thanh, H.V., Carroll, K.C., 2025. Incorporating deep learning into hydrogeological modeling: advancements, challenges, and future directions. J. Geophys. Res. Mach. Learn. Comput. 2. https://doi.org/10.1029/2025JH000703.

Deng, L., Zhang, X., Tao, S., Zhao, Y., Wu, K., Liu, J., 2023. A spatiotemporal graph convolution-based model for daily runoff prediction in a river network with non-Euclidean topological structure. Stoch. Environ. Res Risk Assess. 37, 1457–1478. https://doi.org/10.1007/s00477-022-02352-6.

Duc, L., Sawada, Y., 2023. A signal-processing-based interpretation of the Nash–Sutcliffe efficiency. Hydrol. Earth Syst. Sci. 27, 1827–1839. https://doi.org/10.5194/hess-27-1827-2023.

Dunkerley, D.L., 2019. Rainfall intensity bursts and the erosion of soils: an analysis highlighting the need for high temporal resolution rainfall data for research under current and future climates. Earth Surf. Dyn. 7, 345–360. https://doi.org/10.5194/esurf-7-345-2019.

Dwivedi, D., Poeppl, R.E., Wohl, E., 2025. Hydrological connectivity: a review and emerging strategies for integrating measurement, modeling, and management. Front. Water 7. https://doi.org/10.3389/frwa.2025.1496199.

Fakour, F., Mosleh, A., and Ramezani, R.: A Structured Review of Literature on Uncertainty in Machine Learning & Deep Learning, https://doi.org/10.48550/arXiv.2406.00332, 1 June 2024.

Feng, F., He, X., Zhang, H., Chua, T.-S., 2023. Cross-GCN: enhancing graph convolutional network with kk-order feature interactions. IEEE Trans. Knowl. Data Eng. 35, 225–236. https://doi.org/10.1109/TKDE.2021.3077524.

Gacu, J.G., Monjardin, C.E.F., Mangulabnan, R.G.T., Mendez, J.C.F., 2025. Application of artificial intelligence in hydrological modeling for streamflow prediction in ungauged watersheds: a review. Water 17, 2722. https://doi.org/10.3390/w17182722.

Gonzales-Inca, C., Calle, M., Croghan, D., Torabi Haghighi, A., Marttila, H., Silander, J., Alho, P., 2022. Geospatial artificial intelligence (GeoAI) in the integrated hydrological and fluvial systems modeling: review of current applications and trends. Water 14, 2211. https://doi.org/10.3390/w14142211.

Granata, F., Di Nunno, F., 2025. Pathways for Hydrological Resilience: strategies for adaptation in a changing climate. Earth Syst. Environ. https://doi.org/10.1007/s41748-024-00567-x.

Gu, C., Tang, Q., Zhu, G., Ma, J., Gu, C., Zhang, K., Sun, S., Yu, Q., Niu, S., 2021. Discrepant responses between evapotranspiration- and transpiration-based ecosystem water use efficiency to interannual precipitation fluctuations. Agric. For. Meteorol. 303, 108385. https://doi.org/10.1016/j.agrformet.2021.108385.

Guo, Y., Zhang, Y., Zhang, L., Wang, Z., 2021. Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: a comprehensive review. WIREs Water 8, e1487. https://doi.org/10.1002/wat2.1487.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. 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.

Halder, J., Vystavna, Y., Wassenaar, L.I., 2022. Nitrate sources and mixing in the Danube watershed: implications for transboundary river basin monitoring and management. Sci. Rep. 12, 2150. https://doi.org/10.1038/s41598-022-06224-5.

Herrera, P.A., Marazuela, M.A., Hofmann, T., 2022. Parameter estimation and uncertainty analysis in hydrological modeling. WIREs Water 9, e1569. https://doi.org/10.1002/wat2.1569.

Ho, L., Goethals, P., 2022. Machine learning applications in river research: Trends, opportunities and challenges. Methods Ecol. Evol. 13, 2603–2621. https://doi.org/10.1111/2041-210X.13992.

Hodson, T.O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci. Model Dev. 15, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.

Jahan, A., Rai, N., Khan, M.U., Dar, T., Kumar, S., 2025. Isotopic characterization (δ18O, δD and d-excess) and modelling of stream flow partitioning along an altitudinal gradient in the Satluj River basin (SRB) Himalayan-alluvial plain using end member mixing analysis (EMMA) and associated uncertainty. Acta Geochim. https://doi.org/10.1007/s11631-025-00782-x.

Jia, L., Yen, N., Pei, Y., 2023. Spatial and temporal water quality data prediction of transboundary watershed using multiview neural network coupling. IEEE Trans. Geosci. Remote Sens. 61, 1–16. https://doi.org/10.1109/TGRS.2023.3334291.

Kant, C., Meena, R.S., Singh, S.K., 2025. A critical appraisal on various hydrological and hydrodynamic models. Water Conserv Sci. Eng. 10, 24. https://doi.org/10.1007/s41101-024-00328-x.

Khalili, M.A., Guerriero, L., Pouralizadeh, M., Calcaterra, D., Di Martire, D., 2023. Monitoring and prediction of landslide-related deformation based on the GCN-LSTM algorithm and SAR imagery. Nat. Hazards 119, 39–68. https://doi.org/10.1007/s11069-023-06121-8.

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrol. Earth Syst. Sci. 23, 4323–4331. https://doi.org/10.5194/hess-23-4323-2019.

Koch, F., Prasch, M., Bach, H., Mauser, W., Appel, F., Weber, M., 2011. How will hydroelectric power generation develop under climate change scenarios? a case study in the Upper Danube Basin. Energies 4, 1508–1541. https://doi.org/10.3390/en4101508.

Krajewski, W.F., Ghimire, G.R., Demir, I., Mantilla, R., 2021. Real-time streamflow forecasting: AI vs. Hydrologic insights. J. Hydrol. X 13, 100110. https://doi.org/10.1016/j.hydroa.2021.100110.

Kumar, A., Gaurav, K., Singh, A., Yaseen, Z.M., 2024. Assessment of machine learning models to predict daily streamflow in a semiarid river catchment. Neural Comput. Applic 36, 13087–13106. https://doi.org/10.1007/s00521-024-09748-1.

Laimighofer, J., Melcher, M., Laaha, G., 2022. Low-flow estimation beyond the mean – expectile loss and extreme gradient boosting for spatiotemporal low-flow prediction in Austria. Hydrol. Earth Syst. Sci. 26, 4553–4574. https://doi.org/10.5194/hess-26-4553-2022.
Li, X., Sun, Q.-L., Zhang, Y., Sha, J., Zhang, M., 2024. Enhancing hydrological extremes prediction accuracy: integrating diverse loss functions in Transformer models. Environ. Model. Softw. 177, 106042. https://doi.org/10.1016/j.envsoft.2024.106042.
Li, Y., Yu, D., Liu, Z., Zhang, M., Gong, X., and Zhao, L.: Graph Neural Network for spatiotemporal data: methods and applications, https://doi.org/10.48550/arXiv.2306.00012, 30 May 2023.
Liu, J., Koch, J., Stisen, S., Troldborg, L., Schneider, R.J.M., 2024. A national-scale hybrid model for enhanced streamflow estimation – consolidating a physically based hydrological model with long short-term memory (LSTM) networks. Hydrol. Earth Syst. Sci. 28, 2871–2893. https://doi.org/10.5194/hess-28-2871-2024.
Liu, Z., Coleman, N., Patrascu, F.I., Yin, K., Li, X., Mostafavi, A., 2025. Artificial intelligence for flood risk management: a comprehensive state-of-the-art review and future directions. Int. J. Disaster Risk Reduct. 117, 105110. https://doi.org/10.1016/j.ijdrr.2024.105110.
Lóczy, D., 2015. In: Lóczy, D. (Ed.), Climate and Drainage, in: Landscapes and Landforms of Hungary. Springer International Publishing, Cham, pp. 19–27. https://doi.org/10.1007/978-3-319-08997-3_3.
Lu, D., Konapala, G., Painter, S.L., Kao, S.-C., and Gangrade, S.: Streamflow Simulation in Data-Scarce Basins Using Bayesian and Physics-Informed Machine Learning Models, https://doi.org/10.1175/JHM-D-20-0082.1, 2021.
Lucas, L.V., Brown, C.J., Robertson, D.M., Baker, N.T., Johnson, Z.C., Green, C.T., Cho, S.J., Erickson, M.L., Gellis, A.C., Jasmann, J.R., Knowles, N., Prein, A.F., Stackelberg, P.E., 2025. Gaps in water quality modeling of hydrologic systems. Water 17, 1200. https://doi.org/10.3390/w17081200.
Maier, N., Gimbert, F., Gillet-Chaulet, F., 2022. Threshold response to melt drives large-scale bed weakening in Greenland. Nature 607, 714–720. https://doi.org/10.1038/s41586-022-04927-3.
Marshall, S.R.O., Tran, T.-N.-D., Tapas, M.R., and Nguyen, B.Q.: Integrating artificial intelligence and machine learning in hydrological modeling for sustainable resource management, International Journal of River Basin Management, 0, 1–17, https://doi.org/10.1080/15715124.2025.2478280, n.d.
Mehedi, M.A.A., Khosravi, M., Yazdan, M.M.S., Shabanian, H., 2022. Exploring temporal dynamics of river discharge using univariate long short-term memory (LSTM) recurrent neural network at east branch of delaware river. Hydrology 9, 202. https://doi.org/10.3390/hydrology9110202.
Mekonnen, Z.A., Riley, W.J., Shirley, I.A., Bouskill, N.J., Grant, R.F., 2024. Changes in high-latitude surface energy balance driven by snowpack and vegetation dynamics under warmer climate. Environ. Res. Lett. 20, 014031. https://doi.org/10.1088/1748-9326/ad98aa.
Melsen, L.A., Puy, A., Torfs, P.J.J.F., Saltelli, A., 2025. The rise of the Nash-Sutcliffe efficiency in hydrology. Hydrol. Sci. J. 70, 1248–1259. https://doi.org/10.1080/02626667.2025.2475105.
Mischel, S.A., Zentis, C., Plessow, H., 2025. The BALTEX dataset of the GLobal Runoff Data Centre (GRDC): a dataset of river discharge draining into the Baltic Sea. Front. Environ. Sci. 13. https://doi.org/10.3389/fenvs.2025.1599188.
Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF ensemble prediction system: methodology and validation. Q. J. R. Meteorol. Soc. 122, 73–119. https://doi.org/10.1002/qj.49712252905.
Mumbi, A.W., Li, F., Bavumiragira, J.P., Fangninou, F.F., 2021. Forecasting water consumption on transboundary water resources for water resource management using the feed-forward neural network: a case study of the Nile River in Egypt and Kenya. Mar. Freshw. Res. 73, 292–306. https://doi.org/10.1071/MF21118.
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D.G., Piles, M., Rodríguez-Fernández, N.J., Zsoter, E., Buontempo, C., Thépaut, J.-N., 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data 13, 4349–4383. https://doi.org/10.5194/essd-13-4349-2021.
Nichersu, I., Nichersu, I., Constantinescu, A., Nichersu, A., 2022. A Transdisciplinary Approach Using Danube River Multi-connectivity in Wetland Management. In: Negm, A., Zaharia, L., Ioana-Toroimac, G. (Eds.), in: The Lower Danube River: Hydro-Environmental Issues and Sustainability. Springer International Publishing, Cham, pp. 405–442. https://doi.org/10.1007/978-3-031-03865-5_14.
Oswald, S.E., Angermann, L., Bogena, H.R., Förster, M., García-García, A., Lischeid, G., Paton, E.N., Altdorff, D., Attinger, S., Güntner, A., Hartmann, A., Hendricks Franssen, H.-J., Hildebrandt, A., Kleinschmit, B., Orth, R., Peng, J., Ryo, M., Schrön, M., Wagner, W., Wagener, T., 2024. Hydrology on Solid Grounds? Integration is key to closing knowledge gaps concerning landscape subsurface water storage dynamics. Hydrol. Process. 38, e15320. https://doi.org/10.1002/hyp.15320.
Pastor, I., Tanislav, D., Nedelea, A., Dunea, D., Serban, G., Haghighi, A.T., Sabau, D., Bretcan, P., Pastor, I., Tanislav, D., Nedelea, A., Dunea, D., Serban, G., Haghighi, A.T., Sabau, D., Bretcan, P., 2024. Morphometric analysis and prioritization of sub-watersheds located in heterogeneous geographical units—case study: The Buzău River Basin. Sustainability 16. https://doi.org/10.3390/su16177567.
Peng, H., Shen, S., Zhang, H., Wang, F., Guo, F., Zhang, R., 2025. Enhancing disaster prediction with Bayesian deep learning: a robust approach for uncertainty estimation. Front. Appl. Math. Stat. 11. https://doi.org/10.3389/fams.2025.1653562.
Potdar, A.S., Kirstetter, P.-E., Woods, D., Saharia, M., 2021. Toward predicting flood event peak discharge in ungauged basins by learning universal hydrological behaviors with machine learning. J. Hydrometeorol. 22, 2971–2982. https://doi.org/10.1175/JHM-D-20-0302.1.
Qiu, D., Xu, R., Gao, P., Mu, X., 2024. Effect of vegetation restoration type and topography on soil water storage and infiltration capacity in the Loess Plateau, China. CATENA 241, 108079. https://doi.org/10.1016/j.catena.2024.108079.
Razavi, S., Duffy, A., Eamen, L., Jakeman, A.J., Jardine, T.D., Wheater, H., Hunt, R.J., Maier, H.R., Abdelhamed, M.S., Ghoreishi, M., Gupta, H., Döll, P., Moallemi, E. A., Yassin, F., Strickert, G., Nabavi, E., Mai, J., Li, Y., Thériault, J.M., Wu, W., Pomeroy, J., Clark, M.P., Ferguson, G., Gober, P., Cai, X., Reed, M.G., Saltelli, A., Elshorbagy, A., Sedighkia, M., Terry, J., Lindenschmidt, K.-E., Hannah, D.M., Li, K., Asadzadeh, M., Harvey, N., Moradkhani, H., Grimm, V., 2025. Convergent and Transdisciplinary Integration: on the Future of Integrated Modeling of Human-Water Systems. Water Resour. Res. 61. https://doi.org/10.1029/2024WR038088.
Ren, W., Li, X., Zheng, D., Zeng, R., Su, J., Mu, T., Wang, Y., 2023. Enhancing flood simulation in data-limited glacial river basins through hybrid modeling and multi-source remote sensing data. Remote Sens. 15, 4527. https://doi.org/10.3390/rs15184527.
Sabzipour, B., Arsenault, R., Troin, M., Martel, J.-L., Brissette, F., Brunet, F., Mai, J., 2023. Comparing a long short-term memory (LSTM) neural network with a physically-based hydrological model for streamflow forecasting over a Canadian catchment. J. Hydrol. 627, 130380. https://doi.org/10.1016/j.jhydrol.2023.130380.
Sarker, S., Leta, O.T., 2025. Review of watershed hydrology and mathematical models. Eng 6, 129. https://doi.org/10.3390/eng6060129.
Schiller, H., Miklós, D., Sass, J., 2010. The Danube River and its Basin Physical Characteristics, Water Regime and Water Balance. In: Brilly, M. (Ed.), Hydrological Processes of the Danube River Basin: Perspectives from the Danubian Countries. Springer Netherlands, Dordrecht, pp. 25–77. https://doi.org/10.1007/978-90-481-3423-6_2.
Singh, A., Shafi Dar, S., Bansal, S., Singh, R., and Kumar, N.: A Comprehensive Survey on Node Classification Methods, Applications, and Challenges: From Traditional Approaches to Graph Neural Network and Beyond, https://doi.org/10.2139/ssrn.5286802, 10 June 2025a.
Singh, S., Mishra, K., Chavan, R., Tiwari, H.L., 2025b. Advancements and challenges in hydrological modeling: a comprehensive review. Hydrol. Hydrol. Model. Singap. 423–442. https://doi.org/10.1007/978-981-97-7474-6_32.
Sun, A.Y., Jiang, P., Mudunuru, M.K., Chen, X., 2021. Explore spatio-temporal learning of large sample hydrology using graph neural networks. Water Resour. Res. 57. https://doi.org/10.1029/2021WR030394.
Twaróg, B., 2025. Spatial flows of information entropy as indicators of climate variability and extremes. Entropy 27, 1132. https://doi.org/10.3390/e27111132.
Wang, H., Cao, L., Feng, R., 2021. Hydrological similarity-based parameter regionalization under different climate and underlying surfaces in ungauged basins. Water 13, 2508. https://doi.org/10.3390/w13182508.
Wang, P., Wu, X., Yichen, 2025. Ai-driven approaches to flood risk management: overcoming data bias and enhancing decision-making. Clim. Risk Manag. 50, 100752. https://doi.org/10.1016/j.crm.2025.100752.
Wang, Q., Zhu, S., 2024. Multi-station joint runoff forecasting using graph neural network coupled with spatial connectivity of hydrological stations. Math. Geosci. https://doi.org/10.1007/s11004-024-10167-0.
Westra, S., Fowler, H.J., Evans, J.P., Alexander, L.V., Berg, P., Johnson, F., Kendon, E.J., Lenderink, G., Roberts, N.M., 2014. Future changes to the intensity and frequency of short-duration extreme rainfall. Rev. Geophys. 52, 522–555. https://doi.org/10.1002/2014RG000464.

Williams, G.P., 2025. Friends don't let friends use Nash-Sutcliffe Efficiency (NSE) or KGE for hydrologic model accuracy evaluation: a rant with data and suggestions for better practice. Environ. Model. Softw. 194, 106665. https://doi.org/10.1016/j.envsoft.2025.106665.
Wu, J., Chen, X., Yu, Z., Yao, H., Li, W., Zhang, D., 2019. Assessing the impact of human regulations on hydrological drought development and recovery based on a 'simulated-observed' comparison of the SWAT model. J. Hydrol. 577, 123990. https://doi.org/10.1016/j.jhydrol.2019.123990.
Wu, Z., Cui, N., Zhang, W., Liu, C., Jin, X., Gong, D., Xing, L., Zhao, L., Wen, S., Yang, Y., 2024. Estimating soil moisture content in citrus orchards using multi-temporal sentinel-1A data-based LSTM and PSO-LSTM models. J. Hydrol. 637, 131336. https://doi.org/10.1016/j.jhydrol.2024.131336.
Xu, C., Wang, W., Hu, Y., Liu, Y., 2024a. Evaluation of ERA5, ERA5-Land, GLDAS-2.1, and GLEAM potential evapotranspiration data over mainland China. J. Hydrol. Reg. Stud. 51, 101651. https://doi.org/10.1016/j.ejrh.2023.101651.
Xu, Q., Shi, Y., Bamber, J., Tuo, Y., Ludwig, R., and Zhu, X.X.: Physics-aware Machine Learning Revolutionizes Scientific Paradigm for Machine Learning and Process-based Hydrology, https://doi.org/10.48550/arXiv.2310.05227, 12 July 2024b.
Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, L., Ran, G., 2023. Deep transfer learning based on transformer for flood forecasting in data-sparse basins. J. Hydrol. 625, 129956. https://doi.org/10.1016/j.jhydrol.2023.129956.
Yan, D., Wang, Y., Qin, D., Zhang, J., 2025. Hydrological geography: Theoretical framework, research progress, and future development directions. Geogr. Res. Bull. 4, 186–224. https://doi.org/10.50908/grb.4.0_186.
Yang, T.-H., Liu, W.-C., 2020. A general overview of the risk-reduction strategies for floods and droughts. Sustainability 12, 2687. https://doi.org/10.3390/su12072687.
Yang, Y., Pan, M., Feng, D., Xiao, M., Dixon, T., Hartman, R., Shen, C., Song, Y., Sengupta, A., Delle Monache, L., Ralph, F.M., 2025. Improving streamflow simulation through machine learning-powered data integration and its potential for forecasting in the Western U.S. Hydrol. Earth Syst. Sci. 29, 5453–5476. https://doi.org/10.5194/hess-29-5453-2025.
Yaseen, Z.M., 2023. A new benchmark on machine learning methodologies for hydrological processes modelling: a comprehensive review for limitations and future research directions. Knowl. Based Eng. Sci. 4, 65–103. https://doi.org/10.51526/kbes.2023.4.3.65-103.
Yu, C., Xu, Z., Li, Y., Yang, Y., Cai, Y., Yang, Z., 2023. Rethinking environmental flow management strategies in reservoir operations from an integrated water quality perspective. J. Hydrol. 626, 130196. https://doi.org/10.1016/j.jhydrol.2023.130196.
Yu, M., Huang, Q., Li, Z., 2024a. Deep learning for spatiotemporal forecasting in Earth system science: a review. Int. J. Digit. Earth 17, 2391952. https://doi.org/10.1080/17538947.2024.2391952.
Yu, Q., Tolson, B.A., Shen, H., Han, M., Mai, J., Lin, J., 2024b. Enhancing long short-term memory (LSTM)-based streamflow prediction with a spatially distributed approach. Hydrol. Earth Syst. Sci. 28, 2107–2122. https://doi.org/10.5194/hess-28-2107-2024.
Zhang, F., Wu, Z., Di, D., Wang, H., 2023. Water resources allocation based on water resources supply-demand forecast and comprehensive values of water resources. J. Hydrol. Reg. Stud. 47, 101421. https://doi.org/10.1016/j.ejrh.2023.101421.
Zhang, L., Song, H., Aletras, N., Lu, H., 2022. Node-feature convolution for graph convolutional networks. Pattern Recognit. 128, 108661. https://doi.org/10.1016/j.patcog.2022.108661.
Zhang, Y., Wang, Q.J., Frost, A.J., Trewin, B., Pudashine, J., Velasco-Forero, C., Su, C.-H., Villani, V., 2025. Advancing high-resolution precipitation analysis for operational hydrological applications in australia: data sources, methodologies, and recommendations, WIREs. Water 12, e70037. https://doi.org/10.1002/wat2.70037.
Zhao, X., Wang, H., Bai, M., Xu, Y., Dong, S., Rao, H., Ming, W., 2024. A comprehensive review of methods for hydrological forecasting based on deep learning. Water 16, 1407. https://doi.org/10.3390/w16101407.
Zhou, W., Wang, W., Wang, X., 2025. Research on lightning prediction based on GCN-LSTM model integrating spatiotemporal features. Atmosphere 16, 447. https://doi.org/10.3390/atmos16040447.
Zhu, X., Zhu, L., Guo, J., Liang, S., Dietze, S., 2021. GL-GCN: global and local dependency guided graph convolutional networks for aspect-based sentiment classification. Expert Syst. Appl. 186, 115712. https://doi.org/10.1016/j.eswa.2021.115712.
Zhu, Z., Cai, Y., Dai, X., Bai, X., Yang, Y., Li, X., Yang, Z., 2024. Spatiotemporally contiguous precipitation events across China: Role of atmospheric rivers and tropical cyclones in heavy precipitation. Ecol. Indic. 158, 111426. https://doi.org/10.1016/j.ecolind.2023.111426.