Journal of Hydrology 672 (2026) 135396

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

Enhancing runoff simulation in data-scarce mountainous regions: a coupled SWAT and transferable transformer approach☆

Yi He [a,b,1,*], Rui Yan [b,1], Yanhong Tang [a], Jinyu Zhang [b], Hui Qian [c], Dejing Chen [a], Lei Zhu [a], Xin Cao [a,b,*]

[a] Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China
[b] College of Computer Science, Northwest University, Xi'an, Shaanxi 710127, China
[c] School of Water and Environment, Chang'an University, Xi'an 710054, Shaanxi, China

ARTICLE INFO

This manuscript was handled by Huaming Guo, Editor-in-Chief

Keywords:
Runoff prediction
SWAT
Transfer learning
Trend-seasonal decomposition
Mountainous basin

ABSTRACT

Runoff prediction in mountainous watersheds remains challenging due to complex nonlinear hydro-meteorological interactions, strong spatial heterogeneity, and limited long-term observations. Physically based models such as SWAT provide valuable process representations but are often affected by parameter uncertainty, whereas conventional deep learning models struggle to concurrently capture fast storm responses and slow hydrological dynamics. To overcome these limitations, this study develops the SWAT-HydroTransformer, a physics–data hybrid framework that integrates SWAT-simulated hydrological processes as physical constraints into a multi-scale Transformer architecture. The model incorporates a trend–seasonal decomposition mechanism and the enhanced AutoCorrelation attention module to explicitly separate low-frequency catchment storage signals from high-frequency meteorological fluctuations. The framework was applied to the Shiquan River Basin, China. Compared with standalone SWAT, Long Short Term Memory (LSTM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and the basic Transformer, the HydroTransformer achieved substantially higher predictive skill, yielding Nash–Sutcliffe Efficiency (NSE) values of 0.843 under baseline conditions and 0.786 under non-stationary climatic conditions. Model interpretability analyses using SHapley Additive exPlanations (SHAP) and spectral decomposition further demonstrated strong physical consistency: seasonal components corresponded to precipitation-driven variability, while trend components captured baseflow and storage dynamics. To evaluate regional transferability, a fine-tuning strategy was implemented across three data-limited basins. After transfer learning, NSE values improved to 0.729, 0.826, and 0.634, compared with the corresponding results obtained from direct cross-basin transfer, indicating robust cross-basin adaptability. Overall, the SWAT-HydroTransformer provides an efficient, interpretable, and transferable solution for runoff forecasting in complex mountainous environments.

1. Introduction

Runoff prediction remains a fundamental cornerstone of hydrology and water resources management (Pham et al., 2021). The accuracy of these forecasts directly influences critical decision-making processes, including flood early warning systems, reservoir operations, and water allocation (Chang and Guo, 2020). Moreover, reliable predictions are indispensable for maintaining watershed ecological security, promoting regional sustainable development, and mitigating disaster risks (Eldardiry and Hossain, 2021). Despite its importance, achieving high-precision runoff prediction remains a significant scientific challenge. The runoff generation process is governed by complex, coupled interactions between rainfall, evapotranspiration, soil moisture, and surface–groundwater exchange (Hui et al., 2025). These processes exhibit strong nonlinearity, multi-scale variability, and spatial heterogeneity, making it difficult for traditional empirical models to fully capture the

☆ This article is part of a special issue entitled: 'River Basin Ecohydrological Processes' published in Journal of Hydrology.

* Corresponding authors at: Shaanxi Key Laboratory of Earth Surface System and Environmental Carrying Capacity, College of Urban and Environmental Sciences, Northwest University, Xi'an 710127, China and College of Computer Science, Northwest University, Xi'an, Shaanxi 710127, China.
E-mail addresses: yihe@nwu.edu.cn (Y. He), xin_cao@163.com (X. Cao).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jhydrol.2026.135396

Available online 27 March 2026

underlying dynamic mechanisms (Ausloos et al., 2017). Furthermore, many watersheds suffer from sparse observation networks and inconsistent monitoring records, creating a data-scarcity bottleneck that further complicates modeling efforts (Evin et al., 2024).

These challenges are particularly pronounced in mountainous watersheds. The complex topography-characterized by significant elevation gradients, steep slopes, and deep valleys-results in a highly uneven spatial distribution of landscape characteristics. Rainfall in these regions exhibits intense spatiotemporal variability, leading to distinct regional complexities in runoff generation and concentration (Mahmood et al., 2021). Additionally, logistical constraints in mountainous areas often result in a lack of high-density hydro-meteorological stations, making it difficult to obtain long-term, continuous datasets (Tao et al., 2019). Consequently, runoff prediction in these regions is subject to elevated uncertainty.

While traditional physically-based models offer the advantage of elucidating hydrological mechanisms, they are often data-intensive and highly sensitive to parameterization (Linh et al., 2018). In complex mountainous terrain, these models frequently struggle to balance simulation accuracy with operational feasibility. Conversely, machine learning approaches have become increasingly essential tools in hydrological research due to their robust capability to model high-dimensional, nonlinear relationships (Zhao et al., 2024).

Among data-driven approaches, Long Short-Term Memory (LSTM) networks have been extensively applied to runoff forecasting due to their superior performance in handling time-series data (Chen et al., 2023; Ma et al., 2025; Shi et al., 2025b; Yong et al., 2025). Furthermore, an increasing number of novel machine learning models have been introduced for runoff prediction, including ensemble learning methods such as random forests and gradient boosting trees (Jin et al., 2025; Wang et al., 2025a). These methods provide more flexible and efficient solutions for runoff forecasting in complex mountainous basins. Frank et al. (2023) conducted an hourly runoff prediction using the LSTM networks in the alpine headwater catchment of Goms Valley, Switzerland, and compared the results with those obtained from the MINERVE standard model. The findings indicated that LSTM achieved prediction accuracy that is comparable to, or even superior to, that of traditional physical models in this complex alpine basin. Szczepanek (2022) applied three gradient boosting models—XGBoost, LightGBM, and CatBoost—to the Skawa River basin in the Polish mountains. The average Nash-Sutcliffe Efficiency (NSE) for these models ranged from 0.85 to 0.89. Chen et al. (2025) employed the LSTM model in the Northern Rocky Mountain (NRB) region of the Qinghai-Tibet Plateau (QTP) to investigate the predictive performance of runoff under various input variables. Shi et al. (2025a) developed the VIC-glacier model specifically for application in cold alpine regions. By integrating LSTM networks with autoregressive error correction techniques, the model significantly enhances the accuracy of runoff predictions. Zhang et al. (2025) employed a hybrid interpretable model that combines Convolutional Neural Networks (CNN), LSTM, and attention mechanisms for runoff prediction in the typical river source region of the Qinghai-Tibet Plateau. The model demonstrated NSE values ranging from 0.79 to 0.92. Hameed et al. (2025) employed XGBoost, LSTM, and Random Forest (RF) models to train and predict 20 years (2002–2021) of runoff data from the Lützental watershed in Switzerland. Among these models, the XGBoost model achieved the highest accuracy, with a NSE of 0.797. These findings demonstrate that data-driven methods hold significant potential for characterizing the complex nonlinear rainfall-runoff processes in mountainous regions.

However, while Recurrent Neural Networks (RNNs) like LSTM are effective, they have limitations in parallel processing and capturing extremely long-term dependencies. In recent years, the Transformer architecture and its derivatives have emerged as powerful alternatives for time-series modeling (Koya and Roy, 2024; Sun et al., 2024). Unlike RNNs, the Transformer's self-attention mechanism effectively captures long-range dependencies across the entire sequence. Its multi-head attention allows the model to simultaneously learn diverse temporal patterns, such as short-term fluctuations (e.g., storm responses) and long-term trends (e.g., seasonality and snowmelt dynamics). This capability is particularly critical for mountainous watersheds where hydrological response is driven by a mix of rapid and slow processes (Zhao et al., 2025). Furthermore, the Transformer's parallel computing architecture enables more efficient training on large-scale or multi-site datasets, often yielding better generalization in cross-watershed scenarios compared to classical deep learning models (Demiray et al., 2024).

To address the specific challenges of mountainous hydrology, this study proposes a multi-scale HydroTransformer model. This architecture systematically improves the standard Transformer by incorporating temporal embedding, trend-seasonal decomposition, multi-scale convolutional smoothing, and an enhanced AutoCorrelation mechanism. These modifications allow the model to better adapt to intrinsic hydrological characteristics, such as multi-temporal scales, strong seasonality, and event-driven pulses. We integrated the HydroTransformer with the Soil and Water Assessment Tool (SWAT) framework to leverage process-based insights. The model was developed and trained using data from the Shiquan watershed in the Han River basin, located in the Qinling-Daba Mountains of China. We employed SHapley Additive exPlanations (SHAP) to interpret key runoff drivers within the model. Subsequently, the pre-trained model was evaluated under both direct transfer and fine-tuning scenarios on three typical data-scarce mountainous watersheds within the same region to assess its cross-watershed transferability and generalization.

Specifically, this study aims to answer three primary research questions:

(1) Can the proposed multi-scale HydroTransformer effectively enhance daily runoff simulation accuracy in complex mountainous watersheds?
(2) How do the specific structural designs and mechanisms of the model contribute to its predictive capability and interpretability?
(3) Can this model maintain robust performance in small-to-medium-sized catchments with limited observational data?

2. Materials and methods

2.1. Study area

The Shiquan River Basin is located in southern Shaanxi Province, China, within the middle–upper reaches of the Han River, the largest tributary of the Yangtze River. The basin covers approximately 23,805 square kilometers and is mainly underlain by yellow–brown soils. It spans the Qinling and Daba mountain ranges, with steep terrain in the northern Qinling Mountains and relatively gentler, rounded ridges in the southern Bashan Mountains. The region has a subtropical monsoon humid climate with four distinct seasons. Precipitation is concentrated in summer and autumn, while winter and spring are relatively dry. The long-term mean annual rainfall is about 873.9*mm*, and the multi-year mean temperature is approximately 14.6°C. (*https://www.shiquan.gov.cn/Node-76009.html*). (Fig. 1b).

The Xiangjiaping Basin is located in Xunyang County, Shaanxi Province, China, in the lower reaches of the Xun River, a tributary of the Han River. The basin covers approximately 6,448 square kilometers and is primarily underlain by gray shale and limestone. It lies within a subtropical monsoon climate, characterized by dry, clear winters and markedly wetter summers, resulting in a distinct seasonal contrast. The annual mean temperature exhibits a latitudinal and altitudinal gradient, with lower temperatures at higher latitudes and elevations. Long-term climatic statistics indicate an average annual precipitation of about 850*mm* and a multi-year mean temperature of roughly 15°C (Li et al., 2014). (Fig. 1c).

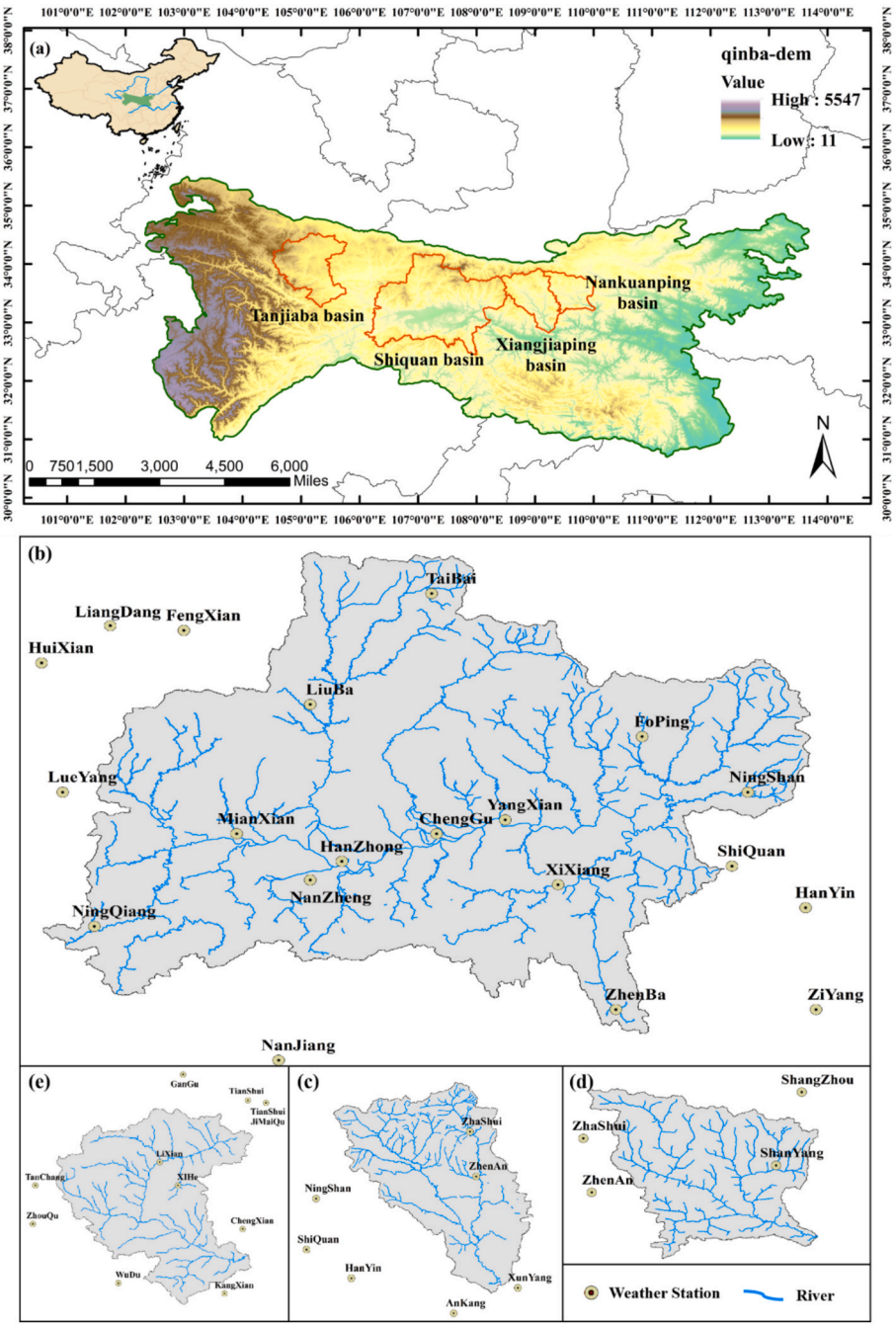

Fig. 1. Geographical overview of the study area. (a) Location of the Qinling-Daba Mountains region in China; (b–e) Basin boundaries, river networks, and meteorological station distribution for the Shiquan, Xiangjiaping, Nankuanping, and Tanjiaba basins, respectively.

The Nankuanping Basin is located in southern Shanyang County, Shangluo City, Shaanxi Province, China, along the middle reaches of the Jinqian River, a tributary of the Han River. The basin covers approximately 3,936 square kilometers and is characterized by continuous mountain ranges and dissected valleys, with elevations generally decreasing from north to southeast. The region represents a shallow low-mountain landscape. It falls within a subtropical monsoon humid climate zone with four distinct seasons, featuring cold, dry winters and hot, rainy summers. Precipitation is concentrated in the summer months. Long-term climatic averages indicate an annual precipitation of about 826*mm* and a mean temperature of approximately 13.1°C (Yang, 2015). (Fig. 1d).

The Tanjiaba Basin is located in Cheng County, Gansu Province, China, within the Jialing River tributary system in the upper Yangtze River. The watershed covers approximately 9,638 square kilometers and is characterized by loess hills and dissected gullies. Positioned in the transition zone between the Loess Plateau and the Qinling Mountains, the basin exhibits complex geological structures, with surface materials dominated by sandy-gravelly soils and generally good vegetation cover. The region has a warm temperate semi-humid climate, featuring mild temperatures and abundant rainfall. Precipitation is concentrated in summer, while winters are cold and damp. Long-term climatic averages indicate an annual precipitation of about 658.7*mm* and a mean temperature of roughly 12.1°C (Lv, 2020). (Fig. 1e).

2.2. Data collection

For the construction of the SWAT model, this study utilized several

datasets: a 30-meter resolution Digital Elevation Model (DEM) (*https://www.resdc.cn/*) for generating river networks, delineating sub-basins, and calculating topographic parameters; remote sensing monitoring data of land use types in China with a 30-meter spatial resolution for the years 1980 and 2005 (*https://www.resdc.cn/*); and the 1 : 1, 000, 000 Harmonized World Soil Database (HWSD) soil data for China (*https://gaez.fao.org/pages/hwsd*). The land use data were reclassified into six categories: cultivated land, forest land, grassland, water bodies, residential land, and unused land. The necessary soil physical parameters were computed using the Soil-Plant-Atmosphere-Water (SPAW) model. The meteorological data (Precipitation (PCP), Temperature (TEMP_min, TEMP_max), Wind Speed (WS), Relative Humidity (RH), Sunshine Radiation (SR)) for the study area, covering the periods from 1972 to 1984 and 2001 to 2016, were obtained from daily observation records. These data were collected from all accessible meteorological stations within and surrounding the study region (the China Meteorological Data Network (*https://data.cma.cn/*)).The daily runoff data utilized in this study were sourced from the Hydrological Yearbook of the People's Republic of China. This dataset includes two observation periods at the Shiquan Hydrological Station (1972-1984 and 2001-2016), and additional records from three other basin outlet hydrological stations: Nankuanping Hydrological Station (2003-2009), Xiangjiaping Hydrological Station (2003-2009), and Tanjiaba Hydrological Station (2007-2013).

2.3. Coupled model construction

This study proposes a coupled framework integrating the SWAT model with the HydroTransformer to enhance daily runoff prediction, particularly in data-scarce regions via transfer learning. The architecture of the SWAT-HydroTransformer is illustrated in Fig. 2.

SWAT is a distributed hydrological model designed for continuous long-term simulations (Philip et al., 2007). For the Shiquan basin, the area was delineated into 35 sub-basins and further subdivided into 630 Hydrological Response Units (HRUs) based on land use, soil, and slope thresholds. In simulations of watershed hydrological processes, variables such as actual evapotranspiration (ET), potential evapotranspiration (PET), groundwater runoff (GWQ), lateral flow (LATQ), percolation (PERC), soil water content (SW), and surface runoff (SURQ) can serve as input variables to enhance Transformer models. The simulation periods (1971-1984 and 2000-2016) included a one-year warm-up period.

Compared to traditional RNNs models, the Transformer model demonstrates enhanced parallel computing capabilities and superior long-term dependency modeling. This enables the efficient capture of multi-scale correlations and periodic features within complex time series data (Su et al., 2025). Building on the Transformer architecture, we developed the HydroTransformer. The model configurations are as follows:

- Input Features: The model utilizes the seven daily hydrological variables from SWAT alongside six meteorological variables. All inputs were organized into a 31-day sliding window (current day and the preceding 30 days) and normalized using mean–variance scaling.
- Target Variable: To eliminate scale effects across different watersheds, the prediction target is defined the ratio of observed daily runoff (m^3/s) to the watershed area.
- Training: The dataset was partitioned into training, validation, and testing sets (6 : 2 : 2 ratio). Model parameters were optimized using the training set, while the validation set was used for model selection and convergence monitoring. The testing set was reserved exclusively for final performance evaluation. To reduce the effects of random weight initialization, the training process was repeated 150 times with different random seeds. The Huber loss function (Huber, 1964) (Huber Loss, Equation (1)) was utilized as the optimization objective. It is a piecewise-defined loss controlled by a threshold parameter δ, behaving quadratically for small errors ($\leq \delta$) and linearly for large errors ($> \delta$), thereby balancing sensitivity to small deviations and robustness to outliers.

$$L_\delta(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2, for |y - f(x)| \leq \delta \\ \delta \bullet \left(|y - f(x)| - \frac{1}{2}\delta \right) for |y - f(x)| > \delta \end{cases} \quad (1)$$

here, δ is the parameter of Huber Loss, y is the true value, and $f(x)$ is the model's predicted value.

- Feature Enhancement: To better capture flow concentration inertia and cumulative precipitation effects, lagged runoff terms (Q_lag1, Q_lag3, Q_lag7) and multi-timescale accumulated precipitation terms (PCP_sum3, PCP_sum7, PCP_sum30) were incorporated (Ma

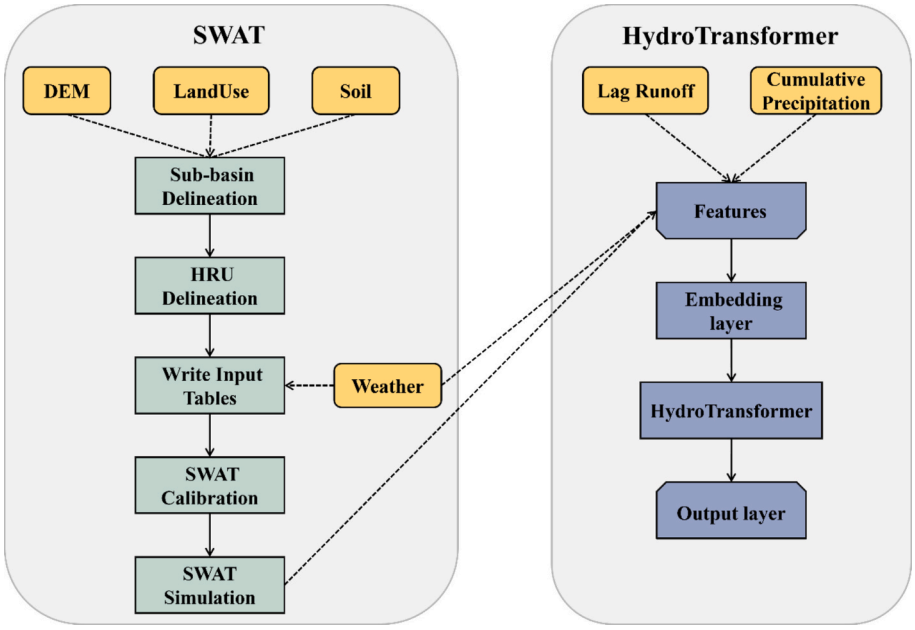

Fig. 2. Architecture of the coupled SWAT-HydroTransformer model.

et al., 2024a). To enable the model to capture seasonal and periodic patterns in the hydrological time series, we included time-encoding features representing daily, weekly, and monthly cycles, generated using sine and cosine transformations (sin_day, cos_day, sin_week, cos_week, sin_month, cos_month).

- Evaluation & Interpretation: Model performance was assessed using the Nash-Sutcliffe Efficiency (NSE, Equation (2)). Furthermore, the SHAP (Lundberg and Lee, 2017) method was employed to quantify feature importance and interpret the model's input–output mechanisms. We also verified the physical consistency of the model-derived trend and seasonal components using Power Spectral Density (PSD) and Pearson correlation analysis (Hyndman and Athanasopoulos, 2018). PSD was estimated using the Welch method to characterize the frequency-domain properties of the decomposed components. In this procedure, each time series was divided into overlapping segments, and the PSD was obtained by averaging the squared magnitudes of the corresponding Fourier spectra. This frequency-domain analysis served as a diagnostic tool to evaluate the separation between higher-frequency variability in the seasonal component and low-frequency variability in the trend component. In addition, Pearson correlation coefficients were calculated between the model-derived trend and seasonal components and the input hydrological and meteorological variables (PET, ET, SW, PERC, SURQ, GWQ, LATQ, PCP, TEMP_min, TEMP_max, RH, WS, and SR) to assess the physical interpretability of the learned representations.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \quad (2)$$

here, n is the sample size, y_i is the observed value, $\hat{y}_i$ is the model predicted value, and $\overline{y_i}$ is the mean of the observed values.

2.4. HydroTransformer model framework

We developed a multi-scale HydroTransformer for hydrological time series forecasting (Fig. 3), building upon the trend-seasonal decomposition and AutoCorrelation mechanisms originally introduced in Autoformer (Wu et al., 2021).

Specifically, while retaining the general decomposition structure of Autoformer, we reinterpret and redesign the trend and seasonal modeling components from a hydrological perspective. The trend branch is designed to represent long-term runoff evolution and cumulative hydrological effects, whereas the seasonal branch focuses on short-term variability and periodic fluctuations driven by climate forcing. To this end, the trend component is modeled using a multi-kernel convolutional network to capture runoff responses across different temporal accumulation scales, and the seasonal component is enhanced by explicitly integrating an annual-cycle constraint into the AutoCorrelation mechanism. These hydrology-oriented adaptations allow the model to better represent the multi-scale accumulation behavior and strong seasonality that characterize hydrological systems.

In the embedding layer, Raw hydro-meteorological sequences are normalized and enriched with temporal embeddings (encoding date, day of year, month, etc.) to inject explicit temporal context (Irani et al., 2025). A sliding window approach with concatenation is used to construct fixed-length time-series samples, in which multivariate hydrological and meteorological variables from consecutive time steps were concatenated within each window. This design allows the model to capture short-term temporal dependencies and cumulative hydrological effects while ensuring a consistent input structure for the Transformer-based framework. The multi-scale decomposition module explicitly separates trend and seasonality in the input sequence through convolutional average pooling at various scales. The trend component is processed by a dual-channel trend network, which consists of

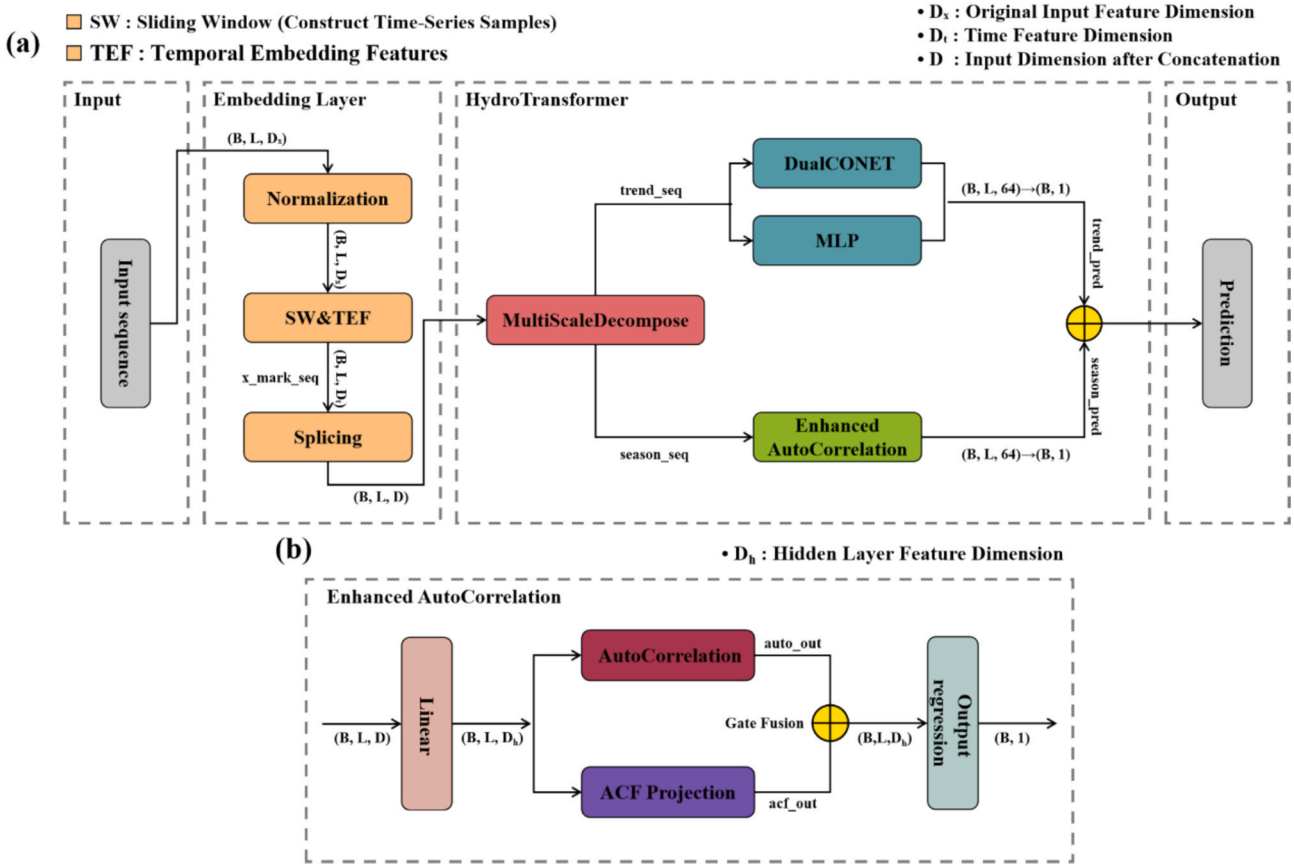

Fig. 3. Framework of the HydroTransformer model. (a) Overall model architecture; (b) Detailed structure of the Seasonal Branch (Enhanced AutoCorrelation).

DualCONET and a Multi-Layer Perceptron (MLP). DualCONET is a multi-scale convolutional module that captures smooth long-term trends and medium-term nonlinear variations through parallel one-dimensional convolutions with different kernel sizes, which are adaptively fused via a gating mechanism. In parallel, the MLP is implemented as a lightweight two-layer fully connected network to model global nonlinear relationships in the trend sequence. Both components are trained jointly in an end-to-end manner, without separate architecture optimization. This structure simultaneously extracts smooth long-term trends and medium-term nonlinear fluctuations (Wang et al., 2025b). The seasonal component is analyzed using an enhanced AutoCorrelation module that utilizes autocorrelation-based delay aggregation and Autocorrelation Function (ACF) projection mechanisms. This module effectively captures complex periodic patterns and temporal dependencies (Su and Guan, 2025), improving the prediction of short-term runoff fluctuations. Finally, the outputs from the trend and seasonal branches are integrated to generate the final daily runoff prediction. This module retains both the gradual evolution characteristics and the rapid variation components of runoff.

2.5. Transfer learning strategy for data-scarce basins

Data scarcity significantly hinders accurate hydrological modeling in many mountainous regions, making it challenging to fully characterize the runoff generation process (Pabasara et al., 2024). Transfer learning, a machine learning approach, facilitates the transfer of existing knowledge from data-rich source domains to data-limited target domains (Zhuang et al., 2021).

To address this, we employed a transfer learning strategy to migrate knowledge from the data-rich Shiquan basin (source domain) to the data-limited Xiangjiaping, Nankuanping, and Tanjiaba basins (target domains). The Xiangjiaping basin and Nankuanping basin exhibit high similarity to the Shiquan basin in terms of spatial location, climatic characteristics, and underlying surface conditions. The Tanjiaba basin is part of the Jialing River system and shows distinct hydrological characteristics. A unified transfer-and-fine-tuning framework was applied to all target basins without basin-specific adaptation, enabling a consistent evaluation of cross-basin transferability under varying degrees of hydrological similarity. Recognizing that runoff generation mechanisms can shift due to climate change and landscape evolution (Liu et al., 2023), we pre-trained two distinct models on the Shiquan basin: one representing a historically wet period and another representing a relatively dry period. These period-specific models were then transferred to the target basins to evaluate their cross-regional applicability and generalization capabilities under diverse climatic conditions.

Specifically, two transfer strategies were implemented:

- Direct Transfer: Two period-specific HydroTransformer models pretrained on the source basin were directly applied to each target basin without additional training.
- Fine-Tuning Transfer: Two period-specific HydroTransformer models pretrained on the source basin were fine-tuned using a subset of observational data from each target basin. The first 15% of the target basin training period was selected as a contiguous chronological block for fine-tuning. During fine-tuning, the lower encoder layers of the pretrained model were frozen, and only the trend-seasonality modeling layers were updated.

To further assess the physical consistency of climatic forcing among basins prior to model transfer, Principal Component Analysis (PCA) was applied to the daily meteorological drivers of all study basins. PCA is a dimensionality reduction technique that transforms correlated variables into orthogonal principal components while retaining the dominant variance structure (Rahman and Rahman, 2020). In this study, daily precipitation, temperature, relative humidity, wind speed, and solar radiation were standardized before analysis. The projection of meteorological variables into a low-dimensional principal component space was used to evaluate the similarity of climatic driving structures among the source and target basins.

3. Result

3.1. Daily runoff simulation in the Shiquan River basin

To evaluate the HydroTransformer model under varying hydrological conditions, we selected two distinct time frames for the Shiquan River basin: Period I (1972–1984) and Period II (2001–2016). The model's performance was benchmarked against the physically-based SWAT model, the Recurrent Neural Network-based LSTM, ensemble learning methods (Random Forest and XGBoost), and the standard Transformer architecture. These models have been extensively applied in hydrological prediction and time series modeling and are therefore considered representative baselines for a fair and systematic performance comparison.

As illustrated in Fig. 4a, the HydroTransformer exhibited a strong correlation between simulated and observed hydrographs in Period I. The model achieved NSE values of 0.860 during training and 0.829 during validation. Visually, the model attempted to characterize runoff processes during dry seasons and captured the majority of flood peak events, although noticeable deviations occurred for extreme peak values, particularly during the validation period. When compared to benchmark models (Table 1), HydroTransformer demonstrated optimal fitting capability on the test set, achieving an NSE of 0.843, significantly outperforming SWAT, LSTM, RF, XGBoost, and the standard Transformer.

In Period II, the simulation efficiency of all models generally declined (Table 1). This reduction may be related to changes in catchment characteristics and intensified human activities, which have been reported to significantly alter vegetation dynamics and associated land-surface conditions in the Qinba Mountains region (Liu et al., 2024), potentially increasing hydrological variability. During Period II, the HydroTransformer achieved an NSE of 0.786, exhibiting better performance than all other data-driven and physically-based models considered. While its ability to replicate certain extreme flood peaks saw a slight reduction compared to Period I (Fig. 4b), it exhibited comparatively stable performance during dry seasons (Fig. 5). This resilience underscores the model's strong generalization capability and robustness in handling observed variations in runoff patterns, including the timing, magnitude, and frequency of flow events seen across different periods and hydrological conditions.

Overall, HydroTransformer demonstrat superior capabilities in simulating daily runoff when compared to traditional mechanistic models and conventional data-driven models across various periods characterized by diverse hydrological conditions.

3.2. Mechanism analysis: rationality and effectiveness

To understand the physical interpretability of the HydroTransformer, we analyzed the spectral characteristics and feature importance of the model's internal components. PSD analysis utilizing the Welch method revealed distinct frequency characteristics for the model's trend and seasonal components (Fig. 6). For the Seasonal Component, energy was concentrated in the high-frequency range ($0-0.50day^{-1}$), with a dominant frequency of approximately $0.0027day^{-1}$, corresponding to a cycle of about 365 days. This aligns with the primary periodicity of precipitation. For the Trend Component, energy was focused in the low-frequency range ($0-0.09day^{-1}$), representing slower variations.

Correlation analysis (Table 2) of the model's decomposed components further confirmed the physical consistency of the trend and seasonal components. The seasonal component showed significant correlations ($p < 0.01$) with fast-response variables like PCP and SURQ.

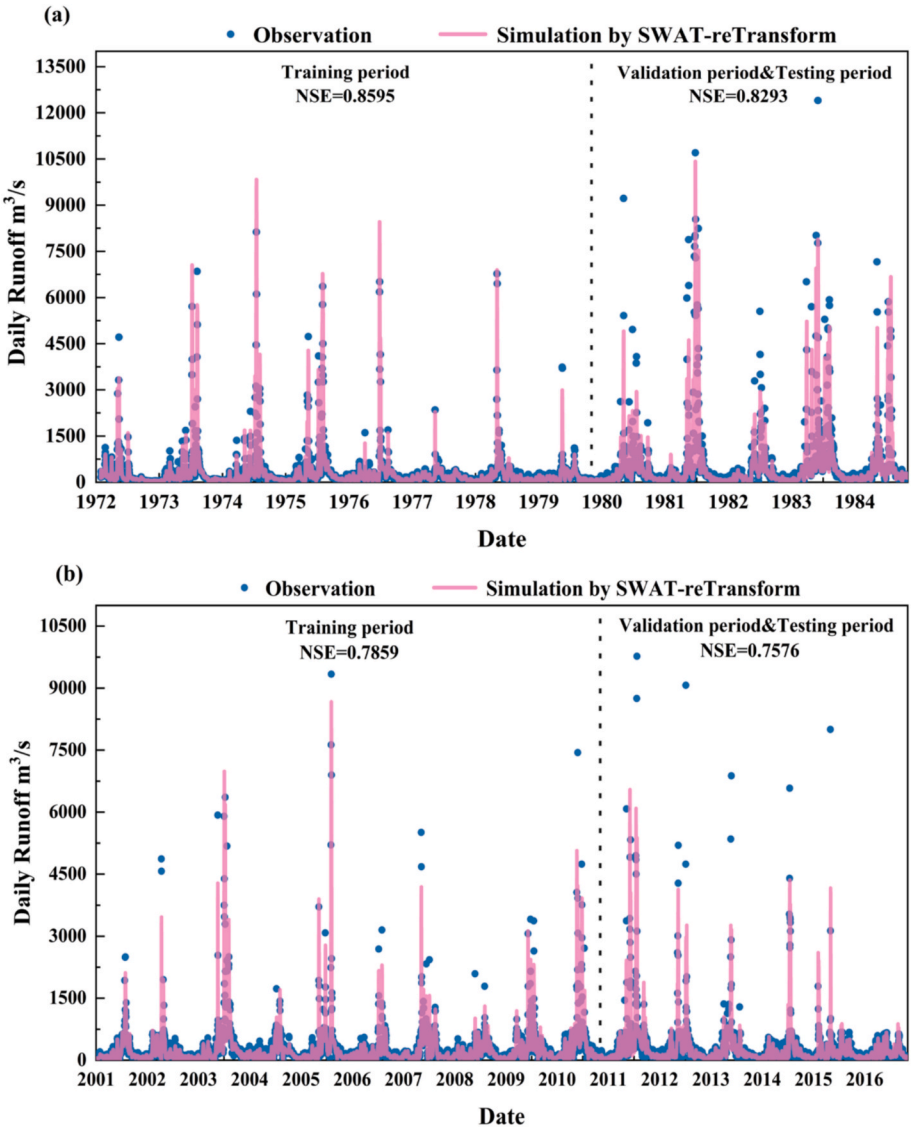

Fig. 4. Comparative evaluation of daily runoff simulations. (a) Daily runoff simulation results for Period I (1972–1984); (b) Daily runoff simulation results for Period II (2001–2016). Blue dots represent observed data; the pink line represents the simulation by the coupled SWAT-HydroTransformer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Comparative performance of hydrological models in the Shiquan basin on the test set.

Model/Period	Period I (Tset NSE)	Period II (Test NSE)
SWAT (SWAT-CUP calibrated)	0.640	0.720
RF	0.700	0.736
XGBoost	0.694	0.743
Transformer	0.726	0.750
LSTM	0.781	0.758
HydroTransformer	0.843	0.786

Note: With the exception of SWAT, all deep learning models were executed 150 times, with the best-performing iteration selected for reporting. The SWAT model was calibrated using SWAT-CUP prior to all experiments, and the same calibrated parameter set was used for both standalone and coupled simulations.

Conversely, the trend component correlated significantly ($p < 0.01$) with slow-response variables such as LATQ, SW, and GWQ. This confirms that the model attempts to distinguish between precipitation-driven high-frequency fluctuations and gradual processes related to watershed storage and baseflow.

We employed SHAP to quantify the marginal contribution of input features (Fig. 7). Across both periods, the dominant predictors were runoff lag terms, reflecting strong temporal memory, together with key meteorological variables such as precipitation, temperature, and radiation. In Period I, feature contributions exhibited a pronounced autocorrelation pattern: Q_lag1 was the strongest predictor (30.00%), followed by SW (6.81%), PCP (5.82%), TEMP_min (5.45%), and Q_lag3 (4.98%). Meteorological factors collectively accounted for only 33.94%. Hydrological statistics indicated a relatively high mean annual runoff ($0.016 m^3/s/km^2$) and flood-peak threshold ($0.051 m^3/s/km^2$), together with a large daily runoff coefficient of variation (2.200), reflecting strong flow fluctuations and frequent peak events. Under these highly variable conditions, the model understandably relied more on short-term flow memory to capture rapid runoff responses. In Period II, Q_lag1 remained the most influential variable (20.76%), but the combined contribution of meteorological factors increased markedly to 45.87%, driven by stronger effects of temperature-, precipitation-, and evapotranspiration-related variables (e.g., TEMP_min, PCP, ET, PET, WS, SR). This period exhibited lower mean annual runoff ($0.012 m^3/s/km^2$), a reduced wet–dry ratio (2247.9→1119.6), and

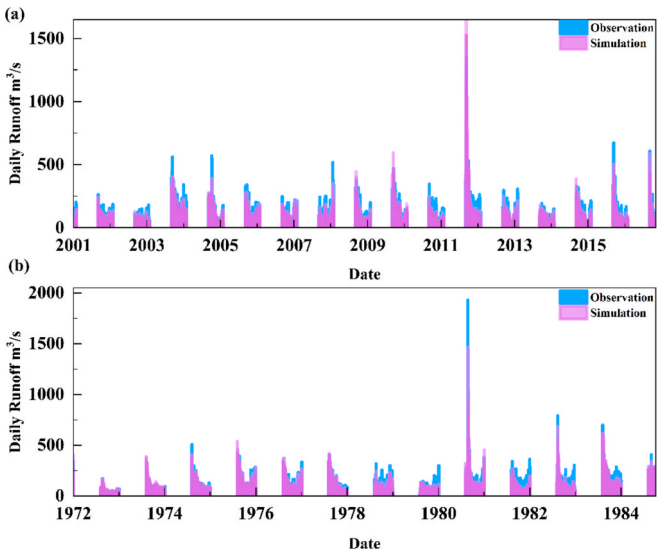

Fig. 5. Comparative evaluation of dry-season daily runoff simulations. (a) Dry-season runoff simulation results for Period I (1972–1984); (b) Dry-season runoff simulation results for Period II (2001–2016). Blue represent observed data; the pink represents the predicted results by the coupled SWAT–HydroTransformer model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

slightly diminished interannual precipitation variability (0.230→0.207), indicating a more stable flow regime. As runoff variability decreased, the predictive power of lagged flow weakened, thereby elevating the role of concurrent meteorological drivers. This shift is consistent with regional climate warming, intensified evapotranspiration, and changes in precipitation distribution.

This analysis demonstrates that the HydroTransformer adaptively shifts its attention mechanism: it prioritizes flow memory during active wet periods and shifts focus to meteorological drivers during drier, evaporation-dominated periods.

3.3. Transfer learning in data-scarce basins

We evaluated the model's cross-regional applicability through two stages: Direct Transfer and Fine-Tuning Transfer. As shown in Table 3, direct transfer performance was highly sensitive to the hydrological similarity between the source (Shiquan) and target basins. For the Xiangjiaping basin, the Period II model achieved substantially better performance ($NSE = 0.684$) than the Period I model ($NSE = 0.355$). In contrast, the Nankuanping basin performed better with the Period I model ($NSE = 0.794$ vs $NSE = 0.695$). The Tanjiaba basin yielded negative NSE values under both transfers (-1.134 and -0.829), indicating that direct transfer fails when runoff generation mechanisms differ significantly.

To address the limitations of direct transfer, we employed a fine-tuning strategy. We froze the lower encoder layers of the pre-trained Shiquan model and updated only the trend-seasonality modeling layers. Fig. 8 illustrates a substantial improvement in performance following fine-tuning. The goodness of fit is quantified using the coefficient of determination (R^2). The NSE values after fine-tuning reached 0.524 and 0.729 in the Xiangjiaping basin, and 0.826 and 0.742 in the Nankuanping basin. Although the Tanjiaba basin exhibited poor performance during direct transfer, its NSE notably increased to 0.546 and 0.634 following fine-tuning.

These results indicate that the HydroTransformer learns generalized hydrological representations that can be rapidly adapted to new regions with minimal data, validating its potential for hydrological forecasting in ungauged or data-scarce mountainous regions.

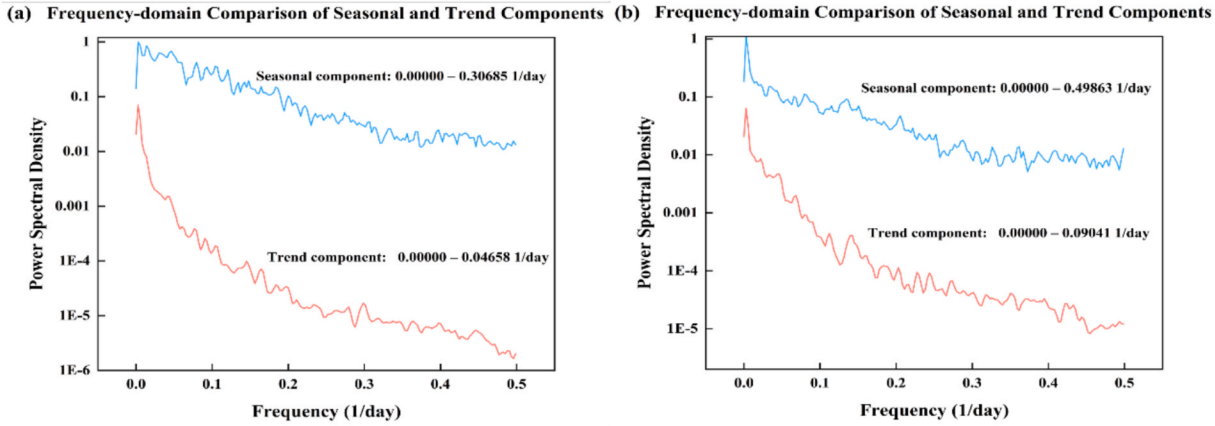

Fig. 6. Frequency domain analysis of model components. (a) Power Spectral Density for Period I (1972–1984); (b) Power Spectral Density for Period II (2001–2016). The blue line indicates the seasonal component (high frequency), and the red line indicates the trend component (low frequency). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Pearson correlation coefficients between model components (trend and seasonal components) and key hydrological variables (Top 5 significant variables, $p < 0.01$).

Season_pred				Trend_pred			
Period I		Period II		Period I		Period II	
PERC	0.482	RH	0.310	LATQ	0.903	LATQ	0.852
SURQ	0.374	PCP	0.307	SW	0.581	GWQ	0.639
LATQ	0.311	TEMP_min	0.286	RH	0.523	PERC	0.472
PCP	0.310	SURQ	0.257	TEMP_min	0.480	SW	0.462
RH	0.180	PERC	0.250	PERC	0.480	RH	0.459

Note: Variables are ranked by the absolute value of their correlation coefficients.

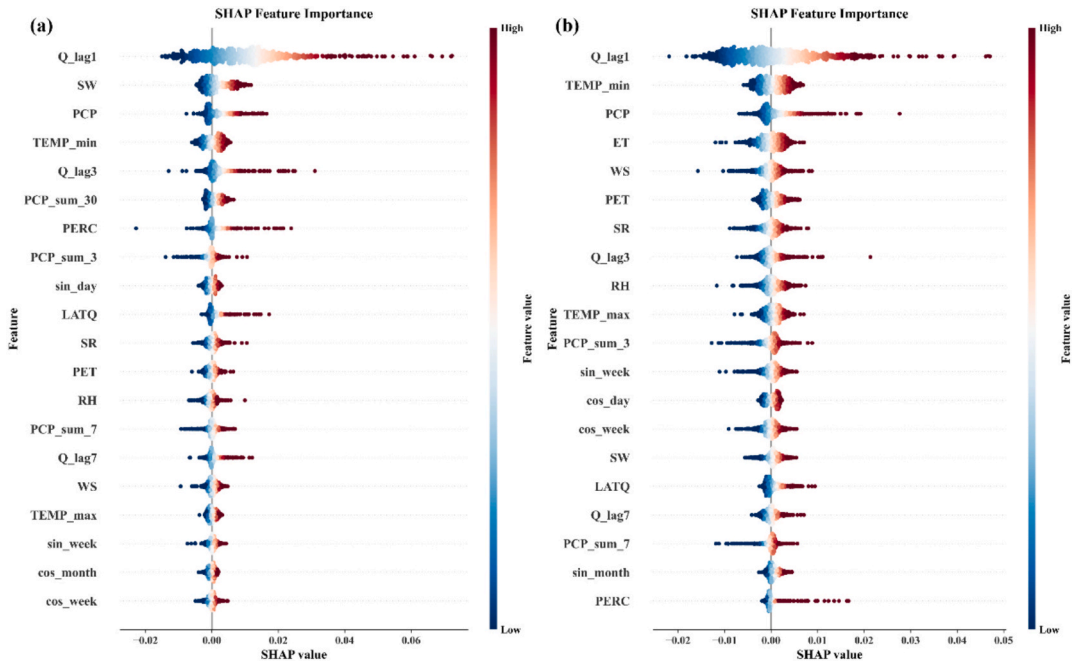

Fig. 7. SHAP beeswarm plots illustrating feature importance. (a) Feature contributions for Period I (1972–1984); (b) Feature contributions for Period II (2001–2016).

Table 3
Performance of direct transfer across basins.

Watershed/Period	Period I (NSE)	Period II (NSE)
the Xiangjiaping Basin	0.355	0.684
the Nankuanping Basin	0.794	0.695
the Tanjiaba Basin	−1.134	−0.829

Note: Target basin data were standardized using source basin statistics.

4. Discussion

4.1. Performance advantages in mountainous basins

In the Shiquan River basin, the HydroTransformer outperformed all benchmark models—including SWAT, LSTM, RF, XGBoost, and the standard Transformer—achieving NSE values of 0.843 in Period I and 0.786 in Period II (as also indicated in the context of Table 1 in the results section). The relatively lower performance in Period II may be partly attributed to limitations in capturing rapid runoff responses under intensified hydrological variability.

The HydroTransformer's primary advantage lies in its architectural ability to decouple and reconstruct multi-scale hydrological processes. Unlike many deep learning models that struggle to balance the capture of rapid flood peaks with the simulation of low-flow dry seasons, our model maintains robustness across both regimes. This stability is attributed to the integration of trend-seasonal decomposition and multi-scale convolutional smoothing. By explicitly modeling temporal statistical patterns, the network concurrently captures short-term storm responses, seasonal periodicities, and multi-year trends. This structural innovation mitigates the forgetting issues often observed in conventional deep models during extreme events or abrupt hydrological regime shifts (Yu et al., 2024). Consequently, the coupled SWAT-HydroTransformer offers a more reliable and interpretable tool for water resource management and flood warning in complex mountainous regions.

4.2. Implications of explicit trend and seasonal representation in runoff modeling

Recent advances in hydrological modeling highlight the importance of explicitly representing trend and seasonal structures in runoff prediction, particularly under nonstationary climatic conditions.

Hybrid SWAT-Transformer frameworks demonstrate that integrating process-based simulations with deep learning can improve prediction accuracy, yet trend and seasonal components are often corrected implicitly rather than structurally separated (Tao et al., 2024). Attention-based neural networks have also been applied to large-sample rainfall-runoff simulation to capture long-range dependencies and basin heterogeneity (Tabas and Samadi, 2024). By modeling cross-temporal interactions, such architectures enhance the representation of low-frequency trends and seasonal persistence, which are essential for robust evaluation across different hydro-climatic regimes. Investigations of groundwater drought dynamics further reveal that hydrological responses involve lagged and multi-scale trend evolution, underscoring the importance of distinguishing long-term variability from intra-annual fluctuations in model assessment (Dikici and Burgan, 2025). Reviews focusing on mountainous hydrology similarly highlight that nonlinear seasonal processes—such as snowmelt and storage-controlled baseflow—require modeling frameworks capable of representing both interannual trends and seasonal asymmetry (Rijal et al., 2025).

Collectively, these studies suggest that improved model evaluation should assess how effectively predictive frameworks capture both low-frequency trend dynamics and seasonal variability, rather than focusing exclusively on overall efficiency indices. In this context, the present study introduces an explicit seasonal-trend decomposition within a Transformer-based architecture, enabling a structurally informed evaluation of interannual and intra-annual variability.

4.3. Feasibility of cross-basin transfer learning

The explicit separation of long-term trends and short-term fluctuations within the HydroTransformer architecture significantly facilitates cross-basin transferability. The trend component effectively encapsulates regional climatic controls, while the enhanced AutoCorrelation

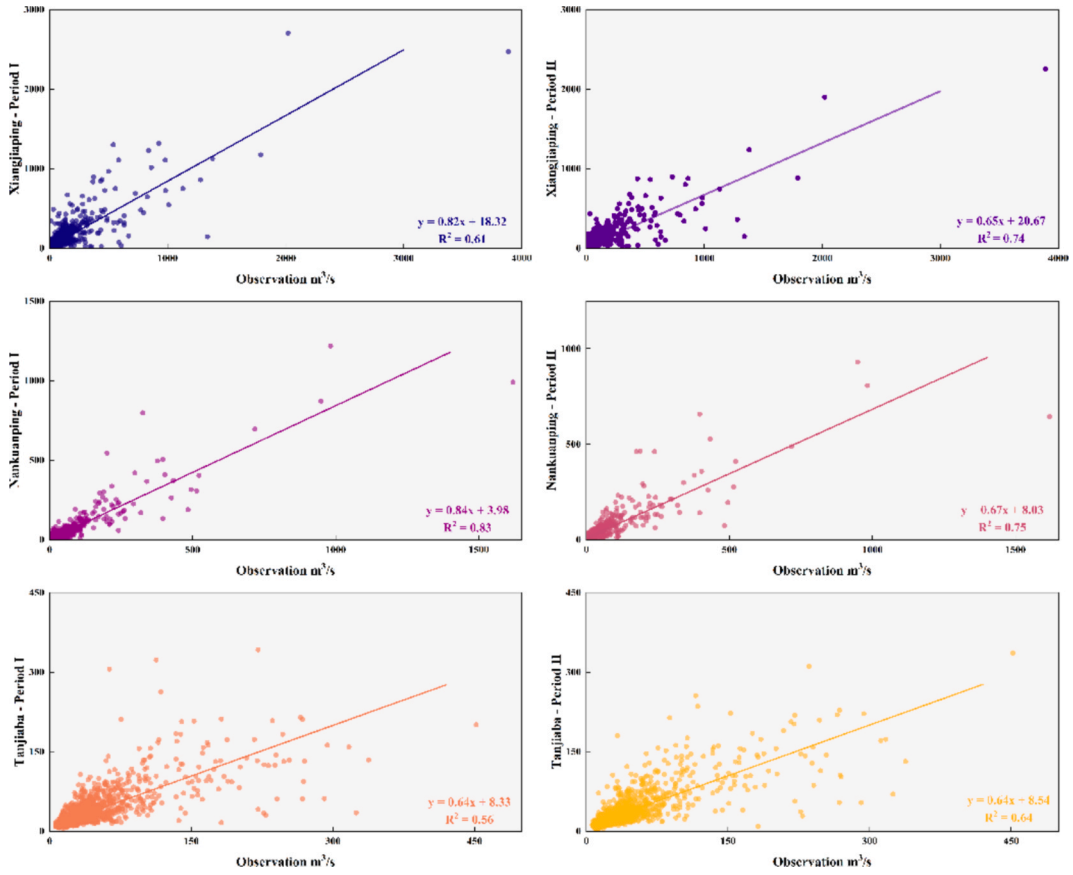

Fig. 8. Correlation analysis of Fine-Tuning Transfer results. Scatter plots compare simulated versus observed runoff after fine-tuning the model on target watersheds.

module captures universal event similarities and teleconnection structures. This allows the model to leverage hydrological response patterns learned from data-rich source basins to improve predictions in data-scarce target basins (Ma et al., 2024b).

To validate the physical basis for this transfer, we conducted PCA on the daily meteorological drivers across all study basins (Fig. 9). The projection of meteorological variables onto a low-dimensional space revealed significant overlap between the source (the Shiquan basin) and transfer basins. This indicates that these basins share a homogeneous climatic driving structure, providing a necessary climatic prerequisite for model transfer.

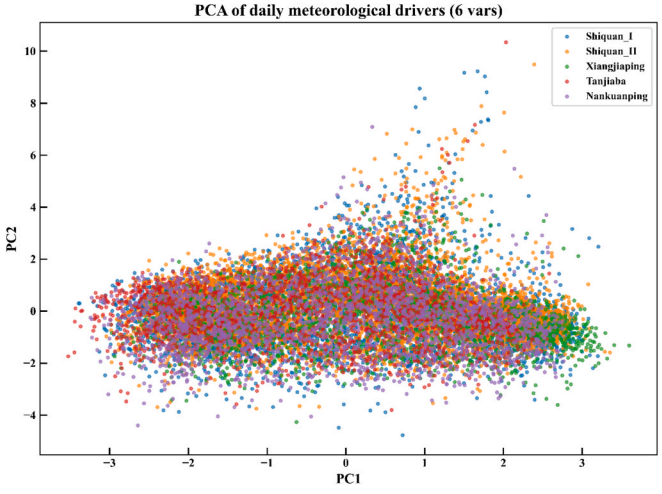

Fig. 9. Principal Component Analysis (PCA) of meteorological drivers. This plot projects daily meteorological variables (precipitation, temperature, humidity, wind speed, radiation) from all basins and periods onto the first two principal components (PC1 and PC2). The significant overlap indicates climatic homogeneity across the regions.

Table 4 summarizes the DEM-derived physical attributes of the source and target basins. All basins are characterized by steep mean slopes ($21-29^{\circ}$) and large elevation ranges ($>2000m$), reflecting similar geomorphological controls on runoff generation and flow concentration in mountainous environments. Although the drainage area of the source basin is larger, the target basins fall within the same mountainous hydrological regime. Importantly, the source basin selection relied exclusively on these DEM-derived physical attributes, ensuring that the transfer learning framework is established under physically consistent conditions rather than hydrological performance.

However, physical and climatic similarity alone does not necessarily

Table 4
DEM-derived physical attributes of the source and target basins.

Watershed	Area (km^2)	Mean Slope ($^{\circ}$)	Elevation Min (m)	Elevation Max (m)	Elevation Range (m)
the Shiquan Baisn	24985.6	23.16	344.0	3557.0	3213.0
the Xiangjiaping Basin	6310.2	28.54	222.0	2948.0	2726.0
the Nankuanping Basin	4083.4	27.11	286.0	2418.0	2132.0
the Tanjiaba Basin	9749.1	21.1	824.0	3542.0	2718.0

Note: All attributes were derived exclusively from DEM data and used to characterize the physical similarity between the source and target basins, without involving observed runoff information.

guarantee optimal transfer performance. To interpret the variability in transfer learning outcomes, we conducted a post-hoc hydrological similarity analysis using a station-based Pearson correlation matrix computed from daily runoff time series (Fig. 10). The results indicate that the Xiangjiaping basin maintains consistently high daily runoff similarity with the Shiquan basin across both periods, while the Nankuanping basin exhibits comparable similarity (0.38 in Period I and 0.39 in Period II). Notably, the observed transfer performance is broadly consistent with these hydrological similarity patterns, with basins exhibiting higher post-hoc hydrological similarity generally achieving better transfer accuracy (Xu et al., 2023). This analysis does not inform source basin selection, but rather provides insight into how basin-specific hydrological behavior influences the effectiveness of transfer learning.

4.4. Limitations and future research

Despite its robust performance, the current model has limitations. The decomposition mechanism relies on fixed-scale convolutional windows. While physically interpretable, this fixed scale may struggle to fully characterize non-stationary processes during extreme events or abrupt structural changes. Additionally, like all data-driven models, the HydroTransformer remains sensitive to the quality of input forcing data; significant biases in meteorological inputs can lead to unstable fitting.

Future research should focus on systematic evaluations across diverse climate zones, particularly in regions influenced by glacial melt or extreme aridity. Promising developmental directions include the integration of Physics-Informed Neural Networks (PINNs) to constrain mass balance, dynamic decomposition scale learning to adapt to changing frequencies, and multi-modal data fusion to incorporate satellite remote sensing data (Liu et al., 2022; Huynh et al., 2025; G.R. et al., 2024; Yang et al., 2021). These advancements would further enhance interpretability and reduce uncertainty in ungauged mountainous catchments.

5. Conclusion

This study developed the SWAT-HydroTransformer, a coupled hydrological model designed to address the challenges of runoff simulation in complex mountainous terrain. The model achieved better simulations in the Shiquan watershed, with NSE values of 0.843 (Period I) and 0.786 (Period II), significantly outperforming standard physical and deep learning benchmarks. Our analysis confirms that the model's superior performance stems from its multi-scale trend-seasonal decomposition architecture. By explicitly separating high-frequency storm responses from low-frequency trends, the model effectively captures the complex dynamics of mountainous hydrology, as evidenced by spectral and SHAP feature analysis. Furthermore, the study demonstrated the efficacy of fine-tuning transfer learning for data-scarce regions. By transferring knowledge from the source basin, the model achieved effective runoff predictions in three ungauged target basins, with optimal NSE values reaching 0.729, 0.826, and 0.634 respectively. These results underscore that successful cross-basin transfer relies on a combination of homogeneous climatic drivers and high hydrological similarity between basins.

CRediT authorship contribution statement

Yi He: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology, Conceptualization. **Rui Yan:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis. **Yanhong Tang:** Software, Formal analysis. **Jinyu Zhang:** Data curation. **Hui Qian:** Formal analysis. **Dejing Chen:** Methodology. **Lei Zhu:** Methodology. **Xin Cao:** Writing – review & editing, Supervision, Data curation, Conceptualization.

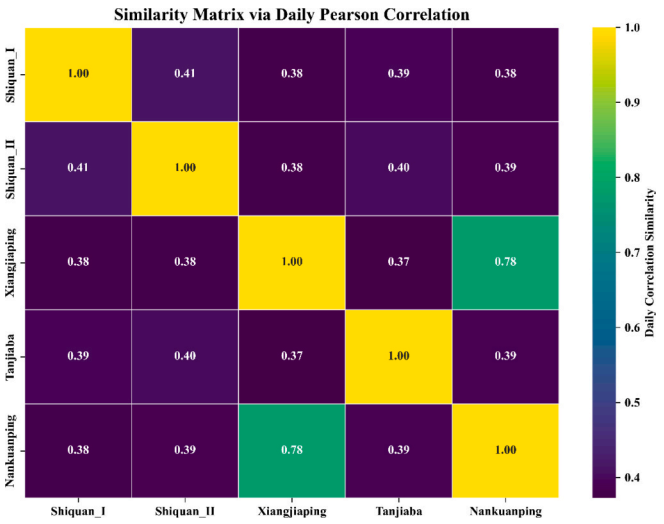

Fig. 10. Hydrological Similarity Matrix. This heatmap displays the average Pearson correlation coefficients between daily runoff series of different basins and periods. Higher values (yellow/green) indicate greater hydrological similarity, which correlates with improved transfer learning performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was jointly supported by the Special Funds of the National Natural Science Foundation of China (Grant No. 42341102) and the National Science and Technology Basic Resource Investigation Program (Grant No. 2017FY100904).

Data availability

Data cannot be made publicly available; readers should contact the corresponding author for details.

References

Ausloos, M., Cerqueti, R., Lupi, C., 2017. Long-range properties and data validity for hydrogeological time series: the case of the Paglia river. Physica A 470, 39–50. https://doi.org/10.1016/j.physa.2016.11.137.

Chang, F.J., Guo, S.L., 2020. Advances in hydrologic forecasts and water resources management. Water 12 (6), 1819. https://doi.org/10.3390/w12061819.

Chen, S.Y., Huang, J.L., Huang, J.C., 2023. Improving daily streamflow simulations for data-scarce watersheds using the coupled SWAT-LSTM approach. J. Hydrol. 622, 129734. https://doi.org/10.1016/j.jhydrol.2023.129734.

Chen, T., Liu, Z., Song, Z.J., Zhang, J.Y., Zhao, W.D., Dong, Q.Y., Jiang, J.X., Zhou, L., Ao, T.Q., 2025. LSTM-based runoff forecasting using multiple variables: a case study of the Nyang River, a typical basin on the Tibetan Plateau. Water 17, 1465. https://doi.org/10.3390/w17101465.

Demiray, B.Z., Sit, M., Mermer, O., Demir, I., 2024. Enhancing hydrological modeling with transformers: a case study for 24-h streamflow prediction. Water Sci. Technol. 89 (9), 2326–2341. https://doi.org/10.2166/wst.2024.110.

Dikici, M., Burgan, H.I., 2025. Exploring the effect of meteorological and hydrological trends on groundwater drought index: the case of Seyhan Basin. Ecohydrology 18 (6), e70110. https://doi.org/10.1002/eco.70110.

Eldardiry, H., Hossain, F., 2021. The value of long-term streamflow forecasts in adaptive reservoir operation: the case of the high aswan dam in the transboundary Nile River Basin. J. Hydrometeorol. 22 (5), 1099–1115. https://doi.org/10.1175/JHM-D-20-0241.1.

Evin, G., Lay, M.L., Fouchier, C., Penot, D., Colleoni, F., Mas, A., Garambois, P., Laurantin, O., 2024. Evaluation of hydrological models on small mountainous catchments: impact of the meteorological forcings. Hydrol. Earth Syst. Sci. 28 (1), 261–281. https://doi.org/10.5194/hess-28-261-2024.

Frank, C., Russwurm, M., Fluixa-Sanmartin, J., Tuia, D., 2023. Short-term runoff forecasting in an alpine catchment with a long short-term memory neural network. Front. Water 5, 1126310. https://doi.org/10.3389/frwa.2023.1126310.
G.R., A.N., S A., El-Shafie, A., Ahmed, A.N., 2024. Enhancing hydrological data completeness: a performance evaluation of various machine learning techniques using probabilistic fusion imputer with neural networks for streamflow data reconstruction. J. Hydrol. 639, 131583. https://doi.org/10.1016/j.jhydrol.2024.131583.
Hameed, M.M., Masood, A., Hamid, A., Elbeltagi, A., Razali, S.F.M., Salem, A., 2025. Forecasting monthly runoff in a glacierized catchment: a comparison of extreme gradient boosting (XGBoost) and deep learning models. PLoS One 20 (5), e0321008. https://doi.org/10.1371/journal.pone.0321008.
Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Statist. 35 (1), 73–101. https://doi.org/10.1214/aoms/1177703732.
Hui, Q., Luo, J.G., Hou, J., Cui, H.H., Zuo, G.G., 2025. Simulated division of flood processes in the composite terrain region based on the multi-layer hydrological process combination model. Sci. Rep. 15, 21805. https://doi.org/10.1038/s41598-025-06343-9.
Huynh, N.N.T., Garambois, P.A., Renard, B., Colleoni, F., Monnier, J., Roux, H., 2025. A distributed hybrid physics–AI framework for learning corrections of internal hydrological fluxes and enhancing high-resolution regionalized flood modeling. Hydrol. Earth Syst. Sci. 29 (15), 3589–3613. https://doi.org/10.5194/hess-29-3589-2025.
Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice, 2nd edition. OTexts, Melbourne, Australia. https://otexts.com/fpp2/, (accessed December 8, 2025.
Irani, H., Ghahremani, Y., Kermani, A., Metsis, V., 2025. Time series embedding methods for classification tasks: a review. Expert. Syst. 42 (11), e70148. https://doi.org/10.1111/exsy.70148.
Jin, C., Xu, Q., Chen, X., Cai, J., Meng, H., 2025. Enhancing runoff prediction accuracy of deep learning model using baseflow separation method and timestamp information. J. Hydrol. 662, 134044. https://doi.org/10.1016/j.jhydrol.2025.134044.
Koya, S.R., Roy, T., 2024. Temporal fusion transformers for streamflow prediction: value of combining attention with recurrence. J. Hydrol. 637, 131301. https://doi.org/10.1016/j.jhydrol.2024.131301.
Li, Z.J., Liu, P., Zhang, W., Chen, X.Z., Deng, C., 2014. Comparative application of the SWAT model and the Xin'anjiang model in the Xunhe River Basin of the Hanjiang River. Water Resour. Res. 3 (4), 307–314. https://doi.org/10.12677/JWRR.2014.34038 (in Chinese). https://www.hanspub.org/journal/paperinformation?paperid=13918, (accessed December 8, 2025).
Linh, H., Rajith, M., Moore, Karen E.B., Owens, Emmet M., Tammo Steenhuis, S., 2018. The effect of input data resolution and complexity on the uncertainty of hydrological predictions in a humid vegetated watershed. Hydrol. Earth Syst. Sci. 22 (11), 5947–5965. https://doi.org/10.5194/hess-22-5947-2018.
Liu, H.D., Li, M.J., Li, T.Q., Wu, L.Y., Zheng, H., 2024. Impacts of intensified human activity on vegetation dynamics in the Qinba Mountains, China. Forests 15 (9), 1561. https://doi.org/10.3390/f15091561.
Liu, Y., Guan, Z.L., Huang, T.T., Wang, C.C., Guan, R.H., Ma, X.Y., 2023. Combined effects of land use/cover change and climate change on runoff in the Jinghe River Basin, China. Atmosphere 14 (8), 1237. https://doi.org/10.3390/atmos14081237.
Liu, Y.Y., Kutz, J.N., Steven Brunton, L., 2022. Hierarchical deep learning of multiscale differential equation time-steppers. Philos. Trans. a Math Phys. Eng. Sci. 380 (2229), 20210200. https://doi.org/10.1098/rsta.2021.0200.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), pp. 4768–4777 https://dl.acm.org/doi/10.5555/3295222.3295230.
Lv, Y.B., 2020. Analysis of Hydrological Characteristics of Liaojiaba hydrological station on Xihan River. Shaanxi Water Resour. 5, 38–44. https://doi.org/10.16747/i.cnki.cn61-1109/tv.2020.05.014 (in Chinese). https://www.doc88.com/p-64661859632335.html, (accessed December 8, 2025).
Ma, K., He, D.M., Liu, S.Y., Ji, X., Li, Y.G., Jiang, H.R., 2024a. Novel time-lag informed deep learning framework for enhanced streamflow prediction and flood early warning in large-scale catchments. J. Hydrol. 631, 130841. https://doi.org/10.1016/j.jhydrol.2024.130841.
Ma, K., Shen, C.P., Xu, Z.Y., He, D.M., 2024b. Transfer learning framework for streamflow prediction in large-scale transboundary catchments: sensitivity analysis and applicability in data-scarce basins. J. Geog. Sci. 34, 963–984. https://doi.org/10.1007/s11442-024-2235-x.
Ma, R.J., An, Q., Liu, L., Cheng, Y.M., Liu, X.C., 2025. Runoff forecast model integrating time series decomposition and deep learning for the short term: a case study in the Weihe River Basin, China. Water 17 (18), 2718. https://doi.org/10.3390/w17182718.
Mahmood, T.H., Putkonen, J., Sobbe, A., 2021. Spatially variable precipitation and its influence on water balance in a headwater alpine basin. Nepal. Water. 13 (3), 254. https://doi.org/10.3390/w13030254.
Pabasara, K., Gunawardhana, L., Bamunawala, J., Sirisena, J., Rajapakse, L., 2024. Significance of multi-variable model calibration in hydrological simulations within data-scarce river basins: a case study in the dry-zone of Sri Lanka. Hydrology 11 (8), 116. https://doi.org/10.3390/hydrology11080116.
Pham, L.T., Luo, L.F., Finley, A., 2021. Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds. Hydrol. Earth Syst. Sci. 25 (6), 2997–3015. https://doi.org/10.5194/hess-25-2997-2021.
Rahman, A.S., Rahman, A., 2020. Application of principal component analysis and cluster analysis in regional flood frequency analysis: a case study in New South Wales, Australia. Water 12 (3), 781. https://doi.org/10.3390/w12030781.
Rijal, M., Luo, P.P., Mishra, B.K.M., Thakuri, S., Zhang, Y.B., Zhao, Y., 2025. Mountainous flood resilience: a comprehensive systematic review of flood analysis methods. Arch. Comput. Meth. Eng. https://doi.org/10.1007/s11831-025-10439-w.
Shi, C., Liu, Q., Bai, Y.G., Yu, Q.Y., Lu, Z.L., Liu, C.S., Cao, B., Ren, L., Li, M., Miao, G., Hu, C.H., 2025a. Improving runoff simulation in cold alpine regions based on VIC-glacier by combining LSTM error correction technology. J. Hydrol. 663, 134251. https://doi.org/10.1016/j.jhydrol.2025.134251.
Shi, W.Y., Wan, X.Y., Zhao, F.Z., Deng, R.X., 2025b. A dual-model framework combining nonlinear autoregressive with exogenous inputs (NARX) and LSTM networks for enhanced daily runoff prediction and error correction. Environ. Model. Software 192, 106570. https://doi.org/10.1016/j.envsoft.2025.106570.
Su, G.Y., Guan, Y.P., 2025. MSDformer: an autocorrelation transformer with multiscale decomposition for long-term multivariate time series forecasting. Appl. Intell. 55, 179. https://doi.org/10.1007/s10489-024-06105-6.
Su, L.Y.L., Zuo, X.M., Li, R., Wang, X., Zhao, H., Huang, B.D., 2025. A systematic review for transformer-based long-term series forecasting. Artif. Intell. Rev. 58, 80. https://doi.org/10.1007/s10462-024-11044-2.
Sun, W., Chang, L.C., Chang, F.J., 2024. Deep dive into predictive excellence: transformer's impact on groundwater level prediction. J. Hydrol. 636, 131250. https://doi.org/10.1016/j.jhydrol.2024.131250.
Szczepanek, R., 2022. Daily streamflow forecasting in mountainous catchment using XGBoost, LightGBM and CatBoost. Hydrology. 9, 226. https://doi.org/10.3390/hydrology9120226.
Tao, C., Xin, L., Liu, S.M., Li, H.Y., Xu, Z.W., Tan, J.L., Zhang, Y., Ren, Z.G., Xiao, L., Deng, J., Jin, R., Ma, M.G., Wang, J., Yang, X.F., 2019. Integrated hydrometeorological, snow and frozen-ground observations in the alpine region of the Heihe River Basin, China. Earth Syst. Sci. Data 11 (3), 1483–1499. https://doi.org/10.5194/essd-11-1483-2019.
Tabas, S.S., Samadi, S., 2024. Chapter 12 - Attention-based deep neural network for rainfall-runoff simulation across the continental United States. Intell. Syst. Earth Environ. Planet. Sci. 315–334. https://doi.org/10.1016/B978-0-443-13293-3.00003-8.
Tao, J.H., Gu, Y.C., Yin, X., Chen, J.L., Ao, T.Q., Zhang, J.Y., 2024. Coupling SWAT and transformer models for enhanced monthly streamflow prediction. Sustainability 16 (19), 8699. https://doi.org/10.3390/su16198699.
Philip, G., Manuel, R., Colleen, G., Jeffrey, A., 2007. The soil and water assessment tool: historical development, applications, and future research directions. Trans. ASABE (Am. Soc. Agric. Biol. Eng.). 50 (4), 1211–1250. https://doi.org/10.22004/AG.ECON.10285.
Wang, W.C., Ye, F.R., Wang, Y.Y., Gu, M., 2025a. A singular spectrum analysis-enhanced BiTCN-selfattention model for runoff prediction. Earth Sci. Inf. 18, 31. https://doi.org/10.1007/s12145-024-01524-y.
Wang, Z.J., Zhao, C.N., Huang, Y.Q., Zhang, J.H., Lan, J.L., 2025b. Tlsam: multi-scale long-short trend fusion model for long-term multi-variable time series prediction. Knowl. Inf. Syst. 67, 11193–11222. https://doi.org/10.1007/s10115-025-02448-9.
Wu, H.X., Xu, J.H., Wang, J.M., Long, M.S., 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. arXiv. 2106, 13008. doi:10.48550/arXiv.2106.13008.
Xu, Y.H., Lin, K.R., Hu, C.H., Wang, S.L., 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.
Yang, L., Meng, X.H., Karniadakis, G.E., 2021. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. J. Comput. Phys. 425 (15), 109913. https://doi.org/10.1016/j.jcp.2020.109913.
Yang, Y.P., 2015. Evaporation coefficient conversion analysis at the nankuanping hydrological station. Shaanxi Water Resour. 2 (2), 2025. https://doi.org/10.3969/j.issn.1673-9000.2015.02.007 (in Chinese). https://xueshu.baidu.com/usercenter/paper/show?paperid=2dce217d5c2392d5b287806396e1f777, (accessed December 8).
Yong, K.Q., Li, M.L., Xiao, P., Gao, B., Zheng, C.X., 2025. Monthly streamflow forecasting for the irtysh river based on a deep learning model combined with runoff decomposition. Water 17 (9), 1375. https://doi.org/10.3390/w17091375.
Yu, Y., Ma, R.Z., Ma, Z.M., 2024. Robformer: a robust decomposition transformer for long-term time series forecasting. Pattern Recogn. 153, 0031–3203. https://doi.org/10.1016/j.patcog.2024.110552.
Zhang, C.C., Zhou, Y.Y., Lu, F., Liu, J.W., Zhang, J.Y., Yin, Z.Y., Ji, M.Y., Li, B.Q., 2025. Assessing the performance and interpretability of the CNN-LSTM-Attention model for daily streamflow forecasting in typical basins of the eastern Qinghai-Tibet Plateau. Sci. Rep. 15, 82. https://doi.org/10.1038/s41598-024-84810-5.
Zhao, T.L., Fang, L.X., Ma, X., Li, X.M., Zhang, C.M., 2025. TFformer: a time–frequency domain bidirectional sequence-level attention based transformer for interpretable long-term sequence forecasting. Pattern Recogn. 158, 110994. https://doi.org/10.1016/j.patcog.2024.110994.
Zhao, X.F., Wang, H.Y., Bai, M.Y., Xu, Y.J., Dong, S.W., Rao, H., Ming, W.Y., 2024. A comprehensive review of methods for hydrological forecasting based on deep learning. Water 16 (10), 1407. https://doi.org/10.3390/w16101407.
Zhuang, F.Z., Qi, Z.Y., Duan, K.Y., Xi, D.B., Zhu, Y.C., Zhu, H.S., 2021. A comprehensive survey on transfer learning. Proc. IEEE 109 (1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555.