https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

Integrating Physical-Based Xinanjiang Model and Deep Learning for Interpretable Streamflow Simulation: A Multi-Source Data Fusion Approach across Diverse Chinese Basins

Zhaocai Wang[1], Nannan Xu[1], Wei Song[1,*], Xingxing Zhang[2], Junhao Wu[3] and Xi Chen[3,*]

[1]College of marine Ecology and Environment, Shanghai Ocean University, Shanghai, 201306, P. R. China

[2]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, P. R. China

[3]State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai, 200062, P. R. China

[*]Corresponding author: Wei Song (wsong@shou.edu.cn; ORCID: 0000-0002-0604-5563); Xi Chen (xchen@geo.ecnu.edu.cn)

Abstract: The simulation of streamflow is a complex task due to its intricate formation process. Existing single models struggle to accurately capture the stochastic, non-stationary, and nonlinear dynamics of basin streamflow in changing environments. This study integrated process-driven hydrological mechanism models with data-driven deep learning models, considering various factors like hydrology, meteorology, environment, and the interconnected effects of upstream and downstream rivers, to form an interpretable hybrid streamflow simulation model. The study collected multiple external variables to better understand the hydrologic system complexity and used the Maximum Information Coefficient (MIC) to analyze their relationship with streamflow. Subsequently, the Xinanjiang (XAJ) model with physical mechanisms was employed, alongside the TCN-GRU model integrating Temporal Convolutional Network (TCN) and Gated Recurrent Unit (GRU), for separate streamflow simulations. Furthermore, a combined method was employed, achieving nonlinear ensemble through Random Forest (RF), resulting in the hybrid XAJ-TCN-GRU model. This model shows promising results in simulating streamflow in four different basins in China, achieving high Nash-Sutcliffe Efficiency (NSE) values with 0.991, 0.971, 0.984, and 0.986 for the Wuding River, Chu River, Jianxi River, and Qingyi River respectively. In terms of streamflow simulation, flood simulating, and interval simulation, this model outperforms other benchmark models. Additionally, the study quantified the contributions of each hydro-meteorological variable to the long-term streamflow trend using mean absolute SHAP values (SHAPABS), Feature Importance (FI), and Permutation Feature Importance (PFI), thereby enhancing the model's external interpretability. The results of this study are of significant importance for optimizing water resource management and mitigating flood disasters.

Keywords: Streamflow simulation; Xinanjiang model; Deep learning; nonlinear ensemble; Physical mechanism; Interpretability analysis

1 Introduction

Water resources are among the most precious natural resources on Earth, and accurate streamflow simulation has become an indispensable component in hydraulic engineering and water resources management (Zhang et al., 2023). Rational and efficient streamflow simulation not only provides solid support for the sustainable development of society and the economy but also has profound implications

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

in flood warning and optimization of water resource scheduling (Shao et al., 2024). However, the complex nature of river hydrological environments renders the formation and dynamics of streamflow highly intricate processes (Liu et al., 2023; Ahmed et al., 2022). Therefore, exploring high-precision and reliable streamflow simulation models has become an urgent and practical issue in the field of hydrology (Thébault et al., 2024).

To date, methodologies for streamflow simulation are generally classified to physical-based (Gebremariam et al., 2014; Bai et al., 2017) and data-driven approaches (Gao et al., 2020; Vilaseca et al., 2023). Physical-based models represent conventional simulation methodologies that describe the physical mechanisms governing the transformation from rainfall to streamflow (Cheng et al., 2020). These approaches utilize mathematical equations and calibrate parameters for diverse locations to effectively simulate streamflow dynamics (Özgen-Xian et al., 2020; Leonarduzzi et al., 2021). Typical physical-based models include Precipitation Runoff Modeling System (PRMS) (Hwang et al., 2011), Soil and Water Assessment Tool (SWAT) (Oruç et al., 2023), and Xinanjiang (XAJ) model (Hao et al., 2015; Jiang et al., 2023). The XAJ model is used for simulating rainfall-runoff processes, which is proposed by Zhao et al. (1992) in China. The model is mainly used in humid and semi-humid regions, and is particularly suitable for flood simulation and water resource management (Lei et al., 2023). Nevertheless, the complexity and interrelations of hydrological systems inevitably introduce uncertainties, particularly concerning model parameter estimation and structure selection (Zuo et al., 2020).

Machine learning models exhibit flexibility in capturing and modeling nonlinear relationships, thus serving as effective tools for streamflow simulation (Zhu et al., 2023). Additionally, the structure and complexity of machine learning models can be adjusted according to specific research needs, allowing them to better adapt to streamflow simulation tasks at different scales and spatiotemporal ranges (Han and Morrison, 2022; Ahmadpour et al., 2022). Machine learning methods encompass various models, such as Random Forest (RF) (Contreras et al., 2021), Decision Tree (DT) (Jehanzaib et al., 2021), and Support Vector Machine (SVM) (Samantaray et al., 2022), among others. When addressing complex problems, machine learning models often increase parameters, complexity of structure, or introduce more features to enhance fitting ability. However, overly complex models may overfit the details and noise in the training data, raising the risk of overfitting.

The field of machine learning continues to evolve, and deep learning, as an important extension of it, has attracted considerable attention. Its core concept is to enable computers to learn and understand more complex abstract concepts by leveraging layered, stacked deep architectures (Chen et al., 2023; Ng et al., 2023). Deep learning models demonstrate strong generalization capabilities in streamflow simulation across different basins and periods, adapting well to diverse hydrological conditions (Xu et al., 2023; Wei et al., 2023). The Temporal Convolutional Network (TCN) significantly enhances the accuracy of streamflow simulation by effectively capturing long-term temporal dependencies within the data (Lin et al., 2020). Unlike traditional methods, TCN's capacity to process sequential data through causal convolutions enables it to model both short- and long-term patterns in streamflow, which is essential for accurately capturing the dynamics of runoff processes. Additionally, its use of residual connections promotes faster and more stable convergence during training, thereby reducing the risk of

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

vanishing gradients. This combination of strengths directly enhances the model's simulation capability, establishing TCN as an effective tool for more reliable and precise streamflow forecasts. The Gated Recurrent Unit (GRU) also plays a crucial role in streamflow simulation by adeptly capturing the sequential dependencies inherent in hydrological data, which is vital for modeling the variability of water flow over time (Zheng et al., 2023). Through its gating mechanism, GRU selectively updates and forgets information, allowing it to retain essential patterns while filtering out noise, thereby improving its accuracy in simulating fluctuating streamflow patterns. This ability to manage time-dependent features positions GRU as well-suited for capturing sudden changes and longer-term dependencies in streamflow data. Deep learning models are often considered "black boxes," with their internal parameters and decision processes challenging to interpret, which presents difficulties in understanding and validating whether the model accurately captures physical mechanisms (Katipoğlu and Sarıgöl, 2023). While deep learning models excel at learning patterns from large datasets, they lack explicit modeling of physical equations and laws. In scenarios where models must make reasonable simulations based on physical mechanisms, this limitation can lead to suboptimal performance.

Kim et al. (2021) pointed out that data-driven models (Artificial Neural Network (ANN) and LSTM) perform better in predicting peak flow conditions, whereas physical-based hydrological models perform better in low flow conditions. Since deep learning models have high predictive accuracy and physical models based on hydrological principles are explainable, some researchers choose to combine them to enhance both accuracy and insight in streamflow prediction (Kurian et al., 2020). Most studies use deep learning models to post-process the outputs of physical-based models (Cho and Kim, 2022; Parisouj et al., 2022). This approach provides relatively reasonable initial information, using physical-based models to generate inputs with physical significance, which helps guide the deep learning model (Granata et al., 2024). However, due to the limitations in assumptions and accuracy of these models, the initial inputs may contain certain errors, which are further amplified and propagated in deep learning models, ultimately affecting the reliability of the simulation results.

In order to maximize the combined benefits of the conceptual rainfall - runoff model and deep learning models for streamflow simulation and to overcome the challenges faced by single models in coping with complex hydrological simulating in environmental conditions, this study aims to develop a robust model for streamflow simulation by effectively combining conceptual rainfall - runoff model with deep learning models. Unlike previous studies that employed sequential coupling, this research focuses on nonlinear ensemble methods (Xuan et al., 2021; Wang and Dong, 2024). By combining the high interpretability features of conceptual rainfall - runoff model with the powerful data mining capabilities of machine learning models, the integrated hybrid model can address issues such as bias complementarity, enhanced generalization ability, and improved simulations of extreme events.

Therefore, this study runs the XAJ model and the TCN-GRU model in parallel and employs a combined approach, utilizing RF to perform nonlinear ensemble of the simulations from both models, successfully constructing a new hybrid model named XAJ-TCN-GRU. Additionally, various interpretability methods were employed to quantitatively analyze the impact of meteorological and hydrological factors on the hydrological processes. To validate the hybrid model's generalization across varied hydrological conditions, it was applied to four basins in China with different hydrological

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

characteristics for daily streamflow simulation. Furthermore, the robustness of the model is confirmed by simulation tests with the addition of "noise" data.

2 Methodologies

This section outlines the methods used in the study. Specific descriptions of calculation of MIC, noise data injection test, flood simulating, and interval simulation can be found in **Supplementary Information 2.**

2.1 Xinanjiang model (XAJ)

The XAJ model employs a three-source approach for streamflow calculation, utilizing surface streamflow routing through the unit hydrograph method, while separate linear storage reservoirs simulate the routing processes of interflow and baseflow (Zhao and Liu, 1995). Its main features include the application of storage excess concepts and Muskingum routing, characterized by subdivision into subunits, multiple sources, and routing stages. Additionally, the XAJ model possesses a simple structure, with parameters that have clear physical meanings and high computational accuracy, which has led to its widespread application in hydrological forecasting (Gong et al., 2021).

The XAJ model encompasses a substantial number of parameters, some of which are highly sensitive; even minor alterations can significantly affect the results, while others display a degree of inertia. **Table 1** presents detailed descriptions of the meanings and sensitivities of 15 parameters, which can be broadly classified four categories based on the model's structure and function: evapotranspiration parameters, flow rate parameters, water source division parameters, and confluence parameters.

Table 1 XAJ model parameter overview

Typology	Parameter	Description	Sensitivity
Evapotranspiration parameters	C	The coefficient of deep evapotranspiration	non-sensitive
	KC	Ratio of potential evapotranspiration to pan evaporation	sensitive
	WLM (mm)	Averaged soil moisture storage capacity of the lower layer	non-sensitive
	WUM (mm)	Averaged soil moisture storage capacity of the upper layer	non-sensitive
Flow rate parameters	B	Exponent of the tension water capacity curve	sensitive
	WM (mm)	Areal mean tension water capacity	sensitive
	IM	Percentage of impervious and saturated areas in the catchment	non-sensitive
Water source division parameters	EX	Exponent of the free water capacity curve	sensitive
	KG	Outflow coefficients of the free water storage to groundwater	sensitive
	KI	Outflow coefficients of the free water storage to interflow	sensitive
	SM (mm)	Areal mean of the free water capacity of the surface soil layer	sensitive
Confluence parameters	k (h)	Storage coefficient of linear reservoirs of Nash unit hydrograph	sensitive
	n	Number of linear reservoirs of Nash unit hydrograph	sensitive
	CG	Recession constants of the groundwater storage	non-sensitive
	CI	Recession constants of the lower interflow storage	non-sensitive

2.2 Temporal convolutional network (TCN)

TCN represents an architectural framework that combines one-dimensional convolutional networks with causal convolutions, efficiently capturing temporal dependencies in data and demonstrating enhanced suitability for addressing time-series challenges (Sun et al., 2024). Dilated convolutional

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

networks enable dilated sampling on the input from the preceding layer, allowing for the extraction of feature information from time-series data characterized by long intervals and non-continuous patterns. In the context of causal convolution, the output at any given moment is solely dependent on the input at that moment and any preceding inputs. TCN's core architecture is composed of dilated causal convolution, as illustrated in **Fig. 1**(a).

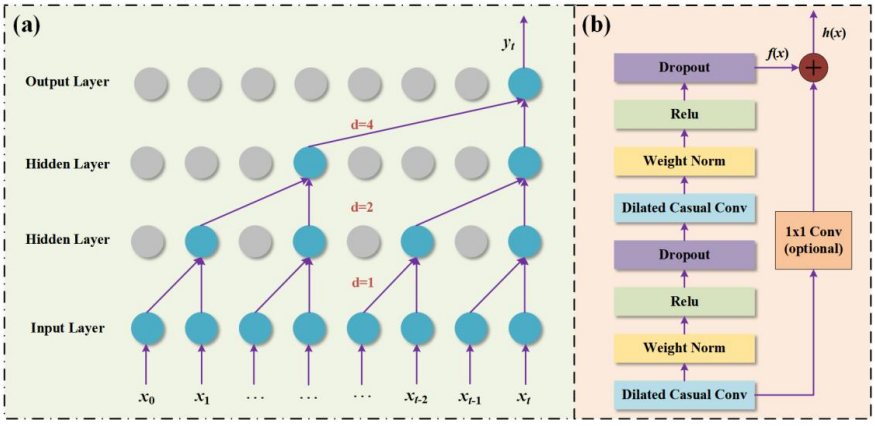


Fig. 1 The TCN model structure: (a) dilated causal convolution, (b)residual module.

The process by which TCNs handle data sequences is analogous to that of dilated convolution. TCNs typically incorporate a greater number of convolutional layers compared to conventional CNNs, continuously adjusting the number of layers. This results in progressively larger dilation factors and convolutional kernels at each layer. However, as the depth of the network increases, the training process becomes increasingly challenging, with the multi-layer backpropagation of error signals potentially leading to issues such as 'gradient vanishing' or 'gradient explosion'. Residual networks have been shown to effectively mitigate these challenges. **Fig.1**(b) illustrates the operation of residual units.

2.3 Gated recurrent unit (GRU)

As a variant of recurrent neural networks (RNNs), GRU effectively solves the long-term memory storage problems and gradient disappearance issues faced by traditional RNNs during backpropagation (Song et al., 2024). **Fig. 2** displays the structure of the GRU. It employs gating mechanisms, comprising an update gate and a reset gate, to govern information flow within the network. The update gate modulates the integration of historical and new information, whereas the reset gate identifies components of past hidden states irrelevant to current computations. Through adaptive memory update and reset mechanisms, GRU effectively models long-term dependencies in sequences.

GRU's computation process can be represented by the following equations:

$$r_t = \sigma(w_r \cdot [h_{t-1}, x_t]) \quad (1)$$

$$z_t = \sigma(w_z \cdot [h_{t-1}, x_t]) \quad (2)$$

$$H_t = \tanh(w_H \cdot [r_t * h_{t-1}, x_t]) \quad (3)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * H_t \quad (4)$$

where r_t represents the reset gate, z_t denotes the update gate, H_t signifies the intermediate state, and h_t corresponds to the GRU output. The weight matrices w_z, w_r and w_H are associated with the update gate, reset gate, and intermediate state, respectively. Here, x_t indicates the current state, h_{t-1}

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

represents the previous hidden state, σ stands for the *sigmoid* activation function, and tanh refers to the hyperbolic tangent function.

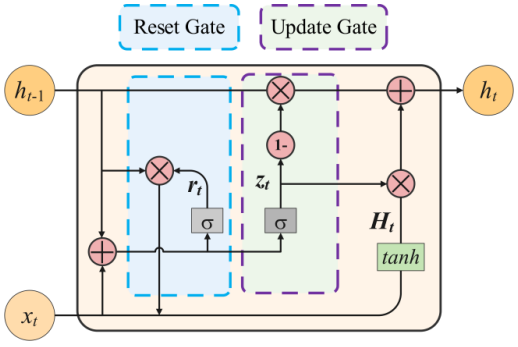


Fig. 2 GRU model internal structure.

2.4 TCN-GRU model

Fig. 3 depicts the architecture of the TCN-GRU model, which consists of input layer, TCN layer, GRU layer, and output layer. Historical streamflow data and highly correlated influencing factors are integrated as input to the TCN layer, which excels in feature extraction, fast convergence, and robustness. Through the TCN layer, input data is processed to produce one-dimensional sequences via dilated causal convolutions and residual connection modules. The GRU model demonstrates strong nonlinear mapping capabilities for time-series data, due to its inherent significant nonlinearity in streamflow time series. Thus, the GRU layer is utilized to learn and model the dynamic feature variation extracted from TCN, effectively capturing temporal correlations among multiple features to enhance simulation accuracy. The output layer generates the final simulated values.

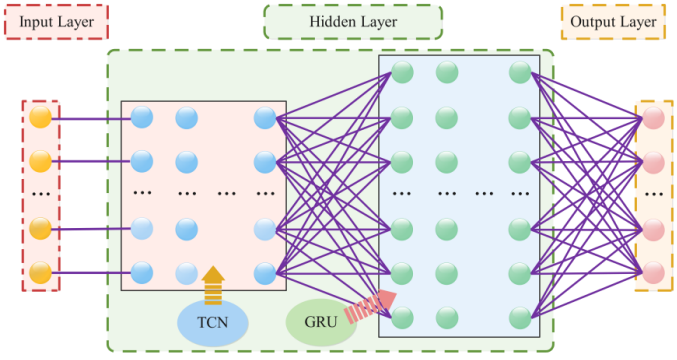


Fig. 3 The framework of the TCN-GRU model.

2.5 Random forest (RF)

RF is an ensemble learning approach that boosts the precision of classification and regression tasks by building and combining the prediction results of multiple decision trees (Qiao et al., 2023). The core principle of this algorithm involves utilizing the bootstrap resampling technique to randomly sample multiple subsets from the original dataset, on which decision trees are constructed for each subset. During the construction of these decision trees, features are randomly selected for splitting, thereby enhancing the model's diversity and generalization capability (Doyle et al., 2023). Ultimately, RF combines the

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

simulation results of the multiple trees through voting or averaging to derive the final classification or regression outcome. Utilizing RF for nonlinear ensemble learning leverages the advantages of its ensemble algorithm, thereby improving simulation accuracy, reducing the risk of overfitting, and effectively managing nonlinear and high-dimensional data (Alnahit et al., 2022; Wu et al., 2023).

2.6 XAJ-TCN-GRU hybrid model

An innovative approach, namely the XAJ-TCN-GRU model, is proposed to integrate the conceptual rainfall - runoff model and deep learning model organically for streamflow simulation. The model includes the following steps:

Step 1: Utilize the XAJ model for initial streamflow simulation, fully leveraging its scientific rigor and reliability based on hydrological theory.

Step 2: Introduce the MIC method for feature selection to optimize the input feature set, thereby enhancing model accuracy and robustness.

Step 3: Employ the TCN-GRU model for further streamflow simulation on the time series data selected through feature selection, leveraging the advantages of deep learning models in capturing temporal relationships.

Step 4: Perform nonlinear ensemble using the RF method to combine the streamflow results simulated separately by the XAJ and TCN-GRU models, with the aim of achieving more accurate and reliable comprehensive simulations through model ensemble.

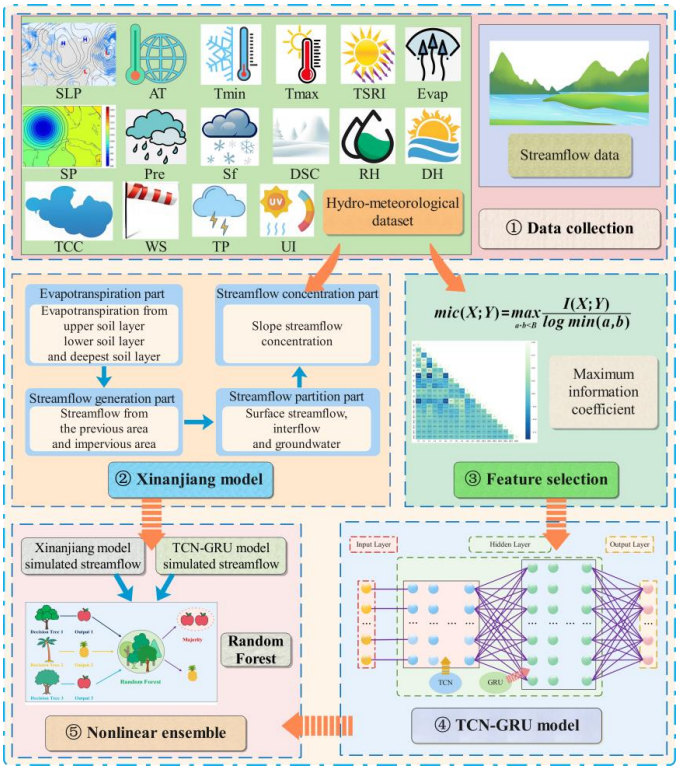


Fig. 4 The technical flowchart of the study.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

The design of this integrated model aims to overcome the limitations of individual models, allowing the conceptual rainfall - runoff model and deep learning model to complement each other, thereby providing a more comprehensive and accurate solution for streamflow simulation. The technical flowchart of the study methodology is depicted in **Fig. 4**.

3 Study area and data description

3.1 Study areas

To validate generalization capabilities of proposed XAJ-TCN-GRU hybrid model, this study conducted experiments across four basins characterized by distinct geographical and hydrological features. These basins include the Wuding River Basin (arid northwestern China), the Chu River Basin (humid eastern coastal region), the Jianxi River Basin (hilly southeastern coastal region), and the Qingyi River Basin (mountainous plateau region in southwestern China). **Fig. 5** presents the geographical information maps of these four basins. Detailed descriptions of these basins can be found in **Supplementary Information 1**. By conducting experiments in basins with varying geographical characteristics and hydrological conditions, it is possible to comprehensively evaluate the adaptability and simulation capability of this model.

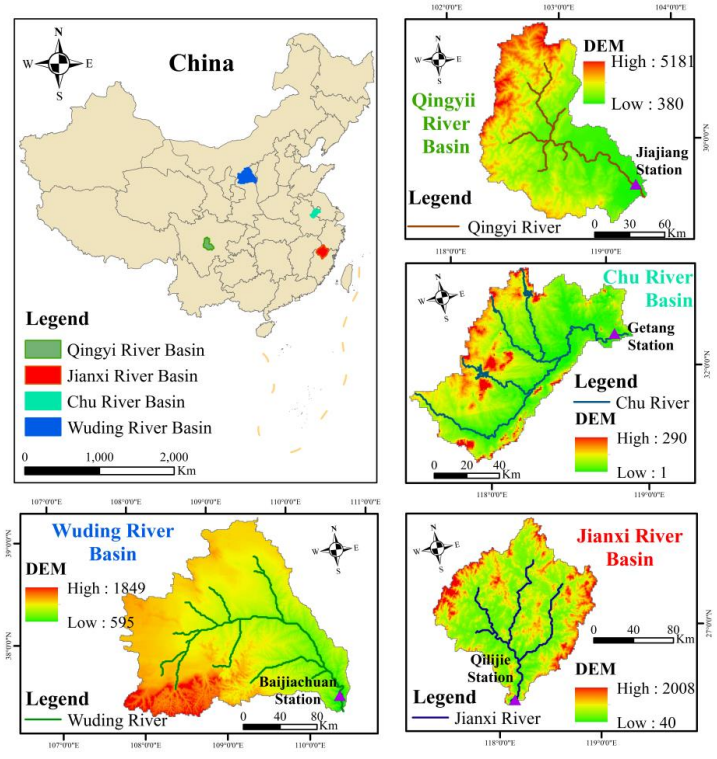


Fig. 5 Geographic information map of Wuding River, Chu River, Jianxi RIver, and Qingyi River basins.

3.2 Data description

In this study, daily observations of 31 meteorological and hydrological variables were collected from multiple stations across the four river basins, as tabulated in **Table 2**. Data acquisition for these

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

variables was conducted via the China Meteorological Science Data Center (http://data.cma.cn/). For hydrological monitoring, the primary control stations in the Wuding River, Chu River, Jianxi River, and Qingyi River basins are Baijiachuan Station, Getang Station, Qilijie Station, and Jiajiang Station, respectively. Daily streamflow data for these stations were retrieved from the measured records in *Hydrological Yearbook of the People's Republic of China*. The entire dataset spans a continuous period from January 1, 2010, to August 31, 2023. Statistical summaries of the streamflow datasets for each basin are presented in **Table 3**, offering a more intuitive characterization of the data.

Table 2 Input variables utilized to integrate the XAJ-TCN-GRU model for daily streamflow simulation.

No	Acronyms	Input variables	Units
1	SLP	Sea level pressure	hPa
2	SP	Surface pressure	hPa
3	AT	2-Meter Average temperature	℃
4	Pre	Precipitation	mm
5	Tmax	2-Meter air temperature - daily max	℃
6	Tmin	2-Meter air temperature - daily min	℃
7	Sf	Snowfall	mm
8	DSC	Depth of snow cover	mm
9	GT	Ground temprature	℃
10	DPT	Dew point temperature	℃
11	RH	Relative humidity	%
12	Evap	Evaporation	mm
13	PET	Potential Evapotranspiration	mm
14	WS	10-meter wind speed	m/s
15	MGWS	Maximum gust wind speed	m/s
16	AGWS	Average gust wind speed	m/s
17	LMS	Latitudinal wind speed	m/s
18	MWS	Meridional wind speed	m/s
19	LCC	Low level cloud cover	/
20	MCC	Medium level cloud cover	/
21	HCC	High level cloud cover	/
22	TCC	Total cloud cover	/
23	NSRI	Net solar radiation intensity	J/m^2
24	TSRI	Total solar radiation intensity	J/m^2
25	DR	Direct radiation	J/m^2
26	DH	Daylight hours	h
27	UI	Ultraviolet intensity	J/m^2
28	TP	Thunderstorm probability	K
29	MTP	Maximum thunderstorm probability	K
30	K	K-index	/

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

31	CUPE	Convective usable potential energy	J/kg

Table 3 Descriptive statistics of daily streamflow values for each basin (unit: m³/s).

Basin	Count	Mean	Max	Min	Std	Skew
Wuding	4991	117.5602	1200.8438	12.5313	135.7014	2.7103
Chu	4991	146.2993	2993.1719	18.5938	207.3180	4.8575
Jianxi	4991	246.5827	2917.1094	47.4375	257.6926	4.2258
Qingyi	4991	545.0607	3683.1406	101.2031	411.4504	2.0973

3.3 Experimental designs

The input for the XAJ model consists of simulated precipitation and evaporation, while the output is simulated streamflow. The dataset is divided into training, validation, and test sets in a 6:2:2 ratio. The SCEM-UA method is used to calibrate the model parameters to ensure prediction accuracy.

The 31 hydrological and meteorological variables listed in **Table 2** were used as feature variables, while the streamflow was used as the target variable to construct the dataset for inputting into the TCN-GRU model. In order to simplify the model, decrease computational costs, and enhance the interpretability of the data, MIC was employed for feature selection on the 31 feature variables. For each basin, 10 highly correlated feature variables with the streamflow were selected, as shown in **Table 4**.

The proposed model was built on a personal computer with a 1.8 GHz Intel i5 processor and 8 GB of memory. The TCN-GRU model was developed using the popular deep learning frameworks Keras and TensorFlow. Additionally, visualization tools like Matplotlib were used to present the model training outcomes visually, ensuring compatibility of the development environment and intuitive analysis of the results. After feature selection via MIC, a new dataset including the selected features and streamflow was constructed. The dataset was illustrated in **Fig. 6**. The training set was used to train the TCN-GRU model, while the validation set was applied for hyperparameter tuning. To improve the performance of deep learning models, a random search algorithm was used to optimize hyperparameters. Firstly, based on existing research results and domain expertise, the range of potential hyperparameter values was defined. Then, the random search algorithm randomly selected hyperparameter combinations and evaluated model performance based on the validation set to systematically explore this range of values. To ensure result stability and reliability, five independent optimization runs were conducted, and performance metrics were averaged. Through this approach, the optimal hyperparameter combination maximizing model performance was efficiently identified. The final hyperparameter configurations for proposed TCN-GRU are listed in **Table S1**. The trained TCN-GRU simulated streamflow by inputting the testing set.

Using the simulated results generated by the XAJ model and TCN-GRU model on the training set as feature inputs, a random forest model is trained to learn how to effectively perform nonlinear ensemble. To reduce the risk of overfitting, the maximum depth is set to limit the complexity of each tree, thereby preventing the model from overfitting the noise in the training data. Next, the simulated results from the XAJ and TCN-GRU models on the test set are input into the trained random forest model for nonlinear ensemble, resulting in the final simulation.

The evaluation metrics adopted in this research include MAE (m³/s), RMSE (m³/s), MAPE (%), NSE, and Kling-Gupta Efficiency (KGE). For assessing interval simulation performance, two metrics

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

are employed: Simulation Prediction Interval Normalised Average Width (PINAW) and Interval Coverage Probability (PICP). A detailed description of these metrics is shown in **Supplementary Information 3**.

Table 4 Predictive input variables obtained using MIC for key feature selection.

Basin	Key feature variables selected through MIC (see Table 2 for variable index numbers)
Wuding	12,10,11,9,5,3,4,13,30,19
Chu	10,5,3,9,1,2,4,12,31,30
Jianxi	1,2,12,10,5,9,3,31,30,21
Wuding	5,10,9,3,4,30,1,2,12,31

4 Results

4.1 simulated results for four basins of the XAJ-TCN-GRU model

The XAJ-TCN-GRU model was used in this study to simulate streamflow under four distinct hydrological conditions across various basins, presenting the simulated results in the form of fitted graphs. As indicated in **Fig. 6**, it is evident that the proposed hybrid model exhibits good fitting performance in all basins, indicating its strong generalization capability. This effectiveness can be attributed to the XAJ model, which incorporates prior knowledge based on physical processes through the modeling of basin hydrological features and equation solving, thereby enhancing both the interpretability and generalization capacity of the model.

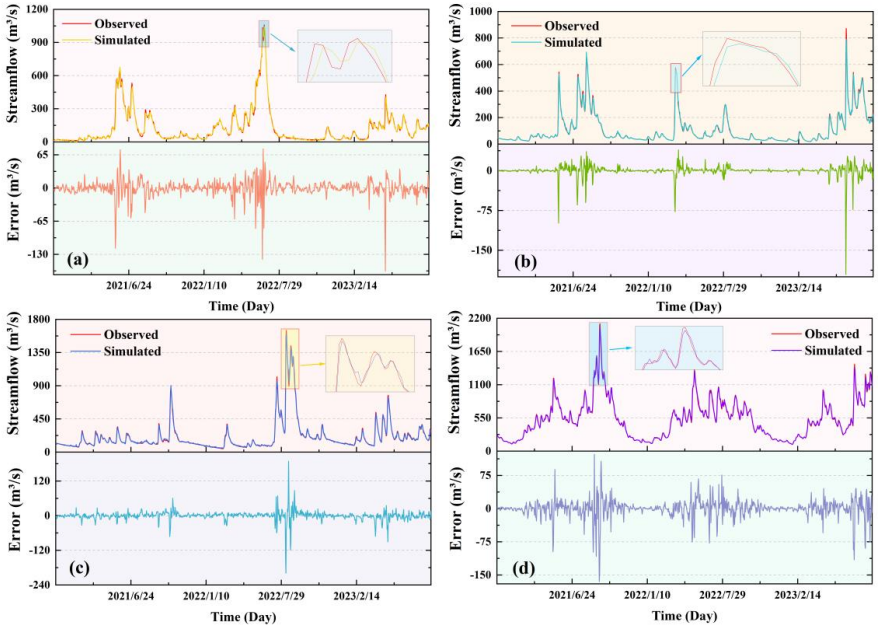


Fig.6 Streamflow simulation fitting results of XAJ-TCN-GRU model for (a) Wuding River (b) Chu River (c) Jianxi River, and (d) Qingyi River basins.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

Notably, in basins with larger streamflow, such as Qingyi River, our model demonstrated precise peak simulation capabilities, presenting a significant advantage over traditional XAJ, TCN, and GRU models. Additionally, we observed relatively gentle simulation errors during periods of lower flow, while errors increased significantly and exhibited noticeable fluctuations during peak periods. Moreover, as the number of peaks increased, the fluctuation of errors also increased. However, overall, the absolute error remained within 200 m^3/s.

In basins with lower streamflow, such as the Wuding River and Chu River, we observed slight delays in simulations. This indicates that the hydrological processes in these basins may be influenced by other complex factors, including terrain, soil type, and vegetation cover, which were not fully accounted for or simulated during model development. Consequently, these factors may affect the model's simulation performance. There are inherent limitations in modeling and simulating complex hydrological processes within specific basins.

4.2 Model performance comparison

To better highlight the simulation capabilities of the proposed model, four baseline models (XAJ, LSTM, TCN, and GRU) and one hybrid model (TCN-GRU) were applied to simulate daily streamflow in the four basins as control models for the study. LSTM is a neural network architecture widely applied in time series forecasting and streamflow simulation. By including LSTM in our comparisons, we can conduct a more comprehensive evaluation of our XAJ-TCN-GRU model's performance, ensuring that our research findings have greater universality and robustness. In addition, to further highlight the advantages of nonlinear ensemble, we conducted two additional experiments. One group used the Linear Regression (LR) method to linearly combine XAJ and TCN-GRU models, resulting in a new model called XAJ-TCN-GRU&LR. The other group used the simulations based on XAJ as inputs for the TCN-GRU model, leading to the construction of the XAJ-Infused TCN-GRU model.

Table 5 provides a detailed overview of the streamflow simulation performance of the eight models in the four basins. Clearly, the XAJ-TCN-GRU model exhibits superior performance across all five metrics. The XAJ-TCN-GRU model exhibits relatively low RMSE and MAE values in each basin, particularly when compared to the traditional conceptual rainfall - runoff model XAJ, resulting in a significant reduction in simulation errors. For instance, in the Wuding River basin, the RMSE and MAE of the XAJ model are 69.212 m^3/s and 64.320 m^3/s, respectively, while the XAJ-TCN-GRU model achieves only 7.032 m^3/s and 3.876 m^3/s for these metrics. Additionally, when compared with individual TCN and GRU models, the integrated TCN-GRU model exhibits higher NSE and KGE values across the four basins. This indicates the TCN-GRU has effectively combined complementary advantages of TCN and GRU, thus improving the model's simulation capability. The NSE values obtained using the TCN-GRU model in the Wuding River, Chu River, Jianxi River, and Qingyi River basins are 0.990, 0.947, 0.965, and 0.976, respectively. Upon nonlinear ensemble with the simulations of the XAJ model, the simulation accuracy significantly improves. The NSE values achieved using the XAJ-TCN-GRU model for streamflow simulation in the four basins are 0.991, 0.971, 0.984, and 0.986, respectively, representing enhancements of 0.10%, 2.53%, 1.97%, and 1.02% relative to the TCN-GRU model. Additionally, compared to the linearly integrated XAJ-TCN-GRU&LR model, the improvements are 1.33%, 1.80%, 1.66%, and 0.62%, respectively. The nonlinear ensemble approach is capable of better capturing intricate

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

nonlinear dependencies within the data, thereby mitigating the risk of error propagation (Wang et al., 2022; Xu et al., 2025). The RF algorithm used for ensemble is less likely to amplify errors compared to linear combinations, as it captures the complex relationships between different model simulations through its decision tree-based structure. Each decision tree in RF is capable of capturing different patterns in the data, and the ensemble process averages out the errors and uncertainties, thereby reducing the impact of potential errors from individual models.

We also found that the simulation results of the XAJ-Infused TCN-GRU model were not satisfactory, with the KGE values for the Wuding River, Chu River, and Qingyi River basins even falling below those of TCN-GRU model. This result indicates that directly inputting the outputs from the physical-based model into the deep learning model may propagate the errors from the former to the latter, leading to a decrease in simulation accuracy. The XAJ-TCN-GRU model effectively leverages the strengths of both conceptual rainfall - runoff and deep learning models and compensates for their potential shortcomings in streamflow simulation through nonlinear ensemble.

Table 5 The evaluation metrics of the six models for streamflow simulation on validation and testing sets.

Basin	Model	Validation set					Testing set				
		RMSE	MAE	MAPE	NSE	KGE	RMSE	MAE	MAPE	NSE	KGE
Wuding	XAJ	65.418	68.461	0.294	0.819	0.726	69.212	64.320	0.308	0.806	0.707
	LSTM	24.284	15.397	0.219	0.973	0.872	26.841	18.376	0.246	0.971	0.867
	TCN	25.671	13.649	0.134	0.979	0.952	20.810	9.220	0.122	0.982	0.975
	GRU	20.923	12.390	0.152	0.980	0.931	17.649	10.248	0.147	0.987	0.954
	TCN-GRU	16.342	8.175	0.068	0.990	0.985	15.935	9.985	0.071	0.990	0.981
	XAJ-TCN-GRU&LR	10.973	6.964	0.042	0.990	0.986	12.613	8.032	0.058	0.986	0.974
	XAJ-Infused TCN-GRU	16.544	11.370	0.056	0.986	0.981	18.901	10.073	0.064	0.983	0.964
	XAJ-TCN-GRU	6.553	3.684	0.039	0.993	0.991	7.032	3.876	0.041	0.991	0.987
Chu	XAJ	70.613	55.165	0.663	0.681	0.674	71.621	53.927	0.693	0.653	0.641
	LSTM	40.638	27.112	0.315	0.897	0.774	38.192	24.694	0.302	0.901	0.797
	TCN	33.852	21.672	0.263	0.916	0.857	30.824	17.360	0.245	0.936	0.897
	GRU	30.994	15.718	0.230	0.948	0.931	29.314	17.954	0.236	0.942	0.925
	TCN-GRU	24.222	13.693	0.100	0.952	0.950	28.029	14.555	0.108	0.947	0.942
	XAJ-TCN-GRU&LR	17.354	10.397	0.067	0.958	0.947	23.972	11.301	0.073	0.951	0.945
	XAJ-Infused TCN-GRU	20.993	12.387	0.081	0.955	0.941	25.394	13.977	0.080	0.949	0.938
	XAJ-TCN-GRU	15.397	6.314	0.038	0.969	0.958	11.368	4.311	0.031	0.971	0.962
Jianxi	XAJ	125.376	98.498	0.693	0.703	0.718	139.150	110.025	0.760	0.668	0.679
	LSTM	56.317	43.197	0.312	0.930	0.864	55.624	41.842	0.308	0.931	0.868
	TCN	45.671	23.994	0.098	0.959	0.951	42.042	18.766	0.093	0.961	0.960
	GRU	32.478	18.453	0.094	0.970	0.948	40.224	22.669	0.121	0.964	0.941
	TCN-GRU	30.226	11.389	0.048	0.972	0.957	39.564	15.947	0.057	0.965	0.955
	XAJ-TCN-GRU&LR	15.397	7.619	0.037	0.985	0.971	22.007	10.243	0.049	0.976	0.963
	XAJ-Infused TCN-GRU	17.308	9.114	0.042	0.978	0.966	32.904	13.354	0.051	0.971	0.958

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

	XAJ-TCN-GRU	12.077	6.356	0.031	0.990	0.986	16.732	8.179	0.040	0.984	0.979
Qingyi	XAJ	103.677	71.594	0.576	0.754	0.704	112.101	88.635	0.634	0.720	0.681
	LSTM	78.642	57.349	0.112	0.935	0.885	78.773	56.548	0.113	0.938	0.882
	TCN	56.555	40.375	0.072	0.971	0.980	50.613	32.678	0.067	0.975	0.962
	GRU	60.348	42.374	0.079	0.969	0.936	55.138	39.466	0.075	0.970	0.947
	TCN-GRU	46.612	27.315	0.047	0.978	0.966	49.626	29.892	0.051	0.976	0.985
	XAJ-TCN-GRU&LR	21.982	13.988	0.027	0.980	0.972	29.370	15.976	0.034	0.972	0.968
	XAJ-Infused TCN-GRU	29.037	23.974	0.033	0.975	0.968	36.691	24.503	0.039	0.970	0.964
	XAJ-TCN-GRU	16.785	9.428	0.018	0.988	0.980	20.862	11.881	0.020	0.986	0.974

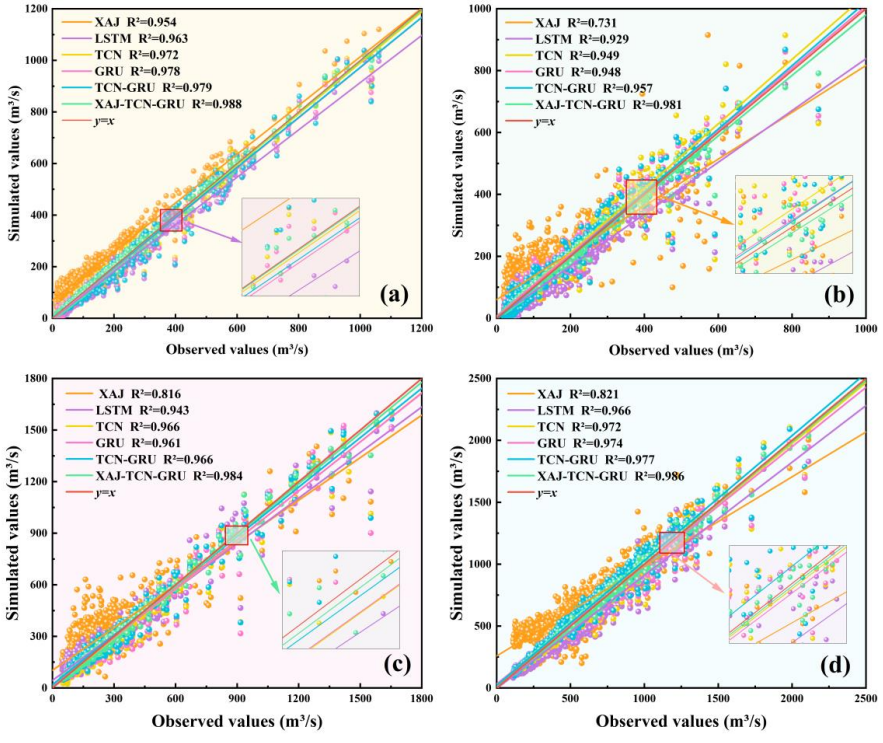


Fig. 7 The regression images of each model on the testing set of (a) Wuding River (b) Chu River (c)Jianxi River, and (d) Qingyi River basins.

Scatter plots provide an intuitive display of the model's explanatory power over data variations, as depicted in **Fig. 7**. Upon analysis, the XAJ model exhibits a tendency to overestimate streamflow during low-flow periods and underestimate it during high-flow periods. In contrast, deep learning models demonstrate superior streamflow simulation capabilities. Notably, the fitting line of the XAJ-TCN-GRU model is closest to the 1:1 line. The R^2 values of this model in the four basins are 0.988, 0.981, 0.984, and 0.986, respectively, representing improvements of 3.56%, 34.20%, 20.59%, and 20.10% compared with the standalone XAJ model. In comparison with the TCN-GRU model, the improvements are 0.92%, 2.51%, 1.86%, and 0.92%, respectively. In the Chu River basin, the R^2 value of the XAJ model is significantly low, indicating that traditional conceptual rainfall-runoff models are not applicable to this

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

basin. Across the four basins, the simulation results of TCN and GRU are comparable, with R^2 values both exceeding 0.94, highlighting the important role of these two models in each basin. The integrated TCN-GRU model effectively combines the advantages of both models, demonstrating good adaptability across all basins and the ability to handle complex hydrological conditions. These findings underscore the broad applicability and superiority of XAJ-TCN-GRU model in diverse basins.

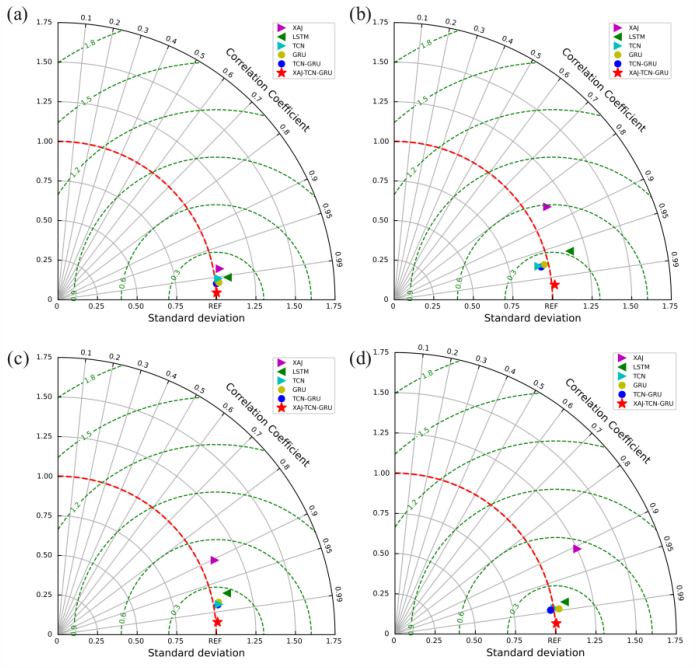


Fig. 8 Taylor diagram illustrating correlation coefficient and the standard deviation difference for XAJ-TCN-GRU model compared with comparative models at (a) Wuding River (b) Chu River (c) Jianxi River, and (d) Qingyi River basins.

The Taylor diagram (Taylor, 2001) is widely employed to evaluate the correlation and simulation accuracy between model predictions and observed data. **Fig. 8** illustrates that different symbols denote various simulation methods, with standard deviation on the horizontal axis and vertical axis, and the dotted line representing RMSE. The optimal simulation accuracy is represented by the center of the horizontal axis in the Taylor diagram. Notably, in **Fig. 8**, the XAJ-TCN-GRU model's predictions for the Wuding River basin are positioned at the optimal location in the Taylor diagram, , and the correlation coefficient is 0.99. Likewise, for the Chu River, Jianxi River, and Qingyi River basins, the XAJ-TCN-GRU model's results outperform those of the comparative models and are closest to the observed values. By contrast, the XAJ model's simulated streamflow shows significant deviations from the observed data. This further highlights the proposed model's advantage in streamflow simulation.

5 Discussion

5.1 Model robustness analysis

To evaluate the robustness and resilience of the model to changes in the input data, we performed noise data injection testing and out-of-context validation.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

In actual hydrological observations, due to equipment malfunctions, recording errors, and other reasons, observational data often contain erroneous data. We treat these data as noise data. Based on the common error range in field data (1%-2%), the noise data injection ratio was set to 2% to ensure the model's robustness and stability under higher error conditions (Wang et al., 2023; Zhao et al., 2024). Specifically, 2% of the data in the training set was replaced with randomly generated noise data to simulate changes in the input data. Subsequently, we used this perturbed dataset to train the models and make streamflow simulations to test the models' responsiveness to noise data situations. Noise data injection helps evaluate how the models perform when facing unknown or noise data situations in the real world and aids in improving the models' performance and generalization capabilities. Additionally, we conducted 10 independent runs for each model and recorded the results of each run, including the model's outputs and performance metrics. By running the model multiple times, we could observe the variations and fluctuations in the models' performance across different runs, thereby enabling a more thorough assessment of its performance and improving the evaluation of its robustness and reliability.

The detailed outcomes of the noise data injection test are recorded in **Supplementary Information 4**. The boxplots in **Fig. 9** provide a detailed representation of the normal distribution of indicators obtained from the 10 runs of each model during the noise data injection test. It is apparent that the deep learning models demonstrate good stability. Among them, the LSTM, GRU, and TCN-GRU models show relatively stable simulation performance across the four basins in the noise data injection test. However, it is noteworthy that the TCN model exhibits significant fluctuations in the tests conducted in the Chu and Jianxi basins. Overall, the XAJ-TCN-GRU model demonstrates the strongest stability and the best simulation performance in the noise data injection test. The XAJ-TCN-GRU model leverages the advantages of integrating conceptual rainfall - runoff and deep learning models, enabling it to capture the physical characteristics of streamflow data while also learning the deeper features of the data, thereby performing exceptionally well in handling noise data.

In the out-of-context validation, we excluded data from wet-year in the training sets of the four basins and used these data separately for model validation. This approach aims to evaluate the model's performance under extreme hydrological conditions (i.e., wet-year) that it has not encountered before, providing a better understanding of its generalization ability and robustness. The out-of-context validation results for the four basins are shown in **Table 6**. By comparing **Tables 5** and **6**, we can see that the model's simulation accuracy declines slightly under wet-year conditions compared to regular years. However, it still maintains a high level of accuracy overall, indicating that the model has a certain degree of generalization ability and can make reasonable simulations under extreme flow conditions. In the Jianxi River and Qingyi River basins, the model's simulation accuracy decreases more significantly during wet-year, likely because these regions experience greater precipitation and flow fluctuations during wet-year. This variability may not be sufficiently captured in the training set, thereby affecting the ability of the model to learn and adapt under these extreme conditions.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

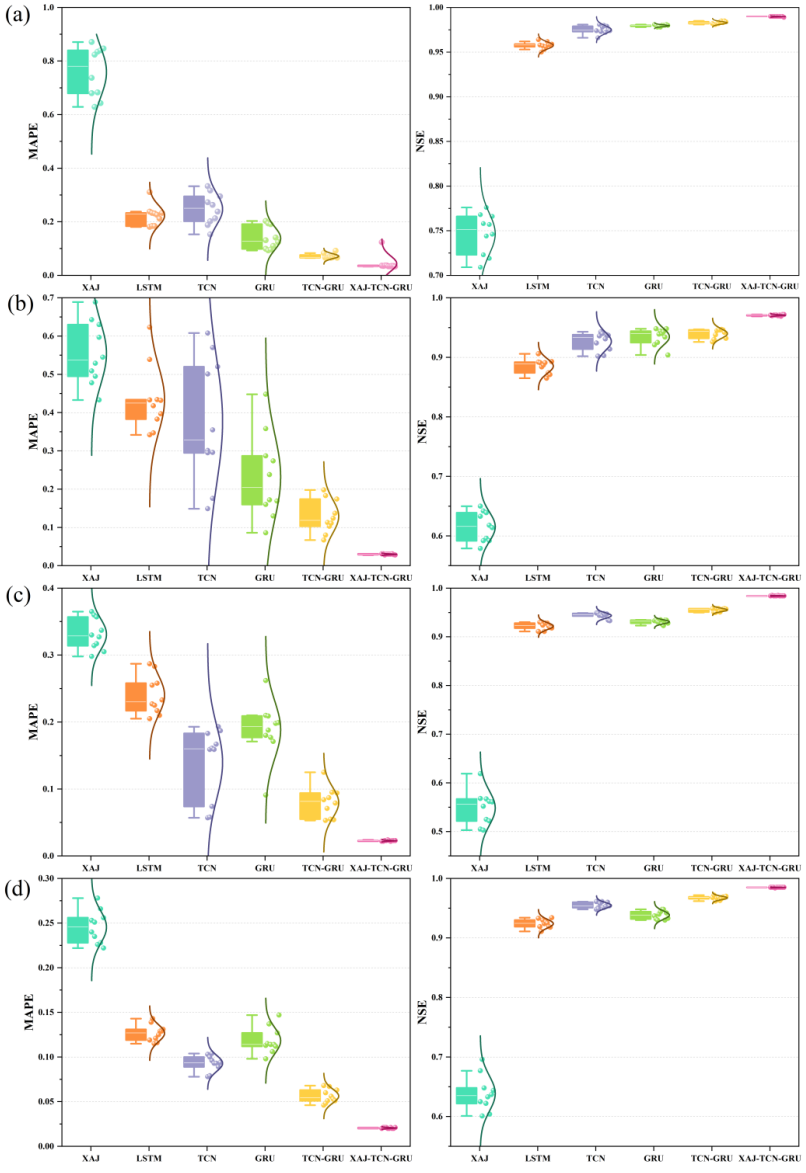


Fig. 9 Box plots of the proposed XAJ-TCN-GRU model and comparative models for simulating streamflow in terms of MAPE and NSE for (a) Wuding River (b) Chu River (c) Jianxi River, and (d) Qingyi River basins.

Table 6 Results of out-of-context validation on the XAJ-TCN-GRU model.

Basin	RMSE	MAE	MAPE	NSE	KGE
Wuding	8.210	5.381	0.049	0.973	0.970
Chu	13.943	6.876	0.041	0.952	0.945

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

Jianxi	20.786	11.763	0.051	0.969	0.940
Qingyi	28.014	17.002	0.048	0.963	0.951

5.2 Flood simulation analysis

Flood simulation is a core component of watershed management and disaster prevention planning. Especially during the flood season, accurate streamflow simulations are paramount for the prompt execution of flood control measures and the rational allocation of water resources. This section investigates streamflow simulations across four distinct basins during both flood and non-flood seasons, providing a comprehensive assessment of their capabilities.

Based on the climatic and geographical conditions, the flood seasons for the Wuding River and Chu River primarily occur from June to September, while for the Jianxi and Qingyi Rivers, they commence slightly earlier, from May to September. These flood seasons represent periods characterized by the most significant variations in streamflow, making them particularly challenging to simulate. According to the data presented in **Table 7**, the XAJ-TCN-GRU model demonstrates exceptional performance in simulating streamflow during the flood seasons across all four basins. Its NSE values for the Wuding River, Chu River, Jianxi River, and Qingyi River are 0.988, 0.968, 0.984, and 0.980, respectively, all of which are close to or exceed 0.98, indicating the model's high accuracy in flood season simulations. Furthermore, during non-flood seasons, the model maintains high NSE values of 0.987, 0.970, 0.980, and 0.987, respectively, further confirming its stability and reliability. It is important to note that although the accuracy of the XAJ-TCN-GRU in simulating streamflow during flood seasons for the Chu River and Qingyi River basins is slightly lower compared to non-flood seasons, this does not indicate poor performance. In fact, the complexities arising from factors such as climate change and topography make it challenging for any simulation model to achieve perfection during flood seasons. Nonetheless, compared with the other five comparative models, the XAJ-TCN-GRU model demonstrates the lowest RMSE, MAE, and MAPE in streamflow simulation during flood seasons across all four basins, highlighting its superiority in flood season simulation.

This study also placed a specific emphasis on the accurate simulation of extreme hydrological events. Periods in each basin where streamflow exceeded four times the standard deviation of the testing set were classified as flood events. Through in-depth hydrological data analysis, a total of eight such flood events were identified across the four basins. **Table 8** presents detailed simulation outcomes of these flood events using different models in the four basins. Clearly, the XAJ-TCN-GRU model exhibits the least simulation bias in all four basins when observing the data in **Table 8**, with remarkably lower RMSE and MAPE values than other models. In the four flood events of the Wuding River basin, the RMSE of XAJ-TCN-GRU was reduced by 30.60%, 42.16%, 66.47%, and 43.01% compared with TCN-GRU. This significant reduction not only highlights the importance of integrating the results of conceptual rainfall - runoff model to enhance the simulation accuracy of hybrid models but also further validates the high accuracy and dependability of the XAJ-TCN-GRU in simulating flood events. The XAJ-TCN-GRU model's in-depth understanding of hydrological processes provides an important foundation for flood peak simulation in areas with limited hydrological data, particularly in regions with sparse observation stations, where the model's portability is especially significant.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

However, it is important to note that while this model provides insights into potential applications for flood peak simulation, our work focuses on streamflow simulation rather than direct flood simulation. Directly using streamflow simulation results for flood peak simulation has limitations. First, streamflow models often rely on coarse spatial and temporal scales, which may overlook critical local features and flood dynamics. Second, inaccuracies in model parameters and input data, particularly under extreme precipitation conditions, can amplify errors in flood peak estimation. These factors highlight the need for further development to improve the accuracy and reliability of flood peak simulations using streamflow models. The results suggest that, although the model's simulation can serve as an approximation of flood peak flows in areas with limited observation stations, further development is needed to improve the reliability of flood simulations. Future research may explore incorporating additional flood-related factors and advanced post-processing techniques to better translate streamflow simulations into accurate flood simulations.

Table 7 Simulation outcomes of the XAJ-TCN-GRU model and comparative models in four basins during flood and non-flood seasons.

Basin	Model	Flood Season					Non-Flood Season				
		RMSE	MAE	MAPE	NSE	KGE	RMSE	MAE	MAPE	NSE	KGE
Wuding	XAJ	69.271	64.966	0.739	0.892	0.806	69.137	63.952	0.714	0.568	0.549
	LSTM	32.717	25.314	0.230	0.976	0.968	23.341	14.880	0.254	0.951	0.943
	TCN	17.518	10.996	0.090	0.983	0.980	14.810	8.229	0.139	0.970	0.972
	GRU	20.058	12.752	0.114	0.971	0.982	17.658	10.255	0.147	0.987	0.964
	TCN-GRU	25.466	13.993	0.068	0.985	0.988	17.962	7.922	0.073	0.971	0.965
	XAJ-TCN-GRU	8.998	5.301	0.036	0.988	0.990	5.703	3.124	0.044	0.987	0.978
Chu	XAJ	85.177	69.672	0.719	0.657	0.597	63.776	46.071	0.683	0.664	0.593
	LSTM	50.564	35.345	0.220	0.885	0.843	30.069	19.286	0.343	0.836	0.743
	TCN	36.581	23.193	0.159	0.909	0.886	22.526	14.420	0.289	0.908	0.837
	GRU	40.883	29.002	0.230	0.925	0.894	21.214	12.387	0.240	0.919	0.884
	TCN-GRU	45.087	26.718	0.126	0.940	0.907	20.078	8.389	0.099	0.927	0.901
	XAJ-TCN-GRU	16.674	7.957	0.037	0.968	0.927	7.362	2.471	0.028	0.970	0.936
Jianxi	XAJ	134.435	106.805	0.615	0.777	0.781	91.986	75.216	0.650	0.658	0.493
	LSTM	61.793	42.369	0.221	0.953	0.928	50.556	41.489	0.374	0.776	0.744
	TCN	51.917	23.662	0.088	0.959	0.936	26.816	15.121	0.098	0.937	0.883
	GRU	54.410	30.448	0.131	0.964	0.948	24.736	16.865	0.113	0.946	0.897
	TCN-GRU	57.847	22.569	0.061	0.967	0.970	42.063	15.961	0.057	0.961	0.934
	XAJ-TCN-GRU	22.484	10.799	0.039	0.984	0.978	10.534	6.229	0.041	0.980	0.961
Qingyi	XAJ	135.246	106.087	0.155	0.788	0.657	105.228	92.937	0.544	0.659	0.604
	LSTM	99.776	72.860	0.095	0.885	0.791	58.095	44.164	0.125	0.940	0.826
	TCN	71.052	46.127	0.056	0.941	0.867	26.295	17.646	0.066	0.978	0.904
	GRU	68.884	48.399	0.088	0.931	0.889	27.406	20.906	0.067	0.980	0.892
	TCN-GRU	77.301	63.413	0.066	0.945	0.902	28.924	21.477	0.048	0.982	0.935
	XAJ-TCN-GRU	29.702	19.063	0.023	0.980	0.964	9.954	6.467	0.017	0.987	0.971

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

Table 8 Simulation results of flood events using the proposed XAJ-TCN-GRU and comparative models for four basins.

Basin	Flood event	XAJ		LSTM		TCN		GRU		TCN-GRU		XAJ-TCN-GRU	
		RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Wuding	2021/5/22 - 2021/5/24	71.073	0.120	60.424	0.103	25.224	0.042	22.668	0.032	16.720	0.030	11.604	0.015
	2021/6/5 - 2021/6/6	21.040	0.038	24.434	0.044	19.277	0.035	10.468	0.245	6.618	0.011	3.828	0.007
	2022/5/28 - 2022/5/29	61.448	0.112	12.086	0.022	21.479	0.039	10.125	0.018	8.897	0.016	2.985	0.005
	2022/6/1 - 2022/6/26	66.882	0.066	65.672	0.066	67.349	0.065	49.185	0.041	42.754	0.036	24.366	0.023
Chu	2023/6/20 - 2023/6/21	132.625	0.170	36.060	0.046	86.269	0.110	78.822	0.101	44.289	0.057	26.699	0.034
Jianxi	2022/8/10 - 2022/8/16	56.266	0.036	82.850	0.055	94.629	0.064	65.441	0.047	48.885	0.036	34.579	0.023
	2022/8/20 - 2022/9/1	196.557	0.131	111.546	0.057	105.329	0.058	92.535	0.058	96.338	0.062	50.825	0.028
Qingyi	2021/9/4 - 2021/9/8	285.960	0.118	241.618	0.088	210.184	0.086	209.760	0.073	177.104	0.077	94.696	0.043

5.3 Interval simulation

Traditional point simulations often prove insufficient to fully account for the inherent uncertainty in simulations, a challenge typically unavoidable in streamflow modeling. To obtain numerical estimates with their associated reliability, utilizing interval simulation methods represents a more reasonable approach. This approach not only provides an average simulation estimate but also captures the uncertainty level inherent in the simulation, thereby enabling more robust and scientifically grounded decision-making. In this section, interval simulation methods based on error modeling are investigated, building on the foundation of point simulations. The methodology adopted follows Song et al. (2015), namely, constructing simulation confidence interval at a specified significance level. The framework integrates the point simulation results from **Sections 4.2** and **4.3**, using the logit distribution function to match the simulation error sequence and performing interval simulation at the designated significance level. Two evaluation metrics are used to assess interval simulation accuracy: PICP and PINAW. At a given significance level, a higher PICP value combined with a lower PINAW value indicates better model performance.

Three distinct significance levels were chosen for each of the four river basins, with corresponding PICP and PINAW results documented in **Table 9**. Interval simulation results of the XAJ-TCN-GRU model at these four sites are visualized in **Figs. 9–11**. Notably, at different levels of significance, the majority of actual observations are located within the shaded area of the simulated interval, indicating that the XAJ-TCN-GRU model exhibits good coverage probabilities across different significance levels. Furthermore, an apparent trend shows that as the significance level increases, the average width of the simulation intervals tends to decrease.

At a 0.05 significance level, the XAJ-TCN-GRU model realizes PICP rates of 94.29%, 95.09%, 95.29%, and 95.09% across the four river basins, respectively. These values show improvements over TCN-GRU model, highlighting the efficacy of integrating the conceptual rainfall-runoff model with the nonlinear ensemble. In terms of PINAW, the XAJ-TCN-GRU model yields values of 0.022, 0.044, 0.034, and 0.034, compared with the traditional conceptual rainfall-runoff model XAJ, which has PINAW values of 0.094, 0.244, 0.186, and 0.228. These results demonstrate a substantial reduction in the average width of simulation intervals for the XAJ-TCN-GRU model. The data in **Table 9** further confirms that the XAJ-TCN-GRU model excels in generating more compact and narrower simulation windows,

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

thereby reduces decision-making risk caused by inaccuracies and uncertainty. This effectively addresses uncertainties and better aligns with practical requirements.

Table 9 Evaluation of the XAJ-TCN-GRU and comparative models in interval simulation for four basins.

Basin	Model	α=0.05		α=0.1		α=0.2	
		PINAW	PICP	PINAW	PICP	PINAW	PICP
Wuding	XAJ	0.094	83.75%	0.073	75.03%	0.048	59.12%
	LSTM	0.072	89.38%	0.056	81.76%	0.037	64.83%
	TCN	0.050	93.39%	0.051	89.78%	0.033	81.96%
	GRU	0.065	92.89%	0.041	89.98%	0.028	79.76%
	TCN-GRU	0.053	93.69%	0.039	90.98%	0.026	82.67%
	XAJ-TCN-GRU	0.022	94.29%	0.017	91.48%	0.011	85.37%
Chu	XAJ	0.244	88.88%	0.190	79.76%	0.125	60.52%
	LSTM	0.133	89.98	0.104	84.67%	0.068	72.65%
	TCN	0.115	93.49%	0.090	89.98%	0.059	79.76%
	GRU	0.109	92.69%	0.086	87.37%	0.056	77.66%
	TCN-GRU	0.108	94.29%	0.084	91.68%	0.055	84.57%
	XAJ-TCN-GRU	0.044	95.09%	0.034	93.29%	0.022	88.38%
Jianxi	XAJ	0.186	84.97%	0.145	74.55%	0.095	50.20%
	LSTM	0.088	92.48%	0.069	89.68%	0.045	79.16%
	TCN	0.080	94.99%	0.063	92.79%	0.041	86.17%
	GRU	0.086	94.69%	0.067	90.78%	0.044	82.06%
	TCN-GRU	0.079	95.19%	0.062	92.28%	0.041	86.97%
	XAJ-TCN-GRU	0.034	95.29%	0.027	93.49%	0.018	89.58%
Qingyi	XAJ	0.228	73.45%	0.178	60.72%	0.117	42.99%
	LSTM	0.101	86.07%	0.079	77.25%	0.052	56.21%
	TCN	0.083	87.80%	0.065	90.78%	0.043	63.83%
	GRU	0.080	93.09%	0.062	88.58%	0.041	73.95%
	TCN-GRU	0.081	94.29%	0.063	90.58%	0.041	79.66%
	XAJ-TCN-GRU	0.034	95.09%	0.027	90.78%	0.018	81.46%

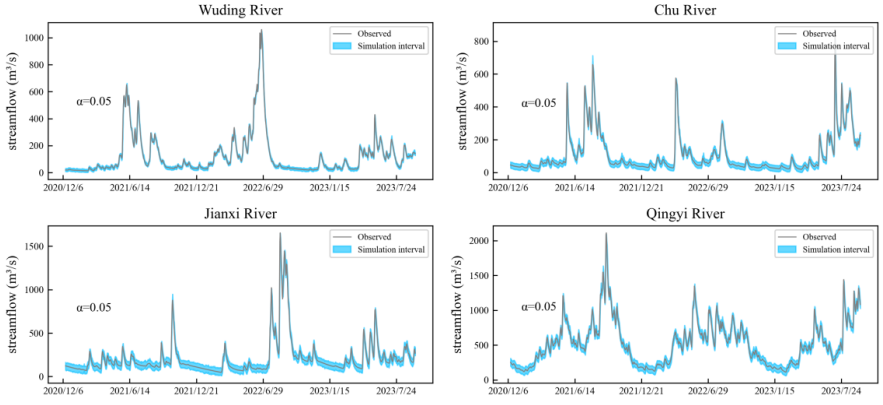


Fig. 10 Interval simulation outcomes of the XAJ-TCN-GRU model for four basins at α=0.05.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

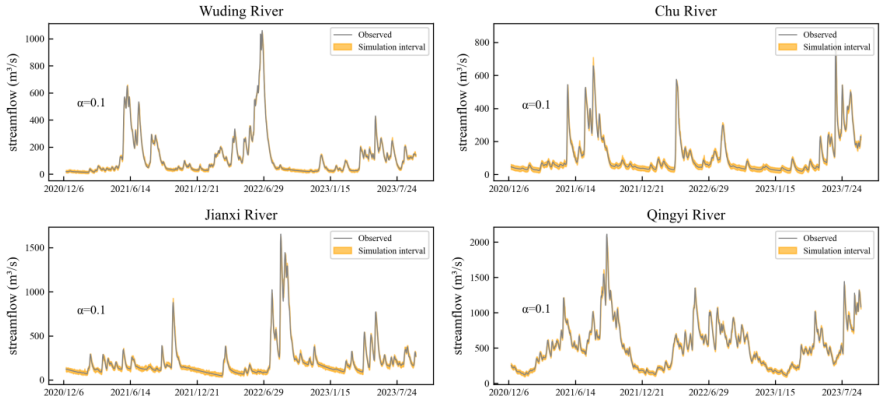


Fig. 11 Interval simulation outcomes of the XAJ-TCN-GRU model for four basins at α=0.1.

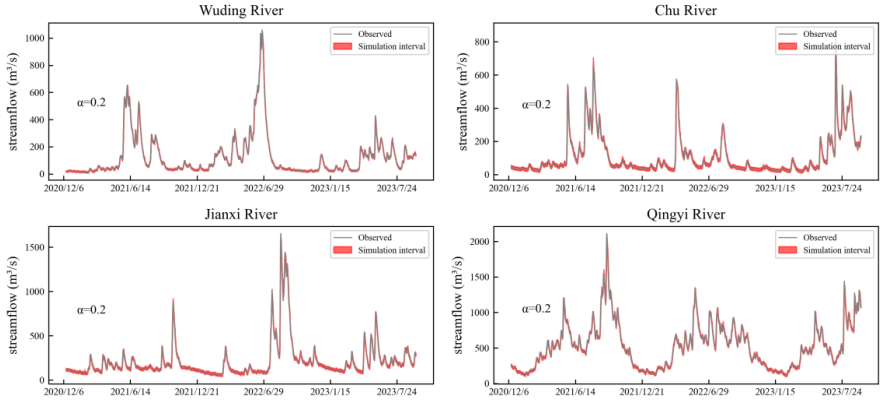


Fig. 12 Interval simulation outcomes of the XAJ-TCN-GRU model at four basins for α=0.2.

5.4 Interpretability of deep learning model

Although the XAJ-TCN-GRU model demonstrates excellent performance across all basins, showcasing strong stability and robustness, the lack of transparency in deep learning models hinders the verification of their simulation logic. To improve the interpretability of the XAJ-TCN-GRU, this study employs three key indicators: mean absolute SHAP values (SHAPABS), Feature Importance (FI), and Permutation Feature Importance (PFI). These indicators can accurately calculate the specific contribute of each variable input to long-term trends on streamflow in the TCN-GRU model, thereby offering the clearer insight of the internal operation mechanism of TCN-GRU model.

Fig. 13 presents density scatter plots of various input variables across the four basins, with the y-axis arranged in descending order according to their contributions to the simulated values per sample. Positive SHAP values indicate contributions that enhance the model's performance, with larger values reflecting more substantial contributions. In cases where multiple samples exhibit identical contribution values, these samples are vertically stacked within the plots. **Fig. 14** provides a comprehensive overview of the importance ranking of hydro-meteorological variables influencing streamflow, as determined by the TCN-GRU. The numerical values of three key indicators for the input variables of the TCN-GRU across the four basins, along with their normalization, are detailed in **Supplementary Information 4**.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

Combining **Fig. 13** and **Fig. 14**, it can be seen that dew point temperature is the most influential feature variable in the basins of Wuding River, Chu River, and Qingyi River. A small difference between the dew point temperature and the surrounding environmental temperature suggests a high water vapor content in the air, indicating high humidity. In high-humidity environments, water vapor is prone to condensation, which may significantly affect hydrological cycles, precipitation processes, and evaporation processes within the basin.

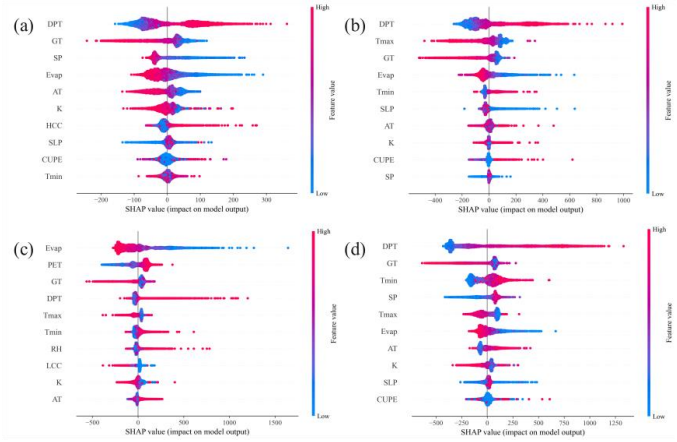
Furthermore, the second influential variables in each basin exhibit significant differences, reflecting the importance of varying hydrological characteristics and climatic conditions for flow simulation. In the Wuding River basin, surface pressure is identified as the second most important variable, which is closely related to the region's arid climate characteristics. In arid environments, surface pressure plays a critical role in the hydrological cycle, influencing the formation of precipitation and the evaporation of water. In contrast, in the Chu River basin, the daily maximum temperature is the second most important variable, indicating that temperature fluctuations in a humid environment may directly affect the evaporation and flow of water bodies. Meanwhile, in the Qingyi River basin, the influence of daily minimum temperature is noted, reflecting how variations in daily minimum temperatures may directly affect the cooling and condensation processes of water bodies, thereby impacting flow. We observed that evaporation also has a significant impact on streamflow in all four basins, with lower evaporation values having a greater impact on streamflow simulation. When evaporation is low, it implies that more water remains in liquid form on the surface or in water bodies, potentially leading to more streamflow.

Fig. 13 shows that the importance rankings of different feature variables vary significantly across the three interpretation methods in the four basins. Dew point temperature generally ranks high in both SHAPABS and PFI across all basins, indicating its high importance in streamflow simulation. However, in FI, the importance of dew point temperature does not always rank at the top. For example, in the Chu River and Jianxi River basins, the FI method places more emphasis on other variables, such as surface pressure and evapotranspiration. This difference suggests that FI may focus more on the impact of local variables on model decisions, while PFI and SHAPABS take into account the overall contribution of variables to simulation performance.

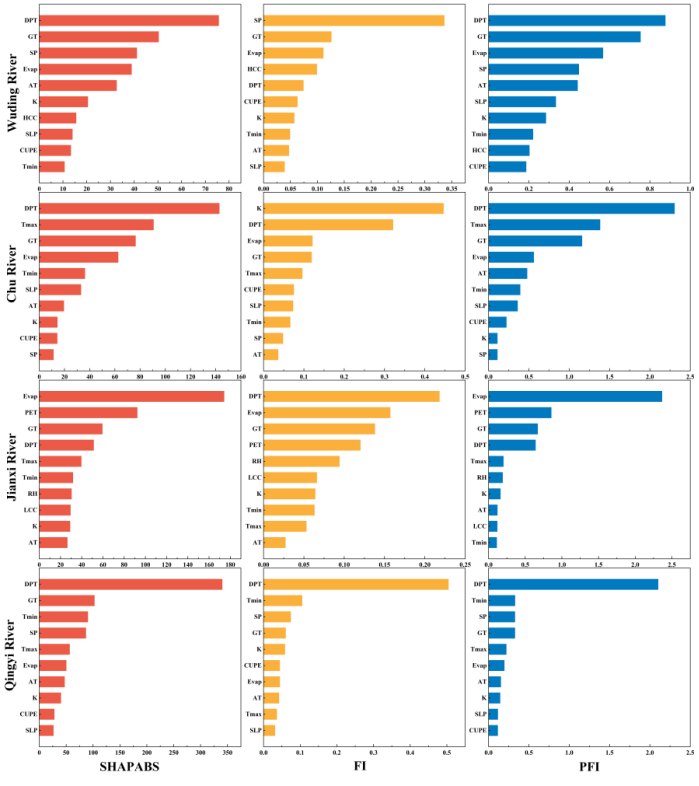
In comparison, SHAPABS and PFI show relatively consistent rankings across most basins, with variables such as dew point temperature and ground temperature maintaining stable positions. This indicates that SHAPABS and PFI are more robust in identifying variable importance, while FI is more influenced by basin characteristics, leading to greater variation in variable rankings across basins.

These metrics offer meaningful insights by elucidating the individual contributions of each input variable to model performance. These techniques aid in identifying the most critical variables during the analysis, thereby deepening understanding of the factors driving model simulations and providing guidance for future research improvements.

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025



572
573 **Fig. 13** SHAP scores for the key external input variables at (a) Wuding River (b) Chu River
574 (c)Jianxi River, and (d) Qingyi River basins.



575
576 **Fig. 14** Bar plots of the SHAPABS, FI, and PFI of input variables based on the TCN-GRU.

577 **6 Conclusion**

578 This study presents an innovative hybrid model for streamflow simulation that effectively integrates
579 the strengths of both conceptual rainfall - runoff and deep learning models. By employing nonlinear
580 ensemble, it addresses the potential shortcomings of each approach in streamflow simulation. The results

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

indicate that our proposed XAJ-TCN-GRU model achieves superior accuracy in streamflow simulation, significantly outperforming other benchmark models. Additionally, the robustness of the XAJ-TCN-GRU model and its exceptional performance in flood simulation and interval simulation are explored, further highlighting the model's broad potential and flexibility in applications of hydrological modelling. In order to comprehensively understand the model's capabilities, SHAPABS, FI, and PFI are introduced as evaluation tools. These tools quantify the contributing factors of different inputs on the long-term trends for streamflow across various basins, thereby enhancing the model's external interpretability.

However, this study has the following limitations in the development of the flood simulation model:

(1) The input variables did not account for lagged features, overlooking the temporal dependencies and historical impacts of flow changes.

(2) This study primarily relied on basic hydrological and meteorological variables for feature selection, without incorporating broader environmental factors (e.g., land use types, vegetation cover). This limitation may prevent the model from fully capturing the multidimensional factors influencing flood variability, potentially reducing simulation accuracy in complex environments.

Future research can build on these limitations to further enhance the model. Overall, this study presents a novel and efficient approach for streamflow simulation, featuring broad application prospects in watershed water management and flood disaster warning. In addition, this study highlights the key value of nonlinear sets in hydrological models, providing important references and insights for subsequent related research.

Acknowledgments

It was supported by the National Natural Science Foundation of China (No. 11701363) and Humanities and Social Sciences Research Planning Fund Program, Ministry of Education, China (No. 24YJAZH167).

CRediT authorship contribution statement

ZW and NX designed the research. NX and XZ collected and preprocessed the data. ZW and NX conducted all the experiments and analyzed the results. JW assisted with the paper's background. NX wrote the first draft of the manuscript with contributions from ZW. WS and XC supervised the study and edited the manuscript.

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.

Data Availability Statement

Streamflow from this study site was obtained from the Hydrological Yearbook of the People's Republic of China and provided by the Shanghai Qingyue Information Technology Service Centre (https://data.epmap.org/page/index); meteorological data were obtained from the China Meteorological Network (https://weather.cma.cn/). Model source code can be obtained from its Github repository (https://github.com/zcwang1028/code.git).

References

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

Ahmadpour, A., Mirhashemi, S., Haghighat jou, P., & Foroughi, F. (2022). Comparison of the monthly streamflow forecasting in Maroon dam using HEC-HMS and SARIMA models. Sustainable *Water Resources Management*, **8**(5), 158. https://doi.org/10.1007/s40899-022-00686-1

Ahmed, A. M., Deo, R. C., Ghahramani, A., Feng, Q., Raj, N., Yin, Z., & Yang, L. (2022). New double decomposition deep learning methods for river water level forecasting. *Science of The Total Environment*, **831**, 154722. https://doi.org/10.1016/j.scitotenv.2022.154722

Alnahit, A. O., Mishra, A. K., & Khan, A. A. (2022). Stream water quality prediction using boosted regression tree and random forest models. *Stochastic Environmental Research and Risk Assessment*, **36**(9), 2661-2680. https://doi.org/10.1007/s00477-021-02152-4

Bai, P., Liu, X., Liang, K., Liu, X., & Liu, C. (2017). A comparison of simple and complex versions of the Xinanjiang hydrological model in predicting runoff in ungauged basins. *Hydrology Research*, **48**(5), 1282-1295. https://doi.org/10.2166/nh.2016.094

Chen, Z., Lin, H., & Shen, G. (2023). TreeLSTM: A spatiotemporal machine learning model for rainfall-runoff estimation. *Journal of Hydrology: Regional Studies*, **48**, 101474. https://doi.org/10.1016/j.ejrh.2023.101474

Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., & Pain, C. C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, **590**, 125376. https://doi.org/10.1016/j.jhydrol.2020.125376

Cho, K., & Kim, Y. (2022). Improving streamflow prediction in the WRF-Hydro model with LSTM networks. *Journal of Hydrology*, **605**, 127297. https://doi.org/10.1016/j.jhydrol.2021.127297

Contreras, P., Orellana-Alvear, J., Muñoz, P., Bendix, J., & Célleri, R. (2021). Influence of random forest hyperparameterization on short-term runoff forecasting in an andean mountain catchment. *Atmosphere*, **12**(2), 238. https://doi.org/10.3390/atmos12020238

Doyle, J. M., Hill, R. A., Leibowitz, S. G., & Ebersole, J. L. (2023). Random forest models to estimate bankfull and low flow channel widths and depths across the conterminous United States. *JAWRA Journal of the American Water Resources Association*, **59**(5), 1099-1114. https://doi.org/10.1111/1752-1688.13116

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *Journal of Hydrology*, **589**, 125188. https://doi.org/10.1016/j.jhydrol.2020.125188

Gebremariam, S. Y., Martin, J. F., DeMarchi, C., Bosch, N. S., Confesor, R., & Ludsin, S. A. (2014). A comprehensive approach to evaluating watershed models for predicting river flow regimes critical to downstream ecosystem services. *Environmental modelling & software*, **61**, 121-134. https://doi.org/10.1016/j.envsoft.2014.07.004

Gong, J., Yao, C., Li, Z., Chen, Y., Huang, Y., & Tong, B. (2021). Improving the flood forecasting capability of the Xinanjiang model for small-and medium-sized ungauged catchments in South China. *Natural Hazards*, **106**, 2077-2109. https://doi.org/10.1007/s11069-021-04531-0

Granata, F., Zhu, S., & Di Nunno, F. (2024). Advanced streamflow forecasting for Central European Rivers: the cutting-edge Kolmogorov-Arnold networks compared to Transformers. *Journal of Hydrology*, **645**, 132175. https://doi.org/10.1016/j.jhydrol.2024.132175

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

661 Han, H., & Morrison, R. R. (2022). Improved runoff forecasting performance through error predictions
662 using a deep-learning approach. *Journal of Hydrology*, **608**, 127653.
663 https://doi.org/10.1016/j.jhydrol.2022.127653
664 Hao, F., Sun, M., Geng, X., Huang, W., & Ouyang, W. (2015). Coupling the Xinanjiang model with
665 geomorphologic instantaneous unit hydrograph for flood forecasting in northeast China. *International*
666 *Soil and Water Conservation Research*, **3**(1), 66-76. https://doi.org/10.1016/j.iswcr.2015.03.004
667 Hwang, Y., Clark, M. P., & Rajagopalan, B. (2011). Use of daily precipitation uncertainties in streamflow
668 simulation and forecast. *Stochastic environmental research and risk assessment*, **25**, 957-972.
669 https://doi.org/10.1007/s00477-011-0460-1
670 Jehanzaib, M., Bilal Idrees, M., Kim, D., & Kim, T. W. (2021). Comprehensive evaluation of machine
671 learning techniques for hydrological drought forecasting. *Journal of Irrigation and Drainage*
672 *Engineering*, **147**(7), 04021022. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001575
673 Jiang, X., Zhang, L., Liang, Z., Fu, X., Wang, J., Xu, J., Zhang, Y., & Zhong, Q. (2023). Study of early
674 flood warning based on postprocessed predicted precipitation and Xinanjiang model. *Weather and*
675 *Climate Extremes*, **42**, 100611. https://doi.org/10.1016/j.wace.2023.100611
676 Katipoğlu, O. M., & Sarıgöl, M. (2023). Improving the accuracy of rainfall-runoff relationship estimation
677 using signal processing techniques, bio-inspired swarm intelligence and artificial intelligence
678 algorithms. *Earth Science Informatics*, **16**(4), 3125-3141. https://doi.org/10.1007/s12145-023-01081-
679 w
680 Kim, T., Yang, T., Gao, S., Zhang, L., Ding, Z., Wen, X., Gourley, J. J., & Hong, Y. (2021). Can artificial
681 intelligence and data-driven machine learning models match or even replace process-driven hydrologic
682 models for streamflow simulation?: A case study of four watersheds with different hydro-climatic
683 regions across the CONUS. *Journal of Hydrology*, **598**, 126423.
684 https://doi.org/10.1016/j.jhydrol.2021.126423
685 Kurian, C., Sudheer, K. P., Vema, V. K., & Sahoo, D. (2020). Effective flood forecasting at higher lead
686 times through hybrid modelling framework. *Journal of Hydrology*, **587**, 124945.
687 https://doi.org/10.1016/j.jhydrol.2020.124945
688 Lei, X., Cheng, L., Ye, L., Zhang, L., KIM, J. S., Qin, S., & Liu, P. (2023). Integration of the generalized
689 complementary relationship into a lumped hydrological model for improving water balance
690 partitioning: A case study with the Xinanjiang model. *Journal of Hydrology*, **621**, 129569.
691 https://doi.org/10.1016/j.jhydrol.2023.129569
692 Leonarduzzi, E., Maxwell, R. M., Mirus, B. B., & Molnar, P. (2021). Numerical analysis of the effect of
693 subgrid variability in a physically based hydrological model on runoff, soil moisture, and slope
694 stability. *Water Resources Research*, **57**(4), e2020WR027326.
695 https://doi.org/10.1029/2020WR027326
696 Lin, K., Sheng, S., Zhou, Y., Liu, F., Li, Z., Chen, H., Xu, C., Chen, J., & Guo, S. (2020). The exploration
697 of a temporal convolutional network combined with encoder-decoder framework for runoff forecasting.
698 *Hydrology Research*, **51**(5), 1136-1149. https://doi.org/10.2166/nh.2020.100

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

699 Liu, S., Qin, H., Liu, G., Xu, Y., Zhu, X., & Qi, X. (2023). Runoff forecasting of machine learning Model
700 based on selective ensemble. *Water Resources Management*, **37**(11), 4459-4473.
701 https://doi.org/10.1007/s11269-023-03566-1
702 Ng, K. W., Huang, Y. F., Koo, C. H., Chong, K. L., El-Shafie, A., & Ahmed, A. N. (2023). A review of
703 hybrid deep learning applications for streamflow forecasting. *Journal of Hydrology*, **625**, 130141.
704 https://doi.org/10.1016/j.jhydrol.2023.130141
705 Oruç, H. N., Çelen, M., Gülgen, F., Öncel, M. S., Vural, S., & Kılıç, B. (2023). Evaluating the effects of
706 soil data quality on the SWAT runoff prediction Performance; A case study of Saz-Cayirova catchment,
707 Turkey. *Urban Water Journal*, **20**(10), 1592-1607. https://doi.org/10.1080/1573062X.2022.2056060
708 Özgen-Xian, I., Kesserwani, G., Caviedes-Voullième, D., Molins, S., Xu, Z., Dwivedi, D., Moulton, J.
709 D., & Steefel, C. I. (2020). Wavelet-based local mesh refinement for rainfall – runoff simulations.
710 *Journal of Hydroinformatics*, **22**(5), 1059-1077. https://doi.org/10.2166/hydro.2020.198
711 Parisouj, P., Mokari, E., Mohebzadeh, H., Goharnejad, H., Jun, C., Oh, J., & Bateni, S. M. (2022).
712 Physics-informed data-driven model for predicting streamflow: A case study of the Voshmgir Basin,
713 Iran. *Applied Sciences*, **12**(15), 7464. https://doi.org/10.3390/app12157464
714 Qiao, X., Peng, T., Sun, N., Zhang, C., Liu, Q., Zhang, Y., Wang, Y., & Nazir, M. S. (2023).
715 Metaheuristic evolutionary deep learning model based on temporal convolutional network, improved
716 aquila optimizer and random forest for rainfall-runoff simulation and multi-step runoff prediction.
717 *Expert Systems with Applications*, **229**, 120616. https://doi.org/10.1016/j.eswa.2023.120616
718 Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander,
719 E. S., Mitzenmacher, M., & Sabeti, P. C. (2011). Detecting novel associations in large data sets.
720 *science*, **334**(6062), 1518-1524. https://doi.org/10.1126/science.1205438
721 Samantaray, S., Das, S. S., Sahoo, A., & Satapathy, D. P. (2022). Monthly runoff prediction at Baitarani
722 river basin by support vector machine based on Salp swarm algorithm. *Ain Shams Engineering Journal*,
723 **13**(5), 101732. https://doi.org/10.1016/j.asej.2022.101732
724 Shao, P., Feng, J., Lu, J., Zhang, P., & Zou, C. (2024). Data-driven and knowledge-guided denoising
725 diffusion model for flood forecasting. *Expert Systems with Applications*, **244**, 122908.
726 https://doi.org/10.1016/j.eswa.2023.122908
727 Song, J., Meng, H., Kang, Y., Zhu, M., Zhu, Y., & Zhang, J. (2024). A method for predicting water
728 quality of river basin based on OVMD-GAT-GRU. *Stochastic Environmental Research and Risk*
729 *Assessment*, **38**(1), 339-356. https://doi.org/10.1007/s00477-023-02584-0
730 Song, Y., Qin, S., Qu, J., & Liu, F. (2015). The forecasting research of early warning systems for
731 atmospheric pollutants: A case in Yangtze River Delta region. *Atmospheric Environment*, **118**, 58-69.
732 https://doi.org/10.1016/j.atmosenv.2015.06.032
733 Sun, P., Wang, J., & Yan, Z. (2024). Ultra-short-term wind speed prediction based on TCN-MCM-EKF.
734 *Energy Reports*, **11**, 2127-2140. https://doi.org/10.1016/j.egyr.2024.01.058
735 Thébault, C., Perrin, C., Andréassian, V., Thirel, G., Legrand, S., & Delaigue, O. (2024). Multi-model
736 approach in a variable spatial framework for streamflow simulation. *Hydrology and Earth System*
737 *Sciences*, **28**(7), 1539-1566. https://doi.org/10.5194/hess-28-1539-2024

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

738 Vilaseca, F., Castro, A., Chreties, C., & Gorgoglione, A. (2023). Assessing influential rainfall–runoff
739 variables to simulate daily streamflow using random forest. *Hydrological Sciences Journal*, **68**(12),
740 1738-1753. https://doi.org/10.1080/02626667.2023.2232356
741 Wang, H., Qin, H., Liu, G., Liu, S., Qu, Y., Wang, K., & Zhou, J. (2023). A novel feature attention
742 mechanism for improving the accuracy and robustness of runoff forecasting. *Journal of Hydrology*,
743 **618**, 129200. https://doi.org/10.1016/j.jhydrol.2023.129200
744 Wang, J., Cheng, Q., & Sun, X. (2022). Carbon price forecasting using multiscale nonlinear integration
745 model coupled optimal feature reconstruction with biphasic deep learning. *Environmental Science and*
746 *Pollution Research*, **29**(57), 85988-86004. https://doi.org/10.1007/s11356-021-16089-2
747 Wang, J., & Dong, Y. (2024). An interpretable deep learning multi-dimensional integration framework
748 for exchange rate forecasting based on deep and shallow feature selection and snapshot ensemble
749 technology. *Engineering Applications of Artificial Intelligence*, **133**, 108282.
750 https://doi.org/10.1016/j.engappai.2024.108282
751 Wei, X., Wang, G., Schmalz, B., Hagan, D. F. T., & Duan, Z. (2023). Evaluate Transformer model and
752 Self-Attention mechanism in the Yangtze River basin runoff prediction. *Journal of Hydrology:*
753 *Regional Studies*, **47**, 101438. https://doi.org/10.1016/j.ejrh.2023.101438
754 Wu, J., Wang, Z., Dong, J., Cui, X., Tao, S., & Chen, X. (2023). Robust runoff prediction with
755 explainable artificial intelligence and meteorological variables from deep learning ensemble model.
756 *Water Resources Research*, **59**(9), e2023WR035676. https://doi.org/10.1029/2023WR035676
757 Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, L., & Ran, G. (2023). Deep transfer learning based on
758 transformer for flood forecasting in data-sparse basins. *Journal of Hydrology*, **625**, 129956.
759 https://doi.org/10.1016/j.jhydrol.2023.129956
760 Xu, Y., Liu, T., Fang, Q., Du, P., & Wang, J. (2025). Crude oil price forecasting with multivariate
761 selection, machine learning, and a nonlinear combination strategy. *Engineering Applications of*
762 *Artificial Intelligence*, **139**, 109510. https://doi.org/10.1016/j.engappai.2024.109510
763 Xuan, W., Shouxiang, W., Qianyu, Z., Shaomin, W., & Liwei, F. (2021). A multi-energy load prediction
764 model based on deep multi-task learning and ensemble approach for regional integrated energy
765 systems. *International Journal of Electrical Power & Energy Systems*, **126**, 106583.
766 https://doi.org/10.1016/j.ijepes.2020.106583
767 Zhang, Y., Ye, A., Analui, B., Nguyen, P., Sorooshian, S., Hsu, K., & Wang, Y. (2023). Comparing
768 quantile regression forest and mixture density long short-term memory models for probabilistic post-
769 processing of satellite precipitation-driven streamflow simulations. *Hydrology and Earth System*
770 *Sciences*, **27**(24), 4529-4550. https://doi.org/10.5194/hess-27-4529-2023
771 Zhao, R. J. (1992). The Xinanjiang model applied in China. *Journal of hydrology*, 135(1-4), 371-381.
772 https://doi.org/10.1016/0022-1694(92)90096-E
773 Zhao, R. J., & Liu, X. R. (1995). The Xinanjiang model. *Computer models of watershed hydrology*., 215-
774 232.
775 Zhao, Y., Huang, Y., Wang, Z., & Liu, X. (2024). Carbon futures price forecasting based on feature
776 selection. *Engineering Applications of Artificial Intelligence*, **135**, 108646.
777 https://doi.org/10.1016/j.engappai.2024.108646

https://doi.org/10.5194/egusphere-2025-2377
Preprint. Discussion started: 24 June 2025

778 Zheng, Z., Ali, M., Jamei, M., Xiang, Y., Karbasi, M., Yaseen, Z. M., & Farooque, A. A. (2023). Design
779 data decomposition-based reference evapotranspiration forecasting model: a soft feature filter based
780 deep learning driven approach. *Engineering Applications of Artificial Intelligence*, **121**, 105984.
781 https://doi.org/10.1016/j.engappai.2023.105984
782 Zhu, S., Wei, J., Zhang, H., Xu, Y., & Qin, H. (2023). Spatiotemporal deep learning rainfall-runoff
783 forecasting combined with remote sensing precipitation products in large scale basins. *Journal of*
784 *Hydrology*, **616**, 128727. https://doi.org/10.1016/j.jhydrol.2022.128727
785 Zuo, G., Luo, J., Wang, N., Lian, Y., & He, X. (2020). Decomposition ensemble model based on
786 variational mode decomposition and long short-term memory for streamflow forecasting. *Journal of*
787 *Hydrology*, **585**, 124776. https://doi.org/10.1016/j.jhydrol.2020.124776