https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

A Physics-Constrained Bayesian Transformer framework for reliable probabilistic flood forecasting across contrasting hydrological regimes

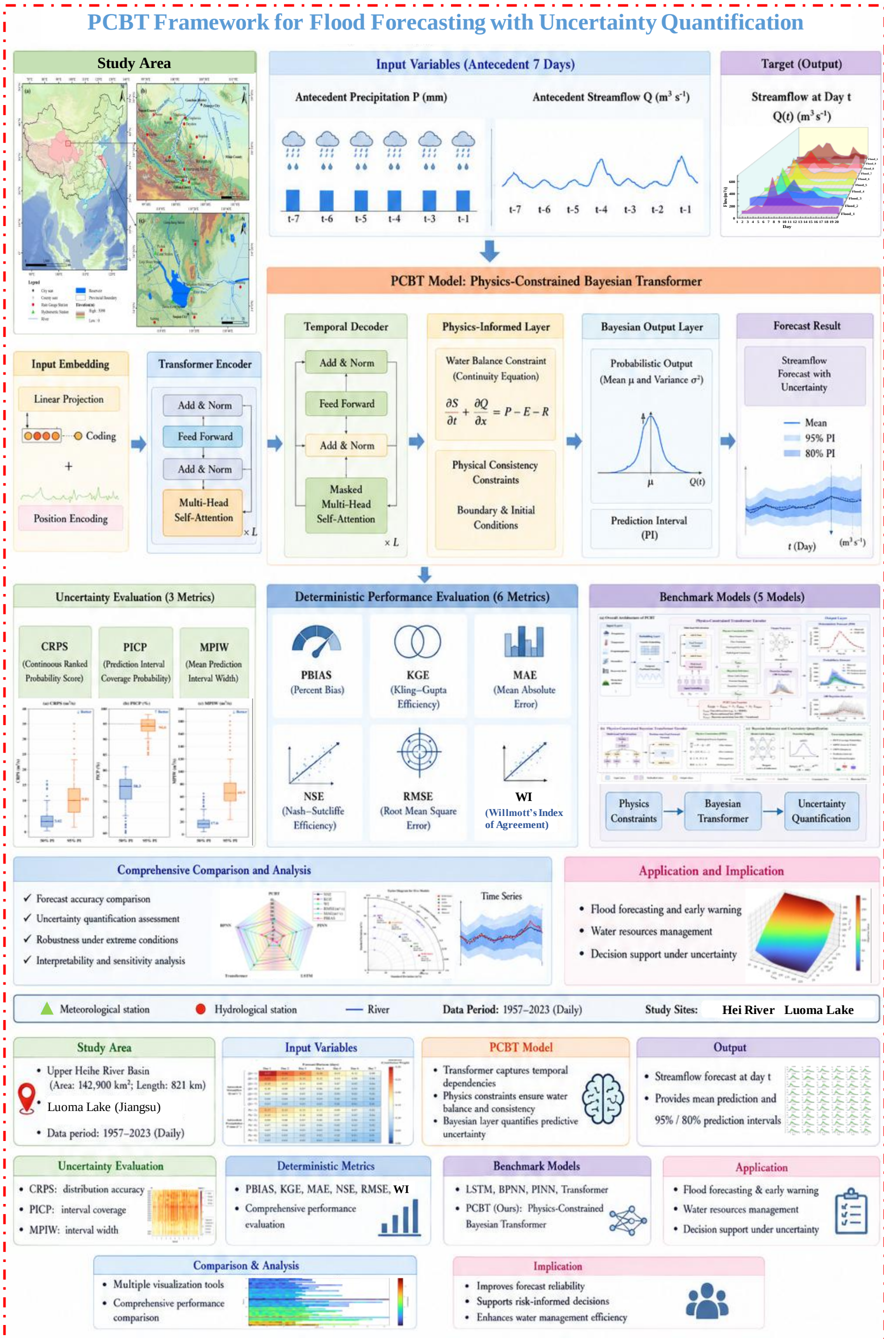
Mingrui Shi[a], Hongyuan Fang[a*]

a. College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou 225008, China

[*] Corresponding authors at: 1. College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou, 225008, China. Tel.: +86 13952794936; e-mail address: 18752731501@163.com. (Hongyuan Fang)

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Graphical abstract:



https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Abstract: Flood forecasting remains challenging due to the nonlinear nature of rainfall–runoff processes, limited observations of extreme events, and inherent uncertainties associated with hydrological prediction. Although deep learning models have demonstrated strong predictive capability, their purely data-driven structures often lack physical consistency and provide limited information regarding forecast reliability. This study develops a Physics-Constrained Bayesian Transformer (PCBT) framework that integrates temporal representation learning, hydrological physical constraints, and Bayesian uncertainty quantification for probabilistic flood forecasting. The proposed framework combines a Transformer encoder to capture long-range dependencies within antecedent precipitation and streamflow sequences, physics-informed constraints based on water balance and flow continuity principles to regulate hydrologically feasible predictions, and a Bayesian inference module to characterize predictive uncertainty through probabilistic forecasting. The framework was evaluated using 50 independent flood events from two hydrologically contrasting watersheds in China: the mountainous Upper Hei River Basin and the lowland Luoma Lake Basin. The two basins represent distinct runoff generation mechanisms and provide a rigorous test of model robustness under heterogeneous hydrological conditions. Results demonstrate that PCBT effectively reproduces flood hydrograph dynamics and improves both deterministic accuracy and probabilistic reliability compared with Long Short-Term Memory (LSTM), Transformer, Physics-Informed Neural Network (PINN), and Backpropagation Neural Network (BPNN) benchmarks. Across independent testing events, PCBT achieved Nash–Sutcliffe Efficiency values exceeding 0.85 while providing well-calibrated prediction intervals for uncertainty representation. The Bayesian component successfully captured increased uncertainty during rapid rising stages and flood peaks, whereas physical constraints improved the consistency between predicted and observed hydrological responses. This study demonstrates that integrating deep temporal learning, hydrological knowledge, and probabilistic inference provides an effective pathway toward reliable flood forecasting under diverse hydrological regimes. The proposed framework offers a transferable approach for uncertainty-aware flood prediction and supports risk-informed water resources management under changing environmental conditions.

Keywords: Probabilistic flood forecasting; Physics-informed machine learning; Bayesian Transformer; Prediction uncertainty; Hydrological constraints; Contrasting basins

1. Introduction

Flood forecasting is a fundamental component of hydrological prediction and water resources management because accurate forecasts provide critical support for flood mitigation, reservoir operation, emergency response, ecological protection, and sustainable water allocation (Montanari et al., 2013; Blöschl et al., 2019; Shen, 2018). Under the combined influences of climate change, intensified human activities, and increasing hydrological variability, extreme flood events have become more frequent and severe in many regions worldwide (IPCC, 2021; Gupta et al., 2014; Muñoz-Castro et al., 2026). Reliable flood forecasting therefore plays a crucial role in reducing flood-related losses and improving watershed resilience. However, flood generation and

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

routing processes are governed by complex nonlinear interactions among precipitation, antecedent moisture conditions, watershed characteristics, and channel hydraulics. These interactions introduce considerable uncertainty into hydrological forecasting and make accurate prediction of flood evolution and peak discharge particularly challenging (Gupta et al., 2014; Kratzert et al., 2018).

To address these challenges, data-driven approaches have been increasingly adopted in hydrological forecasting. Unlike conceptual hydrological models, machine learning algorithms can automatically learn nonlinear relationships directly from observations without requiring explicit descriptions of physical processes (Chen et al., 2025a; Liao et al., 2025; Solanki et al., 2025). Early studies mainly relied on backpropagation neural networks (BPNNs), support vector machines (SVMs), and random forest models for rainfall–runoff simulation (Kumar et al., 2026; Amini et al., 2023). More recently, deep learning techniques have significantly advanced hydrological forecasting by improving the representation of nonlinear temporal dynamics. Among them, Long Short-Term Memory (LSTM) networks have demonstrated remarkable capability in capturing long-term temporal dependencies and have become one of the most widely used deep learning models in hydrology (Gegenleithner et al., 2025; Khatun et al., 2023; Zhu et al., 2025b). Numerous studies have reported that LSTM-based models outperform traditional conceptual models and shallow machine learning methods in streamflow forecasting, rainfall–runoff simulation, and flood prediction tasks (Zou et al., 2023; Chen et al., 2022; Mosaffa et al., 2026). For instance, Gegenleithner et al. (2025) demonstrated that LSTM-based hybrid forecasting frameworks substantially

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

outperformed conventional ARIMA models in operational flood forecasting, particularly for long lead-time predictions. This finding suggests that deep learning architectures possess superior capabilities for extracting long-term temporal dependencies from hydrological time series. Chen et al. (2022) proposed a ConvLSTM framework by integrating convolutional neural networks and LSTM networks to jointly extract spatial and temporal features from hydrological data. Their results demonstrated that incorporating spatial information significantly improved flood arrival time and peak discharge predictions compared with conventional time-series models. Zhu et al. (2023) proposed an optimized model based on LSTM networks for early warning and forecasting of ponding in urban drainage systems. This approach can quickly identify and locate areas of ponding while maintaining high predictive accuracy. Zhu et al. (2025b) proposed a hybrid hydrological forecasting framework that integrates the Xin-an-jiang (XAJ) model and LSTM networks to simultaneously capture physical processes and nonlinear relationships. By combining PSO and Bayesian optimization for parameter calibration, Bayesian inference for uncertainty quantification, and SHAP for model interpretability, the framework significantly improved forecasting accuracy, reliability, and probabilistic prediction capability.

Recently, Transformer architectures have emerged as a promising alternative for hydrological time-series forecasting (Saravana et al., 2026). Originally developed for natural language processing, Transformers employ self-attention mechanisms to capture long-range dependencies without relying on recurrent structures (Ambika et al., 2025; Gao et al., 2024). Compared with LSTM models, Transformers can process long

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

sequences more efficiently and better represent complex interactions among multiple hydrometeorological variables (Pölz et al., 2024; He et al., 2026). Their ability to model long-term hydrological memory has led to growing applications in streamflow forecasting, rainfall–runoff simulation, and regional hydrological prediction (Arsenault et al., 2023; Ling et al., 2026). For instance, Fadillah et al. (2025) proposed a hybrid LSTM–Transformer framework that improved flood forecasting accuracy by jointly capturing local and global temporal dependencies, outperforming conventional machine learning models. Xia et al. (2025) proposed a spatiotemporal Pathformer model that improved multi-step flood forecasting by effectively capturing nonlinear spatiotemporal dependencies, outperforming both LSTM and Transformer models. Ambika et al. (2025) developed a Transformer-based FutureTST model for streamflow forecasting that jointly utilizes historical and future hydrometeorological information. The model achieved robust performance over extended forecast horizons and outperformed traditional process-based hydrological models. Nevertheless, despite their strong predictive capability, Transformer-based models remain purely data-driven. Model training depends entirely on historical observations, and the resulting predictions are not explicitly constrained by hydrological principles (Liu et al., 2026). Consequently, these models may generate physically inconsistent outputs, especially under extrapolation conditions, sparse-data scenarios, or extreme flood events where observations are limited (Pan et al., 2026; Küllahcı et al., 2026).

To improve physical consistency, Physics-Informed Neural Networks (PINNs) have recently attracted considerable attention in scientific machine learning (Raissi et al.,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

2019). PINNs incorporate governing equations directly into neural network training by embedding physical residuals into the loss function, thereby constraining model solutions within physically feasible regions. This framework has achieved remarkable success in fluid dynamics, groundwater modeling, environmental systems, and hydrological simulations (Raissi et al., 2019; Yang et al., 2026; Virupaksha et al., 2026). For instance, Virupaksha et al. (2026) investigated continuous-time, discrete-time, and time-decomposition PINNs for transient groundwater flow modeling. Their results demonstrated that the discrete-time PINN achieved superior accuracy and training efficiency, reducing computational costs while improving performance under complex dynamic conditions. Jia et al. (2025) developed a multi-scale hybrid learning framework that integrates PINNs, neural operators, and convolutional neural networks for hydrodynamic transport modeling. By combining multi-scale feature extraction with hard physical constraints, the framework improved prediction accuracy, robustness, and physical consistency in complex spatiotemporal systems. In hydrology, PINNs have demonstrated advantages in improving model robustness, reducing dependence on large training datasets, and enhancing extrapolation capability under data-scarce conditions (Feng et al., 2022; Read et al., 2019). Recent studies have applied PINNs to flood forecasting, water-level prediction, and rainfall–runoff simulation, showing that physical constraints can significantly improve model interpretability and hydrological consistency (Wei et al., 2026; Lu et al., 2025; Tian et al., 2025). For instance, Wei et al. (2026) proposed a Physics-Informed Gated Transformer framework that integrates soft hydrological constraints, gated residual

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

networks, and variance-constrained learning to improve flood forecasting in data-scarce catchments. The model effectively reduced phase-lag errors and enhanced robustness to noisy inputs, outperforming conventional LSTM, HBV, and standard Transformer models in both forecasting accuracy and spatial generalization. Lu et al. (2025) proposed a PINN framework that integrates physically guided hyper-parameter selection and data augmentation to improve lake water level forecasting. The model effectively addressed data imbalance under extreme conditions and achieved higher accuracy and computational efficiency than conventional deep learning and hydrodynamic models. Tian et al. (2025) enhanced PINNs for solving two-dimensional shallow water equations by integrating dimensional transformation and neuron-wise adaptive activation functions. Their results showed that a hybrid primitive–conservative formulation improved simulation accuracy and stability, highlighting the potential of PINNs for physically consistent flood modeling.

Despite these advances, several limitations remain unresolved. Firstly, most existing PINN applications employ fully connected feedforward neural networks as the backbone architecture. Although such structures effectively enforce physical constraints, they are not specifically designed to learn long-range temporal dependencies embedded in flood evolution processes (Chen et al., 2025a). Consequently, their ability to represent complex temporal dynamics remains limited. Second, most Transformer-based hydrological studies focus primarily on prediction accuracy while neglecting hydrological consistency. As a result, physically unrealistic fluctuations may occur in predicted hydrographs, particularly during flood peaks and

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

recession periods (Zhu et al., 2025c; Adibfar et al., 2025; Liu et al., 2026). Thirdly, uncertainty quantification remains largely overlooked in existing physics-informed flood forecasting studies (Kohanpur et al., 2023). Most models provide deterministic forecasts only, making it difficult to assess prediction reliability and support risk-based decision making (Xu et al., 2022; Zhang et al., 2025). Reliable uncertainty estimation is particularly important for operational flood forecasting because management decisions often depend not only on forecast accuracy but also on the confidence associated with predictions (Xu et al., 2022; Han et al., 2017). Nevertheless, most existing deep learning models generate deterministic forecasts and provide limited information regarding predictive uncertainty. To overcome this limitation, Bayesian inference has been increasingly incorporated into hydrological forecasting frameworks. For instance, Zhu et al. (2025a) proposed a Bayesian multi-model ensemble framework for probabilistic groundwater level forecasting. The framework integrates multiple categories of input variables, including autocorrelation, meteorological, hydrological, and anthropogenic factors, to capture the lagged effects and driving mechanisms governing groundwater dynamics.

Therefore, three critical challenges remain in current flood forecasting research: (1) effectively capturing long-term temporal dependencies in flood evolution. (2) ensuring physical consistency throughout the prediction process, and (3) quantifying predictive uncertainty in a physically meaningful manner. To date, few studies have simultaneously addressed these challenges within a unified framework. Existing approaches generally focus on either temporal representation learning, physics-

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

informed modeling, or uncertainty quantification individually, while comprehensive integration of these components remains limited.

To address these gaps, this study proposes a Physics-Constrained Bayesian Transformer (PCBT) framework for probabilistic flood forecasting. The proposed framework integrates Transformer-based temporal feature extraction, physics-informed hydrological constraints, and Bayesian probabilistic inference within a unified architecture. Specifically, the Transformer encoder is employed to extract high-dimensional temporal representations from antecedent precipitation and streamflow sequences and to capture long-range hydrological dependencies. Physical constraints derived from water balance principles and flow continuity relationships are then incorporated into the loss function through a physics-informed learning strategy, ensuring that model predictions remain consistent with hydrological processes. Furthermore, a Bayesian output layer is introduced to quantify predictive uncertainty and generate probabilistic forecasts in the form of prediction intervals.

The proposed framework is evaluated using 50 independent flood events from two hydrologically contrasting watersheds: the Upper Hei River Basin, which is characterized by mountainous runoff processes, and the Luoma Lake Basin, which represents a typical lowland lake–river system. Comparative analyses are conducted against PINN, Transformer, LSTM, and BPNN models using deterministic and probabilistic evaluation metrics. The objectives of this study are threefold: (1) to develop a physically consistent probabilistic flood forecasting framework by integrating Transformer learning, hydrological constraints, and Bayesian inference; (2)

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

to investigate the contribution of physical constraints and uncertainty quantification to forecasting performance; and (3) to evaluate the robustness and transferability of the proposed framework across distinct hydrological regimes. The findings are expected to provide a new pathway toward reliable, interpretable, and uncertainty-aware flood forecasting for water resources management and flood risk mitigation.

2. Methodology

2.1. Overview

The proposed PCBT framework integrates Transformer-based temporal feature learning, physics-informed hydrological constraints, and Bayesian uncertainty quantification within a unified flood forecasting architecture. This design enables simultaneous improvements in predictive accuracy, hydrological consistency, and probabilistic forecast reliability. Initially, historical precipitation and streamflow observations are processed to construct antecedent input sequences, which capture the temporal evolution of hydrological processes. These sequences are encoded through a Transformer network to extract long-range temporal dependencies, while hydrological constraints derived from water balance principles are incorporated to reduce physically unrealistic predictions. A Bayesian output layer is subsequently applied to estimate predictive uncertainty and generate probabilistic forecasting intervals, providing a comprehensive representation of forecast reliability.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

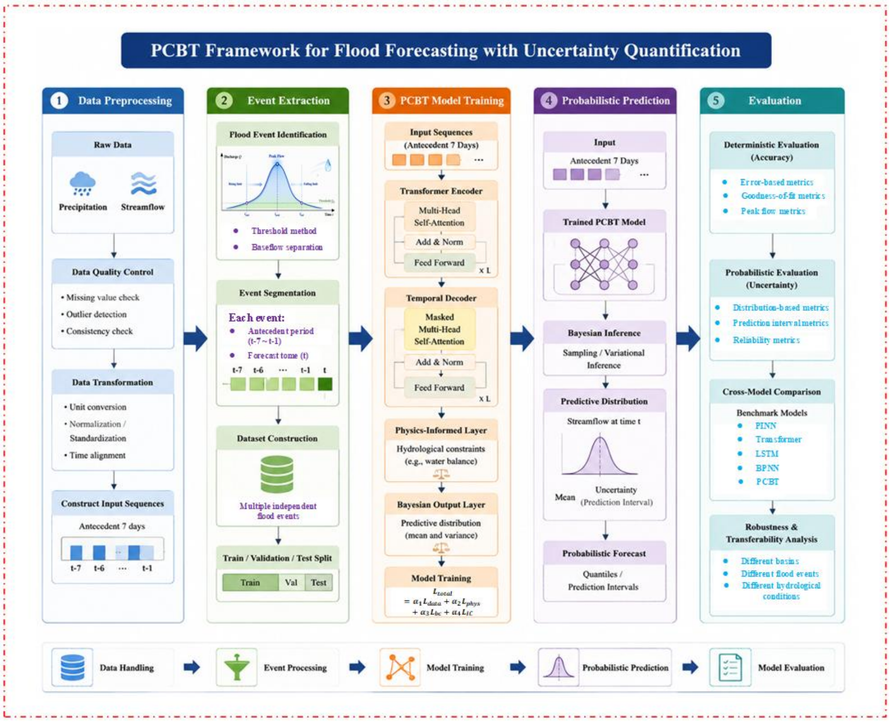


Fig. 1. Research Methodological Framework.

Fig. 1 illustrates the overall technical route of the proposed PCBT framework. Historical precipitation and streamflow records are first processed to construct the input feature set. The PCBT framework then combines transformer-based temporal representation learning with physically informed constraints to improve predictive consistency and hydrological realism. A Bayesian output layer is incorporated to estimate predictive distributions and quantify forecast uncertainty. Model performance is evaluated using both deterministic metrics: Nash–Sutcliffe Efficiency (NSE), Kling–Gupta Efficiency (KGE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Willmott Index of Agreement (WI), and Percent Bias (PBIAS), and probabilistic metrics: Continuous Ranked Probability Score (CRPS), Prediction Interval Coverage Probability (PICP), and Mean Prediction Interval Width (MPIW).

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Finally, the proposed framework is compared with several benchmark models to assess forecasting accuracy, uncertainty representation, robustness, and practical applicability for flood forecasting and water resources management.

2.2. PCBT Framework

2.2.1 Architecture of the PCBT framework

The proposed PCBT framework is developed to provide physically consistent and uncertainty-aware flood forecasting by integrating Transformer-based sequence learning, hydrological process constraints, and Bayesian probabilistic inference within a unified architecture. Unlike conventional deep learning models that rely solely on data-driven representations, PCBT explicitly incorporates hydrological knowledge into the learning process to improve physical realism and generalization capability. The framework consists of three major components: an input representation module for extracting antecedent hydrometeorological information, a physics-constrained Transformer encoder for learning nonlinear temporal dependencies while enforcing hydrological consistency, and a Bayesian output layer for probabilistic prediction and uncertainty quantification. Through the synergistic integration of these components, PCBT simultaneously enhances forecasting accuracy, interpretability, and predictive reliability under diverse hydrological conditions. The overall architecture of the proposed framework is presented in Fig. 2.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

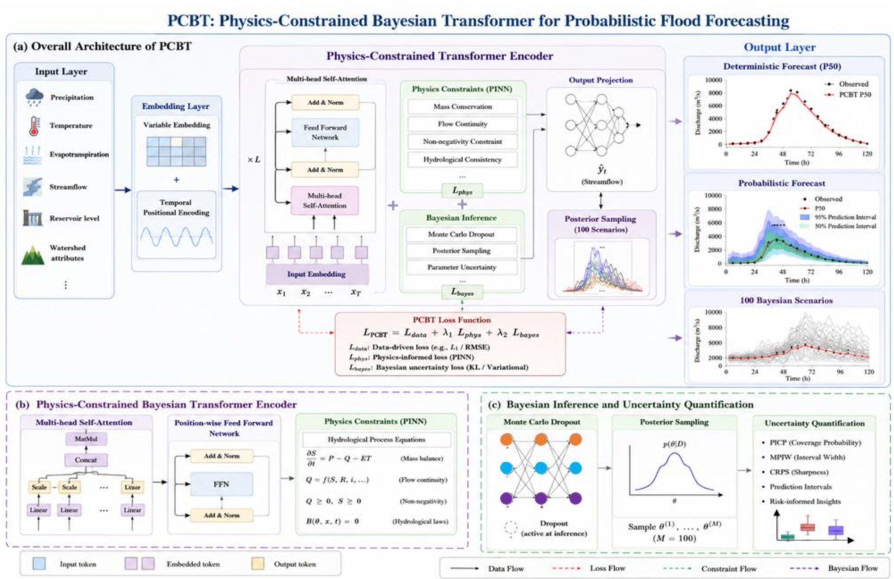


Fig. 2. Overall architecture of the proposed PCBT. (a) The complete framework of PCBT, (b) the internal structure of the physics-constrained transformer encoder, and (c) the Bayesian inference module and uncertainty quantification process.

As shown in Fig. 2, the framework follows a sequential modeling strategy from hydrometeorological input representation to deterministic prediction and probabilistic uncertainty estimation. Historical precipitation, discharge, and watershed-related information are first transformed into embedded feature representations and subsequently processed through a physics-constrained Transformer encoder. Physical constraints derived from hydrological principles are incorporated during training to ensure physically realistic predictions. Bayesian inference is then employed to estimate predictive distributions and generate uncertainty bounds. The resulting framework produces both deterministic forecasts and probabilistic prediction intervals, enabling comprehensive assessment of forecast accuracy and reliability.

(1) Input Layer

The input layer consists of key hydrological variables, including precipitation and

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

discharge. The input data are organized into time-series sequences to preserve temporal dependencies, enabling the model to capture the evolution of hydrological processes under varying rainfall–runoff conditions.

(2) Physics-Constrained Transformer Encoder

The encoder comprises multi-head self-attention and a feedforward network, modeling complex temporal dependencies by progressively stacking multiple layers. Concurrently, physical information is introduced via a PINN module, incorporating the continuity equation for flow, water balance constraints, and boundary conditions to ensure that predictions adhere to hydrological physical laws. This design enables the model to maintain physical consistency and interpretability while learning long-term dependencies.

(3) Bayesian Inference & Uncertainty Quantification

At the output layer, PCBT uses Bayesian inference to sample model parameters, generating multi-scenario (Monte Carlo Sampling) forecast results. Using these forecasts, the model calculates uncertainty metrics such as the prediction interval, CRPS, PICP, and MPIW to systematically quantify the uncertainty in flow forecasts.

(4) Output Layer

The output layer provides both deterministic and probabilistic forecasts. The deterministic forecast displays the median flow rate, while the probabilistic forecast reflects the uncertainty range through multi-scenario simulations. This design allows users to simultaneously obtain the most likely value of the flow rate and its confidence interval, providing a reliable basis for flood early warning and water resource

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

management.

(5) Performance Evaluation and Comparative Analysis

The framework includes uncertainty assessment metrics (CRPS, PICP, MPIW) and deterministic performance metrics (NSE, KGE, MAE, RMSE, PBIAS, WI). It compares these metrics against various benchmark models (LSTM, BPNN, PINN, Transformer) to systematically evaluate the model's performance across different river basins and flood events.

2.2.2. PINN Network

PINN is a class of models that integrates physical constraints into data-driven neural networks (Wong et al., 2026). During training, the system's physical equations, initial conditions, and boundary conditions are embedded into the loss function as constraints, enabling the model to maintain physical consistency while fitting observed data. Typically, a multi-layer feedforward neural network is used as a function approximator. Physical residuals are computed via automatic differentiation and combined with data errors to form the total loss function, thereby optimizing the network parameters and achieving good generalization performance with limited training data.

The PINN model accepts spatio-temporal inputs, including hydrological variables such as discharge and precipitation, as well as other environmental factors. These inputs are first preprocessed through a neural network module, which consists of fully connected layers with nonlinear activation functions (tanh, Swish, or ReLU) to extract high-dimensional feature representations (Jia et al., 2025). The physics-driven module integrates governing hydrological equations, automatic differentiation for derivatives,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

and physical residuals to enforce constraints such as continuity, mass conservation, and boundary conditions. These physically driven outputs are incorporated into the loss function, which combines multiple terms: data-driven loss (L_{data}), physical residual loss (L_{phys}), and boundary and initial condition loss (L_{bc} and L_{IC}). These terms are weighted to form the total loss (L_{total}). During training, the optimizer iteratively updates the model parameters via backpropagation. When L_{total} converges below a preset threshold, the trained model outputs the predicted variables (e.g., flow) along with associated uncertainty quantification results (prediction intervals, Bayesian posterior samples). This design enables the PINN model to learn simultaneously from both data and physical laws, ensuring prediction accuracy while maintaining consistency with hydrological principles. See Fig. S1 in Supplementary material.

2.2.3. Transformer

Transformer adopts an encoder-decoder architecture to learn nonlinear rainfall-runoff relationships from historical hydrometeorological time series (Miao et al., 2026). Through its self-attention mechanism, the Transformer can effectively capture long-term temporal dependencies and complex interactions among multiple variables, thereby better characterizing flood evolution processes and providing more accurate runoff forecasts for the target time period (Fig. S2 in Supplementary material):

(1) Data Entry and Preprocessing

Historical hydrological time series include multi-source information such as flood, precipitation, and temperature. After standardization, the data is fed into the model, and the encoder learns feature representations at different time steps. An independent fully connected layer without activation functions is used to map the input feature vector $X_t \in R^{m \times n}$ to a high-dimensional embedding space, thereby enhancing the model's representational capacity. This "embedding layer" operates on a per-time-step (e.g. one day) basis, using sine and cosine functions to encode temporal position information, as shown in the following formula:

$$E(p, 2i) = \sin \frac{p}{10000^{\frac{2i}{d_m}}} \tag{1}$$

$$E(p, 2i+1) = \cos\left(\frac{p}{10000^{\frac{2i}{d_m}}}\right) \tag{2}$$

where E represents the position code; p represents the time slot position; i represents the unique index; and dm represents the dimension vector.

(2) Encoder

The multi-head attention mechanism is used to capture the interrelationships between key time steps in a time series. This mechanism maps input features to multiple attention subspaces and computes the weights of each attention head in parallel. For each attention head, it generates query (Q), key (K), and value (V) vectors separately. Subsequently, the outputs from each attention head are concatenated and transformed via a linear transformation to obtain the final attention representation, thereby enabling the model to automatically learn and effectively capture complex spatiotemporal dependencies among multiple variables:

$$A(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where A (Q, K, V) represents the attention weight; the vectors Q, K, and V contain complete input sequences of multi-factor information such as flow, water level, and

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

precipitation, which are used to calculate the correlation weights between elements in the sequence; d_k denotes the dimension of the key vector; and T denotes the number of time periods.

(3) Decoder

The features extracted by the encoder are mapped to the predicted flood sequence to enable multi-step flood forecasting. The decoder generates daily flood forecast sequences for use in short-term early warning or water resources management.

2.2.4. Bayesian Uncertainty Quantification

The Bayesian component of the proposed PCBT framework is designed to explicitly quantify predictive uncertainty by representing model outputs as probability distributions rather than deterministic values. This probabilistic formulation enables the model to characterize uncertainty arising from parameter estimation, model structure, and the inherent stochasticity of hydrological processes.

According to Bayes' theorem, the posterior distribution of network parameters θ conditioned on the training dataset D can be expressed as

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{4}$$

where $p(\theta)$ denotes the prior distribution of model parameters, $p(\theta|D)$ is the likelihood function; and $p(D)$ is the marginal likelihood.

The posterior distribution $p(\theta|D)$ combines prior knowledge with observational information and forms the basis for probabilistic prediction.

Let $y = \{y_1, y_2, \dots, y_N\}$ denote the observed streamflow series and $\hat{y} = f_\theta(x)$ represent the model output given the input features x and network parameters θ. The

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

posterior predictive distribution is then defined as

$$p(y|x,D) = \int p(y|x,\theta)p(\theta|D)d\theta \tag{5}$$

where $p(y|x,D)$ represents the likelihood of observations conditioned on the model parameters and $p(\theta|D)$ denotes the posterior parameter distribution.

Because exact evaluation of Eq. (5) is computationally intractable for deep neural networks, Monte Carlo (MC) Dropout is adopted as an efficient variational Bayesian approximation. During inference, dropout layers remain active, allowing repeated stochastic forward passes through the network. This procedure can be interpreted as sampling from an approximate posterior distribution of model parameters (Zhu et al., 2025a).

To approximate this integral efficiently, Monte Carlo dropout is employed, sampling M sets of parameters $\{\theta^{(m)}\}_{m=1}^{M}$ during inference. The posterior predictive mean and variance are estimated as:

$$\hat{\mu}(x) = \frac{1}{M}\sum_{m=1}^{M} f_{\theta(m)}(x) \tag{6}$$

$$\hat{\sigma}^2(x) = \frac{1}{M}\sum_{m=1}^{M}(f_{\theta(m)}(x) - \hat{\mu}(x))^2 \tag{7}$$

These estimates form the basis of probabilistic prediction intervals, e.g., the 95% prediction interval:

$$PI_{95\%}(x) = [\hat{\mu}(x) - 1.96\hat{\sigma}(x), \hat{\mu}(x) + 1.96\hat{\sigma}(x)] \tag{8}$$

The Bayesian module allows explicit quantification of uncertainty arising from both model structure and parameter estimation. Generate multiple prediction samples at each time step to provide robust statistics for CRPS, PICP, and MPIW. This enables rigorous evaluation of the probabilistic performance of the PCBT framework under complex

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

hydrological conditions.

2.3. Loss Function

2.3.1. Construction of loss function

In flood forecasting, peak discharge serves as the core predictive variable, and its modeling and prediction are carried out using the PCBT framework. This framework employs automatic differentiation techniques to accurately calculate the residuals of the governing partial differential equations and combines them with observation errors to form a comprehensive loss function. This loss function is used as the optimization objective and is solved using NSGA-II. During the data-driven learning process, the model strictly adheres to physical laws, providing clear physical constraints for the prediction results, thereby ensuring that flood simulations are both accurate and physically consistent.

(1) Data-driven loss (L_{data})

Measurements were made using the mean square error between observed data (including precipitation, flow, and other monitoring data) and the PINN model prediction:

$$L_{data} = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{Q}_i - Q_i\right|^2 \tag{9}$$

where: $\hat{Q}_i$ is the simulated value of the flow (m^3/s); Q_i is the observed value of the flow (m^3/s); L_{data} represents the loss function for the flood; N is the number of flood samples.

(2) Physical constraint loss (L_{phys})

River discharge is assumed to follow a simplified kinematic-wave-based continuity

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

formulation under the assumption of gradually varied flow:

$$L_{flow} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\partial h(x_i,t)}{\partial t} + c\frac{\partial Q(x_i,t)}{\partial x}\right|^2 \tag{10}$$

where Q denotes discharge (m^3/s), x is spatial position, t is time (h), and c represents the wave propagation velocity. This constraint enforces physically consistent and smooth propagation of flood waves along the river channel.

To further ensure hydrological realism at the basin scale, discharge dynamics are constrained by a simplified rainfall–runoff balance:

$$L_{rr} = \frac{1}{N}\sum_{t=1}^{N}(Q_t - (Q_{t-1} + P_t - \alpha Q_{t-1}))^2 \tag{11}$$

where P_t denotes basin-average precipitation and $\alpha \epsilon [0,1]$ represents a runoff loss coefficient accounting for infiltration, evapotranspiration, and subsurface losses. This formulation provides a simplified but physically interpretable representation of catchment water balance.

The total physical constraint loss is formulated as:

$$L_{phys} = \lambda_1 L_{flow} + \lambda_2 L_{rr} \tag{12}$$

where λ_1 and λ_2 are weighting coefficients balancing river routing dynamics and rainfall–runoff consistency.

(3) Boundary condition loss (L_{bc})

The boundary loss is calculated by comparing the simulated boundary conditions with the actual boundary conditions. The boundary loss function is calculated as follows:

$$Q_{min} \le Q \le Q_{max} \tag{13a}$$

$$\begin{cases} L(Q_{min}) = \frac{1}{N}\sum_{i=1}^{N}\left|Q_{pred}(y_i - Q_{obs}(y_i)\right| \\ L(Q_{max}) = \frac{1}{N}\sum_{i=1}^{N}\left|Q_{pred}(y_i) - Q_{obs}(y_i)\right| \end{cases} \tag{13b}$$

$$L_{bc} = L(Q_{min}) + L(Q_{max}) \tag{13c}$$

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

where $L()$ denotes the boundary loss; $Q_{\mathrm{pred}}(y_i)$ represents the flood at the boundary (m^3/s); $Q_{\mathrm{obs}}(y_i)$ represents the observed flood at the boundary (m^3/s); Q_{min} and Q_{max} are the minimum and maximum values of the boundary flood (m^3/s), respectively.

(4) Initial condition loss (L_{IC})

The initial flow is guaranteed by the following loss:

$$L_{IC} = \frac{1}{N}\sum_{i=1}^{N}\left|Q_{pred}(x_i, t_0) - Q_{obs}(y_i, t_0)\right|^2 \tag{14}$$

where t_0 is the initial time; Q_{pred} represents the predicted flood (m^3/s); Q_{obs} represents the observed flood (m^3/s); and L_{IC} denotes the loss function for flood, respectively.

(5) The total loss function is the weighted sum of the above losses, and is calculated as follows:

$$L_{total} = \alpha_1 L_{data} + \alpha_2 L_{phys} + \alpha_3 L_{bc} + \alpha_4 L_{IC} \tag{15}$$

where α_1, α_2, α_3, and α_4 are weight coefficients representing the importance of each loss term, which can be tuned using cross-validation.

2.3.2. Loss function solution

The workflow of the NSGA-II algorithm for optimizing the total loss function of the PCBT framework is shown in Fig. S3 in Supplementary material. Firstly, the model and loss components (L_{data}, L_{phys}, L_{bc}, L_{IC}, and L_{total}) are defined, and an initial population is randomly generated and evaluated. Secondly, through iterative evolutionary operations including selection, crossover, and mutation, offspring are generated and combined with the parent population. Non-dominated sorting and crowding distance calculation are then applied, and the best individuals are retained to form the next generation (Deng et al., 2025; Chen et al., 2025b). This process repeats

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

until the maximum number of generations is reached, at which point optimal solutions are obtained, from which the final parameter set is selected according to practical performance criteria. The results balance the trade-off between predictive accuracy and physical consistency, and are used for peak discharge forecasting.

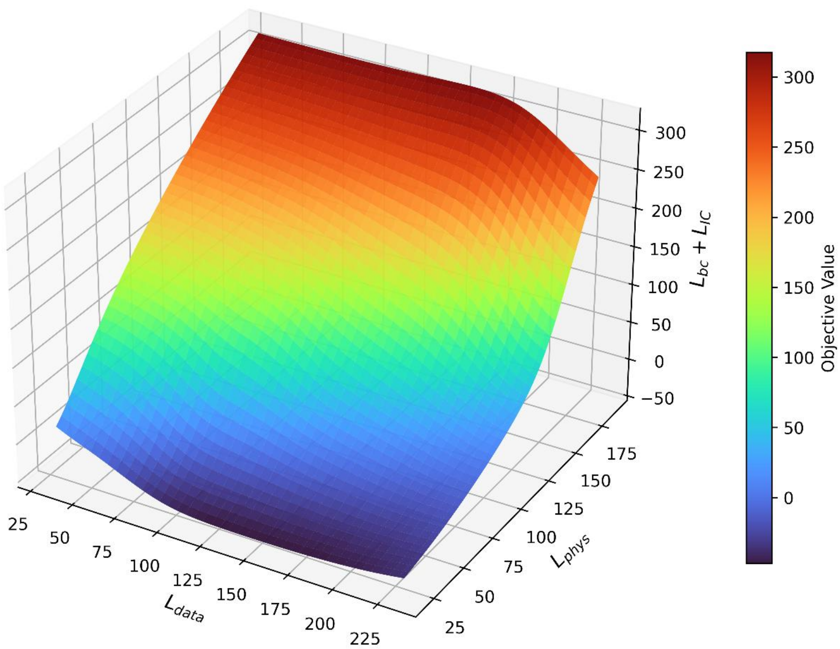


Fig. 3. Pareto-smoothed loss surface of PCBT predictions.

Fig. 3 illustrates the Pareto-smoothed objective surface of the PCBT loss function, revealing the nonlinear trade-off between L_{data} and L_{phys} during NSGA-II optimization. The surface demonstrates stable convergence and highlights the set of non-dominated solutions from which the final parameter configuration is selected, ensuring a balance between predictive accuracy and physical consistency.

2.4. Input selection and parameter determination

To ensure the physical consistency and reproducibility of the proposed PCBT, a systematic framework was developed for input feature selection and antecedent-window determination. Candidate predictors, including precipitation and streamflow variables, were first identified according to hydrological processes governing rainfall–runoff transformation. Pearson correlation analysis (PCC) and cross-correlation functions (CCF) were then employed to quantify the relationships between predictors and target streamflow, thereby determining the effective antecedent period for model inputs (Ahmad et al., 2025). In addition, autocorrelation (ACF) analysis of streamflow series was used to characterize the hydrological memory of the basin and to constrain the maximum length of the input sequence.

Based on the identified hydrological memory, multiple combinations of antecedent windows and forecasting horizons were constructed and used as inputs to the PCBT framework. The Transformer encoder was adopted to capture long-range temporal dependencies among hydrometeorological variables, while physics-informed constraints were incorporated into the learning process to enforce hydrological consistency. A Bayesian output layer was further introduced to characterize predictive uncertainty and generate probabilistic forecasts in the form of prediction intervals.

2.4.1. Variable Filtering

The selection of key forecasting variables was based on hydrological process analysis and statistical correlation analysis (SCA) (Karakoyun et al., 2025). Firstly, candidate variables were selected based on the hydrological mechanisms of the watershed, including precipitation, air temperature, evapotranspiration, and prior flood. Subsequently, PCC is used to quantitatively assess the linear correlation between each variable and flood, and CCF is employed to analyze their lag response characteristics,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

thereby identifying key input variables and their effective lag terms.

(1) PPC

$$\gamma_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{16}$$

where x represents the candidate meteorological variable; y represents flood; γ_{xy} represents the correlation coefficient. When $\gamma_{xy} > \gamma$, the variable is considered to have a significant effect.

(2) Construction of the lagged variable

Building the input based on correlation analysis:

$$x_t^{(k)} = x_{t-k} \tag{17}$$

Denotes the k-day lagged input.

2.4.2. Determination of delay

The maximum lag time reflects the length of a watershed's hydrological memory and is determined jointly using ACF and the rainfall-flood CCF. When the correlation decays to a level outside the range of statistical significance, the corresponding lag time is defined as the effective memory length of the system.

(1) ACF

$$R_{xx}(k) = \frac{\sum_{t=1}^{n-k}(x_t-\bar{x})(x_{t+k}-\bar{x})}{\sum_{t=1}^{n}(x_i-\bar{x})^2} \tag{18}$$

(2) CCF

$$R_{xy}(k) = \frac{\sum_{t=1}^{n-k}(x_t-\bar{x})(y_{t+k}-\bar{y})}{\sqrt{\sum_{t=1}^{n}(x_t-\bar{x})^2}\sqrt{\sum_{t=1}^{n}(y_t-\bar{y})^2}} \tag{19}$$

(3) Definition of Maximum Delay

$$L_{max} = \max\{k \mid R_{xx}(k) > R_c \ or \ R_{xy}(k) > R_c\} \tag{20}$$

where R_c represents the 95% confidence interval threshold. When the correlation falls

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

below the significance level, it is considered that the hydrological influence has diminished. $L_{\max}$ represents the hydrological memory length of the watershed.

2.4.3. Joint Optimization of Lag-Horizon

Based on the determined hydrological memory length, various combinations of maximum lag time and lead time are established, and a joint sensitivity analysis is conducted under physical constraints to determine the optimal parameter configuration.

(1) Parameter Set Definition

$$L\epsilon\{L_1, L_2, \dots, L_m\}, H\epsilon\{H_1, H_2, \dots, H_n\} \tag{21}$$

(2) Physical constraints

$$H \leq L \tag{22}$$

Predictions must not exceed the system's memory capacity.

(3) Joint optimization of the objective function

$$min_{L,H}J(L,H) = \omega_1 RMSE + \omega_2(1 - NSE) + \omega_3|PBIAS| + \omega_4(1 - KGE) \tag{23}$$

(4) Definition of an optimal solution

$$(L^*, H^*) = argminJ(L,H) \tag{24}$$

The optimal lag-horizon combination is selected by minimizing the multi-objective loss function under physical constraints. This framework ensures that input selection and temporal parameters are jointly constrained by hydrological theory and data-driven optimization, improving both interpretability and predictive robustness of the Trans-PINN model.

2.5. Model Training

In this study, a PCBT framework was developed for flood forecasting. Model

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

parameters were optimized by minimizing a composite loss function that simultaneously incorporates data-driven errors and physical constraints, thereby improving predictive accuracy while ensuring adherence to hydrological laws. During training, the Transformer encoder was employed to extract high-dimensional temporal features and capture long-range dependencies among hydrometeorological variables, enhancing the model's representation of rainfall–runoff dynamics. The dataset was divided into training, validation, and testing subsets for parameter learning, hyperparameter tuning, and evaluation of generalization performance, respectively. Model parameters were iteratively updated through gradient-based optimization to minimize discrepancies between predicted and observed streamflow.

2.6. Comparative Analysis

2.6.1. Comparison Model

To comprehensively evaluate the performance of the proposed PCBT, four benchmark models, including PINN, Transformer, LSTM, and BPNN, were selected for comparison. Among them, LSTM is widely used for hydrological forecasting due to its capability to capture long-term temporal dependencies (Fig. S4 in Supplementary material), whereas BPNN represents a conventional data-driven approach for learning nonlinear input–output relationships (Fig. S5 in Supplementary material). Transformer and PINN were further included to assess the contributions of attention-based temporal feature extraction and physics-informed learning, respectively.

To further investigate the contribution of individual model components, an ablation analysis was conducted by comparing the full PCBT framework with its simplified

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

variants. Specifically, the standalone Transformer model was used to evaluate the contribution of physical constraints, whereas the PINN model was adopted to assess the benefits of the Transformer-based temporal representation and Bayesian uncertainty quantification. By systematically removing or simplifying key modules, the independent and synergistic effects of the Transformer architecture, physics-informed constraints, and Bayesian inference were quantified. Through benchmark comparisons and ablation experiments, the effectiveness, robustness, and physical consistency of the proposed PCBT framework, as well as the necessity and complementarity of its constituent components, were comprehensively evaluated.

2.6.2. Performance Metrics

To evaluate the predictive performance of the proposed PCBT, six widely used deterministic metrics, including RMSE, MAE, PBIAS, NSE, KGE, and WI, were employed. Detailed definitions and discussions of these metrics can be found in Shi et al. (2025).

In addition to deterministic evaluation, probabilistic forecasting performance was assessed to quantify predictive uncertainty. Three widely adopted uncertainty metrics, namely the CRPS, PICP, and MPIW, were used to evaluate the reliability, sharpness, and overall quality of probabilistic forecasts. The mathematical definitions of these metrics are presented as follows (Zhu et al., 2025a):

(1) CRPS is used to comprehensively evaluate the discrepancy between the probability forecast distribution and observed values, and it simultaneously reflects both the accuracy of the forecast and its ability to characterize uncertainty. CRPS is

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

defined as the integral of the squared difference between the forecast cumulative distribution function and the observed cumulative distribution function:

$$CRPS(F,y) = \int_{-\infty}^{+\infty} [F(x) - H(x-y)]^2 dx \tag{25a}$$

where $F(x)$ is the cumulative distribution function (CDF) of the predictor variable; y is the observed value; $H(\cdot)$ is the Heaviside step function:

$$H(x-y) = \begin{cases} 0, x < y \\ 1, x \geq y \end{cases} \tag{25b}$$

The smaller the CRPS value, the closer the predicted distribution is to the observed value, and the better the probability prediction performance.

(2) The PICP is used to evaluate the coverage of the prediction interval for the true observed values; it is defined as the proportion of observed values that fall within the prediction interval:

$$PIPC = \frac{1}{N} \sum_{i=1}^{N} c_i \tag{26a}$$

where:

$$c_i = \begin{cases} 1, L_i \leq y_i \leq U_i \\ 0, otherwise \end{cases} \tag{26b}$$

where N represents the sample size; y_i is the i-th observation; L_i and U_i denote the lower and upper bounds of the prediction interval, respectively; c_i is the coverage indicator function.

The closer the PICP is to the preset confidence level (e.g., 95%), the higher the reliability of the prediction interval.

(3) The MPIW is used to measure the width of the prediction interval, i.e., the range of uncertainty in the prediction results, and is defined as:

$$MPIW = \frac{1}{N} \sum_{i=1}^{N} (U_i - L_i) \tag{27}$$

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

where U_i is the upper bound of the prediction interval; L_i is the lower bound of the prediction interval; N is the sample size.

The smaller the MPIW, the narrower the prediction interval and the more concentrated the prediction results. However, an MPIW that is too small may result in reduced coverage; therefore, the quality of the prediction interval should be evaluated in conjunction with the PICP.

3. Case Study

3.1. Study Area

As shown in Fig. 4, the upper reaches of the Hei River (Fig. 4a) basin are located on the northeastern edge of the Qinghai-Tibet Plateau. It is a typical mountainous inland river basin characterized by rugged terrain, significant elevation variations, uneven spatial distribution of precipitation, and the combined effects of snowmelt and rainfall. The mechanisms underlying flood formation in the basin are complex, and its hydrological processes exhibit marked nonlinearity and spatiotemporal heterogeneity (Xue et al., 2025). The Luoma Lake basin (Fig. 4b) is situated in the plain region of the Huai River basin in eastern China and is a typical plain lake basin. The terrain in this area is flat, with a dense river network and significant water storage and regulation by the lake. The flood evolution process is heavily influenced by the connectivity between rivers and lakes as well as hydraulic regulation. Floods are primarily driven by heavy rainfall, and the runoff process is more gradual compared to that in mountainous basins.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

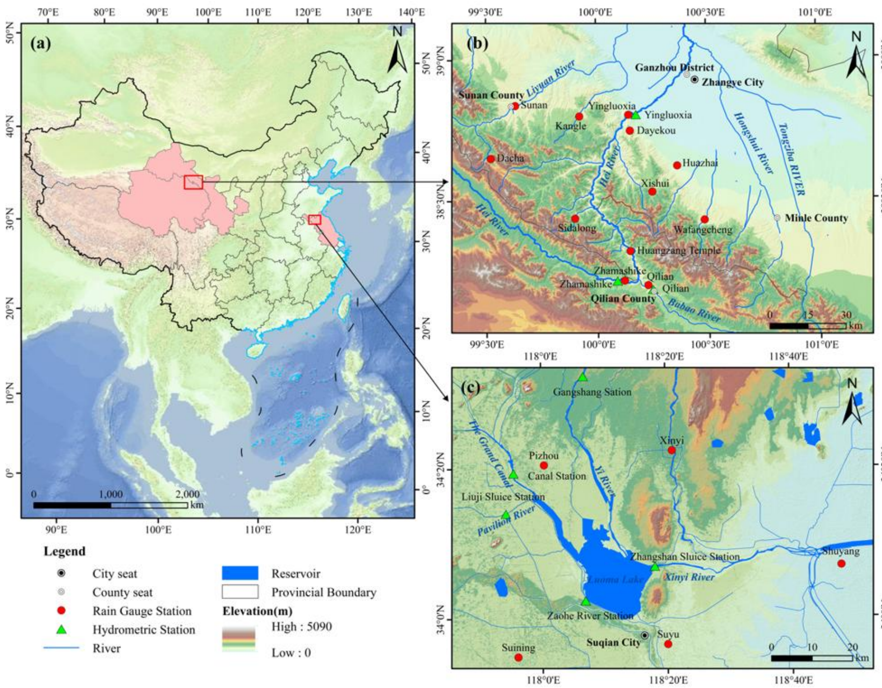


Fig. 4. Study area map (a) is China, (b) is the Upper Hei River, and (c) is the Luoma Lake.

Selecting the upper Hei River basin and the Luoma Lake basin as study areas is of great significance. These two areas represent two distinct and significantly different hydrological environments: a mountainous river basin and a plain lake basin, respectively. The former is characterized by rapid runoff, strong topographic control, and complex runoff generation and discharge mechanisms, while the latter exhibits significant river-lake regulation effects and slow flood propagation processes. By conducting experiments in these two representative basins, we can systematically evaluate the applicability and stability of the PCBT framework under varying topographic conditions, climatic backgrounds, and flood formation mechanisms, thereby verifying its cross-regional generalizability. This design not only aids in assessing the model’s predictive accuracy but also validates the robustness of the

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

physical constraint mechanism and the Bayesian uncertainty quantification method across different hydrological environments, providing a basis for the model's broader application in other basins.

3.2. Data Sources

Hydrometeorological data were collected from two watersheds with distinctly different hydrological characteristics (namely, the upper Hei River and the Luoma Lake) to evaluate the performance and generalizability of the proposed PCBT framework. For the upper Hei River, daily discharge observations from three representative hydrological stations (including Yingluoxia, Qilian, and Zhamashike) were obtained from the Hei River Administration (http://www.hhglj.org/). Corresponding precipitation records were obtained from 12 meteorological stations distributed throughout the basin. The observation period spanned from January 1, 1957, to December 31, 2023 (https://weather.cma.cn/), providing a long-term, continuous record for flood analysis and model development.

For the Luoma Lake, observational data from five key stations were obtained from the Huai River Basin Hydrological Yearbook (http://www.hrc.gov.cn/), including three inflow stations (Gang-Shang, Canal, and Liu-Ji Sluice) and two outflow control stations (Zao-He and Zhang-Shan Sluice). Meteorological data (including precipitation and temperature) were obtained from five national meteorological stations within and around the basin (https://weather.cma.cn/). To characterize the basin's topography, the 30-meter resolution ASTER Global Digital Elevation Model (ASTER GDEM) was obtained from the NASA (https://www.nasa.gov/) and the Geospatial Data Cloud

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

platform. This elevation model data was used to extract topographic information and support hydrological analysis (https://www.gscloud.cn/).

To ensure a rigorous evaluation of model performance, individual flood events were first extracted from the continuous streamflow record using a threshold-based event separation approach. A flood event was identified when streamflow exceeded a predefined threshold Q_{th}, and consecutive exceedances were grouped into a single event. Following this procedure, a total of 50 flood events were identified from the historical streamflow records. Flood events were extracted using a peak-over-threshold approach, in which the thresholds were defined based on the 90-th percentile of daily discharge for each basin. This resulted in threshold values of 120 m^3/s for the Upper Hei River Basin and 1000 m^3/s for the Luoma Lake Basin. To ensure event independence, a minimum inter-event time of 3~5 days was applied between successive flood events.

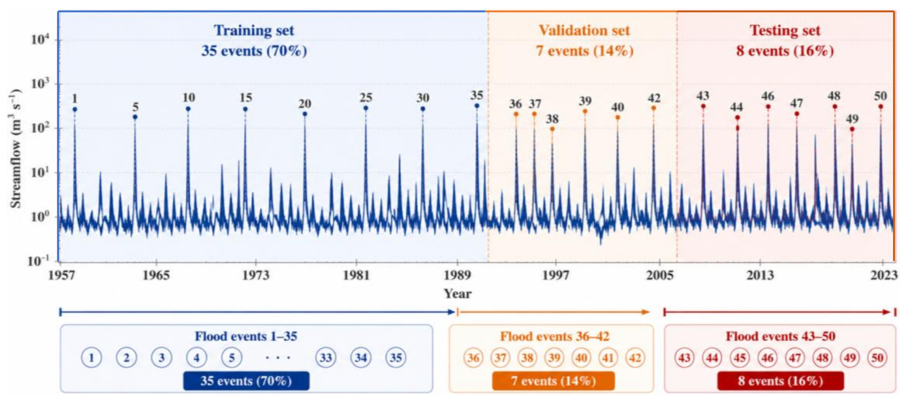


Fig. 5. Dataset partitioning based on flood events. (***Note***: Flood events are identified based on the separation criterion ($Q \geq Q_{th}$) and arranged in chronological order)

Fig. 5 illustrates the temporal distribution of the extracted flood events and the corresponding dataset partitioning strategy. To preserve the chronological characteristics of hydrological processes and avoid information leakage, the events

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

were divided sequentially into training, validation, and testing subsets. Specifically, the first 35 flood events (70%) were used for model training, the subsequent 7 events (14%) were used for hyperparameter tuning and model selection, and the remaining 8 events (16%) were reserved for independent testing. The chronological partitioning strategy enables the model to be evaluated under realistic forecasting conditions, where future flood events are predicted using information derived solely from historical observations. Furthermore, the selected events span a wide range of hydrological conditions and flood magnitudes, providing a robust basis for assessing the predictive skill and generalization capability of the proposed PCBT framework.

It was preprocessed using a normalization method. The normalization formula is as follows:

$$X_s = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{28}$$

Where x and X_s represent the flow values before and after normalization, respectively; $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of the flow data, respectively.

4. Results

The proposed PCBT framework was implemented for 24-h-ahead flood forecasting under contrasting hydrological conditions. The model was driven by antecedent hydrometeorological inputs, including precipitation, streamflow, and derived watershed response features constructed using a sliding time window to capture short- and long-term temporal dependencies. These inputs were embedded into the Transformer-based encoder to extract sequential hydrological dynamics. Specifically, Section 4.1 examines the extraction of flood processes and the sensitivity of input variables. Section 4.2

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

evaluates the predictive performance of PCBT in the Hei River and Luoma Lake basins, and Section 4.3 investigates model comparison across five baseline models using six evaluation metrics, providing a comprehensive assessment of accuracy, stability, and uncertainty quantification.

4.1. Extraction of flood processes and the sensitivity of input variables

4.1.1. Extraction of flood processes

Fig. 6 presents the observed hydrographs of the 50 flood events in Hei river and Luoma lake used in this study. To improve visualization, the events are grouped into five subsets, each containing ten representative floods. Considerable variability can be observed in both flood magnitude and hydrograph shape across the events. Peak discharges range from moderate to extreme flood conditions, while the timing, rising limb, recession characteristics, and flood duration exhibit substantial differences.

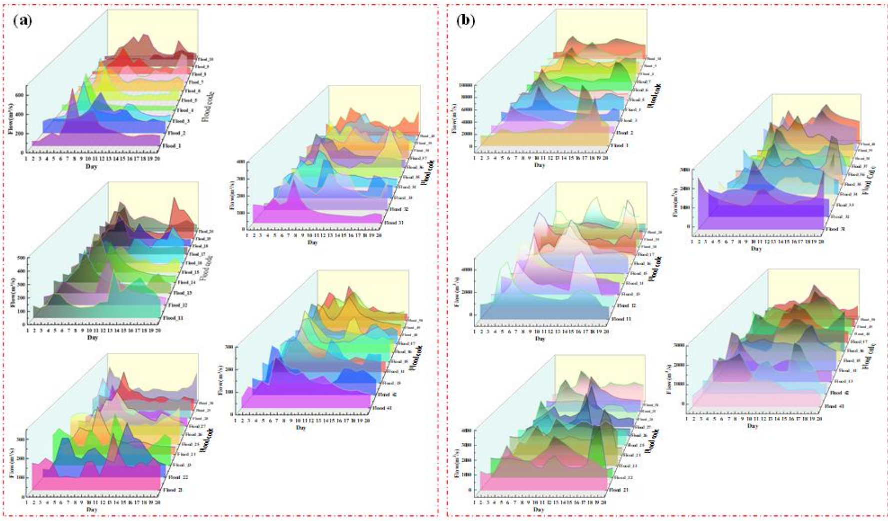


Fig. 6. Composite visualization of 50 observed flood hydrographs in the upper Hei River (a) and Luoma Lake (b).

Several events are characterized by rapid increases in discharge followed by sharp

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

recessions, indicating short-duration and high-intensity runoff responses. In contrast, other events display broader hydrographs with prolonged recession periods, reflecting stronger catchment storage effects and delayed hydrological responses. Multiple hydrographs also exhibit secondary peaks, suggesting the influence of successive precipitation events or complex runoff generation processes.

The large diversity in flood magnitude and temporal dynamics demonstrates the highly nonlinear nature of the rainfall-runoff relationship within the study basin. Such heterogeneity provides a rigorous benchmark for evaluating model robustness and generalization capability. By including flood events spanning a wide range of hydrological conditions, the dataset enables a comprehensive assessment of the proposed PCBT framework under both ordinary and extreme flood scenarios. The observed hydrographs reveal rapid runoff responses and pronounced peak variability, which are typical characteristics of mountain catchments where precipitation events are strongly controlled by topography and exhibit substantial spatial heterogeneity.

4.1.2. Input Sensitivity and Antecedent Information Analysis

Fig. 7 presents the sensitivity weights of antecedent hydrometeorological variables under different forecasting horizons. Antecedent streamflow consistently exhibited greater predictive importance than antecedent precipitation across all forecast horizons. The highest contribution was associated with one-day antecedent discharge (Q_{t-1}), with a sensitivity weight of 0.27 for one-day-ahead forecasting. Variable importance decreased progressively with increasing lag time, indicating that recent hydrological conditions dominate flood predictability in the basin.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

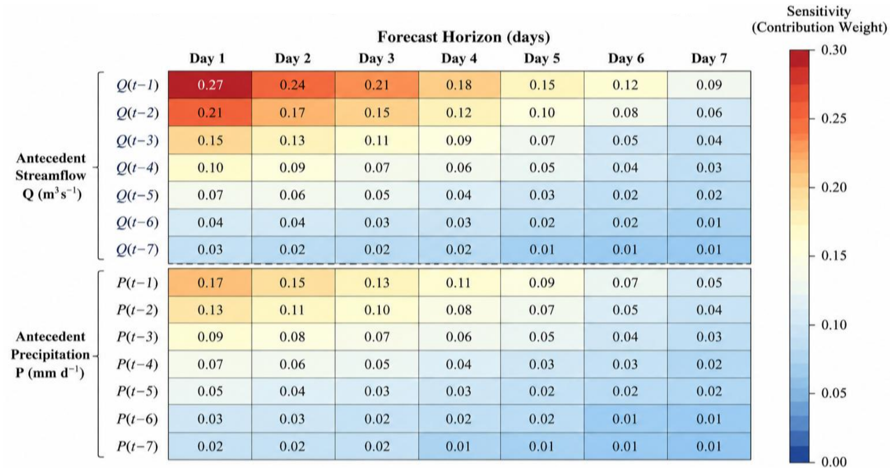


Fig. 7. Sensitivity weights of antecedent hydrometeorological variables across different forecast horizons.

The influence of antecedent precipitation was comparatively weaker but remained important for short-term forecasting. The highest precipitation contribution was observed for P_{t-1}, whereas precipitation information beyond four antecedent days provided limited additional predictive skill. Furthermore, the overall sensitivity of all predictors declined as forecast horizon increased from one to seven days, reflecting the gradual loss of hydrological memory and increasing forecasting uncertainty. These results support the selection of antecedent discharge and precipitation sequences as model inputs and confirm the dominant role of recent streamflow conditions in flood forecasting within the study basin.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Table 1

Architecture and training settings of the proposed PCBT model

Module	Configuration
Input variables	$P(t\text{-}1{:}t\text{-}7)$, $Q(t\text{-}1{:}t\text{-}7)$
Sequence Length	7
Embedding dimension	64
Number of encoder layers	2
Number of decoder layers	2
Multi-head attention	4 heads
Feed-forward network	256 neurons
Activation function	GELU
Bayesian inference	Variational Bayes
Monte Carlo samples	100
Physics-informed constraints	Water balance + continuity equation
Optimizer	Adam
Learning rate	1×10^{-3}
Batch size	32
Epochs	200
Early stopping	20 epochs

Table 1 summarizes the architecture and training configuration of the proposed PCBT framework. Antecedent precipitation and discharge observations from the previous seven time steps, $P(t-1{:}t-7)$ and $Q(t-1{:}t-7)$, are used as model inputs. The Transformer network adopts an embedding dimension of 64 with two encoder layers, two decoder layers, and four attention heads. A feed-forward network containing 256 neurons and the GELU activation function is employed for nonlinear feature extraction. Probabilistic prediction is achieved through variational Bayesian inference with 100 Monte Carlo realizations. Physical constraints are incorporated through the water

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

balance and continuity equations. Model optimization is conducted using the Adam optimizer with a learning rate of 1×10^{-3} and a batch size of 32. Training proceeds for a maximum of 200 epochs, while early stopping with a patience of 20 epochs is used to improve generalization performance and avoid overfitting.

4.2. PCBT forecasting performance

A clear hydrological contrast is observed between the Upper Hei River and Luoma Lake basins. The Upper Hei River, as a mountainous semi-arid catchment, is characterized by steep terrain, rapid runoff generation, and short response time, leading to sharp hydrograph peaks and limited storage effects. In contrast, Luoma Lake represents a lowland lake–river system with strong attenuation and routing effects, resulting in smoother hydrographs and delayed peak responses. These differences lead to distinct error structures across basins, where peak-related errors dominate in the Upper Hei River, while volume and recession biases are more pronounced in Luoma Lake. The contrasting rising and recession limb dynamics further reflect differences in runoff generation mechanisms and catchment storage regulation. The results demonstrate that basin hydrological characteristics play a key role in shaping model performance and uncertainty behavior.

4.2.1. The Upper Hei River basin

Fig. 8 presents the composite rainfall–runoff characteristics of 50 flood events in the Upper Hei River Basin with time aligned at the observed flood peak (0 h). The results show a pronounced pre-event rainfall accumulation, driving a rapid rise in discharge followed by a relatively smooth recession. Across all events, the observed median peak

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

discharge reaches 925 m^3/s, while the PCBT model reproduces a slightly lower median peak of 842 m^3/s, indicating a modest underestimation of extreme flood magnitude. Importantly, the model successfully captures the overall hydrograph shape and variability, with the observed 10~90% discharge range (312~1892 m^3/s) largely enclosed by the predicted interval (365~1642 m^3/s), demonstrating good uncertainty representation. In addition, key event-scale statistics such as runoff volume ($21{,}460\times10^3$ m^3 observed vs. $20{,}185\times10^3$ m^3 simulated) are well preserved, confirming that PCBT maintains physical consistency while achieving reliable probabilistic flood forecasting.

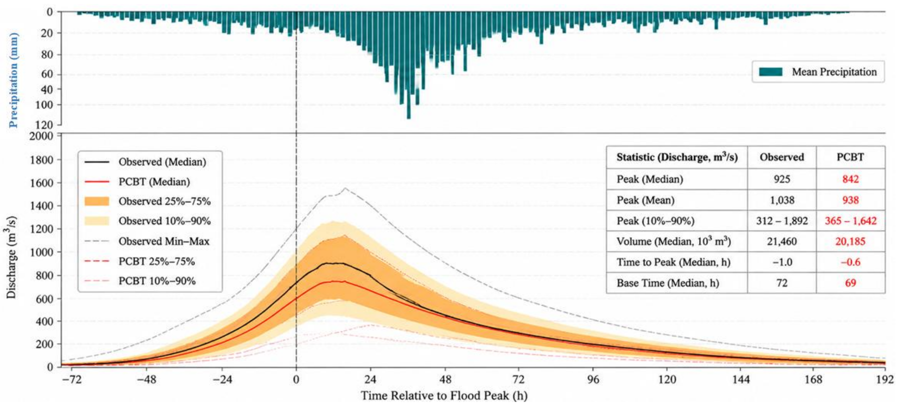


Fig. 8. Composite rainfall runoff characteristics and corresponding PCBT predictions of 50 flood events in Hei River. (***Note***: Time is aligned, so 0 hours is the peak observed time for each event)

Fig. 9 presents the probabilistic flood forecasting results of the PCBT framework for 50 independent flood events in the Hei River. The observed hydrographs are shown as blue dots, while the PCBT median predictions (P_{50}) are indicated by red lines. The 50% prediction interval (P_{25}~P_{75}) and 95% prediction interval ($P_{2.5}$~$P_{97.5}$) are shaded in orange and light blue, respectively. The figure illustrates that the PCBT framework can accurately capture both the central tendency and uncertainty of flood peaks across a

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

range of events, demonstrating its reliability and robustness for hydrological forecasting in the Hei River Basin.

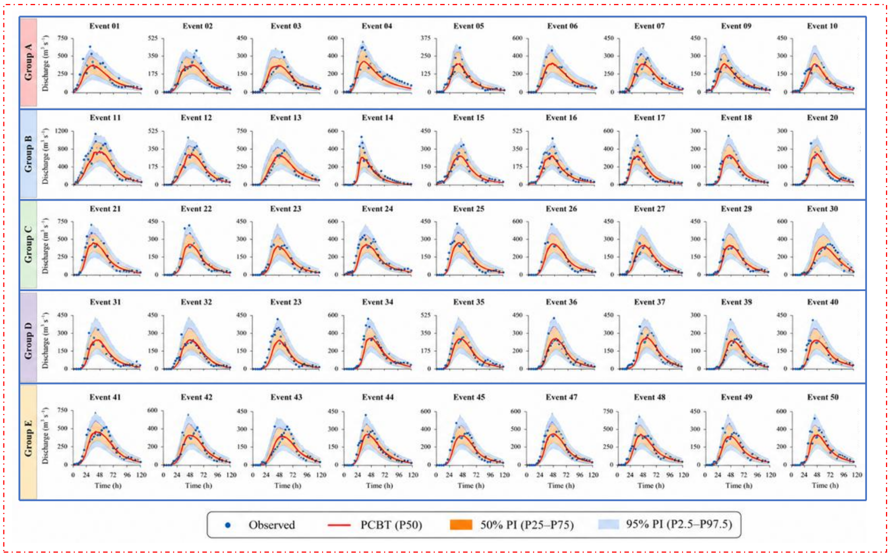


Fig. 9. Probabilistic flood forecasting results of the PCBT framework for 50 independent flood events.

Fig. 10 summarizes the probabilistic forecasting performance of the PCBT framework for 50 flood events in the Hei River Basin. The predictive uncertainty exhibits clear temporal variability among different flood events (Figure 10a). In general, uncertainty is relatively low during the recession stage of flood hydrographs but increases substantially around the rising limb and peak discharge periods, indicating that the model appropriately reflects the higher uncertainty associated with rapid hydrological responses and extreme flow conditions.

The boxplots in Fig. 10b further quantify the uncertainty characteristics under the 50% and 95% prediction intervals. The median CRPS values are approximately 3.62 m^3/s and 9.81 m^3/s for the 50% and 95% Prediction intervals, respectively, suggesting that the predictive distributions remain close to the observed flows despite the increased

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

uncertainty range. The median PICP increases from 75.3% under the 50% PI to 94.0% under the 95% PI, indicating that the observed streamflow is successfully captured within the prediction intervals for most flood events. Notably, the 95% PI coverage approaches the theoretical confidence level, demonstrating good calibration of the Bayesian uncertainty estimates. Meanwhile, the median MPIW increases from 17.6 m^3/s to 66.9 m^3/s, reflecting the expected trade-off between interval coverage and interval sharpness.

Fig. 10c presents the event-wise distributions of forecasting metrics across the 50 flood events. The median MAE and RMSE values are approximately 7~8 m^3/s and 8~10 m^3/s, respectively, indicating relatively small forecasting errors compared with the magnitude of observed flood peaks. The PBIAS values are centered around zero, suggesting the absence of systematic overestimation or underestimation. Furthermore, the NSE and WI values are concentrated near 0.95 and 1.00, respectively, confirming the strong agreement between simulated and observed hydrographs. Although a few extreme events exhibit larger forecasting errors, the overall metric distributions remain highly concentrated, demonstrating the robustness and generalization capability of the PCBT framework across diverse hydrological conditions.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

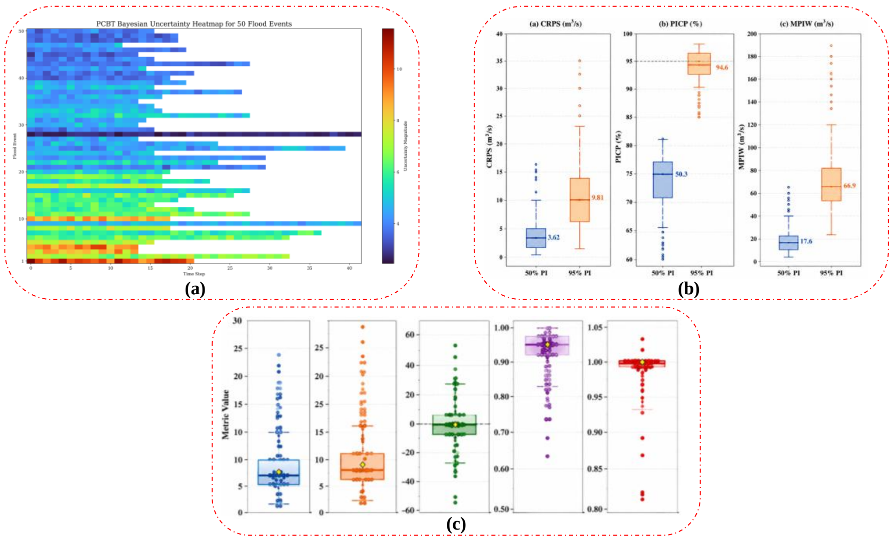


Fig. 10. Probabilistic performance of the PCBT model for 50 flood events in the Hei River. (a) Heatmap of predictive uncertainty across flood events and forecast lead times. (b) Boxplot comparison of CRPS, PICP and MPIW under 50% and 95% prediction intervals. (c) Distribution of event-wise uncertainty statistics across the 50 flood events.

4.2.2. Luoma lake basin

Fig. 11 presents the composite rainfall–runoff characteristics of 50 flood events in the Luoma Lake watershed along with corresponding PCBT framework predictions, serving as an independent evaluation dataset. The time series are aligned such that 0 h corresponds to the observed flood peak for each event. Observed median discharge is indicated by the solid black line, with the 25~75% and 10~90% quantile ranges shaded in dark and light blue, respectively. PCBT predictions are shown in red, with 50% and 90% prediction intervals depicted as dark and light red shaded areas. Key hydrograph statistics, including peak discharge (median: 1372 m^3/s; mean: 1512 m^3/s), volume (median: 86,450×10^3 m^3), time to peak (median: 0 h), and base time (median: 96 h), are summarized in the inset table, alongside PCBT predictions (peak median: 1344 m^3/s; peak mean: 1465 m^3/s; volume median: 84,790 ×10^3 m^3; time to peak: 0.5 h; base time:

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

92 h).

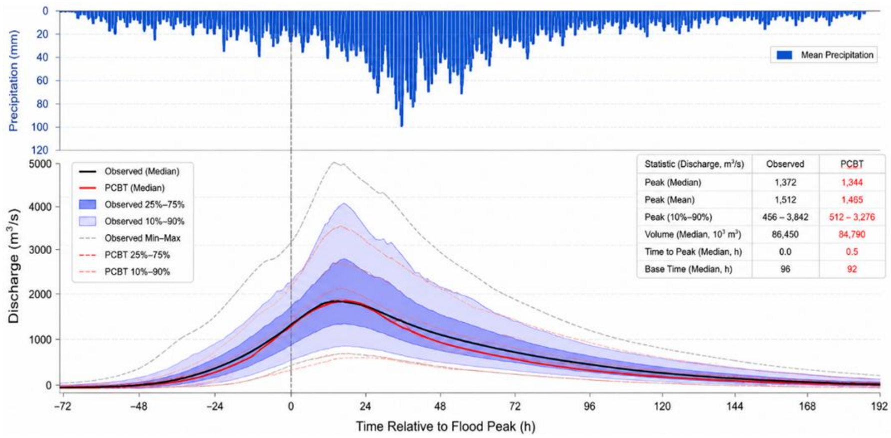


Fig. 11. Composite rainfall runoff characteristics and corresponding PCBT predictions of 50 flood events in Luoma Lake.

Fig. 11 highlights the model's ability to reproduce both central tendencies and event-wise variability of flood hydrographs in a lowland lake environment, providing a benchmark comparison for the Hei River upstream case. The inclusion of Luoma Lake, a representative plain-lake watershed, complements the mountainous Hei River basin, enhancing the generalizability and robustness assessment of the PCBT framework across contrasting hydrological regimes.

Fig. 12 presents the probabilistic streamflow forecasting results of the PCBT framework for 50 flood events in the Luoma Lake Basin. As shown in Fig. 12a, the flood events exhibit substantial variability in magnitude, with peak discharges ranging from approximately 20 to 400 m^3/s. The training (35 events), validation (7 events), and testing (8 events) datasets cover both low- and high-flow conditions, providing a comprehensive basis for model evaluation. Across all events, the median predictions closely follow the observed stream-flows, indicating that the PCBT framework

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

successfully captures the temporal evolution and peak characteristics of flood hydrographs.

The prediction intervals further demonstrate the reliability of the probabilistic forecasts. For most flood events, the observed peak discharges are contained within the corresponding 90% Prediction intervals, suggesting that the Bayesian component effectively characterizes predictive uncertainty. The prediction intervals remain relatively narrow for moderate floods but become wider for larger events, reflecting the increased uncertainty associated with extreme hydrological conditions. Notably, the model maintains satisfactory predictive performance in the testing period, indicating strong generalization capability and robustness under unseen flood events.

Fig. 12b illustrates the spatial–temporal distribution of predictive uncertainty across all flood events and forecast horizons. The uncertainty exhibits clear event-to-event variability, with larger uncertainty generally corresponding to periods of high streamflow. Several major flood events occurring between flood IDs 15~35 show broader prediction intervals, indicating stronger nonlinear rainfall–runoff responses and more complex hydrodynamic processes. In contrast, low-magnitude events are associated with relatively small uncertainty ranges, suggesting that the model is more confident when predicting stable flow conditions.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

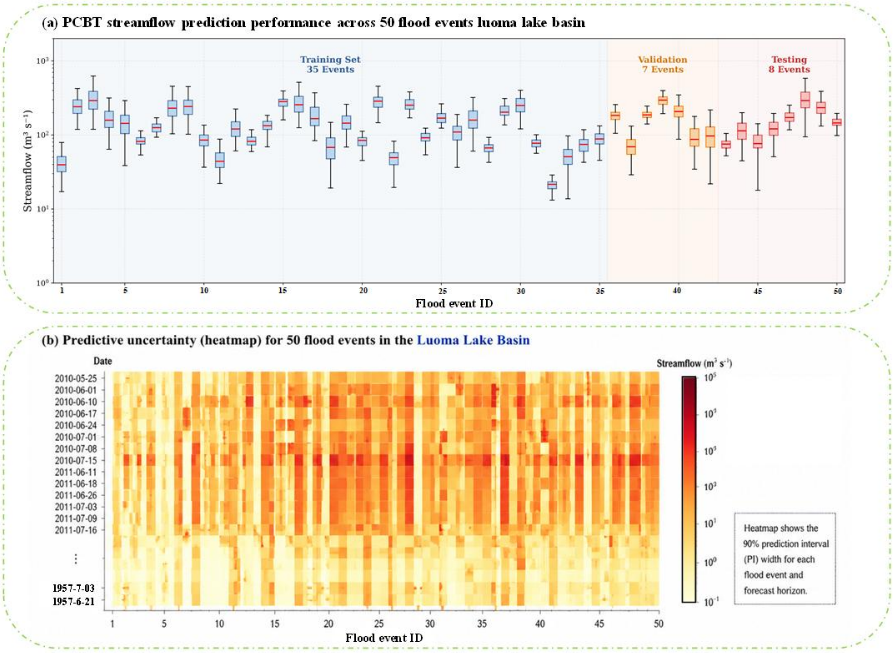


Fig. 12. Probabilistic streamflow prediction and uncertainty analysis of 50 flood events in the Luoma Lake Basin using PCBT.

Fig. 13 presents the deterministic performance of the PCBT framework for 50 flood events in the Luoma Lake, with the dataset partitioned into training (events 1~35, 70%), validation (events 36~42, 14%), and testing (events 43~50, 16%) sets. Six evaluation metrics are reported for each flood event: NSE, KGE, WI, RMSE, MAE, and PBIAS. The median NSE and KGE values for the testing set are 0.86 and 0.84, respectively, indicating strong predictive skill. The WI remains high at 0.91, and RMSE and MAE medians are 31.2 m^3/s and 23.9 m^3/s, respectively. PBIAS shows a slight underestimation tendency with a median of −1.9%. These results demonstrate that PCBT reliably captures the hydrograph dynamics of multiple flood events, with minimal bias and high accuracy across the training, validation, and testing periods.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

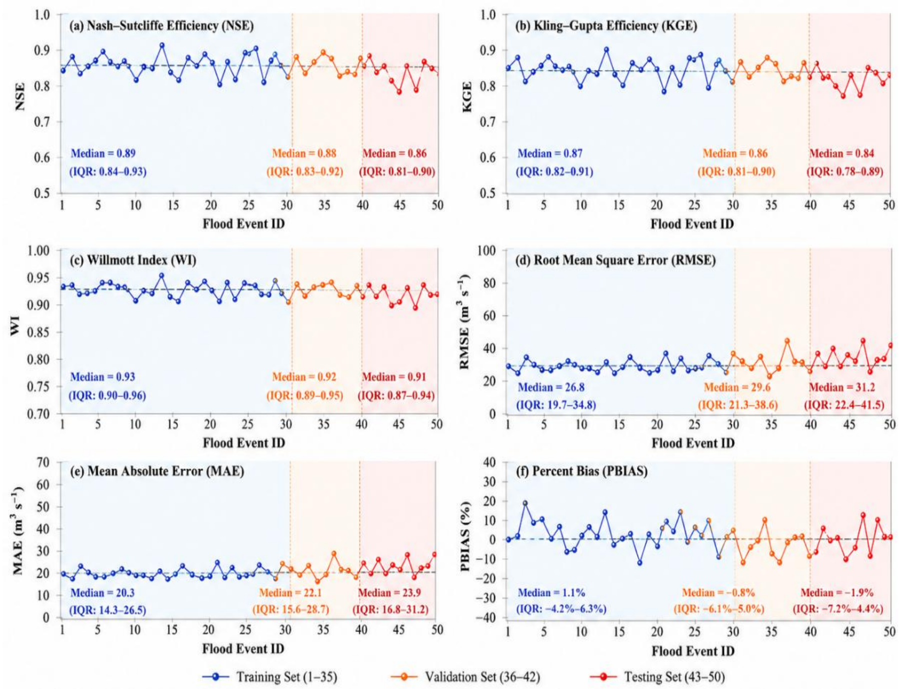


Fig. 13. Performance of PCBT for 50 Flood Events in the Luoma Lake.

Fig. 14 illustrates the distribution of residuals (observed minus predicted streamflow) for the 50 flood events in the Luoma Lake. The residuals are approximately symmetrically distributed around zero, with most falling within ±200 m^3/s, indicating that the PCBT framework provides unbiased predictions across the dataset. Extreme residuals are rare, confirming the model's robustness in capturing both moderate and extreme flood events. Collectively, Figs. 13 and 14 highlight the PCBT framework's capability to accurately reproduce observed flood dynamics and maintain consistent predictive performance across a large set of independent events.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

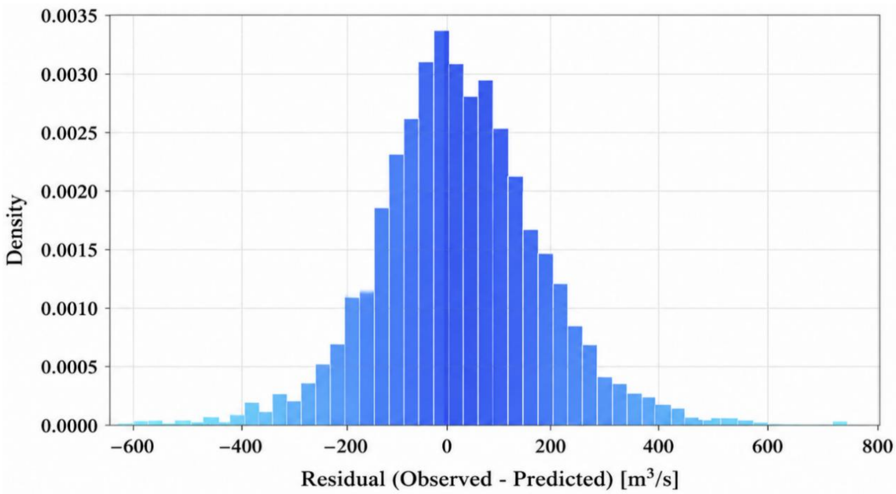


Fig. 14. Error Distribution of PCBT Predictions Luoma Lake Basin.

4.2.3. Overall Performance of PCBT Across Different Datasets

Table 2 summarizes the predictive performance of the PCBT framework across the training, validation, and testing datasets in the Upper Hei River and the Luoma Lake Basin. For the Upper Hei River Basin, the median NSE values were 0.91, 0.85, and 0.86 for the training, validation, and testing datasets, respectively. The corresponding median WI values were 0.91, 0.88, and 0.84. Median RMSE increased from 32.6 m^3/s in the training dataset to 50.5 m^3/s in the testing dataset, while median MAE increased from 22.7 to 33.6 m^3/s. Median PBIAS remained within −7.2%, indicating a slight tendency toward underestimation but no systematic prediction bias. The interquartile ranges of NSE (0.84~0.93, 0.82~0.92, and 0.80~0.92) suggest relatively stable performance across the 50 flood events.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Table 2

Event-wise deterministic performance across training, validation, and testing periods

Watershed	Dataset	Median NSE	25%-75% IQR	10%-90% IQR	Median WI	Median RMSE (m^3/s)	Median MAE (m^3/s)	Median PABIAS
Upper Hei River	Training set	0.91	0.84-0.93	0.82-0.95	0.91	32.6	22.7	−3.6%
	Validation set	0.85	0.82-0.92	0.80-0.94	0.88	41.8	28.9	−5.8%
	Testing set	0.86	0.80-0.92	0.78-0.95	0.84	50.5	33.6	−7.2%
Luoma Lake	Training set	0.89	0.81-0.91	0.79-0.93	0.93	26.8	20.3	1.1%
	Validation set	0.88	0.79-0.89	0.77-0.91	0.92	29.6	22.1	−0.8%
	Testing set	0.86	0.77-0.87	0.75-0.90	0.91	31.2	23.9	−1.9%

For the Luoma Lake Basin, median NSE values are 0.89, 0.88, and 0.86 for the training, validation, and testing datasets, respectively. Median WI remained consistently high at 0.93, 0.92, and 0.91. Median RMSE values were 26.8, 29.6, and 31.2 m^3/s, while median MAE values were 20.3, 22.1, and 23.9 m^3/s. The corresponding median PBIAS values were 1.1%, −0.8%, and −1.9%, indicating that the model reproduced flood volumes with negligible systematic bias. Compared with the Upper Hei River Basin, forecasting errors in the Luoma Lake Basin were generally lower and exhibited smaller variations among different datasets.

Overall, the PCBT framework maintained NSE values greater than 0.85 and WI values greater than 0.84 across all datasets and both basins. The relatively small differences between training, validation, and testing performance indicate good generalization capability and limited overfitting. The consistent performance in both the mountainous Upper Hei River Basin and the low-relief Luoma Lake Basin further

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

demonstrates the robustness and transferability of the proposed framework under contrasting hydrological conditions.

Table 3 presents the uncertainty quantification of PCBT predictions for 50 flood events in the Upper Hei River and Luoma Lake basins, evaluated using three key probabilistic metrics: CRPS, PICP, and MPIW. For the Upper Hei River, the CRPS is 0.093 ± 0.011 m^3/s, PICP is 91.6 ± 1.1%, and MPIW is 656 ± 45 m^3/s, indicating high predictive reliability and well-calibrated uncertainty bounds. In the Luoma Lake basin, the CRPS is slightly higher at 0.118 ± 0.009 m^3/s, PICP is 90.8 ± 1.0%, and MPIW is narrower at 360 ± 40 m^3/s, reflecting more compact prediction intervals due to lower hydrological variability. These results demonstrate that the PCBT framework produces robust probabilistic forecasts, with calibrated uncertainty ranges that capture observed flood variability across both basins.

Table 3
Comparison of uncertainty quantification metrics for PCBT in two basins

Basin	CRPS	PICP(%)	MPIW(m^3/s)
Upper Hei River	0.093±0.011	91.6±1.1	656±45
Luoma Lake	0.118±0.009	90.8±1.0	360±40

4.3. Performance Comparison of Five Models Across Different Basins

All benchmark models, including PINN, LSTM, Transformer, and BPNN, are evaluated under a consistent experimental setting to ensure a fair comparison. Specifically, all models are trained and tested using identical input variables (precipitation and discharge), the same flood event segmentation strategy, and the same training–validation–testing data splits. In addition, hyperparameters of all benchmark models are tuned within comparable search budgets to avoid advantages arising from

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

uneven optimization effort. Under this controlled setup, the proposed PCBT framework is systematically compared against the baseline models using both deterministic and probabilistic evaluation metrics.

Fig. 15 presents a comprehensive comparison of the predictive performance of five models in the upper Hei River. The radar chart summarizes six evaluation metrics, including NSE, KGE, WI, RMSE, MAE, and PBIAS, while the Taylor diagram provides an integrated assessment of correlation coefficient, standard deviation, and centered RMSE. The PCBT framework exhibits the best overall performance among all competing methods. In the Taylor diagram, PCBT achieves the highest correlation coefficient (R = 0.98), the lowest RMSE (13.24 m^3/s), and a standard deviation (SD = 35.34 m^3/s) that closely matches the observed variability. By comparison, PINN yields R = 0.93 and RMSE = 14.58 m^3/s, while LSTM, Transformer, and BPNN show lower correlations of 0.88, 0.79, and 0.74, respectively, together with substantially larger RMSE values. The radar chart further demonstrates that PCBT attains the highest NSE, KGE, and WI values and the lowest error-related metrics, indicating its superior capability in reproducing both flood magnitude and temporal dynamics. These results suggest that the integration of physical constraints, Transformer-based temporal learning, and Bayesian uncertainty quantification substantially improves flood forecasting performance in the mountainous Hei River.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

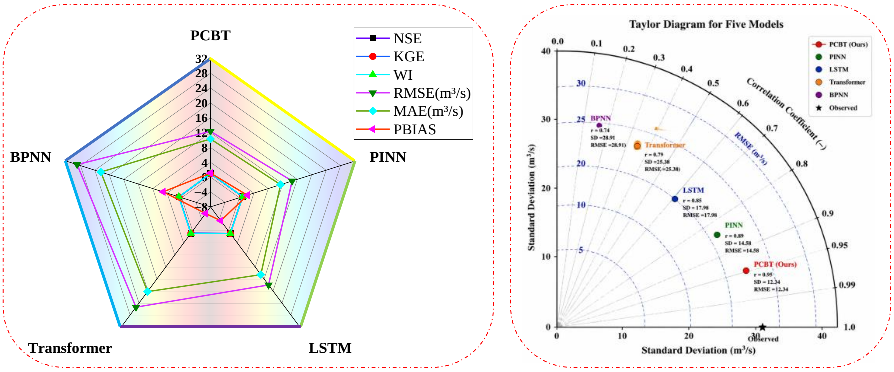


Fig. 15. Comprehensive evaluation of the predictive performance of five models using six evaluation indicators in the upper Hei River.

Fig. 16 shows the corresponding model comparison for the Luoma Lake. Similar patterns can be observed across all evaluation metrics. According to the Taylor diagram, PCBT achieves NSE = 0.95, KGE = 0.94, RMSE = 15.24 m^3/s, MAE = 14.92 m^3/s, and PBIAS = 1.1%, outperforming all benchmark models. The correlation coefficient reaches approximately 0.98, and the model standard deviation is closest to the observed value, indicating a strong ability to reproduce the variability of flood processes. In contrast, the PINN model achieves NSE = 0.89 and KGE = 0.86, whereas LSTM, Transformer, and BPNN exhibit progressively lower predictive skill. The radar chart confirms that PCBT consistently provides the highest efficiency metrics and the lowest prediction errors. The performance gains are particularly evident for RMSE and MAE, demonstrating improved simulation of flood peaks and recession limbs.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

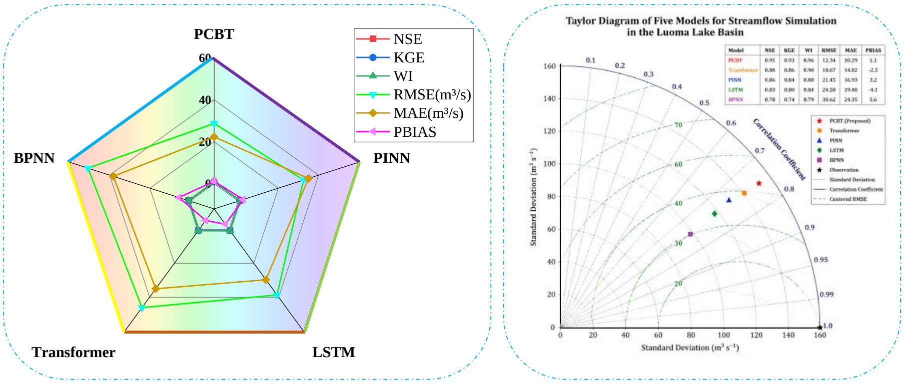


Fig. 16. Comprehensive evaluation of the predictive performance of five models using six evaluation indicators in the Luoma Lake.

A comparison between Figs. 15 and 16 reveals that PCBT maintains superior performance in two hydrologically contrasting environments. The upper Hei River Basin represents a mountainous watershed characterized by rapid runoff concentration and strong topographic controls, whereas the Luoma Lake Basin is a low-relief plain–lake system with relatively slow hydrological responses. The consistent dominance of PCBT across both basins indicates that the proposed framework is capable of capturing complex rainfall–runoff relationships under diverse hydrological conditions, highlighting its robustness, transferability, and general applicability for flood forecasting.

Table 4 summarizes the deterministic performance of PCBT and benchmark models across two contrasting basins using six evaluation metrics (NSE, KGE, WI, RMSE, MAE, and PBIAS). Overall, PCBT consistently achieves the best performance in both basins. In the Upper Hei River Basin, it attains NSE/KGE/WI of 0.95/0.93/0.96 and the lowest errors (RMSE = 12.3 m^3/s, MAE = 10.3 m^3/s), outperforming PINN, LSTM, Transformer, and BPNN. A similar ranking pattern is observed in the Luoma Lake Basin,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

where PCBT also yields the highest accuracy (NSE = 0.90, KGE = 0.88, WI = 0.91) and lowest error magnitudes. Relative to PINN, the second-best model, PCBT improves NSE by 6.7% and 5.9% in the two basins, while reducing RMSE by 15.8% and 13.6%, respectively. These results indicate that PCBT more effectively captures nonlinear rainfall–runoff dynamics and maintains robust predictive skill across heterogeneous hydrological conditions.

Table 4

Detailed quantitative performance metrics of flood forecasting models across two hydrologically contrasting basins, providing numerical basis for Figures 14–15

Watershed	Model	NSE	KGE	WI	RMSE (m^3/s)	MAE (m^3/s)	PBIAS (%)
Upper Hei River	PCBT	0.95	0.93	0.96	12.3	10.3	1.1
	PINN	0.89	0.88	0.92	14.6	11.4	2.2
	Transformer	0.79	0.74	0.82	25.4	20.2	-5.8
	LSTM	0.85	0.83	0.86	17.9	14.5	-3.4
	BPNN	0.74	0.70	0.79	28.9	22.4	5.1
Luoma Lake	PCBT	0.90	0.88	0.91	45.2	38.6	2.0
	PINN	0.85	0.82	0.87	52.3	45.0	3.5
	Transformer	0.75	0.72	0.78	72.1	63.0	-6.5
	LSTM	0.80	0.78	0.82	60.8	52.4	-4.0
	BPNN	0.71	0.68	0.75	80.5	70.2	6.0

5. Discussion

5.1. Physical consistency and predictive skill of the PCBT framework

The results obtained from the Upper Hei River and Luoma Lake basins demonstrate that the PCBT framework provides accurate and physically consistent flood forecasts across a wide range of hydrological conditions. Compared with PINN, LSTM, Transformer, and BPNN models, PCBT consistently achieved the highest NSE, KGE, and WI values while maintaining the lowest RMSE and MAE values (Table 4). In the Upper Hei River, PCBT achieved an NSE of 0.95 and an RMSE of 12.3 m^3/s, representing improvements of 6.7%, 11.8%, 20.3%, and 28.4% over PINN, LSTM,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Transformer, and BPNN, respectively. Similar advantages were observed in the Luoma Lake, where NSE reached 0.90 and RMSE decreased to 45.2 m^3/s.

The superior performance of PCBT can be attributed to the synergistic integration of Bayesian learning, Transformer-based sequence modeling, and physics-informed constraints. Traditional data-driven models such as LSTM and Transformer primarily rely on statistical relationships embedded in historical observations (Jia et al., 2021; Shen et al., 2018). Although these models effectively capture temporal dependencies, they often struggle to maintain physical consistency during extreme flood events. In contrast, the physics-constrained component embedded in PCBT explicitly incorporates hydrological conservation principles into the optimization process. Consequently, the model is capable of learning both data-driven patterns and physically meaningful runoff dynamics simultaneously.

The event-based results shown in Fig. 9 further support this interpretation. Most observed hydrographs are well enclosed within the 95% prediction intervals, while the median predictions accurately reproduce flood rising limbs, peak timing, and recession processes. This indicates that the physical constraints effectively prevent unrealistic predictions during rapidly changing flow conditions. The residual distribution shown in Figure 14 is approximately symmetric around zero, suggesting that systematic bias has been substantially reduced and that prediction errors are primarily associated with random hydrological variability rather than model structural deficiencies.

Another important finding is that predictive skill remains relatively stable across the training, validation, and testing datasets. As shown in Table 2, NSE values in the Upper

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Hei River decrease only from 0.91 to 0.86 between training and testing periods, while Luoma Lake exhibits a similar decline from 0.89 to 0.86. The limited reduction in performance indicates that the proposed framework successfully avoids severe overfitting and maintains strong generalization capability when applied to unseen flood events.

5.2. Uncertainty quantification and reliability of probabilistic forecasts

Reliable uncertainty estimation is essential for operational flood forecasting because decision makers require information regarding both expected flow conditions and associated prediction risks. The Bayesian component of PCBT enables direct quantification of predictive uncertainty through Monte Carlo sampling, generating probabilistic forecasts rather than single deterministic predictions (Willard et al., 2022).

The uncertainty evaluation metrics presented in Table 3 indicate that the probabilistic forecasts are both reliable and informative. For the Upper Hei River basin, the mean PICP reaches 91.6%, while the CRPS remains low at 0.093. Similar results are obtained in Luoma Lake, where PICP equals 90.8% and CRPS equals 0.118. These values suggest that the generated prediction intervals successfully capture most observed flood responses while maintaining relatively sharp forecast distributions.

The uncertainty heatmaps and boxplots shown in Figs. 10 and 12 provide additional insights into forecast behavior. Prediction uncertainty generally increases around flood peaks and during rapid rising stages, reflecting the increased hydrological complexity associated with intense runoff generation and routing processes. Conversely, uncertainty remains relatively low during recession periods when streamflow dynamics

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

are more stable. Such behavior is hydrologically reasonable because runoff generation mechanisms become increasingly nonlinear under high-flow conditions.

The comparison between 50% and 95% prediction intervals further demonstrates the robustness of the probabilistic framework. Although wider prediction intervals naturally produce larger MPIW values, the corresponding increase in PICP confirms improved coverage of extreme events. This trade-off between sharpness and reliability is a common characteristic of probabilistic forecasting systems and indicates that the Bayesian component effectively balances forecast confidence and uncertainty representation.

5.3. Superiority and Cross-Basin Transferability of the PCBT Framework

Recent developments in hydrological forecasting have increasingly emphasized the integration of physical laws, deep sequence learning, and probabilistic inference. PINNs improve physical consistency but often suffer from training instability and limited uncertainty representation (Raissi et al., 2019; Read et al., 2019). Transformer-based models excel at capturing long-range temporal dependencies in hydrological sequences (Kratzert et al., 2022; Muñoz-Castro et al., 2026), yet may generate physically inconsistent predictions under extreme events. Bayesian deep learning enhances predictive reliability through explicit uncertainty quantification (Kendall and Gal, 2017; Zheng et al., 2024), but typically lacks physical constraints.

In contrast, the proposed PCBT framework unifies these three paradigms within a single architecture, enabling complementary interaction between temporal

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

representation learning, physical consistency enforcement, and probabilistic inference. This integration leads to consistent improvements in both accuracy and reliability. Specifically, in the Upper Hei River basin, PCBT achieved NSE = 0.95 and RMSE = 12.3 m^3/s, outperforming PINN (0.89, 14.6 m^3/s), LSTM (0.85, 17.9 m^3/s), Transformer (0.79, 25.4 m^3/s), and BPNN (0.74, 28.9 m^3/s). Similar improvements were observed in Luoma Lake (NSE = 0.90, RMSE = 45.2 m^3/s), confirming that the coupled framework effectively reduces structural prediction errors while preserving generalization ability. Meanwhile, PICP values above 90% and relatively low CRPS further indicate that PCBT achieves a favorable balance between reliability and sharpness, which remains a key challenge in probabilistic flood forecasting.

Beyond predictive accuracy, the framework also demonstrates strong robustness across contrasting hydrological regimes. The Upper Hei River basin is characterized by mountainous terrain and rapid runoff generation, while Luoma Lake represents a lowland lake–river system with pronounced storage and routing effects. Despite these differences, PCBT consistently maintains high performance (NSE > 0.90 and > 0.86 in the two basins, respectively), indicating that the model is not overly dependent on basin-specific statistical patterns. This robustness is primarily attributed to the incorporation of hydrological constraints, which embed physically universal laws into the learning process and enhance generalization capability across watersheds (Feng et al., 2022). Comparative analyses in Figs. 15 and 16 further confirm that PCBT remains closest to the ideal reference in both radar and Taylor space, particularly under complex flood conditions with sharp peaks and rapid hydrograph variation. Overall, these results

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

demonstrate that PCBT provides a unified, physically consistent, and uncertainty-aware framework that significantly advances both predictive performance and cross-basin transferability in flood forecasting.

5.4. Limitations and future research directions

Although the proposed framework demonstrates promising performance, several limitations remain. Firstly, the current model primarily incorporates water balance and continuity constraints. Additional hydrodynamic information, such as momentum equations, channel routing processes, and lake storage dynamics, may further improve physical realism and predictive accuracy (Muñoz-Castro et al., 2026).

Secondly, the study focuses on historical flood events from two representative basins. While the results demonstrate encouraging transferability, broader evaluations across different climatic regions, watershed scales, and hydrological regimes are still required. Future studies should investigate the framework's applicability in highly regulated river systems, arid watersheds, and regions experiencing nonstationary climate conditions (Reichstein et al., 2019; Muñoz-Castro et al., 2026).

Thirdly, uncertainty estimation currently relies on Bayesian inference and Monte Carlo sampling. Although effective, the computational cost increases with the number of samples and forecasting horizons. Future work may explore more efficient probabilistic learning strategies, including ensemble physics-informed models and diffusion-based uncertainty estimation techniques.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

5. Conclusion

Flood forecasting remains challenging because of the strong nonlinearity of rainfall–runoff processes, watershed heterogeneity, and uncertainties associated with extreme flood events. To address these challenges, this study developed a Physics-Constrained Bayesian Transformer (PCBT) framework that integrates Transformer-based sequence learning, physics-informed constraints, and Bayesian probabilistic inference within a unified architecture. The framework was evaluated using 50 independent flood events from two hydrologically contrasting watersheds, namely the Upper Hei River Basin and the Luoma Lake Basin. The major conclusions are as follows.

(1) The proposed PCBT framework successfully achieved physically consistent and highly accurate flood forecasting by integrating Transformer learning with hydrological constraints. Across both study basins, PCBT consistently outperformed PINN, Transformer, LSTM, and BPNN models in deterministic forecasting performance. In the Upper Hei River Basin, PCBT achieved NSE, KGE, and WI values of 0.95, 0.93, and 0.96, respectively, while RMSE and MAE were reduced to 12.3 m^3/s and 10.3 m^3/s. In the Luoma Lake Basin, the corresponding values were 0.90, 0.88, 0.91, 45.2 m^3/s, and 38.6 m^3/s. The results indicate that combining data-driven temporal feature extraction with physical constraints effectively improves the representation of complex rainfall–runoff dynamics and enhances hydrological consistency in flood prediction.

(2) The incorporation of physics-informed constraints and Bayesian uncertainty quantification substantially improved forecasting reliability and predictive robustness. Comparative analyses demonstrated that the physics-constrained architecture reduced

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

physically unrealistic predictions and enhanced flood peak reproduction compared with purely data-driven models. Meanwhile, Bayesian inference enabled explicit characterization of predictive uncertainty, yielding CRPS values of 0.093 ± 0.011 and 0.118 ± 0.009 for the Upper Hei River and Luoma Lake, respectively. The corresponding PICP values reached 91.6% and 90.8%, indicating that the generated prediction intervals effectively captured observed streamflow variability and flood peak uncertainty. These findings confirm that the integration of physical constraints and probabilistic inference contributes substantially to both forecast accuracy and reliability.

(3) The proposed framework exhibited strong robustness and transferability across distinct hydrological regimes. Despite pronounced differences in topography, climate conditions, runoff-generation mechanisms, and storage characteristics between the mountainous Upper Hei River Basin and the lowland lake–river system of Luoma Lake, PCBT maintained stable predictive performance throughout all flood events. Median NSE values remained above 0.85 across training, validation, and testing datasets, while residual distributions were centered near zero and approximately symmetric. These results demonstrate that the framework possesses strong generalization capability and can be effectively transferred across watersheds with contrasting hydrological characteristics.

Overall, the results demonstrate that the synergistic integration of Transformer-based temporal learning, physics-informed hydrological constraints, and Bayesian uncertainty estimation provides a reliable, interpretable, and uncertainty-aware solution for flood forecasting. The proposed PCBT framework offers a promising pathway for

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

improving operational flood forecasting, water resources management, and flood risk mitigation under changing hydroclimatic conditions.

Author Contributions

Mingrui Shi: Writing – Original Draft, Writing – Review & Editing, Conceptualization, Methodology, Software, Visualization, Formal analysis, Data Curation. **Hongyuan Fang:** Conceptualization, Project administration, Investigation, Resources, Funding acquisition, Validation, Supervision.

Conflict of Interest

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

Data Availability Statement

Data will be made available on request.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (52379027), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX25_3991).

References

Adibfar, A., and Davani, H.: A transformer-based downscaling framework for improved precipitation prediction in San Diego County, Journal of Hydrology: Regional Studies, 62, 102738, https://doi.org/10.1016/j.ejrh.2025.102738, 2025.

Ahmad, K., Iqbal, M., Tariq, M. A. U. R., Khan, A. U and Nadeem, A.: Exploring alternate coupling inputs of a data-driven model for optimum daily streamflow prediction in calibrated SWAT-BiLSTM rainfall-runoff modeling, Frontiers in Water, 7, 1558218,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

https://doi.org/10.3389/frwa.2025.1558218, 2025.

Ambika, A. K., Tayal, K., Mishra, V., and Lu, D.: Novel deep learning transformer model for short to sub-seasonal streamflow forecast, Geophysical Research Letters, 52(14), e2025GL116707, https://doi.org/10.1029/2025GL116707, 2025.

Amini, A., Dolatshahi, M., and Kerachian, R.: Effects of automatic hyperparameter tuning on the performance of multi-variate deep learning-based rainfall nowcasting, Water Resources Research, 59(1), e2022WR032789, https://doi.org/10.1029/2022WR032789, 2023.

Arsenault, R., Martel, J. L., Brunet, F., Brissette, F., and Mai, J.: Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models, Hydrology and Earth System Sciences, 27(1), 139-157, https://doi.org/10.5194/hess-27-139-2023, 2023.

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G.: Twenty-three unsolved problems in hydrology (UPH)–a community perspective, Hydrological sciences journal, 64(10), 1141-1158, https://doi.org/10.1080/02626667.2019.1620507, 2019.

Chen, C., Jiang, J., Liao, Z., Zhou, Y., and Wang, H.: A short-term flood prediction based on spatial deep learning network: A case study for Xi County, China, Journal of Hydrology, 607, 127535, https://doi.org/10.1016/j.jhydrol.2022.127535, 2022.

Chen, C., Li, B., Zhang, H., Zhao, M., and Liang, Z.: Performance enhancement of deep learning model with attention mechanism and FCN model in flood forecasting, Journal of Hydrology, 658, 133221, https://doi.org/10.1016/j.jhydrol.2025.133221, 2025a.

Chen, J. X., Wang, S. S., Luo, P. P., Xu, C. Y., and Zhao, H. Y.: Integrating machine learning with NSGA-II to assess the synchronization effects of stormwater disaster hazard and green–blue

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

infrastructure, Journal of Hydrology, 134777, https://doi.org/10.1016/j.jhydrol.2025.134777, 2025b.

Deng, M., Li, Z., Wang, W., Zhang, J., Xiang, K., and Wang, Y.: A multi-objective optimization framework for check dam siting integrating GIS, a hydrological–hydrodynamic coupling model, and NSGA-II, Journal of Hydrology, 134557, https://doi.org/10.1016/j.jhydrol.2025.134557, 2025.

Fadillah, A., and Mursalim, M.: Enhancing Flood Prediction Using Hybrid LSTM-Transformer Deep Learning Approach, Journal of System and Computer Engineering, 6(3), 281-295, https://doi.org/10.61628/jsce.v6i3.2083, 2025.

Feng, D., Liu, J., Lawson, K., and Shen, C.: Differentiable, learnable, regionalized process-based models with multi-physical outputs can approach state-of-the-art hydrologic prediction accuracy, Water Resources Research, 58(10), e2022WR032404, https://doi.org/10.1029/2022WR032404, 2022.

Gao, L., Gu, X., Chen, F., and Wang, J.: Sparse Transformer Network with Spatial-Temporal Graph for Pedestrian Trajectory Prediction, IEEE Access, 12, 144725-144737, https://doi.org/10.1109/ACCESS.2024.3442442, 2024.

Gegenleithner, S., Pirker, M., Dorfmann, C., Kern, R., and Schneider, J.: Long short-term memory networks for enhancing real-time flood forecasts: a case study for an underperforming hydrologic model, Hydrology and Earth System Sciences, 29(7), 1939-1962, https://doi.org/10.5194/hess-29-1939-2025, 2025.

Gupta, H. V., Clark, M. P., and Vrugt, J. A.: Towards a comprehensive assessment of model structural adequacy, Water Resources Research, 50, 797–814,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

https://doi.org/10.1029/2011WR011044, 2024.

Han, S., and Coulibaly, P.: Bayesian flood forecasting methods: A review, Journal of Hydrology, 551, 340-351, https://doi.org/10.1016/j.jhydrol.2017.06.004, 2017.

He, Y., Yan, R., Tang, Y., Zhang, J., and Qian, H.: Enhancing runoff simulation in data-scarce mountainous regions: a coupled SWAT and transferable transformer approach, Journal of Hydrology, 135396, https://doi.org/10.1016/j.jhydrol.2026.135396, 2026.

Jia, L., Yen, N., and Pei, Y.: A physics-coupled deep learning framework for hydrodynamic diffusion modeling in watershed systems: Integrating spatiotemporal networks and environmental constraints, IEEE Access, https://doi.org/10.1109/ACCESS.2025.3542173, 2025.

Jia, X., Willard, J., and Karpatne, A.: Physics-guided machine learning for scientific discovery, Nature Reviews Earth & Environment, 2, 545–557, https://doi.org/10.1145/3447814, 2021.

Karakoyun, Y., Katipoğlu, O. M., and Dogan, A.: Deep learning and adaptive boosting for hydroelectric power prediction using hydro-meteorological data: Insights and feature importance analysis, Engineering Applications of Artificial Intelligence, 158, 111434, https://doi.org/10.1016/j.engappai.2025.111434, 2025.

Kendall, A., and Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision, Advances in neural information processing systems, 30, 2017.

Khatun, A., Chatterjee, C., Sahu, G., and Sahoo, B.: A novel smoothing-based long short-term memory framework for short-to medium-range flood forecasting, Hydrological Sciences Journal, 68(3), 488-506, https://doio.org/10.1080/02626667.2023.2173012, 2023.

Kohanpur, A. H., Saksena, S., Dey, S., Johnson, J. M., Riasi, M. S., and Yeghiazarian, L.: Urban flood modeling: Uncertainty quantification and physics-informed Gaussian processes

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

regression forecasting, Water Resources Research, 59(3), e2022WR033939, https://doi.org/10.1029/2022WR033939, 2023.

Kratzert, F., Klotz, D., and Brenner, C.: Rainfall–runoff modelling using long short-term memory networks, Hydrology and Earth System Sciences, 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., and Herrnegger, M.: Toward learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrology and Earth System Sciences, 26, 4345–4374, https://doi.org/10.5194/hess-23-5089-2019, 2022.

Küllahcı, K., Kartal, E., and Altunkaynak, A.: A Novel Spectral Decomposition Framework for TimesNet: Integrating DCT and Wavelet Transforms for Hydrological Forecasting, Water Resources Management, 40(7), 259, https://doi.org/10.1007/s11269-026-04529-y, 2026.

Kumar, A., Singh, P. K., Kumar, V., Kumar, R. P., and Patel, R.: Comparative Analysis of Machine Learning Models in Groundwater Quality Assessment: A Systematic Review, Water Environment Research, 98(5), e70401, https://doi.org/10.1002/wer.70401, 2026.

Legg, S.: Climate change 2021-the physical science basis, Interaction, 49(4), 44-45, 2021.

Liao, Y., Wang, Z., Yu, H., Gao, W., and Zeng, Z.: Accelerating urban flood inundation simulation under spatio-temporally varying rainstorms using ConvLSTM deep learning model, Water Resources Research, 61(8), e2025WR040433, https://doi.org/10.1029/2025WR040433, 2025.

Ling, M., Guan, Y., Lian, Y., Sun, X., and Gao, Y.: An improved Hydrology-Informed attention LSTM (HIA-LSTM) model for runoff simulation with seasonal snowmelt, Journal of Hydrology, 135231, https://doi.reg/10.1016/j.jhydrol.2026.135231, 2026.

Liu, Z., Zhu, D., Luo, J., and Li, D.: Urban pluvial flood maximum water depth prediction using a

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

transformer-cnn hybrid deep learning model with data augmentation, Water Resources Management, 40(4), 165, https://doi.org/10.1007/s11269-026-04524-3, 2026.

Lu, L., Yan, T., Chen, Y., Wang, H., and Yang, T.: A lake water level prediction method based on data augmentation and Physics-Informed Neural Networks with imbalanced data, Journal of Hydrology, 134660, https://doi.org/10.1016/j.jhydrol.2025.134660, 2025.

Miao, G., Li, W., Liu, C., Hu, C., and Dong, X.: Application of Transformer Hybrid Model in Runoff Process Prediction Based on Frequency Decomposition and Physical Correction of Multi-Source Data, Water Resources Management, 40(8), 370, https://doi.org/10.1007/s11269-026-04702-3, 2026.

Mosaffa, H., Pappenberger, F., Prudhomme, C., Chantry, M., and Rüdiger, C.: A GNN routing module is all you need for LSTM Rainfall–Runoff models, Hydrology and Earth System Sciences, 30(7), 2079-2092, https://doi.org/10.5194/hess-30-2079-2026, 2026.

Muñoz, D. F., Moftakhari, H., and Moradkhani, H.: Quantifying cascading uncertainty in compound flood modeling with linked process-based and machine learning models, Hydrology and Earth System Sciences, 28(11), 2531-2553, https://doi.org/10.5194/hess-28-2531-2024, 2024.

Muñoz-Castro, E., Anderson, B. J., Astagneau, P. C., Swain, D. L., and Mendoza, P. A.: How well do hydrological models simulate streamflow extremes and drought-to-flood transitions, Hydrology and Earth System Sciences, 30(3), 825-848, https://doi.org/10.5194/hess-30-825-2026, 2026.

Pan, X., Hou, J., Li, D., Wang, Y., and Li, X.: A Study on Rapid Dynamic Flood Forecasting in Small Watersheds Using a GNN-Transformer Approach Integrated with Spatial Physical Information, Water Resources Management, 40(4), 173, https://doi.org/10.1007/s11269-025-

04360-x, 2026.

Pölz, A., Blaschke, A. P., Komma, J., Farnleitner, A. H., and Derx, J.: Transformer versus LSTM: A comparison of deep learning models for karst spring discharge forecasting, Water Resources Research, 60(4), e2022WR032602, https://doi.org/10.1029/2022WR032602, 2024.

Raissi, M., Perdikaris, P., and Karniadakis, G. E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics, 378, 686–707, https://doi.org/10.1016/j.jcp.2018.10.045, 2019.

Read, J. S., Jia, X., and Willard, J.: Process-guided deep learning predictions of lake water temperature, Water Resources Research, 55(11), 9173–9190, https://doi.org/10.1029/2019WR024922, 2019.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., and Denzler, J.: Deep learning and process understanding for data-driven Earth system science, Nature, 566(7743), 195-204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Saravana, M. K., Roopa, M. S., Arunalatha, J. S., and Venugopal, K. R.: Transformers for Multivariate Time Series Forecasting: Comprehensive Analysis, Challenges, Research Opportunities and Future Prospects, IEEE Access, https://doi.org/10.1109/ACCESS.2026.3654408, 2016.

Shen, C.: A transdisciplinary review of deep learning research and its relevance for water resources scientists, Water Resources Research, 54, 8558–8593, https://doi.org/10.1029/2018WR022643, 2018.

Shi, M., Fang, H., Xie, Y., Du, H., and Liu, S.: A flood forecasting method coupling the CE-QUAL-

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

W2 and PINN models, Journal of Hydrology, 661, 133699, https://doi.org/10.1016/j.jhydrol.2025.133699, 2025.

Solanki, H., Vegad, U., Kushwaha, A., and Mishra, V.: Improving streamflow prediction using multiple hydrological models and machine learning methods, Water Resources Research, 61(1), e2024WR038192, https://doi.org/10.1029/2024WR038192, 2025.

Tian, Y., Ding, S., Huang, L., Su, G., and Chen, J.: Physics-informed neural networks for solving the two-dimensional shallow water equations with terrain topography and rainfall source terms, Water Resources Research, 61(9), e2025WR040052, https://doi.org/10.1029/2025WR040052, 2025.

Virupaksha, A., Lehmann, F., Hoteit, H., Younes, A., and Fahs, M.: Modeling transient groundwater flow in unconfined aquifers under dynamic conditions using physics-informed neural networks, Water Resources Research, 62(3), e2025WR040754, https://doi.org/10.1029/2025WR040754, 2026.

Wei, S., Wang, X.: Physics-informed gated transformer for robust flash flood forecasting in data-scarce small catchments via variance-constrained learning, Journal of Hydrology, 135706, https://doi.org/10.1016/j.jhydrol.2026.135706, 2026.

Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V.: Integrating scientific knowledge with machine learning for engineering and environmental systems, ACM Computing Surveys, 55(4), 1-37, https://doi.org/10.1145/3514228, 2022.

Wong, J. C., Gupta, A., Ooi, C. C., Chiu, P. H., and Liu, J.: Evolutionary optimization of physics-informed neural networks: Evo-pinn frontiers and opportunities, IEEE Computational Intelligence Magazine, 21(1), 16-36, https://doi.org/10.1109/MCI.2025.3607749, 2026.

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

Xia, T., Zhou, Y., Xu, C. Y., Liu, P., and Luo, Y.: A spatiotemporal pathformer-based deep learning framework for watershed flood forecasting, Water Resources Research, 61(12), e2025WR040193, https://doi.org/10.1029/2025WR040193, 2025.

Xu, B., He, X.: A physics-informed Bayesian storyline approach to assess sediment transport in the Mekong, Water Resources Research, 58(11), e2022WR032681, https://doi.org/10.1029/2022WR032681, 2022.

Xue, H., Wang, Y., Dong, G., Zhang, C., and Lian, Y.: Multi-source meteorological data assessment on daily runoff simulation in the upper reaches of the Hei River, Northwest China, Journal of Hydrology: Regional Studies, 57, 102100, https://doi.org/10.1016/j.ejrh.2024.102100, 2025.

Yang, Y., Zhang, W., Li, F., Wang, H., and Zhang, H.: Fusing dynamic physical constraints with PINN-xLSTM to enhance accuracy and physical consistency in runoff prediction under extreme hydrological events, Journal of Hydrology, 135310, https://doi.org/10.1016/j.jhydrol.2026.135310, 2026.

Zhang, T., Zhang, R., Li, J., and Feng, P.: Deep learning of flood forecasting by considering interpretability and physical constraints, Hydrology and Earth System Sciences, 29(21), 5955-5974, https://doi.org/10.5194/hess-29-5955-2025, 2025.

Zheng, F., Yin, H., Zhang, J., Duan, H. F., and Gupta, H. V.: A Bayesian deep learning approach for video-based estimation and uncertainty quantification of urban rainfall intensity, Journal of Hydrology, 640, 131706, https://doi.org/10.1016/j.jhydrol.2024.131706, 2024.

Zhu, W., Tao, T., Yan, H., Yan, J., and Wang, J.: An optimized long short-term memory (LSTM)-based approach applied to early warning and forecasting of ponding in the urban drainage system, Hydrology and Earth System Sciences, 27(10), 2035-2050,

https://doi.org/10.5194/egusphere-2026-4974
Preprint. Discussion started: 8 September 2026

https://doi.org/10.5194/hess-27-2035-2023, 2023.

Zhu, F., Sun, Y., Han, M., Hou, T., and Zeng, Y.: A robust bayesian multi-machine learning ensemble framework for probabilistic groundwater level forecasting, Journal of Hydrology, 650, 132567, https://doi.org/10.1016/j.jhydrol.2024.132567, 2025a.

Zhu, F., Zhu, O., Han, M., Liu, W., and Guo, X.: A hybrid process-data driven framework for real-time hydrological forecasting with interpretable deep learning, Journal of Hydrology, 134082, https://doi.org/10.1016/j.jhydrol.2025.134082, 2025b.

Zhu, S., Wang, Z., Zhang, W., and Yang, J.: Application of the ResNet-transformer model for runoff prediction based on multi-source data fusion, Water Resources Management, 39(12), 6073-6092, https://doi.org/10.1007/s11269-025-04241-3, 2025c.

Zou, Y., Wang, J., Lei, P., and Li, Y.: A novel multi-step ahead forecasting model for flood based on time residual LSTM, Journal of Hydrology, 620, 129521, https://doi.org/10.1016/j.jhydrol.2023.129521, 2023.